Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 44: 112–129, 2017
doi: 10.1111/sjos.12244

A Semi-parametric Transformation Frailty Model for Semi-competing Risks Survival Data

FEI JIANG
Department of Statistics, University of South Carolina
SEBASTIEN HANEUSE
Department of Biostatistics, Harvard University

ABSTRACT. In the analysis of semi-competing risks data interest lies in estimation and inference with respect to a so-called non-terminal event, the observation of which is subject to a terminal event. Multi-state models are commonly used to analyse such data, with covariate effects on the transition/intensity functions typically specified via the Cox model and dependence between the non-terminal and terminal events specified, in part, by a unit-specific shared frailty term. To ensure identifiability, the frailties are typically assumed to arise from a parametric distribution, specifically a Gamma distribution with mean 1.0 and variance, say, σ^2. When the frailty distribution is misspecified, however, the resulting estimator is not guaranteed to be consistent, with the extent of asymptotic bias depending on the discrepancy between the assumed and true frailty distributions. In this paper, we propose a novel class of transformation models for semi-competing risks analysis that permit the non-parametric specification of the frailty distribution. To ensure identifiability, the class restricts to parametric specifications of the transformation and the error distribution; the latter are flexible, however, and cover a broad range of possible specifications. We also derive the semi-parametric efficient score under the complete data setting and propose a non-parametric score imputation method to handle right censoring; consistency and asymptotic normality of the resulting estimators is derived and small-sample operating characteristics evaluated via simulation. Although the proposed semi-parametric transformation model and non-parametric score imputation method are motivated by the analysis of semi-competing risks data, they are broadly applicable to any analysis of multivariate time-to-event outcomes in which a unit-specific shared frailty is used to account for correlation. Finally, the proposed model and estimation procedures are applied to a study of hospital readmission among patients diagnosed with pancreatic cancer.

Key words: frailty, misspecification, multivariate survival analysis, semi-competing risks, semi-parametric models, transformation models

1. Introduction

In biomedical studies, a patient may experience multiple failure types. One specific setting is where interest lies primarily with some so-called *non-terminal* event, the observation of which is subject to some *terminal* event. Consider, for example, the study of end-of-life quality of care for patients diagnosed with pancreatic cancer. In health policy literature, an important marker of quality of care is readmission to the emergency room following discharge from the hospital at which the diagnosis was given (Vest *et al.*, 2010; Warren *et al.*, 2011). In this setting, readmission is a non-terminal event in the sense that experiencing a readmission event does not preclude the patient from experiencing other relevant events including death. In contrast, death is terminal in the sense that experiencing a death event precludes the patient from subsequently experiencing the non-terminal readmission event, indeed any event. Data arising in this context are often referred to as *semi-competing risks data* (Fine *et al.*, 2001).

Notationally, let T_1 and T_2 be the non-terminal and terminal event times, respectively. Fine *et al.* (2001) showed that while the marginal distribution of T_2 is identifiable from

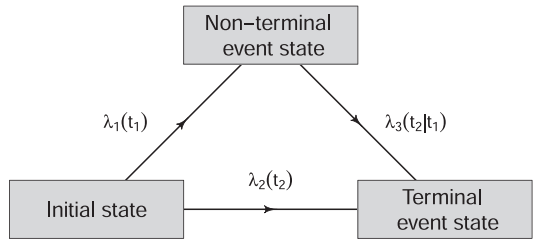

Fig. 1. Multi-state model representation of the semi-competing risks data setting.

semi-competing risks data, the marginal distribution of T_1 is not. Furthermore, the joint distribution of (T_1, T_2) is only identifiable on the upper wedge of the joint support. Towards analyzing semi-competing risks data, two general statistical frameworks have been proposed in the literature. The first imposes models directly on the marginal distributions of T_1 and T_2, with the joint distribution specified via a copula (Fine *et al.*, 2001; Wang, 2003; Jiang *et al.*, 2005; Ghosh, 2006; Peng & Fine, 2007; Lakhal *et al.*, 2008; Hsieh *et al.*, 2008). The second framework views semi-competing risks data as arising from an underlying multi-state model, specifically the illness–death model (Liu *et al.*, 2004; Ye *et al.*, 2007; Zeng & Lin, 2009; Xu *et al.*, 2010; Zeng *et al.*, 2012; Lee *et al.*,2015). Figure 1 provides a representation of this model.

In an illness–death multi-state model, the transitions between the three 'states' are characterized by three intensity or hazard functions, specifically:

$$\begin{aligned}
\lambda_1(t_1) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\Pr(T_1 = t_1+\Delta \mid T_1 \geq t_1, T_2 \geq t_1),\\
\lambda_2(t_2) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\Pr(T_2 = t_2+\Delta \mid T_1 \geq t_2, T_2 \geq t_2),\\
\lambda_3(t_2 \mid t_1) &= \lim_{\Delta\to 0}\frac{1}{\Delta}\Pr(T_2 = t_2+\Delta \mid T_1 = t_1, T_2 \geq t_2).
\end{aligned} \tag{1}$$

In practice, models must be specified for each of these functions. One specification considered by Xu *et al.* (2010) was to model the dependence of these hazard functions on covariates, denoted $\mathbf{X}$, via Cox-type multiplicative models of the form:

$$\begin{aligned}
\lambda_1(t_1 \mid \mathbf{X}, \Gamma) &= \Gamma\lambda_{01}(t_1)\exp(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}),\\
\lambda_2(t_2 \mid \mathbf{X}, \Gamma) &= \Gamma\lambda_{02}(t_2)\exp(\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}),\\
\lambda_3(t_2 \mid t_1, \mathbf{X}, \Gamma) &= \Gamma\lambda_{03}(t_2)\exp(\boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}),
\end{aligned} \tag{2}$$

where $\lambda_{01}(t_1), \lambda_{02}(t_2)$, and $\lambda_{03}(t_2)$ are baseline hazard functions and Γ is a patient-specific random effect or frailty. Intuitively, the inclusion of the Γ frailty terms in the model accounts for dependence between T_1 and T_2 that is induced marginally by covariates not included in $\mathbf{X}$. Because the baseline hazard functions are typically left unspecified, however, identifiability requires imposing a parametric assumption on the distribution of the frailties across the population (Hougaard, 1984; McGilchrist & Aisbett, 1991). If the frailties are assumed to follow a Gamma distribution, as Xu *et al.* (2010) do, the resulting (integrated) likelihood has the advantage of having a closed form expression and is therefore relatively simple to implement and maximize (Nielsen *et al.*, 1992; Kosorok *et al.*, 2004; Su & Wang, 2012). It is well-known, however, that if a parametric random effect/frailty distribution is misspecified, then estimation for other model components is no longer guaranteed to be consistent (White, 1982). This phenomenon has been extensively considered in the literature on related generalized linear mixed

models (Neuhaus *et al.*, 1992; Heagerty & Kurland, 2001; Agresti *et al.*, 2004; Litière *et al.*, 2007; Neuhaus *et al.*, 2011) as well as on shared frailty models (Shih & Louis, 1995; Glidden, 1999; Agresti *et al.*, 2004; Cui & Sun, 2004). To our knowledge, however, potential misspecification of the frailty distribution has not been considered in the context of the analysis of semi-competing risks data.

Towards relaxing distributional assumptions regarding the shared frailty terms in an illness–death model for semi-competing risks data, we consider a novel analysis approach based on transformation models. In the literature, these models are motivated by a desire to move beyond the proportional hazards assumption. Two classes of transformation models for univariate and multivariate survival outcomes have been considered: the first considers transformations of the hazard function (Zeng & Lin, 2007; Zeng *et al.*, 2009) and the second considers transformations of the survival outcomes themselves, a class which includes the well-known accelerated failure time (AFT) model (Pan, 2001). In this paper, we build on the latter, focusing on developing models that consider the impact of covariates directly on the event times (rather than on the hazard function). The proposed framework is developed specifically to permit the frailty distribution to be non-parametric. To balance identifiability, we restrict to a class of flexible parametric transformation functions and error distributions. Towards performing estimation and inference, we derive the semi-parametric efficient score function under the complete data setting (i.e. in the absence of right censoring). To accommodate right censoring, we develop a score imputation-based approach for estimation and inference. Specifically, extending prior work in the univariate setting when covariates are discrete (Wang *et al.*, 2012), the proposed method is developed for the semi-competing risks data where $\mathbf{X}$ can be a mixture of discrete and continuous covariates. For the latter, we derive the asymptotic properties of the resulting estimators and evaluate their small-sample operating characteristics via simulation. Finally, the data are applied to a motivating data set of readmission and death times for patients diagnosed with pancreatic cancer.

2. A semi-parametric transformation model

In statistical practice, the Cox model and AFT models are by far the most extensively used for investigating time-to-event outcomes. Both of these models can be shown to belong to a broader class of linear transformation models (Horowitz, 2010). To see this, let T denote a univariate time-to-event outcome and consider the model $G(T) = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{X} + \xi$, where G is a monotone transformation function and ξ is a random error term. If one sets $G(\cdot) = \log(\cdot)$ and leaves the distribution of ξ arbitrary, then this model is equivalent to the AFT model; if one sets $G(\cdot) = \log \Lambda_0(\cdot)$, with $\Lambda_0(\cdot)$ an arbitrary cumulative hazard function, and assumes that ξ follows an extreme value distribution, then the induced model for the hazard is equivalent to the Cox model.

Extending the univariate transformation model to the semi-competing risks data setting, consider the following semi-parametric linear transformation model for the i^{th} subject:

$$\begin{aligned} G_{\alpha_1}(T_{i1}) &= \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}_i + \xi_{i1}, \\ G_{\alpha_2}(T_{i2}) &= \boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}_i + \xi_{i2}, \text{ if the non-terminal event has not occurred,} \\ G_{\alpha_3}(T_{i2}) &= \boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}_i + \xi_{i3}, \text{ if the non-terminal event has occurred,} \end{aligned} \tag{3}$$

where $\mathbf{X}_i = (X_{i0}, \ldots, X_{ip})$ is a $(p+1)$-dimensional covariate vector with the first entry at 1, $G_{\alpha_l}(\cdot)$ is a transition-specific transformation function, $l = 1, 2, 3$, and ξ_{il} follows a distribution for which the cumulative hazard is defined by $\Gamma_i A_{\kappa_l}(\cdot)$, where $A_{\kappa_l}(\cdot)$ is parametric function

indexed by an unknown finite-dimensional parameter κ_l. As with the univariate transformation model, one can show that the class of models given by (3) includes models previously described in the literature. For example, the Xu *et al.* (2010) model given by (2) is a special case of (3) with $G_{\alpha_l}(\cdot)$ taken to be the log cumulative hazard function for the l^{th} transition and the $A_{\kappa_l}(\cdot) = \exp(\cdot)$ (i.e. each of the ξ_{il} error terms is taken to be distributed according to an extreme value distribution). Letting $\boldsymbol{\theta}_l = (\boldsymbol{\alpha}_l^{\mathrm{T}}, \boldsymbol{\kappa}_l^{\mathrm{T}}, \boldsymbol{\beta}_l^{\mathrm{T}})^{\mathrm{T}}$, we note that the three transition-specific hazard functions corresponding to (3) have the following relatively simple representations:

$$\begin{aligned}
\lambda_{\boldsymbol{\theta}_1}(t_1 \mid \mathbf{X}_i, \Gamma_i) &= \Gamma_i \lambda_{0\boldsymbol{\theta}_1}(t_1 \mid \mathbf{X}_i) = \Gamma_i \dot{A}_{\kappa_1}\{G_{\alpha_1}(t_1) - \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}_i\}\dot{G}_{\alpha_1}(t_1), \\
\lambda_{\boldsymbol{\theta}_2}(t_2 \mid \mathbf{X}_i, \Gamma_i) &= \Gamma_i \lambda_{0\boldsymbol{\theta}_2}(t_2 \mid \mathbf{X}_i) = \Gamma_i \dot{A}_{\kappa_2}\{G_{\alpha_2}(t_2) - \boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}_i\}\dot{G}_{\alpha_2}(t_2), \\
\lambda_{\boldsymbol{\theta}_3}(t_2 \mid t_1, \mathbf{X}_i, \Gamma_i) &= \Gamma_i \lambda_{0\boldsymbol{\theta}_3}(t_2 \mid \mathbf{X}_i) = \Gamma_i \dot{A}_{\kappa_3}\{G_{\alpha_3}(t_2) - \boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}_i\}\dot{G}_{\alpha_3}(t_2),
\end{aligned} \tag{4}$$

where $\dot{A}_{\kappa_l}(u) = dA_{\kappa_l}(u)/du$ and $\dot{G}_{\alpha_l}(t) = dG_{\alpha_l}(t)/dt$. Note that the final equality in the expression for $\lambda_{\boldsymbol{\theta}_3}(\cdot)$ does not depend on t_1. This corresponds to the so-called Markov assumption in which the hazard function for the transition from the non-terminal event to the terminal event is independent of when the non-terminal occurred. While this is a relatively standard assumption in the semi-competing risks literature (e.g. Xu *et al.*, 2010), we return to this point in the Discussion.

As we consider estimation and inference with respect to the unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$, we note that analysts will, in most settings, face two forms of missing data: (i) incomplete observance of the event times due to censoring and (ii) the unobserved latent subject-specific frailties, Γ_i. In the next two sections, we outline our proposed strategy for accommodating both types of missingness, with the specific goal of ensuring estimation and inference that is robust to possible misspecification of the (unknown) frailty distribution.

3. Robust estimation and inference in the absence of censoring

3.1. The full data likelihood

In the absence of censoring, outcome information in the semi-competing risks setting can take on one of two forms: a subject is either observed to experience both the non-terminal and terminal events or solely experiencing the terminal event. To formalize this distinction, let $\mathbf{D}_1 = \{Y_t, I(T_1 \leq T_2), T_2\}^{\mathrm{T}} \equiv \{\min(T_1, T_2), I(T_1 \leq T_2), T_2\}^{\mathrm{T}}$. Based on this notation, the observed outcome vector for a subject that experiences both events will be $\mathbf{d}_1 = \{t_1, 1, t_2\}$; the observed outcome vector for a subject that experiences the terminal event prior to the non-terminal event will be $\mathbf{d}_1 = \{t_2, 0, t_2\}$. Finally, let $\mathbf{F} = (\mathbf{D}_1^{\mathrm{T}}, \mathbf{X}^{\mathrm{T}}, \Gamma)^{\mathrm{T}}$ denote the 'full' data vector (i.e. one in which there is no censoring and the frailty is observable).

Given an i.i.d sample of size n, the full data likelihood is a product of contributions each of the form:

$$\mathcal{L}_1(\mathbf{f}_i \mid \boldsymbol{\theta}, \eta_\Gamma, \eta_{\mathbf{X}}) = \mathcal{L}_1(\mathbf{d}_{i1} \mid \mathbf{x}_i, \gamma_i; \boldsymbol{\theta})\eta_\Gamma(\gamma_i \mid \mathbf{x}_i)\eta_{\mathbf{X}}(\mathbf{x}_i), \tag{5}$$

where η_Γ and $\eta_{\mathbf{X}}$ are the densities of Γ_i and $\mathbf{X}_i$, respectively, and

$$\begin{aligned}
\mathcal{L}_1(\mathbf{d}_{i1} \mid \mathbf{x}_i, \gamma_i; \boldsymbol{\theta}) = {} & \lambda_{\boldsymbol{\theta}_1}(y_{it} \mid \mathbf{x}_i, \gamma_i)^{I(t_{i1} \leq t_{i2})} \exp\{-\Lambda_{\boldsymbol{\theta}_1}(y_{it} \mid \mathbf{x}_i, \gamma_i)\} \\
& \times \lambda_{\boldsymbol{\theta}_2}(y_{it} \mid \mathbf{x}_i, \gamma_i)^{\{1 - I(t_{i1} \leq t_{i2})\}} \exp\{-\Lambda_{\boldsymbol{\theta}_2}(y_{it} \mid \mathbf{x}_i, \gamma_i)\} \\
& \times \lambda_{\boldsymbol{\theta}_3}(t_{i2} \mid y_{i1}, \mathbf{x}_i, \gamma_i)^{I(t_{i1} \leq t_{i2})} \exp\{-\Lambda_{\boldsymbol{\theta}_3}(y_{it}, t_{i2} \mid \mathbf{x}_i, \gamma_i)\}
\end{aligned}$$

with

$$\begin{aligned}
\Lambda_{\boldsymbol{\theta}_1}(t \mid \mathbf{x}_i, \gamma_i) &= \int_0^t \lambda_{\boldsymbol{\theta}_1}(s \mid \mathbf{x}_i, \gamma_i)ds = \gamma_i \int_0^t \lambda_{0\boldsymbol{\theta}_1}(s \mid \mathbf{x}_i)ds, \\
\Lambda_{\boldsymbol{\theta}_2}(t \mid \mathbf{x}_i, \gamma_i) &= \int_0^t \lambda_{\boldsymbol{\theta}_2}(s \mid \mathbf{x}_i, \gamma_i)ds = \gamma_i \int_0^t \lambda_{0\boldsymbol{\theta}_2}(s \mid \mathbf{x}_i)ds, \\
\Lambda_{\boldsymbol{\theta}_3}(u, t \mid \mathbf{x}_i, \gamma_i) &= \int_0^t \lambda_{\boldsymbol{\theta}_3}(s \mid u, \mathbf{x}_i, \gamma_i)ds = \gamma_i \int_u^t \lambda_{0\boldsymbol{\theta}_3}(s \mid \mathbf{x}_i)ds.
\end{aligned}$$

A detailed derivation of this likelihood, based on counting processes, is provided in Section A of the Supporting Information Appendix. Note that the likelihood construction based on the counting process does not require the assumption of latent T_1 when it was truncated by T_2, which may not be the true data-generating mechanism.

3.2. The efficient score function

In practice, because the subject-specific frailties are not observable, one cannot proceed directly with likelihood-based estimation and inference based on (5). A standard way forward is to assume that the Γ_i is i.i.d according to some distribution with the Gamma distribution being a common choice because the induced marginal likelihood has a closed form expression. If this assumption is incorrect, however, the induced marginal likelihood is misspecified and the maximum likelihood estimator is no longer guaranteed to be consistent (White, 1982). Towards ensuring estimation and inference that is robust to possible misspecification of the frailty distribution, we extend an approach developed by Tsiatis & Ma (2004) for measurement error models. Specifically, when Γ is observed, consider the Hilbert space $\mathcal{H}^F$ consisting of all $d_{\boldsymbol{\theta}} \equiv \sum_{l=1}^3 d_{\alpha l} + d_{\kappa l} + p + 1$-dimensional mean-zero, finite variance functions of the full data. For elements $\boldsymbol{\phi}_1(\cdot)$ and $\boldsymbol{\phi}_2(\cdot) \in \mathcal{H}^F$, define the inner product as $< \boldsymbol{\phi}_1, \boldsymbol{\phi}_2 > = E\{\boldsymbol{\phi}_1^{\mathrm{T}}(\mathbf{F})\boldsymbol{\phi}_2(\mathbf{F})\}$. The nuisance tangent space Λ^F is the subspace of $\mathcal{H}^F$, which is the mean square closure of all the elements in the form of $\mathbf{BS}$, where $\mathbf{S}$ is an arbitrary nuisance score function and $\mathbf{B}$ a matrix with $d_{\boldsymbol{\theta}}$ rows (Ma & Zhu, 2012).

When Γ is observed, the nuisance tangent space with respect to the distributions of $\mathbf{X}$ and Γ can be written as

$$\Lambda^F = [\boldsymbol{\phi}(\mathbf{X}, \Gamma) \in \mathcal{H}^F, \text{ such that } E\{\boldsymbol{\phi}(\mathbf{X}, \Gamma)\} = \mathbf{0}].$$

Note the corresponding full data score function

$$\mathbf{U}_{\boldsymbol{\theta}}^F(\mathbf{D}_1; \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{L}_1\{\mathbf{D}_1 \mid \mathbf{X}, \Gamma; \boldsymbol{\theta}\},$$

is orthogonal to Λ^F because of the factorization in (5). Now, because the frailty Γ is not observable, the nuisance tangent space for the observed data, denoted $\mathbf{O} = (\mathbf{D}_1^{\mathrm{T}}, \mathbf{X}^{\mathrm{T}})^{\mathrm{T}}$, is

$$\Lambda = E(\Lambda^F \mid \mathbf{O}) = [E\{\boldsymbol{\phi}(\mathbf{X}, \Gamma) \mid \mathbf{O}\}, \boldsymbol{\phi}(\mathbf{X}, \Gamma) \in \Lambda^F, \text{ such that } E\{\boldsymbol{\phi}(\mathbf{X}, \Gamma)\} = \mathbf{0}], \quad (6)$$

with the corresponding score function:

$$\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) = E\{\mathbf{U}_{\boldsymbol{\theta}}^F(\mathbf{D}_1; \boldsymbol{\theta}) \mid \mathbf{O}\} = \frac{\partial}{\partial \boldsymbol{\theta}} \log \int \mathcal{L}_1\{\mathbf{D}_1 \mid \mathbf{X}, \gamma; \boldsymbol{\theta}\} \eta_\Gamma(\gamma \mid \mathbf{X}) d\gamma. \quad (7)$$

Because the expectation in expression (7) conditions on the observed data, $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta})$ is no longer orthogonal to Λ. One can, however, show that the efficient score function is an element of the space $\Lambda^{\perp}$ defined as

$$\Lambda^{\perp} \equiv [\boldsymbol{\phi}(\mathbf{O}) \text{ such that } E\{\boldsymbol{\phi}(\mathbf{O}) \mid \Gamma, \mathbf{X}\} = \mathbf{0}, a.s.].$$

Noting that there is an unique orthogonal project of $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta})$ onto Λ of the form of $E\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\}$, the efficient score is given by $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) - E\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\}$, where $\mathbf{a}_{\boldsymbol{\theta}}$ can be found by solving the integral equation

$$E[\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) - E\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\} \mid \mathbf{X}, \Gamma] = \mathbf{0}$$

under the correct specification of the unknown η_{Γ}.

Now, consider the, arguably common, situation where η is misspecified. Let η^*_{Γ} denote the misspecified density function (corresponding to the analysts choice of distribution for Γ) and $E^*(\cdot)$ the expectation operator under the incorrect model. Let $\mathcal{H}^*$ denote the Hilbert space induced by the misspecified distribution η^*_{Γ} and define Λ^{*F} and Λ^* analogously to Λ^F and Λ but replacing the expectation $E(\cdot)$ with $E^*(\cdot)$. The following theorem shows that the space orthogonal to the nuisance tangent space is invariant to the specification of η^*_{Γ}.

Theorem 1. For the Hilbert space $\mathcal{H}^*$ induced by a possibly incorrect distribution η^*_{Γ}, the space orthogonal to the nuisance tangent space $\Lambda^* \in \mathcal{H}^*$ is given by

$$\Lambda^{*\perp} = [\boldsymbol{\phi}(\mathbf{O}) \text{ such that } E\{\boldsymbol{\phi}(\mathbf{O}) \mid \Gamma, \mathbf{X}\} = \mathbf{0}, \text{a.s.}],$$

as long as η^*_{Γ} has the same support as the true η_{Γ}.

The proof of this theorem follows the same arguments as those presented in Theorem 1 in Tsiatis & Ma (2004).

Theorem 1 suggests that there exists a function $\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \in \Lambda^F$ such that $E^*\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\}$ is in Λ^*, and $\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) - E^*\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\}$ is orthogonal to Λ^*. That is, there exists a function $\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \in \Lambda^F$ such that

$$E[\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) - E^*\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\} \mid \mathbf{X}, \Gamma] = \mathbf{0},$$

where $\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta})$ is defined in the same way as $\mathbf{U}_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta})$ in (7) but replacing η_{Γ} by η^*_{Γ}. Note, $E^*\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\}$ is uniquely defined for any given η^*_{Γ} because it is the orthogonal projection of $\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta})$ onto Λ^*. From this construction and Theorem 1,

$$\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) = \mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) - E^*\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\} \tag{8}$$

has mean $\mathbf{0}$ even if η^*_{Γ} is misspecified. Furthermore, the solution to $\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}(\mathbf{O}_i; \boldsymbol{\theta}) = \mathbf{0}$ is the semi-parametric efficient estimator of $\boldsymbol{\theta}$ under the true specification for η_{Γ} and when there is no censoring (Tsiatis & Ma, 2004).

3.3. Procedures for obtaining $\boldsymbol{a}_{\boldsymbol{\theta}}$

As is clear from expression (8), in order to evaluate $\mathbf{S}_{\text{eff}}$, one must be able to compute the function $\mathbf{a}_{\boldsymbol{\theta}}$. Here, we describe an algorithm for doing so when the possibly misspecified η^*_{Γ} is taken to be a Gamma distribution, specifically one for which the density is $\eta^*_{\Gamma}(\gamma) = \sigma^{-2/\sigma^2} \gamma^{1/\sigma^2 - 1} \exp(-\gamma/\sigma^2) / \Gamma(1/\sigma^2)$, where $\sigma^2 > 0$ is considered to be known. We restrict attention to this distribution, in part, because it is by far not only the most common assumption made in practice but also because, as mentioned earlier, the induced marginal likelihood has a closed form expression.

Now, it is well-known that any function can be approximated by a step function arbitrarily well as the number of component intervals increases. With this in mind, we assume that for each $\mathbf{X}$, $\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma)$ can be approximated by

$$\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma = \gamma) = \sum_{j=1}^{m} \mathbf{r}_j I(\gamma_j \leq \gamma < \gamma_{j+1}), \tag{9}$$

where $(\gamma_1, \ldots, \gamma_{m+1})$ is a partition of the support of Γ and $\mathbf{r}_j$ is the interval-specific value of an unspecified $d_{\boldsymbol{\theta}}$-dimensional function. In practice, we can find the intervals by first fitting a working model for Γ and then use the quantiles from the estimated frailty distribution as the partition endpoints. Under this specification,

$$E[E^*\{\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma) \mid \mathbf{O}\} \mid \mathbf{X}, \Gamma = \gamma_i] = \sum_{j=1}^{m} \int \mathbf{r}_j \{F_g(\gamma_{j+1} \mid \mathbf{O}) - F_g(\gamma_j \mid \mathbf{O})\} \mathcal{L}_1(\mathbf{d}_1 \mid \mathbf{X}, \Gamma = \gamma_i; \boldsymbol{\theta}) d\mathbf{d}_1,$$

where F_g the posterior distribution of Γ given $\mathbf{O}$ under the possibly misspecified η^*_Γ. The latter can be shown to correspond to a Gamma $[\Delta_1 + 1/\sigma^2 + 1, 1/\sigma^2 + A_{\kappa_1}\{G_{\alpha_1}(Y_t) - \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}\} + A_{\kappa_2}\{G_{\alpha_2}(Y_t) - \boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X}\} + A_{\kappa_3}\{G_{\alpha_3}(T_2) - \boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}\} - A_{\kappa_3}\{G_{\alpha_3}(Y_t) - \boldsymbol{\beta}_3^{\mathrm{T}}\mathbf{X}\}]$ distribution. Define the matrix $\mathbf{R}$ to be the $m \times m$ matrix with $[i, j]^{\text{th}}$ entry given by

$$R_{ij}(\mathbf{X}) = \int \{F_g(\gamma_{j+1} \mid \mathbf{O}) - F_g(\gamma_j \mid \mathbf{O})\} \mathcal{L}_1(\mathbf{d}_1 \mid \mathbf{X}, \Gamma = \gamma_i; \boldsymbol{\theta}) d\mathbf{d}_1.$$

Furthermore, let $\mathbf{b}(\mathbf{X}) = [E^*\{\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) \mid \Gamma = \gamma_1, \mathbf{X}\}^{\mathrm{T}}, \ldots, E^*\{\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) \mid \Gamma = \gamma_m, \mathbf{X}\}^{\mathrm{T}}]^{\mathrm{T}}$, where

$$E^*\{\mathbf{U}^*_{\boldsymbol{\theta}}(\mathbf{O}; \boldsymbol{\theta}) \mid \Gamma = \gamma_i, \mathbf{X}\} = \int \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ \log \int \mathcal{L}_1(\mathbf{d}_1 \mid \mathbf{X}, \gamma; \boldsymbol{\theta}) \eta^*_\Gamma(\gamma \mid \mathbf{X}) d\gamma \right\} \mathcal{L}_1(\mathbf{d}_1 \mid \mathbf{X}, \Gamma = \gamma_i; \boldsymbol{\theta}) d\mathbf{d}_1.$$

One can then obtain the values of $\mathbf{r} = (r_1, \ldots, r_m)^{\mathrm{T}}$ by solving:

$$\mathbf{R}(\mathbf{X})\mathbf{r}^{\mathrm{T}}(\mathbf{X}) = \mathbf{b}^{\mathrm{T}}(\mathbf{X}).$$

Finally, after obtaining $\mathbf{r}_j$, the efficient score function $\mathbf{S}_{\text{eff}}$ can be approximated by

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log \int \mathcal{L}_1(\mathbf{d}_1 \mid \mathbf{X}, \gamma; \boldsymbol{\theta}) \eta^*_\Gamma(\gamma \mid \mathbf{X}) d\gamma \; - \; \sum_{j=1}^{m} \mathbf{r}_j \{F_g(\gamma_{j+1} \mid \mathbf{O}) - F_g(\gamma_j \mid \mathbf{O})\}.$$

4. Robust estimation and inference in the presence of censoring

4.1. The censored data likelihood

In the presence of (right) censoring, $\mathbf{D}_1$ will not be observed for some subjects. To accommodate this, let C denote the censoring time and $\mathbf{D}_2 = \{Y_c, I(T_1 \leq C), C\}^{\mathrm{T}} \equiv \{\min(T_1, C), I(T_1 \leq C), C\}^{\mathrm{T}}$. Based on this notation, the outcome vector for a subject that is censored prior to either of the events being observed would be $\mathbf{d}_2 = \{c, 0, c\}$; the outcome vector for a subject that is observed to experience a non-terminal event but is subsequently censored would have $\mathbf{d}_2 = \{t_1, 1, c\}$. For either of these subjects, the likelihood contribution is of the form:

$$\mathcal{L}_2(\mathbf{d}_{i2}, \mathbf{x}_i, \gamma_i \mid \boldsymbol{\theta}, \eta_\Gamma, \eta_C, \eta_{\mathbf{X}}) = \mathcal{L}_2(\mathbf{d}_{i2} \mid \mathbf{x}_i, \gamma_i; \boldsymbol{\theta}) \eta_\Gamma(\gamma_i \mid \mathbf{x}_i) \eta_C(c_i \mid \mathbf{x}_i) \eta_{\mathbf{X}}(\mathbf{x}_i), \quad (10)$$

where $\eta_C(c_i \mid \mathbf{x}_i)$ is the density for C and

$$\begin{aligned}\mathcal{L}_2(\mathbf{d}_{i2} \mid \mathbf{x}_i, \gamma_i; \boldsymbol{\theta}) &= \lambda_{\boldsymbol{\theta}_1}(y_{ic} \mid \mathbf{x}_i, \gamma_i)^{I(t_{i1} \le c_i)} \exp\{-\Lambda_{\boldsymbol{\theta}_1}(y_{ic} \mid \mathbf{x}_i, \gamma_i)\} \\ &\times \exp\{-\Lambda_{\boldsymbol{\theta}_2}(y_{ic} \mid \mathbf{x}_i, \gamma_i) \\ &\times \exp\{-\Lambda_{\boldsymbol{\theta}_3}(y_{ic}, c_i \mid \mathbf{x}_i, \gamma_i)\}.\end{aligned}$$

4.2. Estimation via non-parametric score imputation

Building on the derivation of the efficient score function for the non-censored observations in Section 3.2, we derive the score function for censored observations based on a score imputation method. Towards this, let $Y_{i2} = \min(T_{i2}, C_i)$, $Y_{i1} = \min(T_{i1}, Y_{i2})$, $\Delta_{i1} = I(T_{i1} \le Y_{i2})$ and $\Delta_{i2} = I(T_{i2} \le C_i)$. Following Wang *et al.* (2012), the overarching strategy is to replace $\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta})$ by its conditional expectation giving the observed data. Specifically, for an individual that experienced a non-terminal event but was censored prior to observation of a terminal event (i.e. Δ_1=1 and Δ_2=0), the unobserved efficient score function is replaced by

$$\mathbf{Q}_1(C, \mathbf{X}; \boldsymbol{\theta}) = E\{\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) \mid T_2 > C > T_1, C, \mathbf{X}\}.$$

For an individual that is censored prior to observation of either event (i.e. Δ_1=0 and Δ_2=0), the unobserved efficient score function is replaced by

$$\mathbf{Q}_2(C, \mathbf{X}; \boldsymbol{\theta}) = E\{\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) \mid \min(T_1, T_2) > C, C, \mathbf{X}\}.$$

Based on these expressions, the resulting censored data score function across all four observed data scenarios is

$$\begin{aligned}\boldsymbol{\psi}(\boldsymbol{\theta}) &= \Delta_2 \mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) + \Delta_1(1-\Delta_2) E\{\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) \mid T_2 > C > T_1, C, \mathbf{X}\} \\ &\quad + (1-\Delta_1)(1-\Delta_2) E\{\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) \mid \min(T_2, T_1) > C, C, \mathbf{X}\}.\end{aligned} \quad (11)$$

Note that because $\boldsymbol{\psi}(\boldsymbol{\theta})$ has mean $\mathbf{0}$ under the true model, estimation based on (11) is consistent. An important practical challenge, however, is that evaluation of the conditional expectation components of $\boldsymbol{\psi}(\boldsymbol{\theta})$ requires specification of the frailty distribution. To resolve this, we propose to use a kernel estimator for $E\{\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) \mid T_2 > C > T_1, C, \mathbf{X}\}$ given by

$$\frac{\sum_{j=1}^n \Delta_{j2} I(T_{j2} > C > T_{j1}) \mathbf{S}_{\text{eff}}(\mathbf{O}_j; \boldsymbol{\theta}) \widehat{\widehat{F}}_C(Y_{j2} \mid \mathbf{X}_j)^{-1} K_{\mathbf{h}}(\mathbf{X} - \mathbf{X}_j)}{\sum_{j=1}^n \Delta_{j2} I(T_{j2} > C > T_{j1}) \widehat{\widehat{F}}_C(Y_{j2} \mid \mathbf{X}_j)^{-1} K_{\mathbf{h}}(\mathbf{X} - \mathbf{X}_j)}, \quad (12)$$

where $\widehat{\widehat{F}}_C$ is a root-n consistent estimator for $1 - F_C$, the survival function for the censoring time, and $K_{\mathbf{h}}(\mathbf{X} - \mathbf{X}_j) = \prod_{k=1}^p K_{h_k}(X_k - X_{jk})$ is a multivariate kernel with a p dimensional bandwidth $\mathbf{h} = (h_k, k = 1, \ldots, p)^{\mathrm{T}}$. Note that we choose the product kernel for $\mathbf{X}$ because it permits the analyst to monitor the bandwidth for each covariate individually which, in turn, facilitates the theoretical derivations and numerical evaluations. For the discrete covariates, we can simply take $K_{h_k}(X_k - X_{jk}) = I(X_k = X_{jk})$, although in practice any sensible multivariate kernel could be used. Similarly, we estimate $E\{\mathbf{S}_{\text{eff}}(\mathbf{O}; \boldsymbol{\theta}) \mid \min(T_1, T_2) > C, C, \mathbf{X}\}$ by

$$\frac{\sum_{j=1}^{n}\Delta_{j2}I\{\min(T_{j1},T_{j2})>C\}\mathbf{S}_{\text{eff}}(\mathbf{O}_j;\boldsymbol{\theta})\widehat{\overline{F}}_C(Y_{j2}\mid\mathbf{X}_j)^{-1}K_{\mathbf{h}}(\mathbf{X}-\mathbf{X}_j)}{\sum_{j=1}^{n}\Delta_{j2}I\{\min(T_{j1},T_{j2})>C\}\widehat{\overline{F}}_C(Y_{j2}\mid\mathbf{X}_j)^{-1}K_{\mathbf{h}}(\mathbf{X}-\mathbf{X}_j)}. \tag{13}$$

Finally, combining expressions (11), (12) and (13), the proposed estimator is obtained as the solution to $\widehat{\boldsymbol{\psi}}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\widehat{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) = \mathbf{0}$, where

$$\begin{aligned}\widehat{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) \;=\; & \Delta_{i2}\mathbf{S}_{\text{eff}}(\mathbf{O}_i;\boldsymbol{\theta}) \;+\; \Delta_{i1}(1-\Delta_{i2})\widehat{E}\{\mathbf{S}_{\text{eff}}(\mathbf{O}_i;\boldsymbol{\theta})\mid T_{i2}>C_i>T_{i1},C_i,\mathbf{X}_i\}\\ & +\;(1-\Delta_{i1})(1-\Delta_{i2})\widehat{E}\{\mathbf{S}_{\text{eff}}(\mathbf{O}_i;\boldsymbol{\theta})\mid\min(T_{i2},T_{i1})>C_i,C_i,\mathbf{X}_i\}.\end{aligned}$$

5. Asymptotic results

To establish asymptotic properties of the estimator proposed on Section 4, we first state the necessary conditions:

- (A1) $E\{\mathbf{S}_{\text{eff}}(\mathbf{O};\boldsymbol{\theta})\} = \mathbf{0}$ has unique solution at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\psi}(\boldsymbol{\theta})$ has a continuous and bounded derivative.
- (A2) In the kernel function $K_{h_k}(x) = h_k K(h_k^{-1}x)$, h_k are bandwidths for covariate $X_k, k = 1,\ldots,p$ and K is an ℓ^{th}-order symmetric kernel function that satisfies $\ell \geq 1$, $\int K(x)dx = 1$, $\int K(x)^2dx < \infty$ and $\int x^2K(x)dx < \infty$. The bandwidths satisfy $nh_k^2 \to \infty$ for each $k = 1,\ldots,p$ and $n\sum_{k=1}^{p}h_k^4 \to 0$.
- (A3) $\mathbf{X}$ has finite second moment.
- (A4) The joint density function for $\mathbf{X}$ is bounded away from zero and infinity on its support.
- (A5) The censoring time is conditionally independent of (T_1, T_2) given $\mathbf{X}$, the effects of which do not depend on time.

For the most part, these conditions are standard. Specifically, condition (A1) is the usual condition for identifiability of model parameters while condition (A2) guarantees convergence of the kernel estimators. Conditions (A3) and (A4) ensure that the asymptotic variance of the estimator is bounded. The first part of condition (A5) is analogous to the usual non-informative censoring assumption in univariate time-to-event analyses. The second part of condition (A5) is an assumption that permits the use of a Cox proportional hazards model for the censoring time:

$$\overline{F}_C(t\mid\mathbf{X}) = \exp\{-\Lambda_{c0}(t)\exp(\boldsymbol{\beta}_{c0}^{\mathrm{T}}\mathbf{X})\}.$$

Under this model, the partial likelihood estimator $\widehat{\boldsymbol{\beta}}_c$, the Breslow baseline hazard estimator $\widehat{\Lambda}_c(t,\widehat{\boldsymbol{\beta}})$ and the corresponding censoring distribution function estimator

$$\widehat{\overline{F}}_C(t\mid\mathbf{X}) = \exp\left\{-\widehat{\Lambda}_c(t,\widehat{\boldsymbol{\beta}})\exp\left(\widehat{\boldsymbol{\beta}}_c^{\mathrm{T}}\mathbf{X}\right)\right\}$$

are all root n consistent to $\boldsymbol{\beta}_{c0}$, $\Lambda_{c0}(t)$ and $\overline{F}_C(t\mid\mathbf{X})$ (Lemma 1). The form of the censoring distribution covers the case when C is assumed to be independent of both $\mathbf{X}$ and $\mathbf{T}$, which is used in Ma & Yin (2010) for an inverse probability weighting-based method. This independent censoring assumption would not affect the theoretic properties of the survival function estimators except that $\boldsymbol{\beta}_{c0} = \mathbf{0}$ is assumed to be known in advanced. It is worthwhile to note that any consistent method can be used here to estimate the censoring distribution.

With the aforementioned conditions/assumptions, the following theorem describes the asymptotic properties of the proposed estimator:

Theorem 2. Let $\mathbf{W}_i = (\mathbf{O}_i^{\mathrm{T}}, C_i)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\theta}}$ the solution to

$$\widehat{\boldsymbol{\psi}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \widehat{\boldsymbol{\psi}}_i(\boldsymbol{\theta}) = \mathbf{0}.$$

Under conditions (A1)–(A5), $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$, where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Furthermore,

$$\begin{aligned} n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = & -\left[E\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}\right]^{-1} \left\{n^{-1/2}\sum_{i=1}^{n}\mathbf{J}_1(\mathbf{W}_i;\boldsymbol{\theta}_0)\right. \\ & + n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\mathbf{J}_2(s,\mathbf{W}_i;\boldsymbol{\theta}_0)dM_{ic}(s) \\ & \left. + n^{1/2}\int_0^{\tau}\mathbf{J}_3(s;\boldsymbol{\theta}_0)ds(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_{c0})\right\} + o_p(1) \end{aligned}$$

has an asymptotic mean-zero normal distribution with variance–covariance matrix, Σ, given by

$$\begin{aligned} & \left[E\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}\right]^{-1}\left[E\{\mathbf{J}_1(\mathbf{W}_i;\boldsymbol{\theta}_0)^{\otimes 2}\} + E\left\{\int_0^{\tau}\mathbf{J}_2(s,\mathbf{W}_i;\boldsymbol{\theta}_0)^{\otimes 2}R_i(s)\lambda_{0c}(s)\exp(\boldsymbol{\beta}_{c0}^{\mathrm{T}}\mathbf{X}_i)ds\right\}\right. \\ & \left. + \left\{\int_0^{\tau}\mathbf{J}_3(s;\boldsymbol{\theta}_0)ds\right\}\Sigma_c^{-1}\left\{\int_0^{\tau}\mathbf{J}_3(s;\boldsymbol{\theta}_0)ds\right\}^{\mathrm{T}}\right]\left(\left[E\left\{\frac{\partial \boldsymbol{\psi}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right\}\right]^{-1}\right)^{\mathrm{T}}, \end{aligned}$$

where $\mathbf{J}_1, \mathbf{J}_2, \mathbf{J}_3$ are defined in Section D of the Supporting Information Appendix.

Inspection of the influence function given in Theorem 2 highlights that the uncertainty comes from three sources: uncertainty associated with the kernel estimator, uncertainty in estimate of the censoring distribution and uncertainty in the estimate $\widehat{\boldsymbol{\beta}}_c$. A proof of this result, along with those of two necessary lemmas, are presented in Appendices C and D. Because calculation of the empirical version of Σ is rather involved, in practice, one could use the bootstrap to obtain standard error estimates and/or confidence intervals (Wang *et al.*, 2012).

6. Simulation studies

6.1. Simulation set-up

To evaluate the small-sample operating characteristics of the proposed methods, we conducted a simulation study. As emphasized in the Introduction, the primary purpose of the patient-specific frailty terms is to account for dependence that is not taken into account by measured covariates included in the model. Our simulation studies are therefore structured to mimic this scenario. Specifically, we considered a data-generating mechanism based on model (3) setting $G_l(\cdot) = \log(\cdot)$, $l = 1, 2, 3$ and $A_l(\cdot) = \exp(\cdot/\kappa_l)$ so that the error terms follow a scaled extreme value distribution with $\kappa_l = 0.5$. For covariates, we considered the two-dimensional vector $\mathbf{X}_i = (1, X_{1i})^{\mathrm{T}}$ and an additional latent variable Z_i associated indirectly with the survival time through X_{1i} and the frailty. More specifically, Z_i is drawn from a standard normal distribution. Given Z_i, a binary X_{1i} was generated as a random Bernoulli draw with probability $\Pr(X_{1i} = 1 \mid Z_i) = \Phi(Z_i)$, with $\Phi(\cdot)$ the cumulative distribution function for the standard normal distribution. Furthermore, a patients frailty was generated from a mixture distribution depending on the value of another variable I_i with probability $\Pr(I_i = 1 \mid Z_i) = \Phi(Z_i)$; if $I_i = 1$, Γ_i was drawn from a Gamma distribution with mean 1 and variance 0.5, denoted here by Gamma(1, 0.5); if $I_i = 0$, Γ_i was drawn from 2 + Gamma(1, 6), a location shifted Gamma

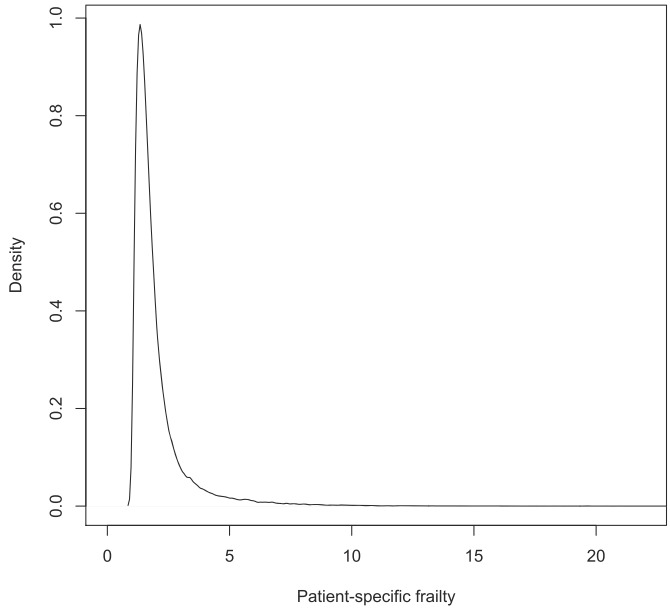

Fig. 2. Graphical representation of the mixture Gamma frailty distribution used in the simulations of Section 6.

distribution. By mixing the two Gamma distributions, the induced marginal distribution has a longer tail than either one of the component distributions; Fig. 2 provides a graphical representation. Note that, although I_i and X_{1i} have the same distribution, they are independent, given the Z_i, which suggests that Γ_i does not account for the dependence explained by $\mathbf{X}_i$. After obtaining multiple samples of $\mathbf{X}_i, \Gamma_i$, survival times based on model (3) were generated using methods described in Jiang & Haneuse (2014) with the 'true' values of $\boldsymbol{\theta}$, denoted $\boldsymbol{\theta}_0$, given in the first column of Table 1. Finally, independent right censoring was introduced by generating a censoring time based on an exponential distribution with the rate parameter chosen to achieve a certain censoring rate.

Based on the previous structure, we generated four sets of 1,000 simulated semi-competing risks data sets, which varied in sample size (n = 100 and 500) and overall censoring rates (0 per cent and 20 per cent). For each of the data sets, we applied the proposed semi-parametric approach described in Section 4. Throughout, to focus the simulations on the misspecification of the distribution of the frailties, we assumed the true transformation and cumulative hazard functions. To approximate the efficient score function, we choose m = 25 equally spaced grid points on the support of the frailty distribution (Section 4.2). Based on preliminary analyses not shown, we found this choice to achieve a reasonable balance between computational tractability and accuracy. Finally, for comparison, we also fit model (3) assuming the frailties arise from a common parametric distribution, specifically a Gamma distribution with mean 1 and variance σ^2.

Finally, we note that the methods described in Sections 2–5 have been implemented in `R` (R Core Team, 2015). The code, together with the scripts for the simulation studies, are freely available via GitHub at `https://github.com/homebovine/semifrailty.git`.

6.2. Results

Results from the simulation study are provided in Table 1. Comparing the results between the proposed semi-parametric transformation (upper half) and the parametric analysis (lower

Table 1. *Comparison of operating characteristics, based on 1,000 simulated samples, between the proposed semi-parametric transformation model for the analysis of semi-competing risks data to a parametric analysis in which the frailty random effects are assumed to arise from a Gamma distribution. Values are as follows:* $\boldsymbol{\theta}_0$ *is the true value of the parameter;* Bias($\widehat{\boldsymbol{\theta}}$) *is the absolute bias for* $\boldsymbol{\theta}$*;* sd($\widehat{\boldsymbol{\theta}}$) *is empirical standard derivation;* MSE *is the empirical mean square error of the estimators*

		Proposed semi-parametric transformation model											
		Censoring rate = 0%						Censoring rate = 20%					
		n = 100			n = 500			n = 100			n = 500		
$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	Bias($\widehat{\boldsymbol{\theta}}$)	sd($\widehat{\boldsymbol{\theta}}$)	MSE	Bias($\widehat{\boldsymbol{\theta}}$)	sd($\widehat{\boldsymbol{\theta}}$)	MSE	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	MSE	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	MSE
κ_1	0.5	0.02	0.076	0.006	0.01	0.048	0.002	0.02	0.076	0.006	0.00	0.050	0.003
κ_2	0.5	0.01	0.082	0.007	0.00	0.052	0.003	0.02	0.087	0.008	0.00	0.053	0.003
κ_3	0.5	0.05	0.123	0.017	0.02	0.077	0.007	0.06	0.138	0.023	0.02	0.084	0.007
β_{10}	−0.5	0.00	0.138	0.019	0.02	0.093	0.009	0.02	0.132	0.018	0.02	0.108	0.012
β_{11}	−1.1	0.01	0.159	0.025	0.02	0.098	0.010	0.02	0.155	0.025	0.04	0.109	0.014
β_{20}	−0.2	0.03	0.159	0.026	0.01	0.106	0.011	0.03	0.146	0.022	0.02	0.120	0.015
β_{21}	−1.1	0.01	0.166	0.028	0.02	0.107	0.012	0.03	0.172	0.030	0.01	0.119	0.014
β_{30}	−0.2	0.05	0.179	0.034	0.04	0.098	0.011	0.06	0.169	0.033	0.04	0.118	0.016
β_{31}	−1.1	0.02	0.172	0.030	0.03	0.104	0.012	0.02	0.172	0.030	0.04	0.120	0.016
		Fully parametric model											
		Censoring rate = 0%						Censoring rate = 20%					
		n = 100			n = 500			n = 100			n = 500		
$\boldsymbol{\theta}$	$\boldsymbol{\theta}_0$	Bias($\widehat{\boldsymbol{\theta}}$)	sd($\widehat{\boldsymbol{\theta}}$)	MSE	Bias($\widehat{\boldsymbol{\theta}}$)	sd($\widehat{\boldsymbol{\theta}}$)	MSE	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	MSE	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	MSE
κ_1	0.5	0.01	0.067	0.005	0.00	0.029	0.001	0.01	0.069	0.005	0.00	0.029	0.001
κ_2	0.5	0.01	0.072	0.005	0.00	0.032	0.001	0.01	0.078	0.006	0.00	0.033	0.001
κ_3	0.5	0.02	0.089	0.009	0.00	0.044	0.002	0.02	0.092	0.009	0.00	0.047	0.002
β_{10}	−0.5	0.42	0.134	0.194	0.43	0.057	0.184	0.42	0.149	0.200	0.42	0.062	0.183
β_{11}	−1.1	0.21	0.167	0.072	0.21	0.071	0.048	0.21	0.174	0.074	0.21	0.073	0.050
β_{20}	−0.2	0.42	0.177	0.210	0.42	0.077	0.182	0.41	0.195	0.206	0.42	0.083	0.183
β_{21}	−1.1	0.19	0.189	0.073	0.21	0.091	0.052	0.20	0.197	0.077	0.21	0.092	0.053
β_{30}	−0.2	0.42	0.142	0.196	0.42	0.063	0.180	0.42	0.150	0.200	0.42	0.066	0.181
β_{31}	−1.1	0.21	0.174	0.075	0.20	0.075	0.045	0.21	0.175	0.074	0.20	0.084	0.048

half), we see that both analyses estimate the κ_l parameters (that index the distribution of the ϵ_{il} terms) with low bias. In contrast, estimates of the regression parameters exhibit substantial bias under the parametric analysis and virtually none under the proposed approach; because both analyses assume the correct functional form for the transformations and the error terms, this difference is solely because of the differences in treatment of the frailty distribution. Comparing the empirical estimates of the standard deviations of the sampling distributions, we see that the estimates based on the proposed approach generally exhibit greater uncertainty. However, the empirical mean squared error indicates that the overall performance of the proposed estimator is substantially better. Finally, across the board, the general conclusions we draw regarding the differences between the proposed approach and a fully parametric analysis do not change as the sample sizes changes nor as the censoring rate changes.

7. Analysis of Medicare data

The methods proposed in this manuscript are applied to a motivating data set consisting of patients diagnosed with pancreatic cancer. Specifically, we consider $n = 621$ Medicare enrollees aged 87 years and older who received a diagnosis for pancreatic cancer during a hospitalization between 01/2005 and 11/2008 and were subsequently discharged (i.e. they did not die during the index hospitalization). For these analyses, the non-terminal event of interest is whether or not the patient was readmitted to the emergency room after having being discharged from the index hospitalization; the terminal event is death. Furthermore, observation time was administratively censored at 90 days because interest typically lies with readmission events soon after a patient is discharged (Lee *et al.*, 2015). Among the $n = 621$ patients in the study sample, 56 patients were observed to experience both events, 34 patients were observed to experience a

Table 2. *Results from the analysis of the Medicare data based on the proposed framework with $\boldsymbol{G}_{\alpha_l}(\boldsymbol{t})$ fixed to be the log-transformation, $l = 1,2,3$. The estimators are the average over 1,000 bootstrap replicates. Standard errors were estimated using the bootstrap, based on 1,000 replicates. 95% confidence intervals are based on the quantiles of the bootstrap estimate of the sampling distribution*

	Proposed semi-parametric model			Fully parametric model		
Parameter, $\boldsymbol{\theta}$[a]	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	95% CI	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	95% CI
Readmission						
κ_1	1.14	0.511	(0.60, 2.92)	0.85	0.086	(0.72, 1.02)
β_{10}	1.41	0.177	(1.08, 2.11)	1.19	0.222	(0.87, 1.63)
β_{11}	−0.35	0.296	(−0.91, 0.38)	−0.45	0.236	(−0.81, 0.09)
Death prior to readmission						
κ_2	0.80	0.146	(0.56, 1.37)	0.84	0.049	(0.76, 0.93)
β_{20}	−0.31	0.232	(−0.68, 0.09)	−0.14	0.121	(−0.33, 0.10)
β_{21}	−0.28	0.186	(−0.71, 0.21)	−0.30	0.168	(−0.59, 0.01)
Death after readmission						
κ_3	0.61	0.176	(0.39, 1.52)	0.91	0.154	(0.66, 1.32)
β_{30}	0.59	0.196	(−0.19, 0.91)	0.69	0.216	(0.31, 1.07)
β_{31}	−1.09	0.421	(−2.36, −0.42)	−0.95	0.465	(−1.69, −0.08)
Frailty variance						
σ^2	–	–	–	1.23	0.223	(0.89, 1.71)

[a] κ_l indexes the scaled exponential distribution for the ξ_{il} error terms; (β_{l0}, β_{l1}) are the regression parameters, l=1,2,3; σ^2 is variance component indexing the Gamma distribution in the fully parametric analysis.

readmission event but were then censored prior to death, 401 patients were observed to experience a death event prior to readmission and 130 patients were censored prior to experiencing either event.

7.1. Analyses

To illustrate the approach, we follow the simulations and consider a single covariate, specifically gender, where $X_{1i} = 0/1$ if the i^{th} patient is male/female. Four sets of analyses are considered. Specifically, we consider both the proposed semi-parametric approach as well as a fully parametric model (i.e. the patient-specific frailties are assumed to arise from a Gamma distribution). For both of these approaches, we consider the transformation functions to be (i) fixed as the log-transformation and (ii) taken from the Box–Cox family of transformations, specifically with $G_{\alpha_l}(t) = (t^{\alpha_l} - 1)/\alpha_l$, $l = 1, 2, 3$. Throughout, as in the simulations, we set $A_{\kappa_l}(\xi) = \exp(\xi/\kappa_l)$ and approximate $\mathbf{a}_{\boldsymbol{\theta}}(\mathbf{X}, \Gamma)$ based expression (9) with $m = 25$ equally spaced grid points. Finally, as is clear from the expression in Theorem 2, the functional form of asymptotic variance–covariance matrix is complex requiring the empirical evaluation of numerous integrals. We, therefore, report bootstrap-based standard error estimates and 95% confidence intervals, based on 1,000 replicates.

7.2. Results

Results from the four sets of analyses are shown in Tables 2 and 3 and in Fig. 3, and Figs B.1 and B.2 in the Supporting Information Appendix. Focusing on the regression parameter for gender (i.e. β_{l1}), all analyses indicate that females diagnosed with pancreatic cancer have lower

Table 3. *Results from the analysis of the Medicare data based on the proposed framework with* $G_{\alpha_l}(t) = (t^{\alpha_l} - 1)/\alpha_l$, *the Box–Cox transformation,* l=1,2,3. *The estimators are the average over 1,000 bootstrap replicates. Standard errors were estimated using the bootstrap, based on 1,000 replicates. 95% confidence intervals are based on the quantiles of the bootstrap estimate of the sampling distribution*

	Proposed semi-parametric model			Fully parametric model		
Parameter, $\boldsymbol{\theta}^a$	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	95% CI	$\widehat{\boldsymbol{\theta}}$	sd($\widehat{\boldsymbol{\theta}}$)	95% CI
Readmission						
κ_1	1.93	0.631	(1.09, 3.31)	1.32	0.178	(1.01, 1.77)
β_{10}	2.79	0.137	(2.51, 3.18)	2.58	0.383	(1.79, 3.43)
β_{11}	−0.38	0.173	(−0.93, −0.01)	−0.55	0.352	(−1.09, 0.26)
α_1	−0.35	0.126	(−0.64, −0.15)	−0.16	0.071	(−0.29, −0.02)
Death prior to readmission						
κ_2	1.40	0.149	(1.15, 1.76)	1.40	0.118	(1.12, 1.59)
β_{20}	0.11	0.204	(−0.45, 0.49)	0.47	0.121	(0.14, 0.69)
β_{21}	−0.30	0.177	(−0.70, 0.15)	−0.34	0.190	(−0.64, 0.06)
α_2	−0.37	0.044	(−0.46, −0.27)	−0.28	0.044	(−0.34, -0.16)
Death after readmission						
κ_3	1.75	0.094	(1.34, 1.91)	1.73	0.317	(1.18, 3.36)
β_{30}	0.28	0.167	(−0.23, 0.55)	0.26	0.399	(−0.51, 1.02)
β_{31}	−1.23	0.156	(−1.74, −0.93)	−1.08	0.728	(−2.90, 0.30)
α_3	0.20	0.089	(0.08, 0.52)	0.33	0.264	(-0.40, 0.74)
Frailty variance component						
σ^2	–	–	–	0.04	0.057	(2.87e-06, 0.72)

[a] κ_l indexes the scaled exponential distribution for the ξ_{il} error terms; (β_{l0}, β_{l1}) are the regression parameters, l=1,2,3; α_l is the Box–Cox power parameter, l=1,2,3; σ^2 is variance component indexing the Gamma distribution in the fully parametric analysis.

risk for readmission and death, both prior to and post readmission, than males. In both tables, the point estimates for the impact of gender on risk of readmission are different between the proposed semi-parametric and parametric analyses; in Table 2, the estimates are $\widehat{\beta}_{11}=-0.35$ and $\widehat{\beta}_{11}=-0.45$ under the two analyses, respectively; in Table 3 the difference is even greater. To aid in the interpretation of these differences, Fig. 3 provides estimates of the conditional survival function (i.e. setting the patient-specific frailty to $\Gamma = 1$) for readmission stratified by gender, from the four analyses; estimates for the corresponding survival functions for death are provided in Section B in the Supporting Information Appendix. Comparing panel (a) to panel (b) and panel (c) to panel (d), we again see a clear indication of differences in the results between the proposed semi-parametric analysis and one that is fully parametric. Collectively, these results indicate that there is some sensitivity to the specification of the frailty distribution, with the proposed semi-parametric analysis likely providing more reliable results.

Comparing the results for the semi-parametric analysis between Tables 2 and 3, we see that while the point estimates for the effect of gender are similar, the standard error estimates are substantially smaller under the model in which a Box–Cox power transformation was used. From Table 3, the results for the power transformation parameters provide a clear indication of a better fit of this model than one in which the transformation is taken as the log function. Further evidence of this is given by the differences in the results for σ^2 between the two tables;

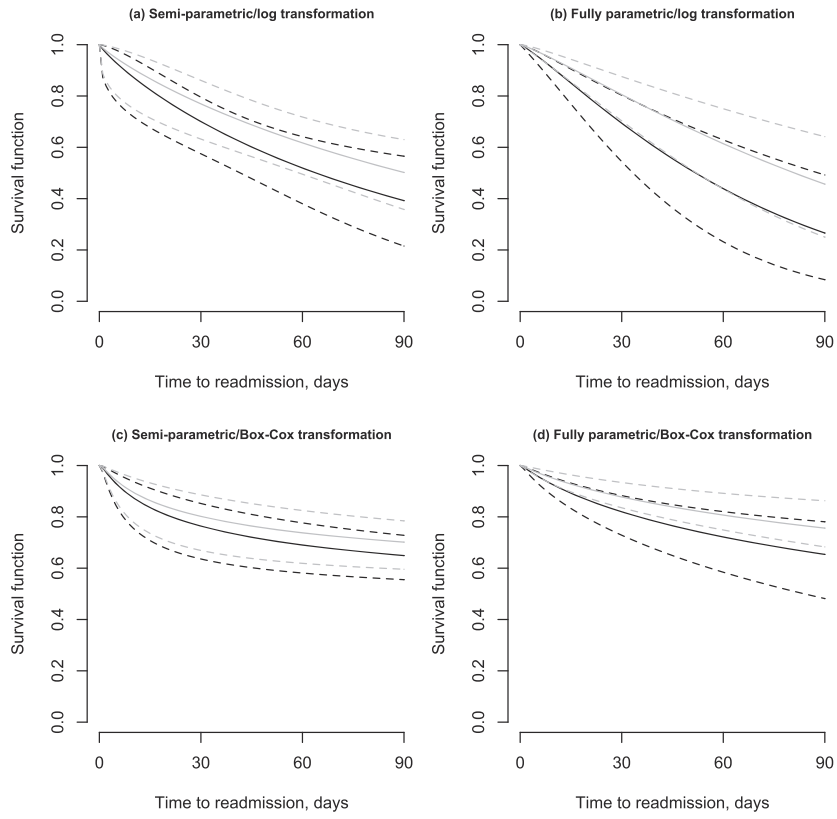

Fig. 3. Estimated conditional survival functions for readmission from four analyses that vary in terms of the transformation functions, $G(\cdot)$, and treatment of the patient-specific frailties. In each panel, black lines correspond to males and grey lines to females. Dashed lines indicate pointwise bootstrap-based 95% confidence bands. Throughout, the patient-specific frailty was set to $\Gamma = 1$.

under the log-transformation, there is substantial residual between-patient variation that is absorbed by the frailties ($\widehat{\sigma}^2$=1.23; 95% CI 0.89, 1.71); under the Cox–Box specification, there is virtually no residual between-patient variation ($\widehat{\sigma}^2$=0.04; 95% CI 2.87e-06, 0.72).

8. Discussion

In this paper, we have proposed a novel semi-parametric transformation model for the analysis of semi-competing risks data. The primary goal of this work was to develop a framework within which the assumption of a Gamma distribution for patient-specific frailties, standard in multi-state models for semi-competing risks data, can be relaxed. To our knowledge, this is the first paper in the semi-competing risks literature that attempts to do so. While the proposed methodology requires the analyst to adopt a *working frailty distribution*, the estimation procedures developed in Sections 3 and 4 yield a consistent estimator that is robust to the specific choice. This robustness comes at a price; however, to ensure identifiability, we restrict attention to specifications for the three hazard functions in the illness–death model that are fully parametric. To mitigate this, we propose a new class of flexible transformation models within which the transformation functions and error distributions can be made as simple or complex as analysts see fit. To accommodate right censoring, we use non-parametric score imputation and show that the resulting estimator is $\sqrt{n}$-consistent and asymptotically normally distributed.

The complexity of semi-competing risks data, coupled with finite sample sizes, typically requires analysts to balance structure and flexibility in their model specifications. Towards this, some parametric structure is invariably imposed either in the way in which covariates influence the underlying risk (i.e. regression parameters) or in terms of distributional assumptions. In principle, analysts can vary these assumptions and examine sensitivity in the results. This will be particularly important in 'data poor' settings where either the event rates are low either due to the overall sample size or if the censoring rate is high. The modelling and estimation framework we propose will be useful in this context, giving analysts a complementary analysis approach in which a standard parametric assumption can be relaxed. While the proposed transformation models are flexible, the proposed estimation procedures can also be used if analysts are interested in restricting to Cox-type or AFT model in order to facilitate interpretation of regression parameters. As with all transformation models, if the $G_{\alpha_l}(\cdot)$ functions are permitted to be more flexible, one can compare risk across sub-populations graphically as we do in Section 7.

Although this paper is motivated by the analysis of semi-competing risks data, the semiparametric transformation model proposed in Section 2 and the estimation procedures derived in Sections 3 and 4 are much more broadly applicable. In particular, the methods could be applied to any analysis of multivariate survival data in which a shared frailty term is used to account for correlation. These include, but are not limited to, the modelling of natural disease history (page 271–300 in Aalen *et al.* (2008) & Yen *et al.* (2010)) and multi-state models for competing risks data (Gorfine & Hsu, 2011).

Finally, a number of extensions of the proposed methods are possible, which should give analysts even greater flexibility. Specifically, we are investigating relaxing the fully parametric specification for the hazard functions by considering estimation/inference for a Cox model for which the baseline hazard is left unspecified. This requires characterizing identifiability when the overall model has two non-parametric components, specifically the frailty distribution and the baseline hazard functions; this is an active area of our own research. Related to this is that while we adopt a Markov assumption for the transition from the non-terminal event to the terminal event (i.e. the hazard $\lambda_{\boldsymbol{\theta}_3}(t_2 \mid t_1)$ in expression (4)), analysts may want to consider other formulations such as the semi-Markov assumption in which one models the so-called sojourn time between the two events (e.g. Xu *et al.* (2010) & Lee *et al.* (2015)). Finally, an additional

important extension will be to permit the observed data to be subject to left truncation and/or interval censoring, which will broaden the scope of applications to which the methods can be applied.

Supporting information

Additional supporting information may be found in the online version of this article at the publishers web site.

References

Aalen, O., Borgan, O. & Gjessing, H. (2008). *Survival and event history analysis: A process point of view*, Springer Science & Business Media, New York.

Agresti, A., Caffo, B. & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Comput. Statist. Data Anal.* **47**, (3), 639–653.

Cui, S. & Sun, Y. (2004). Checking for the gamma frailty distribution under the marginal proportional hazards frailty model. *Statist. Sinica* **14**, (1), 249–267.

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

Ghosh, D. (2006). Semiparametric inferences for association with semi-competing risks data. *Statist. Med.* **25**, (12), 2059–2070.

Glidden, D.V. (1999). Checking the adequacy of the gamma frailty model for multivariate failure times. *Biometrika* **86**, (2), 381–393.

Gorfine, M. & Hsu, L. (2011). Frailty-based competing risks model for multivariate survival data. *Biometrics* **67**, (2), 415–426.

Heagerty, P.J. & Kurland, B.F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88**, (4), 973–985.

Horowitz, J.L. (2010). *Semiparametric and nonparametric methods in econometrics*, Vol. 692, Springer, New York.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**, (1), 75–83.

Hsieh, J.J., Wang, W. & Adam Ding, A. (2008). Regression analysis based on semicompeting risks data. *J. R. Statist. Soc. B* **70**, (1), 3–20.

Jiang, F. & Haneuse, S. (2014). Simulation of semicompeting risk survival data and estimation based on multistate frailty model. *Harvard University Biostatistics Working Paper Series.*

Jiang, H., Fine, J.P, Kosorok, M.R. & Chappell, R. (2005). Pseudo self-consistent estimation of a copula model with informative censoring. *Scand. J. Statist.* **32**, (1), 1–20.

Kosorok, M.R., Lee, B.L. & Fine, J.P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Ann. Statist.* **32**, (4), 1448–1491.

Lakhal, L., Rivest, L.P. & Abdous, B. (2008). Estimating survival and association in a semicompeting risks model. *Biometrics* **64**, (1), 180–188.

Lee, K.H., Haneuse, S., Schrag, D. & Dominici, F. (2015). Bayesian semiparametric analysis of semicompeting risks data: Investigating hospital readmission after a pancreatic cancer diagnosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **64**, (2), 253–273.

Litière, S., Alonso, A. & Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics* **63**, (4), 1038–1044.

Liu, L., Wolfe, R.A. & Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics* **60**, (3), 747–756.

Ma, Y. & Yin, G. (2010). Semiparametric median residual life model and inference. *Canad. J. Statist.* **38**, (4), 665–679.

Ma, Y. & Zhu, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107**, (497), 168–179.

McGilchrist, C.A. & Aisbett, C.W. (1991). Regression with frailty in survival analysis. *Biometrics* **47**, (1), 461–466.

Neuhaus, J.M., Hauck, W.W. & Kalbfleisch, J.D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* **79**, (4), 755–762.

Neuhaus, J.M., McCulloch, C.E. & Boylan, R. (2011). A note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics* **67**, (2), 654–656.
Nielsen, G.G., Gill, R.D, Andersen, P.K. & Sørensen, T.I. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.* **19**, (1), 25–43.
Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Anal.* **7**, (1), 55–64.
Peng, L. & Fine, J.P. (2007). Regression modeling of semicompeting risks data. *Biometrics* **63**, (1), 96–108.
R Core Team. (2015). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.
Shih, J.H. & Louis, T.A. (1995). Assessing gamma frailty models for clustered failure time data. *Lifetime Data Anal.* **1**, (2), 205–220.
Su, Y.R. & Wang, J.L. (2012). Modeling left-truncated and right-censored survival data with longitudinal covariates. *Ann. Statist.* **40**, (3), 1465–1488.
Tsiatis, A.A. & Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91**, (4), 835–848.
Vest, J.R., Gamm, L.D., Oxford, B.A., Gonzalez, M.I. & Slawson, K.M. (2010). Determinants of preventable readmissions in the United States: A systematic review. *Implement. Sci.* **5**, (1), 88.
Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *J. R. Stat. Soc. Ser. B* **65**, (1), 257–273.
Wang, Y., Garcia, T.P. & Ma, Y. (2012). Nonparametric estimation for censored mixture data with application to the Cooperative Huntington's Observational Research Trial. *J. Amer. Statist. Assoc.* **107**, (500), 1324–1338.
Warren, J.L., Barbera, L., Bremner, K.E., Yabroff, K.R., Hoch, J.S, Barrett, M.J., Luo, J. & Krahn, M.D. (2011). End-of-life care for lung cancer patients in the United States and Ontario. *J. Natl. Cancer Inst.* **103**, (11), 853–862.
White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, (1), 1–25.
Xu, J., Kalbfleisch, J.D. & Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics* **66**, (3), 716–725.
Ye, Y., Kalbfleisch, J.D. & Schaubel, D.E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63**, (1), 78–87.
Yen, A.M., Chen, T.H., Duffy, S.W. & Chen, C.D. (2010). Incorporating frailty in a multi-state model: Application to disease natural history modelling of adenoma-carcinoma in the large bowel. *Statist. Methods in Med.* **19**, (5), 529–546.
Zeng, D., Chen, Q., Chen, M.H., Ibrahim, J.G. & Amgen Research Group. (2012). Estimating treatment effects with treatment switching via semicompeting risks models: An application to a colorectal cancer study. *Biometrika* **99**, (1), 167–184.
Zeng, D., Chen, Q. & Ibrahim, J.G. (2009). Gamma frailty transformation models for multivariate survival times. *Biometrika* **96**, (2), 277–291.
Zeng, D. & Lin, D. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *J. R. Statist. Soc. B* **69**, (4), 507–564.
Zeng, D. & Lin, D. (2009). Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* **65**, (3), 746–752.

Received July 2015, in final form February 2016

Fei Jiang, Department of Statistics, University of South Carolina, Columbia, SC, USA.
E-mail: homebovine@gmail.com