Journal of the Indian Society for Probability and Statistics (2024) 25:395–416
https://doi.org/10.1007/s41096-024-00186-9

RESEARCH ARTICLE

A Frailty Model for Semi-competing Risk Data with Applications to Colon Cancer

Elizbeth Ch. Bedia[1] · **Vicente G. Cancho**[1] · **Dipankar Bandyopadhyay**[2]

Accepted: 8 March 2024 / Published online: 13 April 2024

Abstract
In semi-competing risks (which generalizes the competing risks scenario), a subject may experience both terminal and non-terminal events, usually dependent, where the event time to the intermediate non-terminal event (say, tumor recurrence in cancer studies) is subject to censoring by the terminal event (say, death), but not vice-versa. As an alternative to the latent failure time formulation of semi-competing risks with joint survival functions, here, we consider an illness-death (multistate) shared frailty framework, where the dependency between the terminal and non-terminal failure times is incorporated via the power variance frailty between the conditional transition rates that are assumed Markov. Inference is conducted via maximum likelihood. A simulation study is conducted to evaluate the finite sample performance of the model parameters. Finally, we compare and contrast our power variance frailty proposal to known alternatives via application to a colon cancer dataset. Relevant `R` code for implementation of our model is available in GitHub.

Keywords Semi-competing risks · Frailty · Illness-death process · Proportional hazards · Dependent censoring

1 Introduction

In biomedical studies with survival endpoints, the eventual goal is to model the time until occurrence of an event of interest, say death. In some situations, the subjects under study can experience one of the several so-called 'terminal' events, and where the occurrence of an event precludes the subsequent occurrence of any other (Haneuse and Lee 2016). This is the classic 'competing risks' framework (Lau et al. 2009),

✉ Dipankar Bandyopadhyay
dbandyop@vcu.edu

[1] Department of Applied Mathematics and Statistics, University of São Paulo, São Carlos, Brazil

[2] Department of Biostatistics, VCU School of Population Health, Virginia Commonwealth University, One Capitol Square, 7th Floor, 830 East Main Street, PO Box 980032, Richmond, VA 23298-0032, USA

where the time to the event of failure, and the cause of failure/censoring indicator is recorded for each observation. However, in many cancer clinical trials, the occurrence of an intermediate non-terminal event, such as tumor recurrence, is often the event of interest. This non-terminal recurrence may not prevent the subject from death (the terminal event), with the terminal event dependently censoring the non-terminal event, but not vice-versa. This is the semi-competing risks scenario (Fine et al. 2001).

Consider our motivating dataset generated from a randomized, multi-center, concurrently controlled clinical trial (Moertel et al. 1990, 1995) to determine the effectiveness of two adjuvant therapy regimens ('levamisole only', and 'levamisole plus fluorouracil') in improving surgical cure rates for stage III colon cancer. Clinically, a proper evaluation of the survival distribution of patients experiencing intermediate tumor recurrence following a specific treatment regimen post resection is necessary for an informed prediction of the risk of recurrence for other patients, and subsequent adaptive treatment strategies. Yet, this is challenging, given the strong positive correlation (see Fig. 1, left panel) between T_1 (time to tumor recurrence) and T_2 (time to death), with $T_1 \leq T_2$ (data observed only in the upper wedge), which renders the typical independent censoring assumptions in survival analysis (Oakes 1993) and competing-risks (consisting of only terminal events) invalid.

There are two main approaches to analyzing semi-competing risks data. The first approach considers a joint distribution for (T_1, T_2) specified via a copula model in the upper wedge. Within the copula framework, Fine et al. (2001) considered a Clayton copula (Clayton 1978) with two margins, Wang (2003) considered a more general copula setting, while Lakhal et al. (2008) considered an Archimedean copula to estimate the dependency parameter. Related formulation considering time-varying effects of a treatment on the marginal distribution of a non-terminal event was considered in Peng and Fine (2007) and Hsieh and Huang (2012). In addition to semiparametric transformation models (Chen 2012), more recent approaches (Zhou et al. 2016) tackled dependent censoring by first selecting a suitable copula model through an exploratory diagnostic approach (Bandeen-Roche et al. 2005), and then

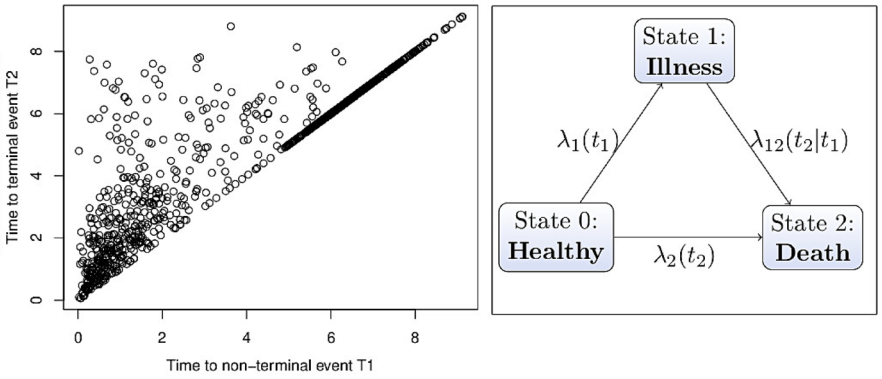

Fig. 1 Correlation plot of T_1 and T_2 for the colon cancer data (left panel). Illness-death model for data with semi-competing risks structure (right panel)

developing an inference procedure to simultaneously estimate the marginal survival function of cancer recurrence, and an association parameter in the copula model. Recently, Peng et al. (2018) considered a semiparametric extension to handle clustered semi-competing risks data. This latent variable formulation via a joint (copula) structure leads to hypothetical interpretation of the marginal distribution of the non-terminal events, and also complicates covariance analysis (Xu et al. 2010). The second approach casts the semi-competing risks problem into the classic illness-death (see, Fig. 1, right panel) compartment framework (Haneuse and Lee 2016), which, in reality, is a special case of multi-state models (Andersen and Keiding 2002). The concept of evaluating semi-competing risks fits naturally into the well-established illness-death paradigm (Xu et al. 2010), where, a patient can either transit to the terminal event either directly, or via the non-terminal event, with the model completely specified via the transition intensity functions for the three distinct transitions. Here, the dependency between T_1 and T_2 is introduced via a shared frailty, or random effects term. While Xu et al. (2010) considered a Gamma frailty, Jiang and Haneuse (2017) extended this to a class of transformation frailty models, allowing a wider range of possible frailty distributions. Within this illness-death framework, there exists other classical (Do Ha et al. 2020; Kim et al. 2019; Lee et al. 2021) and Bayesian (Han et al. 2014; Lee et al. 2015, 2016) frailty-based formulations of semicompeting risks handling various data complications, such as interval-censoring, intermediate missingness, and bias reduction.

In this paper, we consider the power variance function (PVF, Tweedie 1984) as our choice of the frailty, under the illness-death formulation. The PVF is easily tractable (Wienke 2010) due to the closed-form expressions of the marginal survival function, and contains the gamma, inverse Gaussian and positive stable densities as special cases. Choice of the frailty density is crucial in modeling the dependence structure, and misspecified frailty may lead to biased results (Kiche et al. 2019). Although gamma and other frailties have been extensively considered in modeling semicompeting risks, considering a frailty family would be elegant from a generalization perspective, and thereby ascertain whether the fit of these sub-models (that are members of the family) are satisfactory.

The remainder of this article proceeds as follows. After a brief introduction to the PVF density, Sect. 2 presents the general illness-death model, with the corresponding marginal transition rates, joint marginal survival function, leading to its formulation using the PVF frailty. Section 3 develops inference via maximum likelihood, incorporating covariates. In Sect. 4, we apply our model to the colon cancer data. The finite sample performance of the model parameters are evaluated via a small simulation study in Sect. 5. Finally, Sect. 6 concludes, with a discussion. Proof of the theorem presented, as well as detailed derivations of various important results and likelihood construction appear in the accompanying Supplementary Material.

2 Statistical Model

2.1 PVF Density

The PVF density, suggested by Tweedie (1984) and derived independently by Hougaard (1986), is a three-parameter family with parameters $\mu > 0, \sigma > 0$, and $0 < \gamma \leq 1$, with the probability density function $g(z)$ given by:

$$g(z) = e^{-\sigma(1-\gamma)(\frac{z}{\mu}-\frac{1}{\gamma})} \frac{1}{\pi} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(\sigma(1-\gamma))^{k(1-\gamma)} \mu^{k\gamma}}{\gamma^k} \frac{\Gamma(k\gamma+1)}{k!} z^{-k\gamma-1} \sin(k\gamma\pi), \tag{1}$$

The expectation $\mathbb{E}[Z] = \mu$ and variance $\mathrm{Var}[Z] = \mu^2/\sigma$ can be derived from the Laplace transform of the PVF density (Aalen et al. 1992) given by

$$\mathcal{L}_Z(s) = \mathbb{E}[e^{sZ}] = e^{\frac{\sigma(1-\gamma)}{\gamma}\left[1-\left(1+\frac{\mu s}{\sigma(1-\gamma)}\right)^{\gamma}\right]} \tag{2}$$

The PVF density reduces to the gamma ($\gamma \to 0$) and inverse Gaussian ($\gamma = 0.5$) distributions, respectively, while the stable density is a particular case of the PVF under some asymptotic considerations (Wienke 2010). Furthermore, the compound Poisson (cP) distribution (Aalen 1988), which can be constructed as the sum of a Poisson-distributed number of independent and identically distributed (iid) gamma random variables, can be shown to have the same Laplace transform as the PVF model under certain parameterization (Wienke 2010), except for the range of γ, which can be negative in the cP model. Consequently, the density function coincides with the respective function in the PVF model.

2.2 The Illness-Death Framework

Assume n, the number of subjects in our study. Let T_{j1} and T_{j2} be the time to non-terminal and terminal events, respectively, with C_j, the (right) censoring time for the jth subject, $j = 1, \ldots, n$. If the subject dies before the occurrence of the non-terminal event, we define $T_{j1} = \infty$. This specification (Xu et al. 2010) avoids a latent distribution of (T_{j1}, T_{j2}) over the region $t_1 > t_2$. Consider also the p-dimensional covariate vector $\mathbf{x_j} = (x_{j1}, x_{j2}, \ldots, x_{jp})'$ observed for the jth subject. We also assume the censoring time $C_j \perp$ of (T_{j1}, T_{j2}), given $\mathbf{x_j}$. Under this setup, the observed data is denoted by $D = (Y_{j1}, Y_{j2}, \delta_{j1}, \delta_{j2}, \mathbf{x}_j)$, $j = 1, \ldots, n$, where $Y_{j1} = \min\{T_{j1}, Y_{j2}\}$, $\delta_{j1} = I_{\{T_{j1} \leq Y_{j2}\}}$, $Y_{j2} = \min\{T_j, C_j\}$, and $\delta_{j2} = I_{\{T_{j2} \leq C_j\}}$. Also, when $\delta_{j1} = 0$ and $\delta_{j2} = 1$, we have $Y_{j1} = Y_{j2} = T_{j2}$ and $T_{j1} = \infty$. Since $0 \leq Y_{j1} \leq Y_{j2}$, the observations lie on the upper wedge of the two-dimensional graph of (T_{j1}, T_{j2}). For the joint probability model of (T_1, T_2), we consider an absolutely continuous density $f(t_1, t_2), 0 \leq t_1 \leq t_2$ (Fig. 1, left panel). Thus,

$$\int_0^{\infty} \int_{t_1}^{\infty} f(t_1, t_2) dt_2 dt_1 = \mathbb{P}[T_1 < \infty] \leq 1,$$

with the balance of the probability distributed along the line at $t_1 = \infty$ with continuous density $f_\infty(t_2), t_2 > 0$ (Xu et al. 2010).

Now, under the illness-death specification, a subject in the initial state, or state 0 ('full resection of the tumor') can transit directly to state 2 ('death' state), or first to state 1 ('tumor recurrence state'), and then to state 2. This is described in Fig. 1 (right panel). This model is completely specified by the transition or hazard functions for the three distinct transitions: a cause-specific hazard for the illness (tumor) recurrence, $\lambda_1(t_1)$; for death, $\lambda_2(t_2)$; and for death conditional to the time to illness recurrence, $\lambda_{12}(t_2|t_1)$, $0 < t_1 < t_2$. These transition rates are defined as:

$$\begin{aligned}
\lambda_1(t_1) &= \lim_{\Delta\to 0} \frac{\mathbb{P}[T_1 \in [t_1, t_1+\Delta) | T_1 \geq t_1, T_2 \geq t_1]}{\Delta}, \quad t_1 > 0 \\
\lambda_2(t_2) &= \lim_{\Delta\to 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2+\Delta) | T_1 \geq t_2, T_2 \geq t_2]}{\Delta}, \quad t_2 > 0 \\
\lambda_{12}(t_2|t_1) &= \lim_{\Delta\to 0} \frac{\mathbb{P}[T_2 \in [t_2, t_2+\Delta) | T_1 = t_1, T_2 \geq t_2]}{\Delta}, \quad 0 < t_1 < t_2.
\end{aligned} \tag{3}$$

Note, in general, the rate of the terminal event following the occurrence of the non-terminal event at time $T_1 = t_1$, $\lambda_{12}(t_2|t_1)$, can depend on both t_1 and t_2. However, under Markov assumptions, $\lambda_{12}(t_2|t_1)$ depends only on t_2, i.e., $\lambda_{12}(t_2|t_1) = \lambda_{12}(t_2)$. The Markov model is most frequently used because of its simplicity (Meira-Machado et al. 2009). Intuitively, tumor recurrence can occur before the terminal event, but, not vice versa, and the occurrence of the non-terminal event can influence the occurrence of the terminal event, i.e., subject with tumor recurrence may die early. To quantify this (latent) dependency, we consider a shared frailty (or random effect) model (Wienke 2010), where the event times T_1 and T_2 are considered conditionally independent, given the frailty. Denoting this frailty term by a random variable $Z > 0$ with $\mathbb{E}[Z] = 1$, we define conditional transition functions analogous to $\lambda_1(t_1)$, $\lambda_2(t_2)$ and $\lambda_{12}(t_2|t_1)$ as follows:

$$\begin{aligned}
\lambda_1(t_1|Z) &= Z\lambda_{01}(t_1), \quad t_1 > 0, \\
\lambda_2(t_2|Z) &= Z\lambda_{02}(t_2), \quad t_2 > 0, \\
\lambda_{12}(t_2|t_1, Z) &= Z\lambda_{03}(t_2), \quad 0 < t_1 < t_2,
\end{aligned} \tag{4}$$

where $\lambda_{0i}(\cdot)$, $i = 1, 2, 3$ are the baseline hazard functions that can be considered parametric, or non-parametric. The current framework considers the same value of the frailty for the three conditional transition rates. The conditional transition rate for terminal event given that a non-terminal event has occurred, $\lambda_{12}(t_2|t_1, Z)$, is assumed Markov, i.e., $\lambda_{03}(\cdot)$ does not depend on t_1. The conditional explanatory hazard ratio, which characterizes the dependence between T_1 and T_2, is given by $\mathbf{EHR} = \lambda_{12}(t_2|t_1, Z)/\lambda_2(t_2|Z) = \lambda_{03}(t_2)/\lambda_{02}(t_2)$, for $t_2 > t_1$ (Clayton 1978; Xu et al. 2010). Thus, for $\lambda_{02}(t_2) \neq \lambda_{03}(t_2)$ (General model), the dependence of T_2 on T_1 is described both by the conditional (given Z) explanatory hazard ratio, $\lambda_{03}(t_2)/\lambda_{02}(t_2)$, as well as by the common frailty Z, through its unknown parameter, which we denote by $\theta > 0$. When $\lambda_{02}(t_2) = \lambda_{03}(t_2)$, (Restricted model), the dependence of T_1 and T_2 is completely specified by Z. The marginal transition rates can be expressed

as a function of the Laplacian transform $\mathcal{L}_Z(s)$ from the conditional transition rates. This result is presented in Theorem 1 below.

Theorem 1 *Define the conditional transition rates* $\lambda_i(t_i|Z) = Z\lambda_{0i}(t_i)$, $i = 1, 2$, *and* $\lambda_{12}(t_2|t_1, Z) = Z\lambda_{03}(t_2)$, *where* $\lambda_{0i}(\cdot)$, $i = 1, 2, 3$ *are the baseline hazard functions, and* Z *is a continuous random variable* (*frailty*), *with positive support having mean* 1 *and finite variance. If the distribution of* Z *has Laplace transform,* $\mathcal{L}_Z(s)$, *then the marginal transition rates are given by*

$$\text{a. } \lambda_i(t_i) = -\lambda_{0i}(t_i)\frac{d}{ds}\log(\mathcal{L}_Z(s)), \quad s = \Delta_{01}(t_i) + \Delta_{02}(t_i),\ i = 1, 2 \text{ and}$$

$$\text{b. } \lambda_{12}(t_2|t_1) = -\lambda_{03}(t_2)\frac{d}{ds}\log\left(\frac{d}{ds}\mathcal{L}_Z(s)\right),\ s = \Delta_{01}(t_1) + \Delta_{02}(t_1) + \Delta_{03}(t_1, t_2).$$

where $\Delta_{03}(t_1, t_2) = \Delta_{03}(t_2) - \Delta_{03}(t_1)$ *and* $\Delta_{0i}(t_i) = \int_0^{t_i} \lambda_{0i}(t)dt$, $i = 1, 2, 3$. *are baseline cumulative distribution functions.*

Expressions for the conditional survival functions, and proof of Theorem 1 appear in the Supplementary Material (Sections A and B, respectively). Wienke (2010) show that the joint survival function $S(t_1, t_2)$, for shared frailty models can be expressed as the Laplace transform of the frailty distribution evaluated at the cumulative baseline hazard. Along the lines of Xu et al. (2010), we can prove that

$$S(t_1, t_2) = \mathcal{L}_Z(\Delta_{01}(t_1) + \Delta_{02}(t_1) + \Delta_{03}(t_2) - \Delta_{03}(t_1)), \quad 0 < t_1 \le t_2 \tag{5}$$

For the restricted model, the above expression reduces to

$$S(t_1, t_2) = \mathcal{L}_Z(\Delta_{01}(t_1) + \Delta_{02}(t_2)), \quad 0 < t_1 \le t_2 \tag{6}$$

Note, considering the Laplacian of the gamma distribution, the expression in (6) above is the joint survival function proposed by Xu et al. (2010) for the restricted case. Once we know the joint survival function, $S(t_1, t_2)$, we can compute a local measure of association (Clayton 1978; Oakes 1982) between T_1 and T_1, denoted by $\vartheta^\star(t_1, t_2)$, as a function of θ, defined as:

$$\vartheta^\star(t_1, t_2) = \frac{S(t_1, t_2)\frac{\partial^2}{\partial t_1 \partial t_2}S(t_1, t_2)}{\left(\frac{\partial}{\partial t_1}S(t_1, t_2)\right)\left(\frac{\partial}{\partial t_2}S(t_1, t_2)\right)} \tag{7}$$

where ∂ denotes the partial derivatives of the corresponding quantities. Here, $\vartheta^\star(t_1, t_2) = 1(> 1)$ implies independence (positive association) between T_1 and T_2. Under the general and restricted models in (5) and (6), respectively, and assuming the PVF density, $\vartheta^*(t_1, t_2)$ $(0 < t_1 < t_2)$ is well-defined, and takes the form (derivation available in the Supplementary Material, Section C):

$$\vartheta^*(t_1,t_2)=\begin{cases}1+\theta\left(1+\frac{\theta(\Delta_{01}(t_1)+\Delta_{02}(t_1)+\Delta_{03}(t_2)-\Delta_{03}(t_1))}{1-\gamma}\right)^{-\gamma}, & \text{General M.}\\ 1+\theta\left(1+\frac{\theta(\Delta_{01}(t_1)+\Delta_{02}(t_2))}{1-\gamma}\right)^{-\gamma}, & \text{Restricted M.}\end{cases} \tag{8}$$

We have $\lim_{t_i\to\infty}\Delta_{0i}(t_i)=\infty,\ i=1,2,3$. Note, in both cases, $\gamma\longrightarrow 0$ implies $\vartheta^\star(t_1,t_2)=1+\theta$, a constant in (t_1,t_2), which corresponds to the gamma frailty. In this case, the association between T_1 and T_2 increases, or decreases with θ. However, $\vartheta^\star(t_1,t_2)$ decreases with increasing time, with $\vartheta^\star(t_1,t_2)\to 1$ as $t_1\to\infty$ and $t_2\to\infty$. This implies that for a sufficiently large time, T_1 and T_2 are independent, but this does not happen for the gamma model. On the other hand, when $\gamma<0$ (implying the cP model), $\vartheta^\star(t_1,t_2)\to\infty$ as $t_1\to\infty$ and $t_1\to\infty$, i.e., larger t_1 and t_2 leads to greater dependency.

2.3 The Illness-Death PVF Framework

In this section, we consider the illness-death framework for the general model; the corresponding results for the restricted model follows directly by assuming $\lambda_{02}(\cdot)=\lambda_{03}(\cdot)$. In Theorem 1, we consider the frailty Z distributed as a PVF, with $\mathbb{E}[Z]=\mu=1$ and $Var(Z)=\theta=1/\sigma$ (this choice avoids the identifiability problem). Thus, the marginal transition rates are:

$$\begin{aligned}\lambda_i(t_i)&=\lambda_{0i}(t_i)\left(1+\frac{\theta(\Delta_{01}(t_i)+\Delta_{02}(t_i))}{1-\gamma}\right)^{\gamma-1},\ \ t_i>0,\ \ i=1,2.\\ \lambda_{12}(t_2|t_1)&=\lambda_{03}(t_2)\left(1+\frac{\theta(\Delta_{01}(t_1)+\Delta_{02}(t_1)+\Delta_{03}(t_1,t_2))}{1-\gamma}\right)^{-1}\\ &\quad\times\left[\theta+\left(1+\frac{\theta(\Delta_{01}(t_1)+\Delta_{02}(t_1)+\Delta_{03}(t_1,t_2))}{1-\gamma}\right)^{\gamma}\right],0<t_1<t_2.\end{aligned} \tag{9}$$

Note, $\lambda_{12}(t_2|t_1)$, the marginal transition rate from 'recurrence' to 'death', depends on both t_1 and t_2, and is therefore not Markovian, as opposed to its corresponding conditional transition rate, unless Z is constant ($\theta=0$). In this model, not all subjects have the same value of the frailty. Hence, the conditional transition rates are only comparable within subjects sharing frailty. This causes the interpretation of the marginal and conditional transition rates to be different (Lee and Nelder 2004).

As stated before, the PVF density constitutes a family, which includes the Gamma, inverse Gaussian (IG) and positive stable densities (Wienke 2010). In Table 1, we present the marginal transition rates for the Gamma and IG densities; these results were obtained from Eq. (9). The marginal transition rates for the cP model (when $\gamma<0$) are similarly obtained. An interesting property of the cP model is that it allows for a fraction of individuals with zero frailty who never experience the event under study (Wienke 2010). The size of the non-susceptible fraction, p_0, is obtained when $(t_1,t_2)\to\infty$ in $S(t_1,t_2)$, i.e., using the Laplacian (2) in (5) for $\gamma<0$ and assuming $\lim_{t_i\to\infty}\Delta_{0i}(t_i)=\infty, i=1,2,3$, we can show

Table 1 Marginal transition rates for the Inverse Gaussian (IG) and Gamma model

Model	Marginal transition rates	
	$\lambda_i(t_i)\, t_i > 0, i = 1, 2$	$\lambda_{12}(t_2\|t_1), 0 < t_1 < t_2$
IG ($\gamma = 0.5$)	$\frac{\lambda_{0i}(t_i)}{1+2\theta(\Delta_{01}(t_i)+\Delta_{02}(t_i))^{1/2}}$	$\frac{\lambda_{03}(t_2)\theta+\left(1+2\theta(\Delta_{01}(t_1)+\Delta_{02}(t_1)+\Delta_{03}(t_1,t_2))\right)^{1/2}}{\left(1+2\theta(\Delta_{01}(t_1)+\Delta_{02}(t_1)+\Delta_{03}(t_1,t_2))\right)}$
Gamma ($\gamma \to 0$)	$\frac{\lambda_{0i}(t_i)}{1+\theta(\Delta_{01}(t_i)+\Delta_{02}(t_i))}$	$\frac{\lambda_{03}(t_2)(\theta+1)}{1+\theta(\Delta_{01}(t_1)+\Delta_{02}(t_1)+\Delta_{03}(t_1,t_2))}$

$$p_0 = \lim_{\substack{t_1 \to \infty \\ t_2 \to \infty}} S(t_1, t_2) = \exp\left(\frac{1-\gamma}{\theta\gamma}\right). \tag{10}$$

The marginal transition probabilities $\mathbb{P}_{01}(t_1)$, $\mathbb{P}_{02}(t_2)$ and $\mathbb{P}_{12}(t_2|t_1)$ are defined as the probability of being in state 1 at time t_1 given that the previous state 0 was entered at time 0, the probability of being in state 2 at time t_2 given that the previous state (state 0) was entered at time 0, and probability of being in state 2 at time t_2 given that the previous state (state 1) was entered at time t_1, respectively. These quantities are estimated by integrating the conditional transition rates over the possible transition times,

$$\begin{aligned}
\mathbb{P}_{0i}(t_1) &= \int_0^{t_1} \lambda_{01}(t)\left(1+\frac{\theta s_i(t,t_1)}{1-\gamma}\right)^{\gamma-1} \exp\left\{\frac{1-\gamma}{\theta\gamma}\left[1-\left(1+\frac{\theta s_i(t,t_1)}{1-\gamma}\right)^{\gamma}\right]\right\} dt, \quad i = 1, 2. \\
\mathbb{P}_{12}(t_2|t_1) &= \int_{t_1}^{t_2} \lambda_{03}(t)\left(1+\frac{\theta s_3(t,t_1)}{1-\gamma}\right)^{\gamma-1} \exp\left\{\frac{1-\gamma}{\theta\gamma}\left[1-\left(1+\frac{\theta s_3(t,t_1)}{1-\gamma}\right)^{\gamma}\right]\right\} dt
\end{aligned} \tag{11}$$

where $s_1(t,t_1) = \Delta_{01}(t) + \Delta_{02}(t) + \Delta_{03}(t,t_1)$, $s_2(t,t_1) = \Delta_{01}(t) + \Delta_{02}(t)$ and $s_3(t,t_1) = \Delta_{03}(t_1,t)$. The integrals in (11) will be solved numerically, and the appears in proof of this result appear in the Supplementary Material (Section C).

3 Maximum Likelihood Inference

In this section, we develop the inferential procedures for our illness-death model via maximum likelihood (ML). Hereafter we assume a Weibull distribution for hazard baseline function $\lambda_{0i}(t) = \alpha_i\beta_i t^{\beta_i-1}$, for $\alpha_i > 0$ and $\beta_i > 0$, $i = 1, 2, 3$. We incorporate covariates in the conditional transition rates (4) as follows:

$$\begin{aligned}
\lambda_1(t_1|Z,\mathbf{x}) &= Z\lambda_{01}(t_1)\exp(\boldsymbol{\varphi}_1'\mathbf{x}), \quad t_1 > 0 \\
\lambda_2(t_2|Z,\mathbf{x}) &= Z\lambda_{02}(t_2)\exp(\boldsymbol{\varphi}_2'\mathbf{x}), \quad t_2 > 0 \\
\lambda_{12}(t_2|t_1,Z,\mathbf{x}) &= Z\lambda_{03}(t_2)\exp(\boldsymbol{\varphi}_3'\mathbf{x}), \quad 0 < t_1 < t_2
\end{aligned} \tag{12}$$

where $\mathbf{x} = (x_1, \ldots, x_p)'$ is a subject-specific vector of p covariates, $\boldsymbol{\varphi}_i = (\varphi_{i1}, \ldots, \varphi_{ip})'$, $i = 1, 2, 3$ are vectors of coefficients, and Z is the subject-specific shared frailty, distributed independently of $\mathbf{x}$. The corresponding marginal

transition rates with covariates are the same as in (9), where the baseline hazard functions, $\lambda_{0i}(\cdot)$, and the cumulative baseline hazard functions, $\Delta_{0i}(\cdot)$, are multiplied with $\exp(\boldsymbol{\varphi}_i'\mathbf{x})$ for $i = 1, 2, 3$.

The likelihood function is constructed via the conditional transition rates from each of the 4 possible cases, $(\delta_{j1}, \delta_{j2}) = (1,1), (0,1), (1,0), (0,0)$, and then integrating out Z. In this model, we assume the covariate effects are same at all times, and the censoring time C_j is independent of $(T_{j1}, T_{j2}|\mathbf{x}_j)$. Denote $\boldsymbol{\vartheta} = (\gamma, \theta, \boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3)$, the parameter vector to be estimated, where $\boldsymbol{\eta}_i = (\alpha_i, \beta_i)$ is the parameter vector of the Weibull model for the rate $\lambda_{0i}(\cdot)$ and $\boldsymbol{\varphi}_i = (\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{ip})'$, the coefficient vector, $i = 1, 2, 3$. The log-likelihood function for $\boldsymbol{\vartheta}$ given observed data is:

$$
\begin{aligned}
\ell(\boldsymbol{\vartheta}) = & \sum_{j=1}^{n} \left\{ \delta_{j1} \log \lambda_{01}(Y_{j1}) + \delta_{j2}(1-\delta_{j1}) \log \lambda_{02}(Y_{j2}) + \delta_{j1}\delta_{j2} \log \lambda_{03}(Y_{j2}) \right\} \\
& + \sum_{j=1}^{n} \left\{ \delta_{j1} \boldsymbol{\varphi}_1' \mathbf{x}_j + \delta_{j2}(1-\delta_{j1}) \boldsymbol{\varphi}_2' \mathbf{x}_j + \delta_{j1}\delta_{j2} \boldsymbol{\varphi}_3' \mathbf{x}_j \right\} \\
& + \sum_{j=1}^{n} \left\{ \delta_{j1}\delta_{j2} \log \left[\theta + \left(1 + \frac{\theta K_1(\mathbf{x}_j)}{1-\gamma} \right)^{\gamma} \right] + \delta_{j1}(\gamma - 1 - \delta_{j2}) \log \left(1 + \frac{\theta K_1(\mathbf{x}_j)}{1-\gamma} \right) \right\} \\
& + \sum_{j=1}^{n} \left\{ \log \left[\frac{\delta_{j1}(1-\gamma)}{\theta\gamma} \left(1 - \left(1 + \frac{\theta K_1(\mathbf{x}_j)}{1-\gamma} \right)^{\gamma} \right) \right] \right\} \\
& + \sum_{j=1}^{n} \left\{ (\gamma - 1)\delta_{j2}(1-\delta_{j1}) \log \left(1 + \frac{\theta K_2(\mathbf{x}_j)}{1-\gamma} \right) \right\} \\
& + \sum_{j=1}^{n} \left\{ \log \left[\frac{(1-\delta_{j1})(1-\gamma)}{\theta\gamma} \left(1 - \left(1 + \frac{\theta K_2(\mathbf{x}_j)}{1-\gamma} \right)^{\gamma} \right) \right] \right\}
\end{aligned}
\tag{13}
$$

where $K_1(\mathbf{x}_j) = (\Delta_{01}(Y_{j1}) e^{\boldsymbol{\varphi}_1'\mathbf{x}_j} + \Delta_{02}(Y_{j1}) e^{\boldsymbol{\varphi}_2'\mathbf{x}_j} + \Delta_{03}(Y_{j1}, Y_{j2}) e^{\boldsymbol{\varphi}_3'\mathbf{x}_j})$, and $K_2(\mathbf{x}_j) = (\Delta_{01}(Y_{j1}) e^{\boldsymbol{\varphi}_1'\mathbf{x}_j} + \Delta_{02}(Y_{j1}) e^{\boldsymbol{\varphi}_2'\mathbf{x}_j})$, $j = 1, \ldots, n$. The details on the derivation appear in the Supplementary Material (Section D). In the restricted model, note that $\lambda_{02}(Y_{j2}) = \lambda_{03}(Y_{j2})$, for $j = 1, 2, \ldots, n$. However, to guarantee that $\boldsymbol{EHR} = 1$, we also consider $\boldsymbol{\varphi}_2 = \boldsymbol{\varphi}_3$. Thus, the expression in (13) is reduced to the likelihood function of the restricted model.

The ML estimator of $\boldsymbol{\vartheta}$ ($\widehat{\boldsymbol{\vartheta}}$, say) can be obtained via numerical optimization of $\ell(\boldsymbol{\vartheta})$, utilizing the `optim` function in `R`. Although all parameters can be estimated this way, here, we adopt a more computationally efficient profile likelihood approach for γ. The maximization of $\ell(\boldsymbol{\vartheta})$ is a two-step procedure. In general, it is reasonable to expect the parameter γ to belong to the interval $(-\tau, 1)$, $\tau > 0$, since $0 \le \gamma \le 1$ represents the PVF model, while $\gamma < 0$ in the cP model. Hence, the first step involves specifying γ to take values in this interval, and determine the corresponding ML estimates $\widetilde{\theta}(\gamma)$, $\widetilde{\boldsymbol{\eta}}_1(\gamma)$, $\widetilde{\boldsymbol{\eta}}_2(\gamma)$, $\widetilde{\boldsymbol{\eta}}_3(\gamma)$, $\widetilde{\boldsymbol{\varphi}}_1(\gamma)$, $\widetilde{\boldsymbol{\varphi}}_2(\gamma)$, and $\widetilde{\boldsymbol{\varphi}}_3(\gamma)$, and the corresponding (maximized) log-likelihood function $\ell_{\max}(\gamma)$. In the second step, the log-likelihood function $\ell_{\max}(\gamma)$ is maximized, and $\hat{\gamma}$ is obtained. The ML estimates of θ, $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$, $\boldsymbol{\eta}_3$,

$\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \boldsymbol{\varphi}_3$ are, respectively, given by, $\hat{\theta} = \widetilde{\theta}(\gamma)$, $\widehat{\boldsymbol{\eta}}_1 = \widetilde{\boldsymbol{\eta}}_1(\gamma)$, $\widehat{\boldsymbol{\eta}}_2 = \widetilde{\boldsymbol{\eta}}_2(\gamma)$, $\widehat{\boldsymbol{\eta}}_3 = \widetilde{\boldsymbol{\eta}}_3(\gamma)$, $\widehat{\boldsymbol{\varphi}}_1 = \widetilde{\boldsymbol{\varphi}}_1(\gamma)$, $\widehat{\boldsymbol{\varphi}}_2 = \widetilde{\boldsymbol{\varphi}}_2(\gamma)$ and $\widehat{\boldsymbol{\varphi}}_3 = \widetilde{\boldsymbol{\varphi}}_3(\gamma)$. Under suitable regularity conditions (Maller and Zhou 1996), it can be shown that the asymptotic distribution of the MLE $\widehat{\boldsymbol{\vartheta}} = (\hat{\theta}, \widehat{\boldsymbol{\eta}}_1, \widehat{\boldsymbol{\eta}}_2, \widehat{\boldsymbol{\eta}}_3, \widehat{\boldsymbol{\varphi}}_1, \widehat{\boldsymbol{\varphi}}_2, \widehat{\boldsymbol{\varphi}}_3)$ is multivariate normal with mean vector $\boldsymbol{\vartheta}$ and covariance matrix $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\vartheta}}) = \left\{ -\frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^\top} \right\}^{-1} = \{-J(\boldsymbol{\vartheta})\}^{-1}$, evaluated at $\boldsymbol{\vartheta} = \widehat{\boldsymbol{\vartheta}}$. The elements of the observed matrix $J(\boldsymbol{\vartheta})$ are obtained numerically from the Hessian matrix, using the `optim` function. An asymptotic confidence interval with significance level α for each parameter ϑ_r is given by $\left(\hat{\vartheta}_r - z_{\alpha/2} \sqrt{\widehat{\Sigma}^{r,r}}, \hat{\vartheta}_r + z_{\alpha/2} \sqrt{\widehat{\Sigma}^{r,r}} \right)$, where $\widehat{\Sigma}^{r,r}$ is the rth diagonal element of $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\vartheta}})$ estimated at $\widehat{\boldsymbol{\vartheta}}$, for $r = 1, \ldots, p + \dim(\boldsymbol{\varphi}) + 1$, $\dim(\cdot)$ denotes the dimension of the parametric space, and $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

4 Application: Colon Cancer Data

We now apply the illness-death model to the dataset generated from the stage III colon cancer clinical trial. The data appears as the `colon` data in the R `survival` package (Therneau 2015). Here, patients who had curative-intent resections of cancer were assigned to one of the (1) observation only, (2) Levamisole (Lev) only, and (3) Lev + Fluorouracil (5-FU). Here, after the full surgical resection of tumor, 929 patients were followed for 5 years or more (median follow up = 6.5 years). The scientific objective here is to evaluate covariate and treatment effects on the rates of the terminal and non-terminal events, and the dependence between these events. After deleting subjects with incomplete data and missing observation times, we have a subset of $n = 888$ patients with approximately 50% of censoring. For each patient, we have t_{j1}: time to tumor recurrence (in years), t_{j2}: time to death (in years), and x_j: treatment (Observation only, Lev, Lev + 5-FU), $j = 1, \ldots, 888$. R code for implementing our model are available in the `GitHub` link: https://github.com/bandyopd/PVF-Semicompeting.

We consider a Weibull baseline hazard, specified by $\lambda_{0i}(t) = \alpha_i \beta_i t^{\beta_i - 1}$, $i = 1, 2, 3$, with the term incorporating covariates in (13) as $\phi_{ij} = \exp(\varphi_{i1} x_{j1} + \varphi_{i2} x_{j2})$, $j = 1, \ldots, n$, where the covariate (treatment) effects are defined via dummy variables as:

$$x_1 = \begin{cases} 1, & \text{if Levamisole;} \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad x_2 = \begin{cases} 1, & \text{if Levamisole+5-FU;} \\ 0, & \text{otherwise.} \end{cases}$$

We now fit our illness-death model with the PVF frailty, with the particular cases: Inverse Gaussian (IG, $\gamma = 0.5$), and Gamma ($\gamma \to 0$). The parameter γ was estimated via profile likelihood; see Fig. 2 for the respective plots corresponding to the full and the restricted model. The ML estimates (MLE), standard errors (SE), the corresponding confidence intervals (95% CI) are listed in Table 2, for both the general

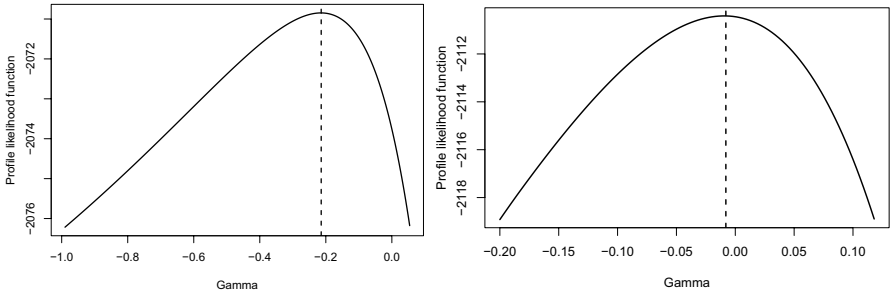

Fig. 2 Profile log-likelihood function of γ. General model (left panel), Restricted model (right panel)

and restricted models. In both cases, we observe a strong evidence of dependence between time to recurrence, and death (revealed by the estimate of θ). Note, for the three frailty models, $\hat{\theta}$ is smaller in the general model, indicating that part of the dependency between T_1 and T_2 is captured by the varying baseline hazards. We also observe that although Lev alone has no significant effect on the risk of recurrence, the combination (Lev + 5-FU) does. In addition, both regimens do not have any significant effect on the risk of death, with or without tumor recurrence, as revealed from both models.

Model comparisons were conducted using the popular AIC/BIC criteria and presented in Table 3. For the restricted model, Gamma frailty provides a (marginally) better fit, also revealed in Table 2 (since, $\hat{\gamma}$ estimate is very close to zero.) Note, although the log-likelihoods for PVF and Gamma were almost identical, the ΔBIC (difference between the BICs) was more prominent (compared to ΔAIC), yet not exceeding the ≥ 10 rule of thumb (Anderson and Burnham 2004) to generate some enthusiastic support for the Gamma model. For the general model, both ΔAIC and Δ BIC were < 10, implying considerably less support for the Gamma model. Although both models can be chosen, we consider the PVF/cP model in our subsequent analysis. The estimated proportion of patients remaining disease free, i.e., $\hat{p_0}$ (given in (10)) is 0.2561, which implies that after a long period of time, approximately 25% of patients might experience neither tumor recurrence nor death from colon cancer after complete tumor resection surgically.

In Fig. 3 (left panel), we plot the conditional explanatory hazard ratio (HR) including covariates, given by $\text{EHR} = \frac{\lambda_{03}(t_2)}{\lambda_{02}(t_2)} \exp\left[(\boldsymbol{\varphi}_3^T - \boldsymbol{\varphi}_2^T)\mathbf{x}\right], t_2 > t_1$. Here, EHR describes how the risk of death changes over time, given that the tumor recurrence occurred at time t_1 (Lee et al. 2015). In particular, EHR > 1 implies tumor recurrence has an effect on the hazard of death due to colon cancer. Although EHR does not depend on t_1 due to it's Markov structure, it's interpretation depends on the condition $t_2 > t_1$, for all t_1 fixed. For a better visualization, we zoomed in on the EHR curve (see Fig. 3, right panel). We observe that the EHR for patients under the competing regimens (Observation only, Lev only, and Lev + 5-FU) is 1 at approximately $t = 19.5, 33.5$ and 38 years, respectively. This implies that at those instants, the time to tumor recurrence has no effect on death.

Table 2 Analysis of Colon cancer data for the PVF, Inverse Gaussian (IG) and Gamma frailty models

	Covariates	Parameters	PVF			IG			Gamma		
			MLE	SE	95% CI	MLE	SE	95% CI	MLE	SE	95% CI
General Model			Transition rate for recurrence								
		γ	− 0.214	–	–	–	–	–	–	–	–
		θ	4.164	1.04	(2.12; 6.20)	9.151	2.96	(3.36; 14.94)	6.364	0.61	(5.18; 7.55)
		β_1	1.764	0.12	(1.54; 1.99)	1.169	0.07	(1.03; 1.31)	1.875	0.11	(1.66; 2.09)
		$\log(\alpha_1)$	− 0.443	0.24	(− 0.90; 0.02)	− 0.555	0.20	(− 0.95; − 0.16)	− 0.150	0.21	(− 0.56; 0.26)
	x_{Lev}	φ_{1_1}	0.023	0.28	(− 0.53; 0.57)	− 0.054	0.18	(− 0.40; 0.30)	0.025	0.27	(− 0.50; 0.55)
	$x_{\text{Lev+5-FU}}$	φ_{1_2}	− 0.585	0.29	(− 1.15;− 0.02)	− 0.827	0.19	(− 1.20; − 0.45)	− 0.747	0.28	(− 1.30; − 0.19)
			Transition rate for death								
		β_2	2.543	0.22	(2.11 2.97)	1.636	0.17	(1.30; 1.97)	2.597	0.23	(2.15; 3.04)
		$\log(\alpha_2)$	− 3.676	0.41	(− 4.47; − 2.88)	− 3.654	0.39	(− 4.42; − 2.89)	− 3.313	0.39	(− 4.09; − 2.54)
	x_{Lev}	φ_{2_1}	− 0.193	0.50	(− 1.18; 0.79)	− 0.298	0.44	(− 1.17; 0.57)	− 0.207	0.50	(− 1.18; 0.77)
	$x_{\text{Lev+5-FU}}$	φ_{2_2}	− 0.217	0.47	(− 1.14; 0.71)	− 0.389	0.41	(− 1.19; 0.41)	− 0.385	0.47	(− 1.31; 0.54)
			Transition rate from recurrence to death								
		β_3	1.933	0.15	(1.64; 2.23)	1.785	0.14	(1.52; 2.05)	2.222	0.14	(1.94; 2.50)
		$\log(\alpha_3)$	− 1.590	0.22	(− 2.02; − 1.16)	− 1.136	0.20	(− 1.53; − 0.75)	− 1.578	0.20	(− 1.98; − 1.18)
	x_{Lev}	φ_{3_1}	0.138	0.29	(− 0.43; 0.71)	0.099	0.21	(− 0.32; 0.52)	0.173	0.27	(− 0.36; 0.70)
	$x_{\text{Lev+5-FU}}$	φ_{3_2}	0.188	0.30	(− 0.41; 0.78)	− 0.022	0.24	(− 0.49; 0.44)	0.076	0.29	(− 0.49; 0.64)

Table 2 (continued)

	Covariates	Parameters	PVF			IG			Gamma		
			MLE	SE	95% CI	MLE	SE	95% CI	MLE	SE	95% CI
Restricted Model			Transition rate for recurrence								
		γ	− 0.01	–	–	–	–	–	–	–	–
		θ	8.92	1.32	(6.35; 11.50)	5.02*	0.63*	(3.79; 6.24)*	9.26	0.63	(8.02; 10.50)
		β_1	2.44	0.10	(2.24; 2.65)	1.74	0.06	(1.61; 1.86)	2.45	0.1	(2.26; 2.65)
		$\log(\alpha_1)$	0.43	0.26	(− 0.08; 0.93)	0.99	0.60	(0.18; 2.17)	0.46	0.24	(− 0.01; 0.92)
	x_{Lev}	φ_{1_1}	0.11	0.32	(− 0.51; 0.74)	− 0.12	0.21	(− 0.53; 0.30)	0.11	0.32	(− 0.52; 0.74)
	$x_{\text{Lev+5-FU}}$	φ_{1_2}	− 0.67	0.34	(− 1.34; 0.01)	− 1.22	0.23	(− 1.68; − 0.76)	− 0.69	0.34	(− 1.36; − 0.02)
			Transition rate for death								
		β_2	2.94	0.13	(2.69; 3.19)	2.31	0.09	(2.14; 2.48)	2.95	0.12	(2.72; 3.19)
		$\log(\alpha_2)$	− 1.88	0.24	(− 2.35; − 1.41)	− 0.71	0.59	(− 1.87; 0.45)	− 1.85	0.23	(− 2.31; − 1.40)
	x_{Lev}	φ_{2_1}	0.22	0.32	(− 0.41; 0.85)	− 0.06	0.22	(− 0.50; 0.38)	0.22	0.33	(− 0.42; 0.86)
	$x_{\text{Lev+5-FU}}$	φ_{2_2}	0.01	0.35	(− 0.67; 0.69)	− 0.77	0.24	(− 1.24; − 0.30)	− 0.01	0.35	(− 0.69; 0.67)

*Estimates are for $\log(\theta)$

Table 3 Log-likelihood, AIC and BIC for the fitted models

Model	General model			Restricted model		
	Log-likelihood	AIC	BIC	Log-likelihood	AIC	BIC
PVF/cP	− 2070.847	4167.695	4230.538	− 2110.411	4240.823	4289.164
IG	− 2135.397	4296.793	4359.637	− 2304.823	4627.645	4671.152
Gamma	− 2073.759	4173.517	4236.361	− 2110.445	4238.890	4282.397

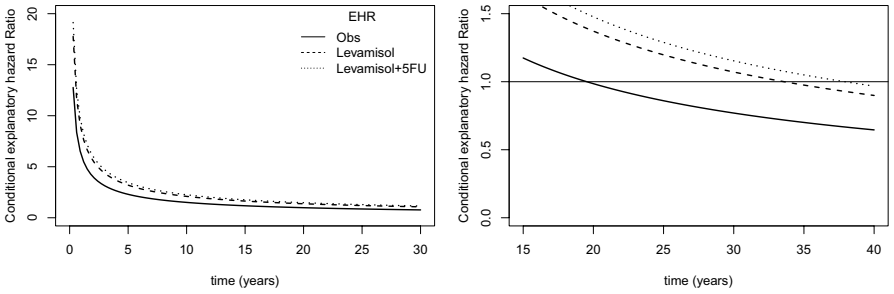

Fig. 3 Estimated (conditional) hazard ratio (**EHR**) for each treatment

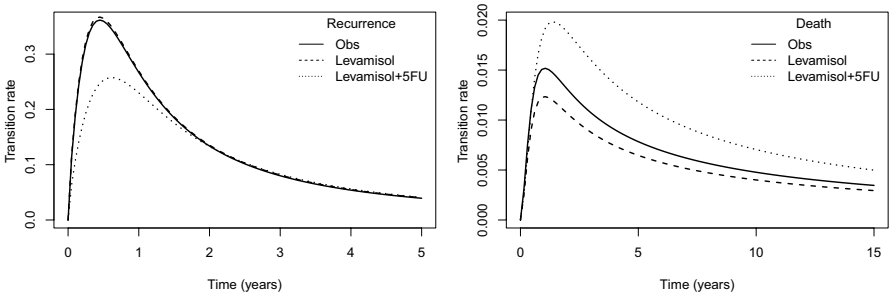

Fig. 4 Estimated marginal transition rates for the transition to recurrence, $\hat{\lambda}_1(t)$ (left panel) and transition to death, $\hat{\lambda}_2(t)$ (right panel), stratified by regimen type

In Fig. 4, we plot the marginal transition rates, $\hat{\lambda}_1(t)$ and $\hat{\lambda}_2(t)$, stratified by treatments. We observe that in the first year, the patients following Lev+ 5-FU regimen have a lower risk of tumor recurrence, compared to patients in the other two groups. This finding is consistent with Table 2, where we observe a significant effect only for the Lev + 5-FU group on tumor recurrence. However, the transition rate to death (from $\hat{\lambda}_2(t)$) is higher with Lev + 5-FU, although Table 2 reveals regimen types do not have significant effects on death.

The estimated transition rates for death after tumor recurrence is plotted in Fig. 5. Note, this transition rate depends on the time to death t_2, in addition to the time since tumor recurrence, and the plots varies with recurrence times t_1. If we

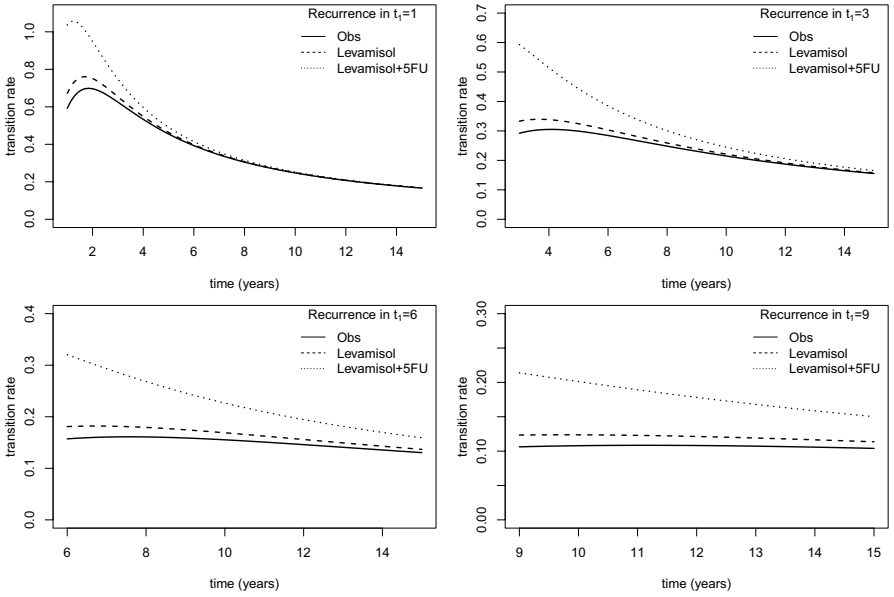

Fig. 5 Estimated marginal transition rates to death given that a patient has had tumor recurrence in t_1, $\hat{\lambda}_1(t_2|t_1)$, stratified by the explanatory variable treatment, for $t_1 = 1, 3, 6$ and 9 years

compare these transition rates with the transition rates for death in Fig. 4 (right panel), the curves without tumor recurrence (Fig. 4 right panel) lie significantly lower than the curves with tumor recurrence (Fig. 5), implying strong evidence of the effect of the non-terminal event (tumor recurrence) on the terminal event (death), and the dependence between them. We can infer that a patient with tumor recurrence has a higher risk of death than another without tumor recurrence. However, we also observe that as t_1 increases, the transition rates to death after tumor recurrence decreases. This implies the effect of tumor recurrence to be less prominent for patients who experience tumor recurrence at the end of the study, than patients who experienced it early.

Finally, Fig. 6 plots the transition probabilities presented in Sect. 2.3. From the plot of $\hat{\mathbb{P}}_{01}(t_1)$ (for tumor recurrence, see upper left), we observe that the probability is lowest for the Lev + 5-FU group, which corroborates with the finding in Table 2. For these patients, the first few post-surgical months following full tumor resection are critical, as revealed by the increasing probability of recurrence till time t, beyond which it decreases. However, the corresponding probabilities for death, i.e., $\hat{\mathbb{P}}_{02}(t_2)$ (see upper right panel), are increasing for all time points, and treatments. Interestingly, this probability is the largest for the Lev + 5-FU group. The plot in the lower left panel compares $\hat{\mathbb{P}}_{01}(t_1)$ to $\hat{\mathbb{P}}_{02}(t_1)$. We observe that a patient is more likely to experience recurrence than death within the first 2.5 years post resection, beyond which the situation reverses. Finally, Fig. 6 (lower

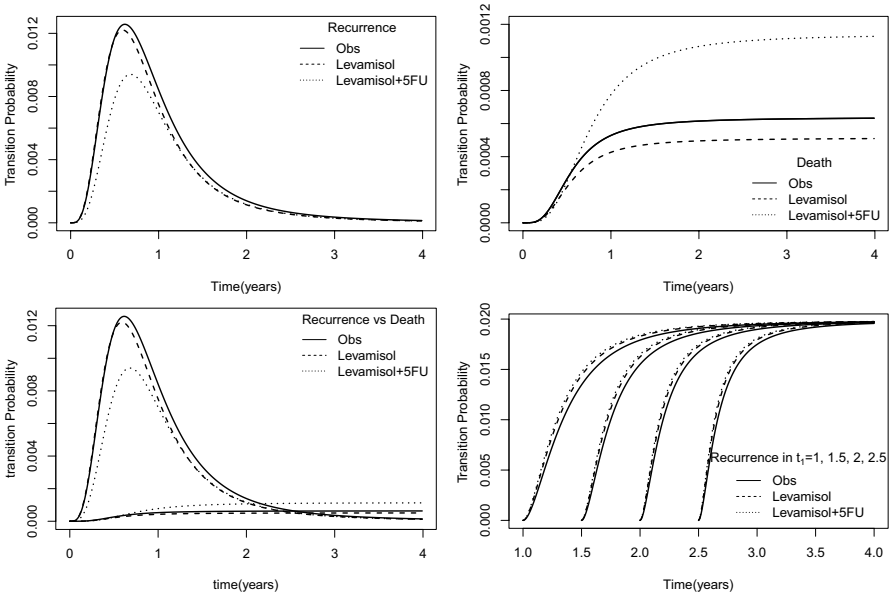

Fig. 6 Estimated marginal transition probabilities for recurrence($\hat{\mathbb{P}}_1(t)$, upper left), death ($\hat{\mathbb{P}}_2(t)$, upper right), their comparison plot (with $\hat{\mathbb{P}}_1(t)$, lower left), and death given patient had tumor recurrence at t_1 ($\hat{S}_{12}(t|t_1)$, lower right), stratified by the explanatory variable treatment regimen. $\hat{S}_{12}(t|t_1)$ is plotted for various values of t_1

right) plots $\hat{\mathbb{P}}_{12}(t|t_1)$, the transition probability to death, given that the patient was in the recurrence state at time t_1, for various choices of t_1. Comparing with $\hat{\mathbb{P}}_{02}(t_2)$ (upper right), we infer that a patient experiencing tumor recurrence at time t_1 is more likely to progress to death, compared to another patient without recurrence. The dependence between the recurrence and death is now manifested through the transition probabilities.

5 Simulation Study

In this section, we present a small simulation study to evaluate the finite sample performance of the ML estimates of the model parameters. We generate semi-competing risks data, following the algorithm of Selle (2016) as follows:

1. Generate $Z \sim PVF(\gamma, 1, \frac{1}{\theta})$, for $\gamma = 0.5$.
2. *Simulating the first event time* t_1 : Denote $p_1 = \mathbb{P}$[not having any transitions before $t_1|Z$], and generate $u_1 \sim$ Uniform(0,1). Equating the expression for p_1 to u_1, we solve for t_1.

3. *Simulating the second event time* t_2: Denote $p_2 =$ conditional probability of staying in state 1 until t_2, and generate $u_2 \sim$Uniform(0,1). Equating the expression for p_2 to u_2, we solve for t_2.
4. Denote $p = \mathbb{P}$[The transition at time**is to state 1**], and generate $u \sim$Uniform(0,1). If $u \leq p$ then $Y_1 = t_1$, $\delta_1 = 1$ and $Y_2 = t_2$, else, $Y_1 = Y_2 = t_1$ and $\delta_1 = 0$.
5. *Simulating the censoring time* C: Generate $u_3 \sim$Uniform(0,1). If $u_3 < 0.5$, then $C \sim$Uniform(v, w), else $C = w$.
6. If $C > Y_2$, then $\delta_2 = 1$; else if $C \geq Y_1$ and $C \leq Y_2$, then $Y_2 = C$ and $\delta_2 = 0$, else $Y_1 = Y_2 = C$ and $\delta_1 = \delta_2 = 0$.

In Step 2, the probability of not having any transitions before time t is given in (A.1), and considering the conditional model with covariates given in (12), we have $p_1 = \exp(-Z[\Delta_{01}(t)e^{(\boldsymbol{\varphi}_1'\mathbf{x})} + \Delta_{02}(t)e^{(\boldsymbol{\varphi}_2'\mathbf{x})}])$. Similarly, if there is a transition to state 1, the second event time t_2 is simulated in step 3, where the conditional probability of staying in state 1 until t_2 is given in (A.2). From (12), we have $p_2 = \exp(-Z[\Delta_{03}(t_1, t_2)e^{\boldsymbol{\varphi}_3'\mathbf{x}}])$. It follows that the probability of going to state 1 at time t (step 4) is $p = \lambda_{01}(t)e^{\boldsymbol{\varphi}_1'\mathbf{x}}/(\lambda_{01}(t)e^{\boldsymbol{\varphi}_1'\mathbf{x}} + \lambda_{02}(t)e^{\boldsymbol{\varphi}_2'\mathbf{x}})$. In step 5, the censoring time was simulated from a mixture distribution, i.e., from a uniform distribution on (1.5,3) with probability 0.5, and a point mass at 3 with probability 0.5. This restricts the average percentage of censored observations between 10% to 20%. The covariates x_{1j} and x_{2j} were generated, respectively, from a Bernoulli (0.5), and uniform (0, 1) distribution, for $j = 1, \ldots, n$. The baseline hazard, once again, follows Weibull (α_i, β_i), $i = 1, 2, 3$. For the restricted model, we consider the coefficient vectors $\boldsymbol{\varphi_1} = \boldsymbol{\varphi_2} = \boldsymbol{\varphi_3} = (1, 1)$, the Weibull parameters $\log\alpha_1 = \log\alpha_2 = \log\alpha_3 = 1$, and $\beta_1 = \beta_2 = \beta_3 = 1$, such that $\boldsymbol{EHR} = 1$. For the general model, we consider $\boldsymbol{\varphi_1} = \boldsymbol{\varphi_2} = \boldsymbol{\varphi_3} = (1, 1)$, $\log\alpha_1 = \log\alpha_2 = 1$, $\log\alpha_3 = 1.25$ and $\beta_1 = \beta_2 = \beta_3 = 1$, leading to $\boldsymbol{EHR} > 1$. Finally, we consider the frailty parameter $\theta = 1$.

For the simulations, we consider sample sizes $n = 250, 500$ and 1000. For each configuration, we conduct $N = 1000$ replications to calculate the averages of the MLEs, denoted by $\bar{\boldsymbol{\vartheta}}$. We compute the standard deviations $SD = \sqrt{\sum_{k=1}^{N}(\hat{\boldsymbol{\vartheta}}_k - \bar{\boldsymbol{\vartheta}})/N}$, where $\hat{\boldsymbol{\vartheta}}_k$ is the MLE vector in the kth simulation, $k = 1, \ldots, N$, root mean squared errors $RMSE = \sqrt{\sum_{k=1}^{N}(\hat{\boldsymbol{\vartheta}}_k - \boldsymbol{\vartheta}^{(\mathbf{0})})^2/N}$ where $\boldsymbol{\vartheta}^{(\mathbf{0})}$ is the initial value vector of parameters, the average standard error (MSE), where $SE_k = \sqrt{\text{diag}(\widehat{\boldsymbol{\Sigma}}_k(\widehat{\boldsymbol{\vartheta}}_k))}$, $k = 1, \ldots, N$, where $\text{diag}(\widehat{\boldsymbol{\Sigma}}_k(\widehat{\boldsymbol{\vartheta}}_k))$ is the diagonal of the kth estimated variance and covariances matrix, the empirical 95% coverage probabilities (CP) for the model parameters, and bias. The simulation results listed in Tables 4 and 5 for the general and restricted models, respectively, reveal that the MLEs are close to the true values, the bias, RMSE and standard errors decrease as sample size increases and the empirical CPs are closer to the nominal coverage level as sample size increases, which are all expected if the underlying estimation scheme is working correctly to produce consistent and asymptotically normal estimates. The simulation study was also repeated for $\theta = 0.5 < 1$ and $\theta = 1.15 > 1$, and the results obtained were very similar.

Table 4 Simulation results for the general model ($\boldsymbol{EHR} > 1$)

Sample size	Rates	Par	Truth	Mean	BIAS	MSE	SD	RMSE	CP
$n = 250$		$\log\theta$	0.0	− 0.182	− 0.182	0.932	0.823	0.843	99.4
	$\boldsymbol{\lambda_1}$	$\log\beta_1$	0.0	− 0.003	− 0.003	0.122	0.110	0.110	95.4
		$\log\alpha_1$	1.0	0.995	− 0.005	0.480	0.433	0.433	95.3
		φ_{11}	1.0	0.997	− 0.003	0.251	0.243	0.243	94.8
		φ_{12}	1.0	1.009	0.009	0.402	0.387	0.387	94.5
	$\boldsymbol{\lambda_2}$	$\log\beta_2$	0.0	− 0.006	− 0.006	0.122	0.108	0.108	95.0
		$\log\alpha_2$	1.0	0.986	− 0.014	0.480	0.418	0.418	94.6
		φ_{21}	1.0	1.002	0.002	0.251	0.251	0.251	94.9
		φ_{22}	1.0	1.004	0.004	0.402	0.394	0.394	95.7
	$\boldsymbol{\lambda_{12}}$	$\log\beta_3$	0.0	− 0.013	− 0.013	0.222	0.211	0.211	94.4
		$\log\alpha_3$	1.25	1.266	0.016	0.381	0.359	0.359	95.5
		φ_{31}	1.0	0.993	− 0.007	0.326	0.321	0.321	95.5
		φ_{32}	1.0	1.035	0.035	0.493	0.501	0.502	94.7
$n = 500$		$\log\theta$	0	− 0.058	− 0.059	0.641	0.592	0.594	97.8
	$\boldsymbol{\lambda_1}$	$\log\beta_1$	0.0	0.004	0.004	0.091	0.084	0.084	95.2
		$\log\alpha_1$	1.0	1.030	0.031	0.364	0.341	0.342	94.7
		φ_{11}	1.0	0.999	− 0.001	0.180	0.175	0.175	96.6
		φ_{12}	1.0	0.992	− 0.008	0.286	0.287	0.286	94.1
	$\boldsymbol{\lambda_2}$	$\log\beta_2$	0.0	0.002	0.002	0.091	0.085	0.085	94.6
		$\log\alpha_2$	1.0	1.024	0.024	0.364	0.340	0.340	94.2
		φ_{21}	1.0	1.000	0.000	0.180	0.179	0.179	95.2
		φ_{22}	1.0	0.997	− 0.003	0.285	0.288	0.288	94.3
	$\boldsymbol{\lambda_{12}}$	$\log\beta_3$	0.0	0.001	0.001	0.165	0.155	0.155	94.5
		$\log\alpha_3$	1.25	1.279	0.029	0.280	0.273	0.275	95.5
		φ_{31}	1.0	1.002	0.002	0.237	0.216	0.216	96.1
		φ_{32}	1.0	1.017	0.017	0.352	0.350	0.350	94.3
$n = 1000$		$\log\theta$	0	0.010	0.010	0.448	0.417	0.417	96.7
	$\boldsymbol{\lambda_1}$	$\log\beta_1$	0.0	0.010	0.010	0.066	0.062	0.063	95.5
		$\log\alpha_1$	1.0	1.034	0.034	0.264	0.251	0.253	96.1
		φ_{11}	1.0	1.010	0.010	0.128	0.126	0.126	95.8
		φ_{12}	1.0	1.013	0.013	0.202	0.201	0.201	95.5
	$\boldsymbol{\lambda_2}$	$\log\beta_2$	0.0	0.006	0.007	0.066	0.062	0.063	94.6
		$\log\alpha_2$	1.0	1.030	0.030	0.264	0.254	0.256	96.0
		φ_{21}	1.0	1.006	0.006	0.128	0.125	0.125	96.4
		φ_{22}	1.0	1.009	0.009	0.202	0.206	0.206	94.8
	$\boldsymbol{\lambda_{12}}$	$\log\beta_3$	0.0	0.008	0.008	0.119	0.111	0.112	94.5
		$\log\alpha_3$	1.25	1.274	0.023	0.200	0.188	0.189	96.2
		φ_{31}	1.0	1.014	0.014	0.170	0.168	0.168	95.0
		φ_{32}	1.0	1.015	0.015	0.249	0.248	0.248	95.1

Table 5 Simulation results for the restricted model ($\boldsymbol{EHR} = 1$)

Sample size	Rates	Par	Truth	Mean	BIAS	Mean SE	SD	RMSE	CP
$n = 250$		$\log\theta$	0.0	− 0.151	− 0.151	0.865	0.761	0.776	99.2
	$\boldsymbol{\lambda_1}$	$\log\beta_1$	0.0	− 0.001	− 0.001	0.117	0.102	0.102	96.6
		$\log\alpha_1$	1.0	1.002	0.002	0.464	0.413	0.412	96.1
		φ_{11}	1.0	0.998	− 0.002	0.249	0.241	0.241	95.1
		φ_{12}	1.0	1.011	0.011	0.401	0.389	0.389	94.6
	$\boldsymbol{\lambda_2}$	$\log\beta_2$	0.0	− 0.002	− 0.002	0.117	0.103	0.103	95.7
		$\log\alpha_2$	1.0	1.008	0.008	0.463	0.416	0.416	95.0
		φ_{21}	1.0	0.994	− 0.006	0.249	0.242	0.242	95.2
		φ_{22}	1.0	0.999	− 0.001	0.401	0.391	0.391	96.4
	$\boldsymbol{\lambda_{12}}$	$\log\beta_3$	0.0	− 0.004	− 0.004	0.209	0.194	0.194	95.3
		$\log\alpha_3$	1.0	1.023	0.023	0.367	0.352	0.353	96.5
		φ_{31}	1.0	0.999	− 0.001	0.321	0.321	0.321	94.8
		φ_{32}	0.5	1.023	0.492	0.503	0.023	0.503	94.0
$n = 500$		$\log\theta$	0.0	− 0.025	− 0.025	0.597	0.559	0.559	97.7
	$\boldsymbol{\lambda_1}$	$\log\beta_1$	0.0	0.009	0.008	0.087	0.082	0.083	94.7
		$\log\alpha_1$	1.0	1.049	0.049	0.350	0.336	0.340	95.4
		φ_{11}	1.0	1.003	0.003	0.178	0.176	0.176	96.3
		φ_{12}	1.0	0.989	− 0.011	0.285	0.287	0.287	94.0
	$\boldsymbol{\lambda_2}$	$\log\beta_2$	0.0	0.006	0.006	0.087	0.082	0.082	94.9
		$\log\alpha_2$	1.0	1.039	0.039	0.349	0.329	0.331	95.4
		φ_{21}	1.0	1.009	0.009	0.178	0.176	0.177	95.7
		φ_{22}	1.0	1.000	0.000	0.285	0.286	0.286	94.8
	$\boldsymbol{\lambda_{12}}$	$\log\beta_3$	0.0	0.009	0.009	0.154	0.145	0.145	94.2
		$\log\alpha_3$	1.0	1.037	0.037	0.267	0.269	0.271	95.3
		φ_{31}	1.0	1.008	0.008	0.232	0.209	0.209	96.4
		φ_{32}	0.5	1.018	0.018	0.351	0.348	0.348	94.8
$n = 1000$		$\log\theta$	0.0	0.024	0.024	0.418	0.391	0.392	96.2
	$\boldsymbol{\lambda_1}$	$\log\beta_1$	0.0	0.010	0.010	0.063	0.058	0.059	95.8
		$\log\alpha_1$	1.0	1.038	0.038	0.252	0.241	0.244	96.2
		φ_{11}	1.0	1.009	0.009	0.127	0.122	0.122	95.7
		φ_{12}	1.0	1.012	0.012	0.201	0.195	0.195	95.9
	$\boldsymbol{\lambda_2}$	$\log\beta_2$	0.0	0.008	0.008	0.063	0.060	0.060	95.2
		$\log\alpha_2$	1.0	1.034	0.034	0.251	0.243	0.246	95.9
		φ_{21}	1.0	1.008	0.008	0.127	0.123	0.124	96.3
		φ_{22}	1.0	1.011	0.011	0.201	0.204	0.205	94.5
	$\boldsymbol{\lambda_{12}}$	$\log\beta_3$	0.0	0.012	0.012	0.111	0.104	0.105	94.9
		$\log\alpha_3$	1.0	1.029	0.029	0.189	0.183	0.185	96.2
		φ_{31}	1.0	1.014	0.014	0.165	0.161	0.161	95.2
		φ_{32}	0.5	1.013	0.013	0.248	0.243	0.244	95.4

6 Conclusion

In this paper, we consider modeling semi-competing data that arises in cancer clinical trials via the shared frailty illness-death framework. Our central contribution is to avoid the latent failure time approach (which comes with well-explored limitations), and explore the illness-death specification via the PVF frailty, which is a flexible and general class containing the gamma, inverse Gaussian and positive stable densities. The convenient Laplace transform of the PVF allows mathematically tractable representations of the hazards, transition probabilities, and survival functions. The simulation study indicate adequate finite sample performance with increasing sample sizes.

The `R` package `SmoothHazard` (Touraine et al. 2013) also fits an illness-death model with both Weibull (parametric), or penalized M-splines (semi-parametric) specifications of baseline hazard for arbitrarily censored survival data. To compare and contrast, our approach incorporates a dependency structure between recurrence time and death time using a shared frailty between conditional transition rates, which acts as a multiplicative effect on baseline hazard rates. This dependency is characterized by the frailty parameter θ.

Our current proposal of a (parametric) Weibull baseline hazard is certainly from the context of achieving computational stability. Future extensions may include various popular and flexible choices of the baseline hazard from the non- and semi-parametric toolbox (Ibrahim et al. 2001), such as a piecewise-constant, and study its impact both on parameter estimation and computational gain. Depending on the dataset, the current setup can also be extended to include a higher-dimensional covariate space, with a principled selection of those covariates (Chapple et al. 2017). Furthermore, the conditional transition rate for terminal event given that a non-terminal event was specified as a Markov model, however, a semi-Markov (Lee et al. 2015) specification is also possible. Inference may also be explored under a Bayesian paradigm. These will be considered elsewhere.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41096-024-00186-9.

Acknowledgements The authors acknowledge funding support from federal grants P30CA016059 and R21DE031879 from the United States National Institutes of Health.

Funding Funding was provided by Foundation for the National Institutes of Health.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aalen OO (1988) Heterogeneity in survival analysis. Stat Med 7:1121–1137
Aalen OO et al (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. Ann Appl Probab 2:951–972
Andersen PK, Keiding N (2002) Multi-state models for event history analysis. Stat Methods Med Res 11:91–115
Anderson D, Burnham K (2004) Model selection and multi-model inference, 2nd edn. Springer, New York
Bandeen-Roche K et al (2005) A diagnostic for association in bivariate survival models. Lifetime Data Anal 11:245–264
Chapple AG, Vannucci M, Thall PF, Lin S (2017) Bayesian variable selection for a semi-competing risks model with three hazard functions. Comput Stat Data Anal 112:170–185
Chen Y-H (2012) Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. Lifetime Data Anal 18:36–57
Clayton DG (1978) A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65:141–151
Do Ha I, Xiang L, Peng M, Jeong J-H, Lee Y (2020) Frailty modelling approaches for semi-competing risks data. Lifetime Data Anal 26:109–133
Fine JP, Jiang H, Chappell R (2001) On semi-competing risks data. Biometrika 88:907–919
Han B, Yu M, Dignam JJ, Rathouz PJ (2014) Bayesian approach for flexible modeling of semicompeting risks data. Stat Med 33:5111–5125
Haneuse S, Lee KH (2016) Semi-competing risks data analysis: accounting for death as a competing risk when the outcome of interest is nonterminal. Circ Cardiovasc Qual Outcomes 9:322–331
Hougaard P (1986) Survival models for heterogeneous populations derived from stable distributions. Biometrika 73:387–396
Hsieh J-J, Huang Y-T (2012) Regression analysis based on conditional likelihood approach under semi-competing risks data. Lifetime Data Anal 18:302–320
Ibrahim JG, Chen M-H, Sinha D (2001) Bayesian survival analysis. Springer, New York
Jiang F, Haneuse S (2017) A semi-parametric transformation frailty model for semi-competing risks survival data. Scand J Stat 44:112–129
Kiche J, Ngesa O, Orwa G (2019) Misspecification of frailty random effects in a clustered survival data. J Stat Econ Methods 8:1–25
Kim J, Kim J, Kim SW (2019) Additive-multiplicative hazards regression models for interval-censored semi-competing risks data with missing intermediate events. BMC Med Res Methodol 19:1–14
Lakhal L, Rivest L-P, Abdous B (2008) Estimating survival and association in a semicompeting risks model. Biometrics 64:180–188
Lau B, Cole SR, Gange SJ (2009) Competing risk regression models for epidemiologic data. Am J Epidemiol 170:244–256
Lee C, Gilsanz P, Haneuse S (2021) Fitting a shared frailty illness-death model to left-truncated semi-competing risks data to examine the impact of education level on incident dementia. BMC Med Res Methodol 21:1–13
Lee KH, Dominici F, Schrag D, Haneuse S (2016) Hierarchical models for semicompeting risks data with application to quality of end-of-life care for pancreatic cancer. J Am Stat Assoc 111:1075–1095
Lee KH, Haneuse S, Schrag D, Dominici F (2015) Bayesian semiparametric analysis of semicompeting risks data: investigating hospital readmission after a pancreatic cancer diagnosis. J Roy Stat Soc Ser C (Appl Stat) 64:253–273
Lee Y, Nelder JA (2004) Conditional and marginal models: another view. Stat Sci 19:219–238
Maller RA, Zhou X (1996) Survival analysis with long-term survivors. Wiley, New York
Meira-Machado L, de Uña-Álvarez J, Cadarso-Suarez C, Andersen PK (2009) Multi-state models for the analysis of time-to-event data. Stat Methods Med Res 18:195–222

Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Goodman PJ, Ungerleider JS, Emerson WA, Tormey DC, Glick JH et al (1990) Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. N Engl J Med 322:352–358

Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Tangen CM, Ungerleider JS, Emerson WA, Tormey DC, Glick JH et al (1995) Fluorouracil plus Levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: a final report. Ann Intern Med 122:321–326

Oakes D (1982) A model for association in bivariate survival data. J R Stat Soc Ser B (Methodol) 44:414–422

Oakes D (1993) A note on the Kaplan–Meier estimator. Am Stat 47:39–40

Peng L, Fine JP (2007) Regression modeling of semicompeting risks data. Biometrics 63:96–108

Peng M, Xiang L, Wang S (2018) Semiparametric regression analysis of clustered survival data with semi-competing risks. Comput Stat Data Anal 124:53–70

Selle ML (2016) Modelling of semi-competing risks using the illness-death model with shared frailty. Master's thesis, NTNU

Therneau TM (2015) A package for survival analysis in S. https://CRAN.R-project.org/package=survival. Version 2.38

Touraine C, Gerds TA, Joly P et al (2013) The smoothhazard package for r: fitting regression models to interval-censored observations of illness-death models. Copenhagen: University of Copenhagen, Department of Biostatistics (Research Report 13/12)

Tweedie M (1984) An index which distinguishes between some important exponential families. In: Statistics: applications and new directions: proceedings of the indian statistical institute golden jubilee international conference (eds. J. K. Ghosh and J. Roy), 579–604. Calcutta, India: Indian Statistical Institute

Wang W (2003) Estimating the association parameter for copula models under dependent censoring. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65:257–273

Wienke A (2010) Frailty Models in Survival Analysis. CRC Press

Xu J, Kalbfleisch JD, Tai B (2010) Statistical analysis of illness-death processes and semicompeting risks data. Biometrics 66:716–725

Zhou R, Zhu H, Bondy M, Ning J (2016) Semiparametric model for semi-competing risks data with application to breast cancer study. Lifetime Data Anal 22:456–471