

A mixture model for bivariate interval-censored failure times with dependent susceptibility

SHU JIANG

*Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: s64jiang@uwaterloo.ca*

RICHARD J. COOK

*Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada*

Summary

Interval-censored failure times arise when the status with respect to an event of interest is only determined at intermittent examination times. In settings where there exists a sub-population of individuals who are not susceptible to the event of interest, latent variable models accommodating a mixture of susceptible and nonsusceptible individuals are useful. We consider such models for the analysis of bivariate interval-censored failure time data with a model for bivariate binary susceptibility indicators and a copula model for correlated failure times given joint susceptibility. We develop likelihood, composite likelihood, and estimating function methods for model fitting and inference, and assess asymptotic-relative efficiency and finite sample performance. Extensions dealing with higher-dimensional responses and current status data are also described.

Keywords: copula, estimating functions, interval-censored, multivariate, nonsusceptible, two-stage-estimation

The final publication of this article: Shu Jiang and Richard J. Cook, A mixture model for bivariate interval-censored failure times with dependent susceptibility, *Statistics in Biosciences* (2020), 12: 37–62 is available at Springer via <https://doi.org/10.1007/s12561-020-09270-7>.

1 INTRODUCTION

1.1 BACKGROUND

Individuals with chronic conditions are routinely followed prospectively in registries in order to learn about the disease course. In many settings, the progression status of each individual is only ascertainable upon detailed clinical or radiological examination. Since individuals can only be assessed at periodic examination times (e.g. when they attend a clinic) disease progression times are interval-censored (Sun et al., 2006).

With complex disease processes, there may be two or more disease-related progression times of interest. In registries of individuals with diabetes, for example, interest may lie in the onset of diabetic retinopathy and diabetic nephropathy (Al-Kateb et al., 2008). Similarly, in systemic lupus erythematosus, a particularly complex multifaceted autoimmune disease, attention may be directed at the development of important neurological, cardiovascular, or renal complications (Gladman and Urowitz, 1987). In the slightly different context of paired organ systems (e.g. vision, hearing), one may wish to model progression for each organ of the system. Sun et al. (2019) develop models for the time to blindness in the left and right eyes in a cohort of individuals with age-related macular degeneration (AREDS Group, 1999). In children with otitis media, an inflammation of the inner ear, tubes are inserted to drain fluid and mitigate risk of total hearing loss, so here interest lies in the time to failure of tubes (Le and Lindgren, 1996).

Statistical methods have been developed in recent years for bivariate interval-censored failure time data (Betensky and Finkelstein, 1999; Kim and Xue, 2002; Sun et al., 2006; Cook et al., 2008b), and bivariate current status (Wang and Ding, 2000). The need to accommodate dependencies between associated failure times may be *(i)* to advance scientific understanding, *(ii)* to facilitate joint statements about two or more disease features, or *(iii)* simply to ensure valid inference.

Despite careful use of available data to characterize multivariate disease processes, residual unexplained variability often remains between individuals. Mixture models are often adopted to model this residual variation, with so-called “cure-rate” models particularly appealing if there appears to be a sub-population of non-susceptible individuals. Following the early work by Farewell (1977) and Farewell (1982) on mixture models for right-censored data, further developments have accommodated nonparametric approaches to estimation of the failure time distribution (Taylor, 1995) and semiparametric models (Sy and Taylor, 2000; Peng and Dear, 2000; Peng, 2003). Chatterjee and Shih (2001) proposed a bivariate model for modeling familial association in disease susceptibility and onset times when failure times are subject to right-censoring. Kim (2016) considered inference for a cure rate model with bivariate interval-censored data via an approximate likelihood.

We develop a bivariate cure rate model for interval-censored failure times and address the setting where the event times are interval-censored. As in Chatterjee and Shih (2001), we use an odds ratio to model the association between latent susceptibility indicators and a copula model for the dependence between the failure times for events to which individuals are susceptible. This formulation enables one to separately consider dependence in the susceptibility and dependence in the failure times among individuals who are susceptible to the corresponding events. The three frameworks for inference include maximum likelihood, composite likelihood and two-stage estimation, and the use of weighted second-order estimating functions. The different methods are developed, implemented and compared to provide a foundation for the further development of methods which may be best addressed using one of these approaches.

The remainder of this paper is as follows. In the following subsection, we describe the data from the University of Toronto Psoriatic Arthritis Cohort that motivates this work. In Section 2, we define notation, formulate the marginal models and describe the association structure. Maximum likelihood and a two-stage estimation procedure are described in Sections 3.1 and 3.2, respectively, and an estimating function approach is developed based on weighted second-order estimating functions in Section 3.3. It is well-known that the two-stage procedure will be less efficient but we study the asymptotic relative efficiency of the estimators from the estimating function approach in Section 3.4. The findings from simulation studies are reported on in Section 4.1 and an application involving the incidence of damage in hand and foot joints in patients with psoriatic arthritis is given in Section 4.2. Extensions to accommodate higher dimensional failure times and current status data are described in Section 5 and general remarks are given in Section 6.

1.2 JOINT DAMAGE IN THE EXTREMITIES IN PSORIATIC ARTHRITIS

The Center for Prognosis Studies in Rheumatic Disease maintains a registry of patients with psoriatic arthritis, an autoimmune condition which features skin and joint involvement. The joint involvement can be manifested as inflammation in and around the joints of the wrist, hand, knee, ankle, foot, lower back, and neck; this inflammation may ultimately lead to damage in affected joints. Patients in the clinical registry are followed according to a standard protocol in which clinic visits and radiographic assessments are scheduled annually and biannually, respectively. At each clinic visit, patients are assessed by a physician to determine the level of damage in each of 64 joints. The level of the damage is measured according to a validated modified Steinbrocker scoring system (Rahman et al., 1998) where a score of 0 represents no damage, and scores from 1 to 5 represent progressively more advanced stages of damage; states 4 and 5 represent the most severe form of damage.

Due to the critical role of hands for functions of daily living (e.g. dressing, opening jars, typing, etc.) and foot joints for maintaining full mobility, physicians are interested in the onset time distribution for damage in the hand and foot joints; time here is measured from the time of diagnosis of psoriatic arthritis. There is little information on how often patients develop damaged joints in these locations, nor is there for the distribution of times to damage. The model we describe in the next section facilitates joint modeling of the susceptibility for damage in these two locations, and for the times to the development of damage in each location given joint susceptibility. We adopt a marginal modeling approach so that inferences from the joint model can be carried out for the individual (e.g. hand or foot) damage processes naturally following model fit, and so that the dependence parameters are functionally independent of the parameters characterizing the marginal distributions.

To study this, we restrict attention to the 28 joints of the hands and 12 joints in the feet. Individuals are considered damage-free in the hands when all hand joints are in state 0, but are considered to have developed damage in the hands when one or more hand joint enters state 1; the time to the onset of foot damage is similarly defined. Modeling the onset of damage in individual joints is possible in principle but we do not consider this here since the focus is on the bivariate setting; extensions to illustrate how this could be done are outlined in Section 5.1.

We consider data from a sample of n independent individuals labeled $i = 1, \dots, n$. We let T_{i1} denote the time to the onset of damage in the hands and T_{i2} the time to the onset of damage in the feet for individual i , $i = 1, \dots, n$. Because individuals are only assessed periodically for joint damage, T_{i1} and T_{i2} are interval-censored; that is, we only know the times of damage that lie between the last negative assessment and the first positive assessment. Let $\mathcal{B}_{ij} = [l_{ij}, r_{ij}]$ denote the censoring interval for T_{ij} and $\mathcal{B}_i = \mathcal{B}_{i1} \times \mathcal{B}_{i2}$ be the censoring region for $T_i = (T_{i1}, T_{i2})'$, where $r_{ij} \rightarrow \infty$ if individual i was not known to experience the type j event resulting in right-censored event time. Figure 1 contains a plot of the censoring regions for eight sample individuals in the clinical registry. Of the 657 individuals in the sample there are 440 (67.0%) with $r_{i1} < \infty$, 460 (70.0%) with $r_{i2} < \infty$, and 367 (55.9%) with both $r_{i1} < \infty$ and $r_{i2} < \infty$ which are the number of individuals known definitively to have developed damage in the hand joints, foot joints, and both the hand and foot joints, respectively. Out of the 657 individuals in the sample, 275 (41.9%) were female. The first, second and third quartiles of the age distribution were 29, 39 and 50 years of age, respectively.

We return to this example in Section 4.2.

2 NOTATION AND MODEL FORMULATION

Let $Z_{ij} = 1$ if individual i is susceptible to a type j event and $Z_{ij} = 0$ otherwise, $j = 1, 2$, and let $X_i = (1, X_{i1}, \dots, X_{i,p-1})'$ denote a $p \times 1$ covariate vector. If $g_1: [0, 1] \rightarrow (-\infty, \infty)$ is a monotonic 1-1 differentiable link function, a binary regression model for $Z_{ij}|X_i$ is specified by setting $g_1(\mu_{ij}) = X_i' \beta_j$ where β_j is a $p \times 1$ vector of regression coefficients, $j = 1, 2$; we let $\beta = (\beta_1', \beta_2')'$. The

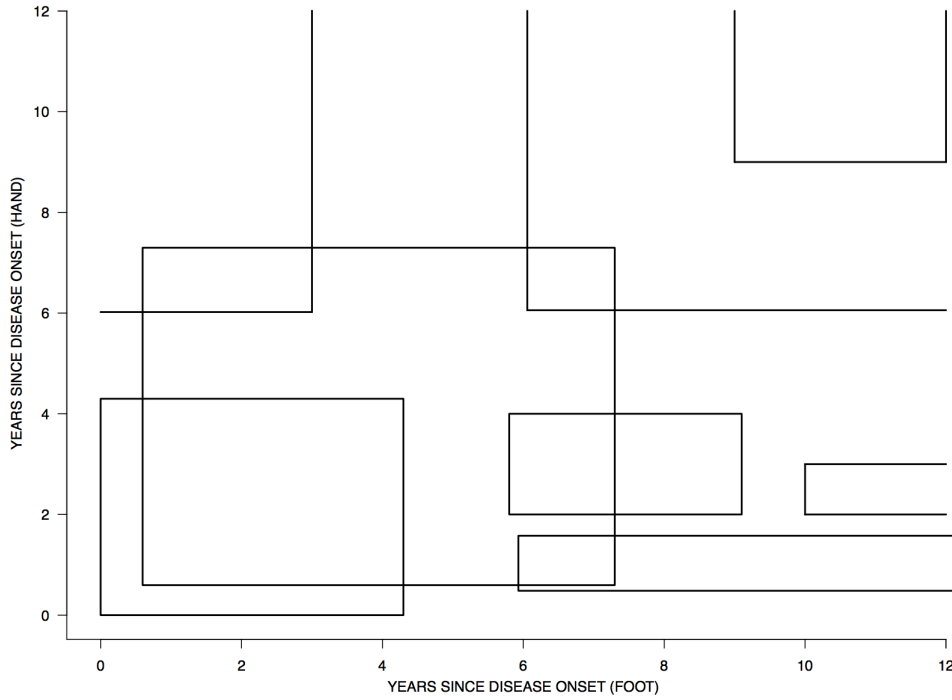


Figure 1: Censoring regions for the time to joint damage in the hands and feet in a sample of eight individuals from the University of Toronto Psoriatic Arthritis Cohort.

association between Z_{i1} and Z_{i2} given X_i can be modeled by the odds ratio

$$\psi_i = \frac{P(Z_{i1} = 1, Z_{i2} = 1|X_i)P(Z_{i1} = 0, Z_{i2} = 0|X_i)}{P(Z_{i1} = 1, Z_{i2} = 0|X_i)P(Z_{i1} = 0, Z_{i2} = 1|X_i)}, \quad (1)$$

as in Lipsitz and Laird (1991). If $g_2: [0, \infty) \rightarrow (-\infty, \infty)$ is a monotonic 1 – 1 differentiable function and $\mu_{ij} = E(Z_{ij}|X_i)$, $j = 1, 2$, a second-order regression model can be specified by setting $g_2(\psi_i) = X_i' \gamma$ where γ is a vector of regression coefficients. The resulting joint distribution $P(Z_i|X_i; \theta_1)$ is indexed by $\theta_1 = (\beta', \gamma)'$.

We let $T_{ij} > 0$ denote the time of a type j event for individual i and let $T_{ij} \rightarrow \infty$ if $Z_{ij} = 0$. The marginal distribution of T_{ij} given $(Z_{ij} = 1, X_i)$ is denoted by $P(T_{ij} \geq t_j|Z_{ij} = 1, X_i) = \mathcal{F}_j(t_j|X_i; \lambda_j)$ and indexed by a $q_j \times 1$ finite dimensional parameter λ_j , $j = 1, 2$, with $\lambda = (\lambda_1', \lambda_2)'$. This conditional distribution could take the form of any conventional location-scale model (Lawless, 2003), parametric proportional hazards model (Broström, 2012), or weakly parametric proportional hazards models in which baseline hazard functions are piecewise-constant (Friedman et al., 1982); we consider the later in Section 4.

We construct the bivariate survivor function

$$P(T_{i1} \geq t_1, T_{i2} \geq t_2|Z_i = (1, 1)', X_i) = \mathcal{F}_{12}(t_1, t_2|X_i), \quad t_1, t_2 \geq 0, \quad (2)$$

using a copula function $C(u_1, u_2; \phi)$ where $0 \leq u_j \leq 1$, $j = 1, 2$, and ϕ is the dependence parameter (Nelsen, 2006). Specifically the joint survivor function for $T_i|Z_i = (1, 1), X_i$ is indexed by $\theta_2 = (\lambda', \phi)'$ and obtained by setting

$$\mathcal{F}_{12}(t_1, t_2|X_i; \theta_2) = C(\mathcal{F}_1(t_1|X_i; \lambda_1), \mathcal{F}_2(t_2|X_i; \lambda_2); \phi).$$

For the class of Archimedean copulas we can write

$$C(u_1, u_2; \phi) = G^{-1}(G(u_1; \phi) + G(u_2; \phi); \phi),$$

where $G:[0, 1] \rightarrow [0, \infty]$ is a continuous, strictly decreasing and convex generator function with $G(1; \phi) = 0$. Within this family of models Kendall's τ is computed as

$$\tau = 1 + 4 \int_0^1 \frac{G(u; \phi)}{G'(u; \phi)} du . \quad (3)$$

(Nelsen, 2006). Since the survivor function is monotonic in time (3) also characterizes the dependence between T_{i1} and T_{i2} given $Z_i = (1, 1)'$, X_i .

The joint model constructed here has the appealing feature that given $Z_i = (1, 0)'$, X_i

$$P(T_{i1} \geq t_1, T_{i2} \geq t_2 | Z_i = (1, 0)', X_i) = \mathcal{F}_{12}(t_1, 0 | X_i) = \mathcal{F}_1(t_1 | X_i) ,$$

and given $Z_i = (0, 1)'$, X_i

$$P(T_{i1} \geq t_1, T_{i2} \geq t_2 | Z_i = (0, 1)', X_i) = \mathcal{F}_{12}(0, t_2 | X_i) = \mathcal{F}_2(t_2 | X_i) .$$

We discuss the implications of relaxing this condition in Section 6. If $Z_i = (0, 0)'$ then $P(T_{i1} \geq t_1, T_{i2} \geq t_2 | Z_i = (0, 0)', X_i) = 1$ for any finite (t_1, t_2) .

3 METHODS FOR ESTIMATION AND INFERENCE

We consider the setting where the failure status of individuals is only intermittently observed at assessment times which are assumed to satisfy the sequential ignorability conditions of Hogan et al. (2004) as described by Grüger et al. (1991). This assumption enables one to focus only on the censoring intervals $\mathcal{B}_{ij} = [l_{ij}, r_{ij}]$ whereby $T_{ij} \in \mathcal{B}_{ij}$, $j = 1, 2$. The censoring region for $T_i = (T_{i1}, T_{i2})'$ is then $\mathcal{B}_i = \mathcal{B}_{i1} \times \mathcal{B}_{i2}$ and the observed data are denoted by $D_i = (\mathcal{B}_i, X_i)$ for each individual i in the sample of n individuals, $i = 1, \dots, n$.

Here, we consider the contributions to the likelihood given the latent susceptibility indicators. If $Z_i = (1, 1)$ for example, then the expression $P(T_i \in \mathcal{B}_i | Z_i = (1, 1), X_i; \theta_2)$ is

$$\mathcal{F}_{12}(l_{i1}, l_{i2} | X_i) - \mathcal{F}_{12}(l_{i1}, r_{i2} | X_i) - \mathcal{F}_{12}(r_{i1}, l_{i2} | X_i) + \mathcal{F}_{12}(r_{i1}, r_{i2} | X_i) .$$

Note however that $P(T_i \in \mathcal{B}_i | Z_i = (1, 0), X_i; \theta_2) = 0$ when $\delta_{i2} = I(r_{i2} < \infty) = 1$ since $T_{i2} \rightarrow \infty$ when $Z_{i2} = 0$, but otherwise (i.e. when $r_{i2} < \infty$)

$$P(T_{i1} \in \mathcal{B}_{i1}, T_{i2} > l_{i2} | Z_i = (1, 0), X_i; \theta_2) = \mathcal{F}_1(l_{i1} | X_i) - \mathcal{F}_1(r_{i1} | X_i) ;$$

the expressions for $P(T_i \in \mathcal{B}_i | Z_i = (0, 1), X_i; \theta_2)$ for $\delta_{i1} = 1$ and $\delta_{i1} = 0$ are similarly obtained. Finally, if $Z_i = (0, 0)$ then $P(T_i \in \mathcal{B}_i | Z_i = (0, 0), X_i; \theta_2) = 0$ if either $r_{i1} < \infty$ or $r_{i2} < \infty$, but otherwise $P(T_i \in \mathcal{B}_i | Z_i = (0, 0), X_i; \theta_2) = 1$ since the event times are taken to be infinite in this case.

Next, we consider estimation and inference based on maximum likelihood, composite likelihood and two-stage estimation, and estimating functions in Sections 3.1, 3.2 and 3.3, respectively. We investigate relative efficiency in Section 3.4.

3.1 MAXIMUM LIKELIHOOD

Here, we partition $\theta = (\beta', \gamma', \lambda', \phi)'$ in terms of $\theta_1 = (\beta', \gamma')'$, the parameter vector governing the distribution of $Z_i | X_i$, and $\theta_2 = (\lambda', \phi)'$, the parameter vector governing the distribution of $T_i | Z_i, X_i$. Since Z_i is latent in general, the observed data likelihood contribution for individual i is

$$L_i(\theta) \propto \sum_{z_i \in \mathcal{Z}} P(Z_i = z_i | X_i; \theta_1) P(T_i \in \mathcal{B}_i | Z_i = z_i, X_i; \theta_2) , \quad (4)$$

where $\mathcal{Z} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ is the sample space for Z_i . While (4) could be maximized directly an expectation-maximization algorithm (Dempster et al., 1977) is particularly appealing when fitting finite mixture models. If Z_i were known, the complete data likelihood contribution from individual i is

$$\mathcal{L}_i(\theta) \propto \prod_{z_i \in \mathcal{Z}} [P(Z_i = z_i | X_i; \theta_1) P(T_i \in \mathcal{B}_i | Z_i = z_i; X_i; \theta_2)]^{I(Z_i = z_i)}. \quad (5)$$

If $\mathcal{S}_{i1}(Z_i | X_i; \theta_1) = \partial \log P(Z_i | X_i; \theta_1) / \partial \theta_1$ and $\mathcal{S}_{i2}(T_i \in \mathcal{B}_i | Z_i, X_i; \theta_2) = \partial \log P(T_i \in \mathcal{B}_i | Z_i, X_i; \theta_2) / \partial \theta_2$ are the complete data score functions, the contributions to the observed data score functions for θ_1 and θ_2 from individual i are

$$\begin{aligned} S_{i1}(\theta) &= E\{\mathcal{S}_{i1}(Z_i | X_i; \theta_1) | D_i; \theta\} \\ S_{i2}(\theta) &= E\{\mathcal{S}_{i2}(T_i \in \mathcal{B}_i | Z_i, X_i; \theta_2) | D_i; \theta\} \end{aligned} \quad (6)$$

respectively, where the expectation is with respect to $Z_i | D_i$ based on

$$P(Z_i = z_i | D_i; \theta) = \frac{P(T_i \in \mathcal{B}_i | Z_i, X_i; \theta_2) P(Z_i | X_i; \theta_1)}{\sum_{z_i \in \mathcal{Z}} P(T_i \in \mathcal{B}_i | Z_i, X_i; \theta_2) P(Z_i | X_i; \theta_1)}.$$

We let $\mathcal{S}_i(\theta) = (\mathcal{S}'_{i1}(\theta_1), \mathcal{S}'_{i2}(\theta_2))'$ and $S_i(\theta) = (S'_{i1}(\theta), S'_{i2}(\theta))'$ be the full complete and observed data score vectors satisfying

$$S_i(\theta) = \sum_{z_i \in \mathcal{Z}} P(Z_i = z_i | D_i; \theta) \mathcal{S}_i(\theta).$$

If θ^r is the estimate of θ at the r th iteration, then θ^{r+1} is obtained by solving

$$\begin{aligned} S_1(\theta_1; \theta^r) &= \sum_{i=1}^n E\{\mathcal{S}_1(Z_i | X_i; \theta_1) | D_i; \theta^r\} = 0, \\ S_2(\theta_2; \theta^r) &= \sum_{i=1}^n E\{\mathcal{S}_2(T_i \in \mathcal{B}_i | Z_i, X_i; \theta_2) | D_i; \theta^r\} = 0. \end{aligned}$$

These steps are repeated iteratively until $|\theta^{r+1} - \theta^r| < \epsilon$ where ϵ is a specified convergence criterion. The observed information matrix can be obtained based on Louis (1982) who showed that

$$I_i(\theta) = E[\mathcal{J}_i(\theta) | D_i] - E[\mathcal{S}_i(\theta) \mathcal{S}'_i(\theta) | D_i] + S_i(\theta) S'_i(\theta),$$

where $\mathcal{J}_i(\theta) = -\partial \mathcal{S}_i(\theta) / \partial \theta'$ is the block-diagonal complete data information matrix. The observed information matrix $I(\theta)$ for the full sample is then given by $I(\theta) = \sum_{i=1}^n I_i(\theta)$.

3.2 COMPOSITE LIKELIHOOD AND TWO-STAGE ESTIMATION

The joint model of Section 2 lends itself to a two-stage estimation procedure in the spirit of Shih and Louis (1995). To this end, we consider a different partition for θ as $(\alpha'_1, \alpha'_2)'$, where $\alpha_1 = (\beta', \lambda)'$ parameterizes the marginal event type-specific processes and $\alpha_2 = (\gamma', \phi)'$ contains the dependence parameters.

The “working independence” composite likelihood used for stage 1 estimation is

$$L_1(\alpha_1) = \prod_{i=1}^n L_{i1}(\alpha_1) \propto \prod_{i=1}^n \prod_{j=1}^2 L_{ij1}(\alpha_1), \quad (7)$$

where $L_{ij1}(\alpha_1)$ is proportional to

$$[\mu_{ij}P(T_{ij} \in \mathcal{B}_{ij}|Z_{ij} = 1, X_i)]^{\delta_{ij}} [\mu_{ij}P(T_{ij} \in \mathcal{B}_{ij}|Z_{ij} = 1, X_i) + 1 - \mu_{ij}]^{(1-\delta_{ij})} ,$$

and $\delta_{ij} = I(r_{ij} < \infty)$ indicates whether individual i is known to have experienced a type j event, $j = 1, 2$. Let $\tilde{\alpha}_1 = (\tilde{\beta}', \tilde{\lambda}')'$ denote the value of α_1 maximizing (7). At stage 2 the dependence parameter α_2 is estimated by inserting $\tilde{\alpha}_1$ into the observed data likelihood (4) to obtain

$$L_{i2}(\tilde{\alpha}_1, \alpha_2) \propto \sum_{z_i \in \mathcal{Z}_i} P(Z_i = z_i|X_i; \tilde{\beta}, \gamma)P(T_i \in \mathcal{B}_i|Z_i = z_i, X_i; \tilde{\lambda}, \phi) , \quad (8)$$

and maximizing

$$L_2(\tilde{\alpha}_1, \alpha_2) = \prod_{i=1}^n L_{i2}(\tilde{\alpha}_1, \alpha_2) ,$$

with respect to α_2 to obtain $\tilde{\alpha}_2$. We let $\tilde{\theta} = (\tilde{\alpha}'_1, \tilde{\alpha}'_2)'$ denote the estimate from this two-stage procedure.

If $\bar{S}_{i1}(\alpha_1) = \partial \log L_{i1}(\alpha_1)/\partial \alpha_1$ and $\bar{S}_{i2}(\alpha) = \partial \log L_{i2}(\tilde{\alpha}_1, \alpha_2)/\partial \alpha_2$, we can write $\bar{S}_i(\theta) = (\bar{S}'_{i1}(\alpha_1), \bar{S}'_{i2}(\alpha))'$. Standard estimating function theory (Boos and Stefanski, 2013) gives

$$\sqrt{n}(\tilde{\theta} - \theta) \rightarrow \text{MVN}(0, \bar{\mathcal{A}}^{-1}(\theta)\bar{\mathcal{B}}(\theta) [\bar{\mathcal{A}}^{-1}(\theta)]') ,$$

where $\bar{\mathcal{A}}(\theta) = E\{-\partial \bar{S}_i(\theta)/\partial \theta\}$ and $\bar{\mathcal{B}}(\theta) = E\{\bar{S}_i(\theta)\bar{S}'_i(\theta)\}$. The matrices $\bar{\mathcal{A}}(\theta)$ and $\bar{\mathcal{B}}(\theta)$ are estimated empirically by

$$\bar{A}(\tilde{\theta}) = -n^{-1} \sum_{i=1}^n \left. \frac{\partial \bar{S}_i(\theta)}{\partial \theta'} \right|_{\theta=\tilde{\theta}} \quad \text{and} \quad \bar{B}(\tilde{\theta}) = n^{-1} \sum_{i=1}^n \bar{S}_i(\theta)\bar{S}'_i(\theta) \Big|_{\theta=\tilde{\theta}}$$

which can be used to obtain robust sample variance estimates.

3.3 WEIGHTED SECOND-ORDER ESTIMATING EQUATIONS

Here, we describe a set of weighted second-order estimating functions which have the form of expected complete data estimating functions one would use if Z_i was known. Given their use of second-order moments, they may be expected to yield more efficient estimators than the two-stage estimators of Section 3.2. Moreover, they introduce the opportunity to consider robustness and efficiency trade-offs through specification of different component matrices. In what follows we partition θ as $(\theta'_1, \theta'_2)'$ as in Section 3.1.

The estimating function we propose for θ_1 is of the form $U_1(\theta) = \sum_{i=1}^n U_{i1}(\theta)$ with

$$U_{i1}(\theta) = \sum_{z_i \in \mathcal{Z}} P(Z_i = z_i|D_i; \theta) \left[H'_{i1}(\theta_1) \Sigma_{i1}^{-1}(\theta_1) \begin{pmatrix} Z_i - \mu_i \\ W_{i1} - \omega_{i1} \end{pmatrix} \right] , \quad (9)$$

where $Z_i = (Z_{i1}, Z_{i2})'$, $\mu_i = E(Z_i|X_i; \beta) = (\mu_{i1}, \mu_{i2})'$, $W_{i1} = Z_{i1}Z_{i2}$ and $\omega_{i1} = E(W_{i1}|X_i; \theta_1)$. The derivative matrix is given by

$$H_{i1}(\theta_1) = \begin{pmatrix} \partial \mu_i / \partial \beta' & 0 \\ \partial \omega_{i1} / \partial \beta' & \partial \omega_{i1} / \partial \gamma' \end{pmatrix} ,$$

and

$$\Sigma_{i1}(\theta_1) = \begin{pmatrix} \text{cov}(Z_i|X_i) & \text{cov}(Z_i, W_{i1}|X_i) \\ \text{cov}(W_{i1}, Z_i|X_i) & \text{var}(W_{i1}|X_i) \end{pmatrix} ,$$

which may be viewed as a complete data 3×3 covariance matrix. The term in square brackets in (9) has the form of a second-order estimating function developed by Prentice and Zhao (1991) for fitting marginal regression models to correlated binary data, but (9) differs in that it is a weighted sum of such terms since $(Z'_i, W_{i1})'$ is unknown.

The set of estimating functions for θ_2 has a similar form and is given by $U_2(\theta) = \sum_{i=1}^n U_{i2}(\theta)$ where

$$U_{i2}(\theta) = \sum_{z_i \in \mathcal{Z}} P(Z_i = z_i | D_i; \theta) \left[H'_{i2}(\theta_2) \Delta_i \Sigma_{i2}^{-1}(\theta_2) \begin{pmatrix} Y_i - \eta_i \\ W_{i2} - \omega_{i2} \end{pmatrix} \right], \quad (10)$$

$Y_{ik} = I(T_{ik} \in \mathcal{B}_{ik})$, $Y_i = (Y_{i1}, Y_{i2})'$, $\eta_i = E(Y_i | Z_i, X_i; \lambda)$, $W_{i2} = Y_{i1} Y_{i2}$ and $\omega_{i2} = E(W_{i2} | Z_i, X_i; \theta_2)$. The matrices are given by

$$H_{i2}(\theta_2) = \begin{pmatrix} \partial \eta_i / \partial \lambda' & 0 \\ \partial \omega_{i2} / \partial \lambda' & \partial \omega_{i2} / \partial \phi \end{pmatrix}, \quad \Delta_i = \begin{pmatrix} z_{i1} & 0 & 0 \\ 0 & z_{i2} & 0 \\ 0 & 0 & z_{i1} z_{i2} \end{pmatrix},$$

and

$$\Sigma_{i2}(\theta_2) = \begin{pmatrix} \text{cov}(Y_i | Z_i, X_i) & \text{cov}(Y_i, W_{i2} | Z_i, X_i) \\ \text{cov}(W_{i2}, Y_i' | Z_i, X_i) & \text{var}(W_{i2} | Z_i, X_i) \end{pmatrix}.$$

The derivative matrix $H_{i2}(\theta_2)$ and covariance matrix $\Sigma_{i2}(\theta_2)$ are analogous to those of (9) but their precise forms are determined by the specification of the copula function for the failure times. If Z_i were known, the matrix Δ_i would ensure that the appropriate elements of this estimating function contribute information about the parameters in the marginal failure time distributions and association parameters; as Z_i is unknown we again take the conditional expectation over the possible values of Z_i .

Letting $U_i(\theta) = (U'_{i1}(\theta), U'_{i2}(\theta))'$ the estimating equations are defined by setting $\sum_{i=1}^n U_i(\theta) = 0$ which can be solved for θ directly, or iteratively in the spirit of the EM algorithm of Section 3.1. In the latter case if θ^r is an estimate at the r th iteration we insert it into $P(Z_i | D_i; \theta)$ in (9) and (10) and simultaneously solve the respective equations for θ_1^{r+1} and θ_2^{r+1} , respectively. As is the case with the EM algorithm this iterative procedure is repeated until $|\theta^{r+1} - \theta^r| < \epsilon$ for some specified convergence criterion ϵ . Algorithms for implementing GEE2 could be exploited for estimation of θ_1 at each step by the creation of a dataframe for a pseudo-sample containing multiple lines per individual corresponding to each possible realization of Z_i with weights based on $P(Z_i | D_i; \theta^r)$. For (10), the derivative and covariance matrices are unique to this setting since the moments are determined by the copula function and the observation times so specialized coding is required. A similar situation is described by Kor et al. (2013) and Zhong and Cook (2016) for interval-censored data and Tolusso and Cook (2009) for current status data. Note that when solving the set of estimating equations in the complete data setting, robustness may be achieved if a diagonal $H_i(\theta)$ matrix is used since in this case one is not attempting to draw any information about the marginals from the association. To simplify calculations, we use diagonal $H_i(\theta)$ and $\Sigma_i(\theta)$ matrices in our simulations and in the application that follows.

Subject to correct specification of the conditional moments, (9) and (10) are unbiased estimating functions, so the estimators $\check{\theta}_1$ and $\check{\theta}_2$ solving the corresponding estimating equations are consistent, with $\check{\theta} = (\check{\theta}'_1, \check{\theta}'_2)'$ having an asymptotic normal distribution

$$\sqrt{n}(\check{\theta} - \theta) \rightarrow N(0, \bar{\mathcal{A}}^{-1}(\theta) \bar{\mathcal{B}}(\theta) [\bar{\mathcal{A}}^{-1}(\theta)]'), \quad (11)$$

with $\bar{\mathcal{A}}(\theta) = E[-\partial U_i(\theta) / \partial \theta']$ and $\bar{\mathcal{B}}(\theta) = E[U_i(\theta) U_i'(\theta)]$. Again in a given sample we obtain estimates

$$\bar{\mathcal{A}}(\check{\theta}) = -n^{-1} \sum_{i=1}^n \frac{\partial U_i(\theta)}{\partial \theta'} \Big|_{\theta=\check{\theta}} \quad \text{and} \quad \bar{\mathcal{B}}(\check{\theta}) = n^{-1} \sum_{i=1}^n U_i(\theta) U_i'(\theta) \Big|_{\theta=\check{\theta}}$$

which can be used for inference.

3.4 A STUDY OF ASYMPTOTIC RELATIVE EFFICIENCY

Here, we investigate the asymptotic relative efficiency of the maximum likelihood estimator and the estimator based on the estimating function approach of Section 3.3 where the large sample variance is given in (11). We consider a single binary covariate X_{i1} with $P(X_{i1} = 1) = P(X_{i1} = 0) = 0.5$ and adopt a logit link for the binary regression model so $g_1(\mu) = \log(\mu/(1 - \mu))$. We set $\beta_{j1} = \log 1.5$ and solved for β_{j0} so that $P(Z_{ij} = 1|X_{i1} = 1; \beta_j) = 0.66$, $j = 1, 2$. We considered values of $\log \psi$ ranging from 0 to $\log 3$ in increments of 0.1. For the event times, we take $T_{ij} \perp X_i | Z_{ij}$ and specify

$$P(T_{ij} \geq t | Z_{ij} = 1, X_i; \lambda_j) = \exp(-(\lambda_{j1}t)^{\lambda_{j2}}),$$

where $\lambda_j = (\lambda_{j1}, \lambda_{j2})'$. We let $\lambda_{j2} = 1$ and determine λ_{j1} such that $P(T_{ij} < 1 | Z_{ij} = 1, X_i) = 0.9$ for $j = 1, 2$ where 1 is an administrative censoring time. A Clayton copula is used for the bivariate failure time distribution with Kendall's τ set to 0.3 or 0.6. To generate the censoring intervals we let $\{N_i(s), 0 < s\}$ denote a counting process for the assessments which arise from a Poisson process with rate $\rho = 10$ giving an average of 10 visits over $(0, 1]$. For the model to be fitted, a piecewise-constant hazard was adopted for each component failure time model, with cut-points at 0.12 and 0.3 chosen to ensure roughly an equal number of expected events in each interval.

In this framework, we evaluate the (i) expected information matrix based on the observed data likelihood (4) as $E\{-\partial^2 \log L(\theta) / \partial \theta' \partial \theta\}$ where $L(\theta) = \prod_{i=1} L_i(\theta)$, and (ii) the asymptotic covariate of the estimator $\check{\theta}$ from the estimating function approach by evaluating $\bar{\mathcal{A}}^{-1}(\theta) \bar{\mathcal{B}}(\theta) [\bar{\mathcal{A}}^{-1}(\theta)]' / n$ based on Section 3.3. Figure 2 contains plots of the asymptotic standard errors of the estimators for selected parameters for a sample of $n = 20,000$ as a function of the odds ratio ψ . We select β_{11} , λ_{11} , ψ and τ and plot the results for the maximum likelihood estimator and estimator based on the estimating functions for $\tau = 0.3$ and 0.6.

The plots reveal that the estimators based on the estimating function approach are highly efficient with asymptotic standard errors tracking those from maximum likelihood estimation very well. There is generally little trend in the precision as a function of ψ with the greatest trend for the estimation of ψ itself. Moreover, the asymptotic standard error of the estimator of Kendall's τ is most sensitive to τ .

4 EMPIRICAL STUDIES AND APPLICATION

4.1 SIMULATION STUDIES

Here, we report on simulation studies conducted to gain insight into the finite sample performance of the three methods of analysis introduced in Section 3. We again consider a binary covariate $X_i \sim \text{Bern}(0.5)$ and logit link giving

$$\text{logit}(P(Z_{ij} = 1 | X_i; \beta_j)) = \beta_{j0} + \beta_{j1} X_i, \quad j = 1, 2.$$

We set $\beta_{j1} = \log 1.5$ and $P(Z_{ij} = 1 | X_i = 1; \beta_j) = 0.66$ and consider $\log \psi = \log 1.5$ and $\log 3$. For the event times, we specify

$$P(T_{ij} \geq t | Z_{ij} = 1, X_i; \lambda_j) = \exp(-(\lambda_{j1}t)^{\lambda_{j2}}),$$

with $\lambda_{j2} = 1$, $j = 1, 2$ and specify λ_{j1} such that $P(T_{ij} < A | Z_{ij} = 1, X_i) = 0.9$ where A is the administrative censoring time; we set $A = 1$ without loss of generality. In the application, interest primarily lies in the effect of covariates on susceptibility, rather than the actual time of damage, so while the developments in Section 3 allow for this dependence, we assume $T_i \perp X_i | Z_i$ here. We use a Clayton copula for the bivariate failure time distribution with Kendall's τ set to 0.3 or 0.6. Let

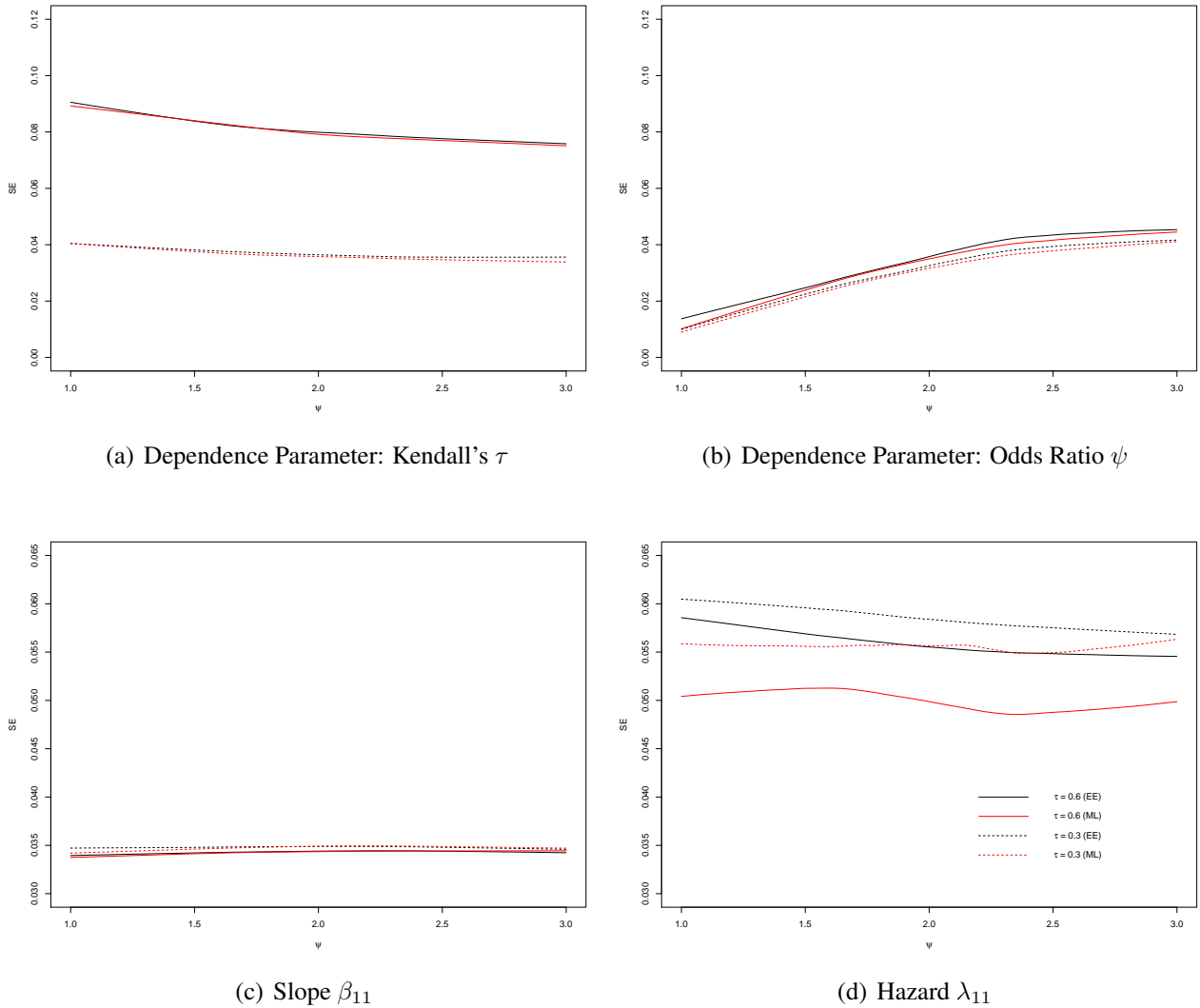


Figure 2: Plot of the asymptotic standard errors for selected parameters based on maximum likelihood and the estimating functions of Section 3.3 with $\tau = 0.3$ and $\tau = 0.6$; sample size $n = 20,000$.

$\{N_i(s), 0 < s\}$ denote the counting process for the assessments which we take to be a Poisson process with rate $\rho = 10$ and 20 to correspond to an average of 10 and 20 visits over $(0, 1]$, respectively. In the analyses, a piecewise-constant baseline hazard was adopted for each component failure time model with cut-points at 0.12 and 0.3 chosen to ensure roughly an equal expected number of failure times in the respective intervals.

The results presented in Tables 1 and 2 are from the analysis of 500 simulated samples of $n = 1000$ individuals each with an average of 10 and 20 visits, respectively; we comment here on the results of Table 1. We see that empirical biases of all methods are negligible, there is excellent agreement between the empirical and average robust standard errors, and the empirical coverage probability are close to the nominal 95% level. The estimators from the two-stage procedure are less efficient than those from maximum likelihood, but the estimators from the weighted estimating equation approach are remarkably efficient, consistent with the findings from the asymptotic study of Section 3.4. Findings from Table 2 are similar with smaller standard errors as expected.

Table 1: Empirical performance of estimators for 500 simulations with 1000 individuals per simulation under a three-piece piecewise constant hazards bivariate mixture model with $E\{N_i(A)\} = 10$ where $A = 1$ is the administrative censoring time.

	$\psi = 1.5$								$\psi = 3$							
	$\tau = 0.3$				$\tau = 0.6$				$\tau = 0.3$				$\tau = 0.6$			
	EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP
MAXIMUM LIKELIHOOD																
β_{10}	0.005	0.141	0.143	94.9	<0.001	0.126	0.124	95.4	0.003	0.140	0.137	94.3	-0.002	0.128	0.128	94.5
β_{11}	0.017	0.168	0.167	94.9	0.014	0.164	0.162	94.1	0.011	0.158	0.160	96.4	0.010	0.156	0.156	95.3
β_{20}	0.010	0.139	0.142	96.8	0.009	0.128	0.135	97.0	0.004	0.138	0.141	96.6	-0.004	0.126	0.126	96.2
β_{21}	0.015	0.158	0.159	95.3	0.012	0.153	0.157	95.9	0.021	0.156	0.157	95.3	0.020	0.153	0.156	95.0
$\log \psi$	0.002	0.205	0.206	95.0	-0.020	0.224	0.219	95.7	0.016	0.204	0.205	96.4	-0.008	0.205	0.206	94.8
$\log \lambda_{11}$	-0.013	0.116	0.114	93.1	-0.011	0.102	0.105	95.0	-0.011	0.115	0.112	93.4	-0.010	0.104	0.105	94.6
$\log \lambda_{12}$	-0.014	0.117	0.115	94.1	-0.011	0.108	0.104	93.7	-0.021	0.117	0.114	93.5	-0.016	0.106	0.105	94.5
$\log \lambda_{13}$	-0.002	0.186	0.188	96.3	<0.001	0.156	0.153	94.8	0.001	0.172	0.174	95.1	0.004	0.145	0.142	93.2
$\log \lambda_{21}$	-0.009	0.109	0.111	95.2	-0.011	0.099	0.100	95.3	-0.007	0.113	0.112	93.9	-0.010	0.102	0.100	93.9
$\log \lambda_{22}$	-0.015	0.119	0.118	94.7	-0.007	0.106	0.104	94.9	-0.021	0.115	0.116	96.4	-0.011	0.099	0.102	96.7
$\log \lambda_{23}$	-0.007	0.175	0.173	93.9	-0.012	0.153	0.151	93.9	-0.010	0.169	0.167	93.2	-0.006	0.143	0.146	95.0
$\log \phi$	<0.001	0.183	0.185	96.1	0.005	0.114	0.118	95.7	0.002	0.182	0.182	94.0	0.012	0.113	0.115	95.3
WEIGHTED ESTIMATING EQUATIONS																
β_{10}	0.006	0.145	0.144	95.0	< 0.001	0.128	0.130	95.4	0.003	0.140	0.140	96.2	-0.002	0.129	0.128	94.6
β_{11}	0.017	0.168	0.163	94.4	0.014	0.165	0.157	93.6	0.011	0.159	0.159	95.6	0.010	0.156	0.156	95.2
β_{20}	0.008	0.137	0.136	96.0	0.007	0.129	0.132	96.6	0.003	0.138	0.142	96.6	0.004	0.126	0.128	96.2
β_{21}	0.014	0.157	0.162	97.0	0.012	0.153	0.157	96.2	0.022	0.156	0.159	95.4	0.020	0.153	0.157	95.4
$\log \psi$	0.002	0.206	0.200	94.8	-0.017	0.226	0.215	95.8	0.015	0.205	0.201	94.4	-0.007	0.207	0.202	95.6
$\log \lambda_{11}$	-0.014	0.121	0.123	93.4	-0.012	0.112	0.107	94.4	-0.013	0.120	0.120	93.2	-0.011	0.115	0.117	93.4
$\log \lambda_{12}$	-0.016	0.118	0.120	94.6	-0.012	0.112	0.113	95.0	-0.021	0.118	0.119	95.8	-0.019	0.113	0.114	95.6
$\log \lambda_{13}$	-0.004	0.191	0.188	94.2	0.001	0.164	0.164	94.2	0.003	0.172	0.174	94.2	0.006	0.148	0.150	95.8
$\log \lambda_{21}$	-0.010	0.114	0.114	93.8	-0.014	0.111	0.108	94.8	-0.009	0.117	0.114	93.8	-0.010	0.111	0.107	93.0
$\log \lambda_{22}$	-0.014	0.120	0.121	95.0	-0.006	0.111	0.113	96.4	-0.021	0.117	0.119	96.6	-0.012	0.104	0.103	96.7
$\log \lambda_{23}$	-0.003	0.173	0.174	95.0	-0.007	0.159	0.157	94.2	-0.008	0.171	0.176	94.4	-0.005	0.147	0.151	95.2
$\log \phi$	-0.011	0.184	0.185	96.2	-0.009	0.115	0.115	95.4	-0.006	0.183	0.179	95.8	0.002	0.112	0.111	95.4
TWO-STAGE																
β_{10}	0.023	0.185	0.183	95.2	0.023	0.185	0.184	95.0	0.017	0.169	0.170	96.7	0.018	0.170	0.170	96.4
β_{11}	0.026	0.184	0.184	94.4	0.026	0.184	0.184	94.6	0.018	0.161	0.167	96.1	0.017	0.162	0.167	96.0
β_{20}	0.014	0.155	0.157	96.6	0.022	0.170	0.176	96.0	0.013	0.157	0.159	96.5	0.015	0.172	0.175	96.6
β_{21}	0.018	0.160	0.157	94.6	0.019	0.159	0.162	96.6	0.027	0.165	0.169	95.3	0.027	0.165	0.171	94.4
$\log \psi$	0.001	0.205	0.205	95.8	-0.040	0.270	0.261	96.6	0.023	0.213	0.216	96.7	-0.011	0.221	0.221	96.0
$\log \lambda_{11}$	-0.021	0.131	0.127	94.6	-0.021	0.131	0.125	93.6	-0.020	0.131	0.132	95.3	-0.019	0.128	0.123	94.4
$\log \lambda_{12}$	-0.024	0.128	0.131	96.0	-0.024	0.128	0.132	95.8	-0.029	0.127	0.123	93.7	-0.030	0.128	0.132	95.6
$\log \lambda_{13}$	-0.020	0.223	0.215	93.8	-0.020	0.223	0.217	93.8	-0.012	0.208	0.205	94.9	-0.013	0.209	0.216	94.6
$\log \lambda_{21}$	-0.013	0.123	0.123	95.4	-0.019	0.124	0.126	95.0	-0.013	0.124	0.123	95.1	-0.018	0.125	0.125	95.0
$\log \lambda_{22}$	-0.017	0.125	0.132	95.8	-0.013	0.127	0.132	96.4	-0.026	0.126	0.123	93.6	-0.022	0.117	0.132	96.8
$\log \lambda_{23}$	-0.008	0.194	0.205	95.8	-0.017	0.214	0.219	96.4	-0.016	0.202	0.206	96.7	-0.020	0.216	0.217	95.2
$\log \phi$	-0.029	0.190	0.203	96.6	-0.042	0.136	0.151	96.0	-0.026	0.185	0.195	96.9	-0.034	0.120	0.138	96.6

Table 2: Empirical performance of estimators for 500 simulations with 1000 individuals per simulation under a three-piece piecewise constant hazards bivariate mixture model with $E\{N_i(A)\} = 20$ where $A = 1$ is the administrative censoring time.

	$\psi = 1.5$								$\psi = 3$							
	$\tau = 0.3$				$\tau = 0.6$				$\tau = 0.3$				$\tau = 0.6$			
	EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP
MAXIMUM LIKELIHOOD																
β_{10}	0.012	0.133	0.137	95.9	0.007	0.121	0.124	95.2	0.014	0.127	0.128	95.3	0.011	0.118	0.120	94.3
β_{11}	0.007	0.155	0.157	95.2	0.006	0.152	0.153	95.3	0.005	0.148	0.147	94.4	0.004	0.144	0.142	94.0
β_{20}	0.013	0.130	0.134	96.2	0.013	0.123	0.123	93.9	0.008	0.127	0.125	92.6	0.003	0.118	0.119	95.3
β_{21}	0.008	0.153	0.153	95.4	0.009	0.153	0.153	95.9	0.016	0.152	0.156	97.0	0.016	0.152	0.153	96.0
$\log \psi$	0.003	0.180	0.181	96.6	-0.013	0.192	0.194	95.0	0.024	0.185	0.187	97.0	0.007	0.188	0.189	96.7
$\log \lambda_{11}$	-0.008	0.098	0.102	94.8	-0.007	0.090	0.089	93.2	-0.005	0.094	0.096	96.7	-0.003	0.088	0.090	95.9
$\log \lambda_{12}$	-0.013	0.106	0.109	96.2	-0.010	0.093	0.095	94.1	-0.010	0.099	0.097	93.3	-0.008	0.091	0.093	95.3
$\log \lambda_{13}$	-0.011	0.162	0.164	95.3	-0.008	0.139	0.136	93.7	-0.010	0.152	0.151	94.2	-0.009	0.130	0.129	95.0
$\log \lambda_{21}$	-0.008	0.097	0.099	95.0	-0.006	0.087	0.087	93.8	-0.012	0.094	0.092	94.0	-0.006	0.088	0.090	96.1
$\log \lambda_{22}$	-0.005	0.104	0.102	94.1	-0.004	0.092	0.092	94.7	-0.006	0.105	0.104	95.1	-0.006	0.092	0.092	95.3
$\log \lambda_{23}$	-0.008	0.153	0.154	94.7	-0.012	0.136	0.136	95.5	-0.018	0.150	0.152	95.2	-0.012	0.127	0.130	95.8
$\log \phi$	-0.003	0.150	0.151	96.7	0.006	0.096	0.098	95.3	<0.001	0.145	0.150	96.7	0.005	0.092	0.092	95.3
WEIGHTED ESTIMATING EQUATIONS																
β_{10}	0.009	0.133	0.132	95.8	0.006	0.123	0.122	95.4	0.012	0.128	0.128	96.4	0.008	0.119	0.119	95.7
β_{11}	0.007	0.154	0.155	95.6	0.005	0.152	0.153	95.2	0.006	0.147	0.154	95.8	0.004	0.145	0.152	96.1
β_{20}	0.013	0.131	0.130	95.6	0.014	0.127	0.122	94.2	0.007	0.128	0.129	96.0	0.002	0.119	0.119	95.9
β_{21}	0.009	0.155	0.155	96.8	0.010	0.154	0.153	96.6	0.016	0.151	0.155	96.8	0.015	0.152	0.152	95.9
$\log \psi$	0.004	0.181	0.183	95.2	-0.012	0.194	0.196	96.4	0.022	0.186	0.190	97.0	0.008	0.188	0.189	94.9
$\log \lambda_{11}$	-0.006	0.101	0.098	94.6	-0.005	0.097	0.093	95.0	-0.005	0.097	0.097	94.5	-0.004	0.094	0.092	94.3
$\log \lambda_{12}$	-0.011	0.107	0.105	94.4	-0.009	0.101	0.098	94.0	-0.009	0.102	0.103	94.6	-0.007	0.097	0.097	94.5
$\log \lambda_{13}$	-0.008	0.161	0.157	93.2	-0.006	0.142	0.137	94.2	-0.007	0.152	0.151	94.0	-0.005	0.135	0.131	94.7
$\log \lambda_{21}$	-0.008	0.102	0.097	93.6	-0.008	0.096	0.093	93.4	-0.012	0.097	0.097	94.6	-0.007	0.095	0.092	95.3
$\log \lambda_{22}$	-0.005	0.106	0.104	95.8	-0.007	0.099	0.098	94.2	-0.002	0.107	0.103	93.8	-0.005	0.099	0.097	95.5
$\log \lambda_{23}$	-0.008	0.157	0.155	93.6	-0.013	0.145	0.136	93.4	-0.017	0.152	0.152	93.2	-0.009	0.130	0.132	95.5
$\log \phi$	-0.010	0.150	0.152	95.4	-0.005	0.097	0.096	95.6	-0.007	0.146	0.149	96.4	-0.003	0.093	0.093	95.7
TWO-STAGE																
β_{10}	0.011	0.145	0.143	96.4	0.011	0.146	0.144	96.0	0.021	0.147	0.146	96.8	0.021	0.147	0.144	96.4
β_{11}	0.008	0.156	0.158	96.4	0.008	0.156	0.158	96.2	0.008	0.153	0.158	96.8	0.008	0.153	0.159	96.8
β_{20}	0.021	0.147	0.146	95.6	0.032	0.167	0.163	95.0	0.016	0.147	0.147	96.0	0.010	0.141	0.144	96.2
β_{21}	0.011	0.158	0.160	96.8	0.016	0.161	0.163	96.8	0.018	0.153	0.156	97.0	0.016	0.154	0.158	96.4
$\log \psi$	-0.002	0.181	0.190	95.6	-0.029	0.218	0.226	96.8	0.029	0.190	0.204	96.8	0.002	0.196	0.210	95.6
$\log \lambda_{11}$	-0.007	0.105	0.103	95.4	-0.007	0.105	0.104	95.6	-0.009	0.101	0.104	95.0	-0.009	0.101	0.104	94.4
$\log \lambda_{12}$	-0.011	0.113	0.111	95.0	-0.011	0.113	0.112	95.4	-0.013	0.111	0.111	95.4	-0.013	0.111	0.112	94.4
$\log \lambda_{13}$	-0.007	0.175	0.175	94.6	-0.007	0.176	0.178	95.2	-0.014	0.177	0.177	94.4	-0.014	0.178	0.177	94.6
$\log \lambda_{21}$	-0.011	0.110	0.105	94.0	-0.015	0.108	0.106	93.6	-0.016	0.105	0.104	95.2	-0.010	0.104	0.104	95.2
$\log \lambda_{22}$	-0.009	0.113	0.113	95.4	-0.016	0.116	0.115	95.4	-0.006	0.117	0.112	94.0	-0.009	0.109	0.112	95.4
$\log \lambda_{23}$	-0.014	0.180	0.178	95.4	-0.029	0.197	0.191	94.0	-0.024	0.180	0.179	96.4	-0.011	0.172	0.178	96.0
$\log \phi$	-0.021	0.150	0.153	97.4	-0.023	0.189	0.182	95.6	-0.019	0.148	0.155	96.7	-0.020	0.157	0.162	95.8

4.2 CO-OCCURRENCE OF DAMAGE IN THE EXTREMITIES IN PSORIATIC ARTHRITIS

We consider data from $n = 657$ patients, each assessed at multiple visits to determine the level of radiological damage in the two disease processes (i.e. hands and feet). Every joint is considered to be at state 0 at the time of disease onset. In the regression model we include sex as a covariate with male as the reference level. After inspecting the quantiles of the distribution of the assessment times, we consider a piecewise-constant hazards model with a single cut-point at 5 years post-onset so the hazard is constant over the interval $[0, 5)$ and $[5, \infty)$. We apply maximum likelihood, two-stage estimation, and the estimating equations approach of Section 3 and report the results in Table 3.

Table 3: Estimates and standard errors from fitting a two-piece piecewise constant hazards bivariate mixture model for the onset of damage in the hands and feet based on the three methods of Section 3.

	ML		Two-Stage		WGEE	
	EST.	SE	EST.	SE	EST.	SE
<i>Parameters for $Z_i X_i$</i>						
β_{10}	1.622	0.191	1.656	0.203	1.518	0.184
β_{11}	-0.361	0.252	-0.367	0.258	-0.355	0.237
β_{20}	1.938	0.245	1.690	0.212	1.681	0.199
β_{21}	-0.559	0.281	-0.487	0.240	-0.492	0.241
$\log\psi$	2.133	0.351	2.051	0.314	2.023	0.294
<i>Parameters for $T_i Z_i, X_i$</i>						
$\log\lambda_{11}$	-0.924	0.199	-0.851	0.211	-0.778	0.215
$\log\lambda_{12}$	-1.520	0.080	-1.566	0.209	-1.512	0.117
$\log\lambda_{21}$	-0.855	0.184	-0.623	0.131	-0.618	0.124
$\log\lambda_{22}$	-1.719	0.088	-1.403	0.231	-1.503	0.126
$\log\phi$	0.487	0.186	0.444	0.237	0.495	0.163

The results in Table 3 show broad agreement between the three methods of analysis in terms of the magnitude of the effects. For the maximum likelihood method, there is a trend towards females having lower risk of being susceptible to damage in hand joints compared to males ($OR = \exp(-0.361) = 0.70$; 95% CI: 0.43, 1.14; $p = 0.151$), but this lower susceptibility for females is statistically significant for the foot joints ($OR = \exp(-0.559) = 0.57$; 95% CI: 0.32, 0.99; $p = 0.047$). Moreover, based on maximum likelihood estimates, the odds ratio characterizing the association between the susceptibility indicators for the hand and foot joints is strong at 8.44 with a 95% CI (3.23, 13.65) suggesting that those who develop damage in the hands are also at high risk of damage in the foot joints.

We now turn to the interpretation of the parameters of the failure time component of the model. From Table 3, we can see there is a lower hazard for damage onset after the first five years following disease onset. Moreover, among individuals who are susceptible to damage in both locations the association in the damage onset times in the hands and feet is estimated $\hat{\tau} = 0.45$ with a 95% CI (0.21, 0.69); thus the onset times of damage in the two extremities are moderately correlated.

To explore the extent and nature of the information in the association parameters in this dataset we plot the profile relative likelihood functions for ψ and ϕ in Figure 3 and give a contour plot of the profile relative likelihood for $(\psi, \phi)'$ in Figure 4. For the latter, for example, if we write $\theta = (\alpha'_1, \alpha'_2)'$

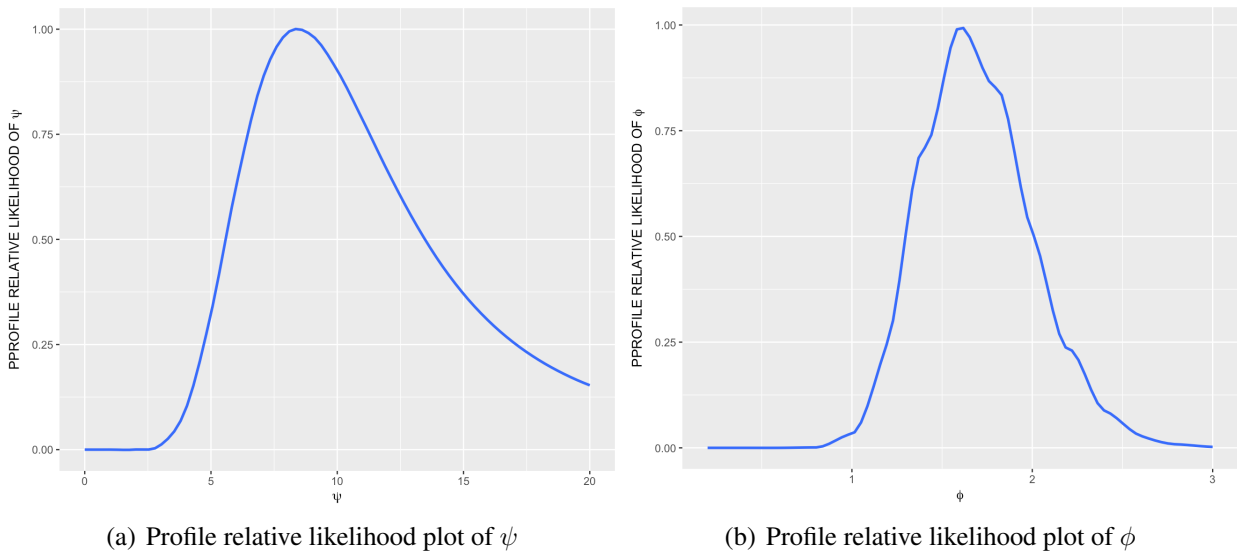


Figure 3: One-dimensional profile relative likelihood plots of ϕ and ψ for onset of damage in the hands and feet.

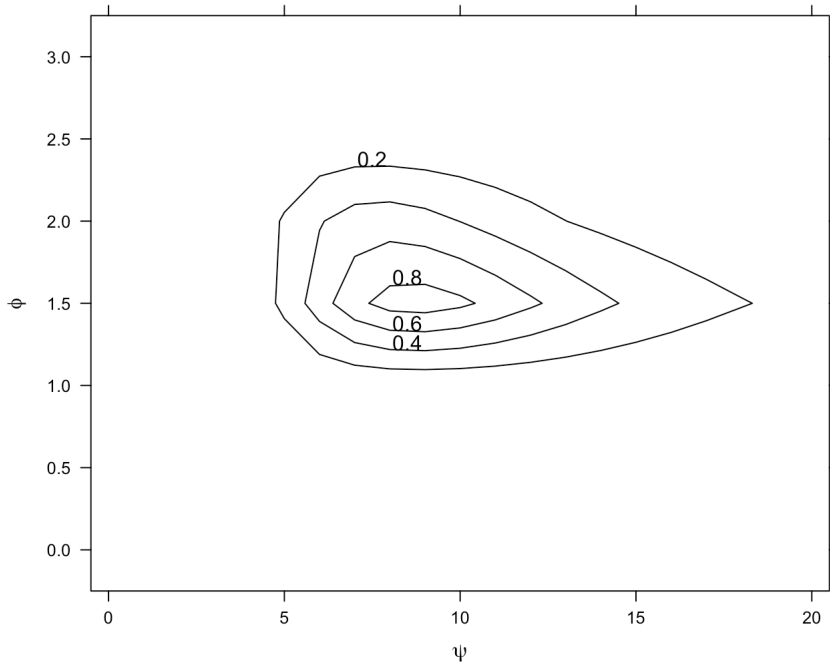


Figure 4: Two-dimensional profile relative likelihood contour plot of ϕ and ψ for onset of damage in the hands and feet.

as in Section 3.2 where $\alpha_1 = (\beta', \lambda)'$ and $\alpha_2 = (\gamma, \phi)'$ with $\gamma = \log \psi$, then if $\tilde{\alpha}_1(\alpha_2)$ is the value maximizing (4) for fixed α_2 , the profile relative likelihood is given by

$$PRL(\alpha_2) = \frac{L(\tilde{\alpha}_1(\alpha_2), \alpha_2)}{L(\hat{\alpha}_1, \hat{\alpha}_2)},$$

and the p -level relative likelihood contours are obtained by solving for $\alpha_2 = (\gamma, \phi)$ satisfying $PRL(\alpha_2) = p$ for $p = 0.20, 0.40, 0.60$ and 0.80 . The approach is similar for the single parameter profile relative likelihood plots where the functions can be plotted directly. We can see that despite

the fact that ψ measures the dependence between two latent variables, the (profile) likelihood is relatively well behaved for this parameter. Likewise, despite the fact that ϕ is relevant only for individuals who are susceptible for both interval-censored events, the profile likelihood suggests it is reasonable to base inference on ϕ or alternatively Kendall's τ . In summary, the contour of the profile likelihood for (ψ, ϕ) does not suggest an identifiability or estimability problem for these parameters and it appears that the dependence parameter estimates are not highly correlated.

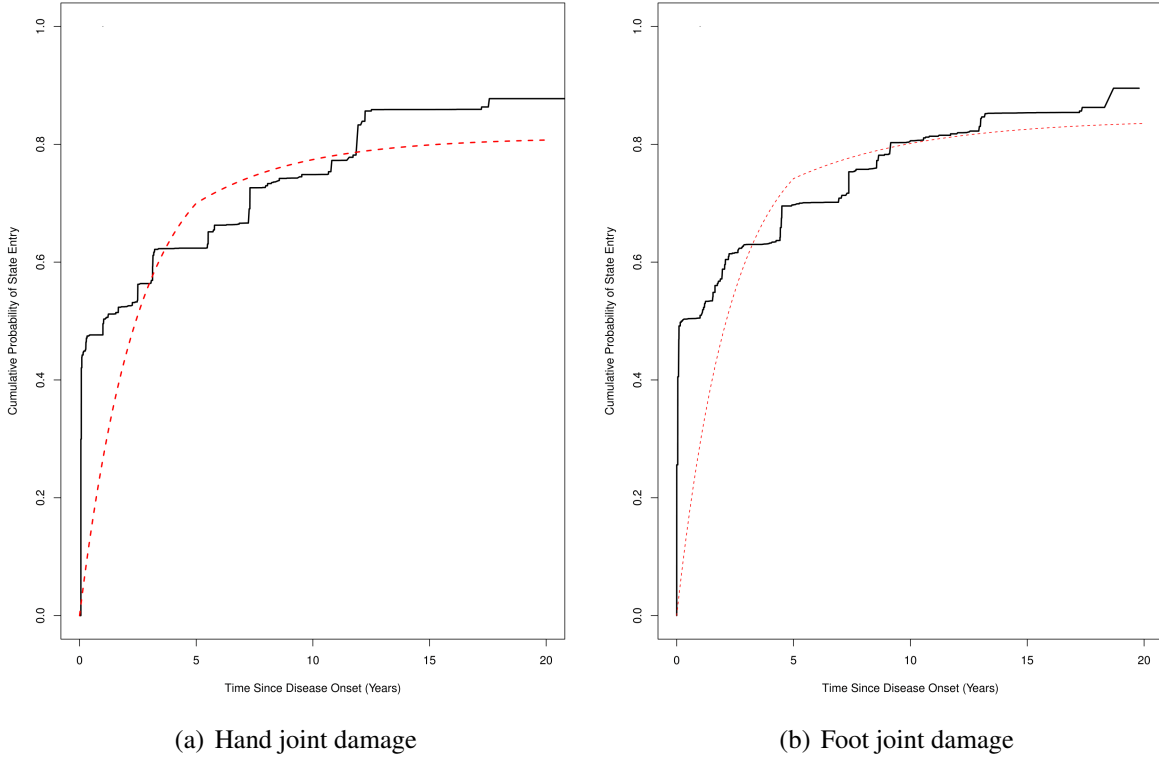


Figure 5: Estimates of the marginal probability of damage in the hands (panel (a)) and feet (panel (b)) joints based on the fitted model with a piecewise constant hazard along with a nonparametric estimate obtained by the pooled adjacent violators algorithm.

Finally to check to see if the marginal features of the fitted model are calibrated with the raw data, we plot the estimated cumulative distribution functions in Figure 5 from the fitted model with piecewise-constant hazards (averaging over the latent susceptibility indicator and the covariate gender) along with nonparametric estimates of the marginal onset time distributions for the hand and foot joints (Turnbull, 1976). The estimates of the marginal cumulative sub-distribution functions from the fitted model are obtained by noting

$$P(T_{ij} < t) = 1 - E_{X_i} \{1 - \mu_{ij} + \mu_{ij} \mathcal{F}_j(t)\}, \quad j = 1, 2,$$

evaluating the functions at the maximum likelihood estimates, and computing the expectation with respect to X_i based on the sample covariate distribution. That is, we plot

$$\widehat{P}(T_{ij} < t) = 1 - \frac{1}{n} \sum_{i=1}^n \{1 - \widehat{\mu}_{ij} + \widehat{\mu}_{ij} \widehat{\mathcal{F}}_j(t)\},$$

where $\widehat{\mu}_{ij} = \exp(\widehat{\eta}_{ij}) / (1 + \exp(\widehat{\eta}_{ij}))$ and $\widehat{\eta}_{ij} = \widehat{\beta}_{j0} + \widehat{\beta}_{j1} X_i$. Again, we see good agreement which we take to mean the model is offering a reasonable fit to the data.

5 SOME EXTENSIONS

5.1 DEALING WITH HIGHER DIMENSIONAL DATA

In some settings, interest may lie in modeling multivariate interval-censored data with large J while accommodating nonsusceptibility and studying the association between the marginal processes. This would be the case if the damage process was modeled at the individual joint level in the psoriatic arthritis data, for example. In this case, the multivariate binary model for the susceptibility indicators naturally extends, where one may let

$$\psi_{ijk} = \frac{P(Z_{ij} = 1, Z_{ik} = 1|X_i)P(Z_{ij} = 0, Z_{ik} = 0|X_i)}{P(Z_{ij} = 1, Z_{ik} = 0|X_i)P(Z_{ij} = 0, Z_{ik} = 1|X_i)}$$

denote the pairwise odds ratio for type j and k susceptibility indicators and omit higher orders of dependence. These odds ratios may be different for each possible pair j and k but we note that while there are several models that can be considered, there are constraints on the admissible odds ratios with multivariate binary data (Liang et al., 1992). A second order regression model may be specified by setting $\log \psi_{ijk} = v'_{ijk}\gamma$ where v_{ijk} is a vector of parameters associated with event types j and k for individual i , $i = 1, \dots, n$.

For the failure times, we can likewise let τ_{jk} denote Kendall's τ for the association between T_{ij} and T_{ik} given $Z_{ij} = Z_{ik} = 1$. A multivariate Gaussian copula function (Nelsen, 2006) is natural to use in this setting since under this model the pairwise dependencies may differ. The likelihood, two-stage and estimating function approaches of Section 3 may all be employed to deal with higher dimensional failure time models. The risk of misspecifying the dependence structure seems greater with a higher dimensional response, so the possible robustness of the two-stage and estimating function approaches make them appealing. In particular, if we let $\alpha_1 = (\beta', \lambda)'$ where now $\beta = (\beta'_1, \dots, \beta'_J)'$ is a $p_J \times 1$ vector and $\lambda = (\lambda'_1, \dots, \lambda'_J)'$ is a $\sum_{j=1}^J q_j \times 1$ vector, then (7) becomes

$$L_{i1}(\alpha_1) \propto \prod_{j=1}^J L_{ij1}(\alpha_1),$$

which can be maximized by separate maximization of the J functions. With a Gaussian copula, (8) will be more difficult to compute and the maximization will be over a higher dimension association parameter, but in principle the same steps can be carried out as in Section 3.2.

The estimating equation approach of Section 3.3 extends very naturally. In (9), the vector W_{i1} would consist of all $J!/((J-2)!2)$ pairwise products of the elements in Z_i and ω_{i1} would be of the same dimension. The dimensions of $H_{i1}(\theta_1)$ and $\Sigma_{i1}(\theta_1)$ would likewise increase but the basic structure of (9) covers the higher dimensional case, and (10) is modified naturally as well.

5.2 ANALYSIS OF BIVARIATE CURRENT STATUS DATA

Current status data represents a special case of the intermittent observation scheme of Section 3 where individuals are examined only once, yielding disease onset times that are either left- or right-censored. Here, we let C_{ij} denote the examination time for individual i for event type j and suppose we observe $Y_{ij} = I(t_{ij} < C_{ij})$, $j = 1, 2$. The data from a sample of size n then takes the form $\{(C_{ij}, Y_{ij}), j = 1, 2, X_i, i = 1, \dots, n\}$. In the absence of a nonsusceptible subpopulation, the pooled adjacent violators algorithm (PAVA) may be used to obtain the maximum likelihood estimate of the survivor function or onset time distribution (Ayer et al., 1955). Lam and Xue (2005) and Cook et al. (2008a) describe an EM algorithm for fitting mixture models accommodating a nonsusceptible fraction to univariate current status data. Here, we consider the use of this algorithm in the two-stage approach of Section 3.2; in stage 1 we use the PAVA in an EM algorithm in order to estimate

$\mathcal{F}_j(t) = P(T_j \geq t | Z_{ij} = 1)$ and β_j in model for $P(Z_{ij} = 1 | X_i; \beta_j)$, $j = 1, 2$, and in stage 2 the association parameters are estimated. The details of the algorithm are as follows.

Let $\alpha_1 = (\beta', \mathcal{F}_j(\cdot), j = 1, 2)'$ denote the parameters which are estimated in the first stage as before, where here this includes the infinite dimensional survivor function estimates for the two failure times. The complete date composite (independence) likelihood in stage 1, treating Z_i as observed, can be written

$$\mathcal{L}_1(\alpha_1) \propto \prod_{i=1}^n \prod_{j=1}^2 ([1 - \mathcal{F}_j(C_{ij})] \mu_{ij})^{z_{ij} y_{ij}} ([\mathcal{F}_j(C_{ij}) \mu_{ij}]^{z_{ij}} [1 - \mu_{ij}]^{1-z_{ij}})^{1-y_{ij}}$$

giving

$$\begin{aligned} \log \mathcal{L}_1(\alpha_1) = \sum_{i=1}^n \sum_{j=1}^2 \left\{ z_{ij} \log \mu_{ij} + (1 - z_{ij}) \log(1 - \mu_{ij}) \right. \\ \left. + z_{ij} [y_{ij} \log(1 - \mathcal{F}_j(C_{ij})) + (1 - y_{ij}) \log \mathcal{F}_j(C_{ij})] \right\}. \end{aligned} \quad (12)$$

With α_1^r the estimates at the r th iteration of the EM algorithm, we define

$$\begin{aligned} Q_{1ij}(\beta_j; \alpha_1^r) &= \eta_{ij}^r \log \mu_{ij} + (1 - \eta_{ij}^r) \log(1 - \mu_{ij}) \\ Q_{2ij}(\mathcal{F}_j(\cdot); \alpha_1^r) &= \eta_{ij}^r [y_{ij} \log(1 - \mathcal{F}_j(C_{ij})) + (1 - y_{ij}) \log \mathcal{F}_j(C_{ij})] \end{aligned}$$

as the elementary contributions to the functions to be optimized, where

$$\eta_{ij}^r = P(Z_{ij} = 1 | Y_{ij} = 0, D_i; \alpha_1^r) = \frac{[\mathcal{F}_j^r(C_{ij})] \mu_{ij}^r}{[\mathcal{F}_j^r(C_{ij})] \mu_{ij}^r + (1 - \mu_{ij}^r)}, \quad j = 1, 2.$$

To consider nonparametric estimation of the survivor functions we let $C_{(j1)} < C_{(j2)} < \dots < C_{(jK_j)}$ denote the K_j unique ordered inspection times for the type j event, and $d_{jk} = \sum_{i=1}^n I(C_{ij} = C_{(jk)}) Y_{ij}$ denote the total number of individuals found to be positive at $C_{(jk)}$. The PAVA can be used to optimize

$$Q_{2j}(\mathcal{F}_j(\cdot); \alpha_1^r) = \sum_{i=1}^n Q_{2ij}(\mathcal{F}_j(\cdot); \alpha_1^r)$$

at each step of an EM algorithm. Specifically, the number of individuals estimated to be at risk at the k th inspection time for process j at the r th iteration is estimated to be $\hat{\Delta}_{jk}^r = \sum_{i=1}^n I(C_{ij} = C_{(jk)}) (Y_{ij} + (1 - Y_{ij}) \eta_{ij}^r)$. An isotonic regression of $(d_{j1}/\hat{\Delta}_{j1}^r, \dots, d_{jK}/\hat{\Delta}_{jK}^r)'$ with weights $(\hat{\Delta}_{j1}^r, \dots, \hat{\Delta}_{jK}^r)'$ gives the updated estimate

$$\hat{\mathcal{F}}_j^{(r+1)}(C_{(jk)}) = \max_{u \leq k} \min_{v \geq k} \left(\frac{\sum_{l=u}^v \sum_{i=1}^n d_{jk}}{\sum_{l=u}^v \hat{\Delta}_{jk}^r} \right).$$

Upon obtaining estimates from all component marginal distributions in stage 1, the estimates are substituted into (4) as before, and estimates of the association parameters can be obtained by maximizing the version of (8) based on the Gaussian (or other multivariate) copula. Such methods can be useful in the study of the joint susceptibility and co-development of two or more chronic diseases using baseline data from large cohort studies. Incorporation of survey weights may also be warranted if complex sampling designs are employed for selection of individuals.

6 DISCUSSION

In this paper, we have developed flexible methods for modeling bivariate interval-censored data which accommodates the possibility that some individuals will be nonsusceptible to one or more of the conditions of interest. This framework enables one to decompose dependence measures into one component for susceptibility and one component for the dependence in the failure times given joint susceptibility. There are numerous applications where this framework can yield useful insight into disease processes. In diabetes, some individuals do not experience significant complications, some develop nephropathy, some develop retinopathy, and some develop both. In settings where genes play a role in susceptibility for the different complications the mover-stayer formulation seems appropriate, particularly when assessment and censoring times are random and making definitive classification of individuals difficult. The model may also be used in the analysis of data from family studies when interest lies in modeling within-family dependence to gain insight into the genetic basis for disease (Laird and Lange, 2006). Here, it may be quite natural to examine the effect of genetic markers on the susceptibility indicators and model the aggregation in terms of odds ratios (Liang and Beaty, 1991).

There are several directions for further development of the proposed model and approaches to estimation and inference. First, one could consider estimating the joint densities of the failure times using local likelihood (Hjort and Jones, 1996; Li et al., 1997). Braun and Stafford (2016) describe how to obtain smooth estimates of multivariate failure time distributions based on interval-censored data; this approach will not yield parsimonious measures of dependence between the marginal failure times, but the density estimates can be used to estimate any desired features of the joint distribution. Alternatively, smoothed estimates of baseline hazard functions could be obtained via local likelihood (Betensky et al., 1999). This approach would be most easily done using a two-stage estimation procedure of Section 3.2 and could be adopted for estimation of the baseline hazard if covariates were considered in the failure time model component.

We introduced the idea of extending the model and methods to deal with multivariate processes and these extensions may be useful for settings where the components have different marginal distributions as described here, but also for the setting where the marginal distributions are the same but dependence arises because of clustering. These models can be fitted in the likelihood approach by introducing constraints (i.e. $\beta_j = \beta$ and $\mathcal{F}_j(t) = \mathcal{F}(t)$, $j = 1, \dots, J$). The two-state estimation procedure could likewise be adapted by imposing these constraints in the first stage but more involved adaptations would be required for the estimating function approach. These modifications, however, are relatively straightforward and the subject of ongoing research with a view to application to modeling the onset of damage in individual joints in individuals affected with arthritis. Finally, we have indicated that we aim to model covariate effects on the latent susceptibility indicators but model the onset time distribution to deal with the variable duration of follow-up of individuals in the motivating study. If there is interest in modeling covariate effects on the failure times, parametric or semiparametric proportional hazards models can be considered for the failure times given susceptibility indicators; having several options for the framework for estimation is appealing given each such extension will raise different computational and inferential challenges.

ACKNOWLEDGEMENTS

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN 155849 and 04207) and the Canadian Institutes for Health Research (FRN 13887). Richard Cook is a Mathematics Faculty Research Chair at the University of Waterloo. The authors thank Drs. Dafna Gladman and Vinod Chandran for helpful discussions regarding the research at the Centre for Prognosis Studies in Rheumatic Disease at the University of Toronto.

REFERENCES

- Al-Kateb, H., Boright, A., Mirea, L., Xie, X., Sutradhar, R., Mowjoodi, A., Bharaj, B., Liu, M., Kucksa, J., Arends, V., Steffes, M., Cleary, P., Sun, W., Lachin, J., Thorner, P., Ho, M., KcKnight, A., Maxwell, A., Savage, D., Kidd, K., Kidd, J., Speed, W., Orchard, T., Miller, R., Sun, L., Bull, S., Paterson, A., and the Diabetes Control and Complications Trial/ Epidemiology of Diabetes Interventions and Complications Research Group (2008). Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: The diabetes control and complications trial/epidemiology of diabetes interventions and complications genetics study. *Diabetes*, 57:218–228.
- AREDS Group (1999). The age-related eye disease study (AREDS): Design implications AREDS report no. 1. *Controlled Clinical Trials*, 20:573.
- Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *The Annual of Mathematical and Statistics*, 26:641 – 647.
- Betensky, R. and Finkelstein, D. (1999). A non-parametric maximum likelihood estimator for bivariate interval censored data. *Statistics in Medicine*, 18:3089 – 3100.
- Betensky, R. A., Lindsey, J. C., Ryan, L. M., and Wand, M. P. (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics*, 55:238 – 245.
- Boos, D. D. and Stefanski, L. (2013). *Essential Statistical Inference: Theory and Methods*. Springer.
- Braun, W. J. and Stafford, J. E. (2016). Multivariate density estimation for interval-censored data with application to a forest fire modelling problem. *Environmetrics*, 27:345 – 354.
- Broström, G. (2012). *Event History Analysis with R*. CRC Press.
- Chatterjee, N. and Shih, J. (2001). A bivariate cure-mixture approach for modeling familial association in diseases. *Biometrics*, 57:779 – 786.
- Cook, R. J., White, B. J., Yi, G. Y., and Lee, K. A. (2008a). Analysis of a nonsusceptible fraction with current status data. *Statistics in Medicine*, 27:2715 – 2730.
- Cook, R. J., Zeng, L., and Lee, K. A. (2008b). A multistate model for bivariate interval-censored failure time data. *Biometrics*, 64:1100 – 1109.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39:1 – 38.
- Farewell, V. T. (1977). A model for a binary variable with time censored observations. *Biometrika*, 64:43 – 46.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041 – 1046.
- Friedman, M. et al. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, 10(1):101–113.
- Gladman, D. and Urowitz, M. (1987). Morbidity in systemic lupus erythematosus. *The Journal of Rheumatology*, Supplement 14:223–226.

- Grüger, J., Kay, R., and Schumacher, M. (1991). The validity of inferences based on incomplete observations in disease state models. *Biometrics*, 47:595–605.
- Hjort, N. and Jones, M. (1996). Locally parametric nonparametric density estimation. *Annals of Statistics*, 24:1619 – 1647.
- Hogan, J. W., Roy, J., and Korkontzelou, C. (2004). Handling drop-out in longitudinal studies. *Statistics in Medicine*, 23:1455 – 1497.
- Kim, M. Y. and Xue, X. (2002). The analysis of multivariate interval-censored survival data. *Statistics in Medicine*, 21:3715 – 3726.
- Kim, Y.-J. (2016). Cure rate model with bivariate interval censored data. *Communications in Statistics – Simulation and Computation*, 46:7116–7124.
- Kor, C., Cheng, K., and Chen, Y. (2013). A method for analyzing clustered interval-censored data based on cox’s model. *Statistics in Medicine*, 32:822–832.
- Laird, N. and Lange, C. (2006). Family-based designs in the age of large-scale gene association studies. *Nature Reviews Genetics*, 7:385–394.
- Lam, K. F. and Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika*, 92:573 – 586.
- Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*, 2nd edition. John Wiley and Sons, Hoboken.
- Le, C. and Lindgren, B. (1996). Duration of ventilating tubes: A test for comparing two clustered samples of censored data. *Biometrics*, 52:328–334.
- Li, L., Watkins, T., and Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, 24:531 – 542.
- Liang, K. Y. and Beaty, T. (1991). Measuring familial aggregation by using odds-ratio regression models. *Genetic Epidemiology*, 8:361–370.
- Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54:3 – 40.
- Lipsitz, S. R. and Laird, N. M. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, 78:153 – 160.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*, 44:226–233.
- Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer, New York.
- Peng, T. (2003). Fitting semiparametric cure models. *Computational Statistics and Data Analysis*, 41:481 – 490.
- Peng, Y. and Dear, K. B. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56(1):237–243.
- Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, 47:825–839.

- Rahman, P., Gladman, D. D., Cook, R. J., Zhou, Y., Young, G., and Salonen, D. (1998). Radiological assessment in psoriatic arthritis. *British Journal of Rheumatology*, 37:760 – 765.
- Shih, J. and Louis, T. A. (1995). Inference on the association parameter in copula models for bivariate survival data. *Biometrics*, 51:1384 – 1399.
- Sun, L., Wang, L., and Sun, J. (2006). Estimation of the association for bivariate interval-censored failure time data. *Scandinavian Journal of Statistics*, 33:637 – 649.
- Sun, T., Liu, Y., Cook, R., Chen, W., and Ding, Y. (2019). Copula-based score test for bivariate time-to-event data, with application to a genetic study and amd progression. *Lifetime Data Analysis*, 25:546–568.
- Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox proportional hazards cure model. *Biometrics*, 56:227 – 236.
- Taylor, J. M. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, pages 899–907.
- Tolusso, D. and Cook, R. J. (2009). Second-order estimating equations for the analysis of clustered current status data. *Biostatistics*, 10:756 – 772.
- Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 38(3):290–295.
- Wang, W. and Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika*, 87:879 – 893.
- Zhong, Y. and Cook, R. (2016). Augmented composite likelihood for copula modeling in family studies under biased sampling. *Biostatistics*, 17:437–452.