

# Dependence modeling for multi-type recurrent events via copulas

JOOYOUNG LEE

*Department of Statistics and Actuarial Science,  
University of Waterloo, Waterloo, ON, N2L 3G1, Canada  
E-mail: j463lee@uwaterloo.ca*

RICHARD J. COOK

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

## Summary

When several types of recurrent events may arise, interest often lies in marginal modeling and studying the nature of the dependence structure. In this paper, we propose a multivariate mixed-Poisson model with the dependence between events accommodated by type-specific random effects which are associated through use of a Gaussian copula. Such models retain marginal features with a simple interpretation, reflect the heterogeneity in risk for each type of event, and provide insight into the dependence between the different types of events. Semiparametric inference is proposed based on composite likelihood to avoid high dimensional integration. An application to a study of nutritional supplements in malnourished children is given in which the goal is to evaluate the reduction in the rate of several different kinds of infection.

*Keywords:* composite likelihood, copula, EM algorithm, frailty, multi-type recurrent events

**This is the peer reviewed version of the following article: Jooyoung Lee and Richard J. Cook, Dependence modeling for multi-type recurrent events via copulas, *Statistics in Medicine* (2019), 38(21): 4066–4082 which has been published in final form at <https://doi.org/10.1002/sim.8283>.**

## 1 INTRODUCTION

### 1.1 OVERVIEW

In many chronic diseases individuals are at risk of several distinct types of potentially recurring events. For example, individuals with asthma are at risk of different types of recurrent exacerbations which are distinguished by the results of cellular analysis of sputum samples (Jayaram et al., 2006). Individuals with the autoimmune disease systemic lupus erythematosus experience recurrent flares of symptoms which may arise in respiratory, neurological, digestive, and other organ systems (Gladman et al., 2002). In neurological studies, recurrent headaches can be sub-classified according to formal criteria to better understand the nature of the underlying chronic condition (Silberstein et al., 2005). Lastly in public health studies in developing countries, children are at risk of different kinds of recurrent

infections (Lemaire et al., 2011) with the different infections leading to different risks for long-term impairment and morbidities. A natural goal in all such settings is to carry out a marginal analysis by estimating the rate of onset for each type of infection. Robust variance estimates can often be obtained to account for the dependence between the different types of events if global statements are desired about covariate effects. However, when the different types of events arise due to an underlying condition (e.g. a compromised immune system due to malnutrition, exposure to contaminated areas, etc.) it can be helpful to better understand the nature of the association between the different types of events. This can be achieved by joint modeling of a multivariate recurrent event process.

Ng and Cook (1999) consider optimal estimating functions for dependence modeling and study the mixed Poisson setting as a special case in the spirit of Nadeau and Lawless (1998). Cai and Schaubel (2004) developed semi-parametric marginal models for multi-type recurrent events data with a view to robust inference through use of sandwich variance estimates. Chen et al. (2012) developed additive marginal models, proved the consistency of estimators, and derived their asymptotic distribution under a working independence assumption. Cook et al. (2010) proposed a bivariate mixed Poisson model with a view to studying the impact of event dependent censoring. Mazroui et al. (2013) also considered bivariate frailty models for two types of recurrent events and death based on a piecewise-constant baseline hazard; methods of estimator and inference based on likelihood and penalized likelihood were developed. While the methods and asymptotic theory developed in these papers are typically given for the general setting of more than two types of events, only two distinct types of events are considered most often in applications. Frequentist analyses based on flexible multivariate frailty models can be challenging to implement when there are more than two types of events, particularly for semiparametric methods.

We propose a model based on a mixed-Poisson formulation and propose methods for fitting and inference based on multivariate random effect distributions constructed via copula functions. This structure is appealing in that it enables separate modeling of heterogeneity and dependence and offers a natural basis for composite (e.g. “pairwise”) likelihood methods and consequent avoidance of the computational burden of full likelihood based inference. We also investigate an even more computationally convenient two-stage estimation procedure based on pairwise likelihood in which marginal models are fitted for each type of event at the first stage, and the dependence parameters are estimated at the second stage. Large sample theory is developed for both of these approaches.

The remainder of the paper is organized as follows. In the next sub-Section we provide a brief review of composite likelihood. In Section 2 we introduce notation, provide details on the model formulation, and give the full and composite likelihoods. An expectation-maximization algorithm is given in Section 3 for semiparametric analysis based on multiplicative rate function models; details on variance estimation are given in Section 3 and the Appendix A. Section 4 reports on simulation studies investigating the finite sample properties of the simultaneous and two-stage estimation procedure based on pairwise composite likelihood and an application is given in Section 5 on a motivating study on the effect of iron supplementation on the occurrence of different types of infections in malnourished children. Concluding remarks are made in Section 6.

## 1.2 REVIEW OF COMPOSITE LIKELIHOOD

We let  $\theta$  denote a  $p \times 1$  parameter of interest. In modeling multivariate data or in other settings involving complex dependence structures, the full likelihood may be complex or too computationally demanding to work with. As an alternative to full likelihood, Lindsay (1988) propose using a composite likelihood defined as

$$CL(\theta; y) = \prod_{j=1}^J L_j(\theta; y)^{w_j}, \quad (1)$$

a weighted product of marginal or conditional likelihood contributions  $L_j(\theta; y)$ ; this may be viewed as an extension of the concept of pseudo-likelihood (Besag, 1974). Each term  $L_j(\theta; y)$  is determined by the selection of  $\{A_1, \dots, A_J\}$ , a set of marginal or conditional events (Varin et al., 2011), where  $L_j(\theta; y) \propto f(y \in A_j; \theta)$ . Varin (2008) provided the excellent review of composite likelihood in different fields and classified composite likelihood contributions as based on conditional or marginal likelihoods. Composite conditional likelihoods are based on the product of conditional densities given conditions which the analyst specifies, whereas the latter is constructed from marginal densities.

As in ordinary likelihood, the composite likelihood score equations are based on unbiased estimating functions under mild regularity conditions. The maximum composite likelihood estimator  $\hat{\theta}$ , obtained by solving the  $p \times 1$  equation  $S(\theta) = \partial \log CL(\theta; y) / \partial \theta = 0$ , is therefore consistent for  $\theta$ . Then,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{A}^{-1}(\theta) \mathcal{B}(\theta) [\mathcal{A}^{-1}(\theta)]') \quad (2)$$

as  $n \rightarrow \infty$ , where  $\mathcal{A}(\theta) = E(-\partial S(\theta) / d\theta')$  and  $\mathcal{B}(\theta) = E(S(\theta) S'(\theta))$  are  $p \times p$  matrices and  $\mathcal{G}(\theta) = \mathcal{A}'(\theta) \mathcal{B}^{-1}(\theta) \mathcal{A}(\theta)$  is the Godambe information matrix (Godambe, 1960).

## 2 LIKELIHOOD AND COMPOSITE LIKELIHOOD FORMULATION

### 2.1 NOTATION AND MODEL SPECIFICATION

Suppose individuals are at risk of  $J$  types of recurrent events and let  $T_{ijk}$  denote the time of the  $k$ th occurrence of the type  $j$  event for individual  $i$  in a sample of size  $n$ ,  $k = 1, \dots, j = 1, \dots, J$ ,  $i = 1, \dots, n$ ; see Figure 1. We let  $N_{ij}(t)$  record the cumulative number of type  $j$  events experienced by individual  $i$  over  $(0, t]$  and let  $\{N_{ij}(u), 0 < u\}$  denote the corresponding counting process. Then  $\Delta N_{ij}(t) = N_{ij}(t + \Delta t^-) - N_{ij}(t^-)$  records the number of such events over  $[t, t + \Delta t)$  and  $\lim_{\Delta t \downarrow 0} \Delta N_{ij}(t) = dN_{ij}(t)$  indicates whether a type  $j$  event occurred at time  $t$  for individual  $i$ . If  $N_i(s) = (N_{i1}(s), \dots, N_{iJ}(s))'$  then  $\{N_i(u), 0 < u\}$  is the multivariate counting process. A  $p \times 1$  vector of fixed covariates, possibly unique for type  $j$  events, is denoted by  $\mathbf{X}_{ij}$  and we let  $\mathbf{X}_i = (\mathbf{X}'_{i1}, \dots, \mathbf{X}'_{iJ})'$ . The history for type  $j$  events is  $H_{ij}(t) = \{N_{ij}(s), 0 \leq s < t, \mathbf{X}_{ij} = \mathbf{x}_{ij}\}$  and the full history of all types of events is written as  $H_i(t) = \{H_{i1}(t), \dots, H_{iJ}(t), 0 \leq s < t\}$ .

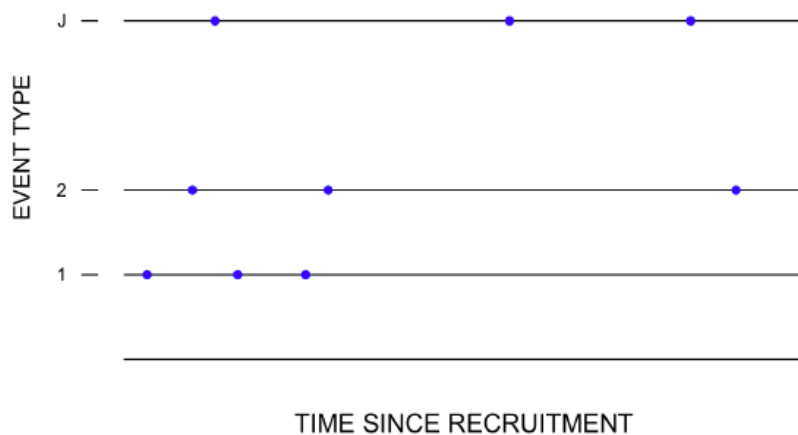


Figure 1: Timeline diagrams for  $J$  different recurrent event processes and a common censoring time

The complete intensity function for a type  $j$  event for individual  $i$  is

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{ij}(t) = 1 | H_i(t))}{\Delta t} = \lambda_{ij}(t | H_i(t)),$$

$j = 1, \dots, J, i = 1, \dots, n$ .

Suppose each individual in a sample of  $n$  independent individuals is to be followed over  $(0, A]$  where  $A$  is an administrative censoring time. To accommodate possible early study withdrawal we define a random censoring time  $C_i^\dagger$  for individual  $i$ , assumed to be conditionally independent of the event processes given  $X_i$ , and let  $C_i = \min(C_i^\dagger, A)$  be the net censoring time and  $Y_i(s) = I(s \leq C_i)$ ,  $i = 1, \dots, n$ . We then let  $d\bar{N}_{ij}(s) = Y_i(s)dN_{ij}(s)$ ,  $d\bar{N}_i(s) = (d\bar{N}_{i1}(s), \dots, d\bar{N}_{iJ}(s))'$ , and  $\bar{N}_{ij}(t) = \int_0^t d\bar{N}_{ij}(s)$  which is the number of type  $j$  events observed over  $(0, t]$ . We let  $\bar{H}_{ij}(t) = \{\bar{N}_{ij}(s), Y_i(s), 0 \leq s < t, \mathbf{X}_{ij} = \mathbf{x}_{ij}\}$  and define the observed history for type  $j$  events and  $\bar{H}_i(t) = \{\bar{H}_{i1}(t), \dots, \bar{H}_{iJ}(t), 0 \leq s < t\}$ .

We adopt a marginal model for each type of recurrent event based on a mixed Poisson process where individual-level random effects act multiplicatively on a rate function to account for variation in risk between individuals that is not explained by the available covariate. This framework also accommodates a dependence between event counts over disjoint intervals (Lawless, 1987; Klein, 1992). If  $U_{ij}$  is a random effect for type  $j$  events for individual  $i$ , let  $\mathbf{U}_i = (U_{i1}, \dots, U_{iJ})'$ ,  $i = 1, \dots, n$ . For a mixed Poisson process subject to independent right censoring, the intensity for a type  $j$  events given  $U_{ij} = u_{ij}$  is taken to be

$$\lim_{\Delta t \downarrow 0} \frac{P(\Delta \bar{N}_{ij}(t) = 1 | \bar{H}_i(t), \mathbf{U}_i = \mathbf{u}_i)}{\Delta t} = u_{ij} Y_i(t) \lambda_{j0}(t; \boldsymbol{\alpha}_j) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j)$$

where  $u_i$  is a realization of  $U_i$ ,  $\lambda_{j0}(t; \boldsymbol{\alpha}_j)$  denotes the baseline event rate function, and covariates are assumed to have a multiplicative effect. This implies that  $\{N_{ik}(s), 0 < s\} \perp U_i^{-k} | U_{ik}$  where  $U_i^{-k}$  is the vector of random effects excluding the element  $U_{ik}$ . We let  $\boldsymbol{\theta}_j = (\boldsymbol{\alpha}'_j, \boldsymbol{\beta}'_j)'$  and  $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_J)'$ .

We assume that  $\mathbf{U}_i$  are independent and identically distributed across the sample of  $n$  individuals. We consider a log-normal distribution of  $U_{ij}$  with  $E(U_{ij}) = 1$  and  $\text{var}(\log(U_{ij})) = \sigma_j^2$ . We let  $G_j(u_{ij}; \sigma_j)$  denote the cumulative distribution function; let  $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_J)'$ .

Consider a  $J$ -dimensional random variable  $\mathbf{V} = (V_1, \dots, V_J)'$  with  $V_j \sim \text{unif}(0, 1)$ ,  $j = 1, \dots, J$ , and  $P(V_1 < v_1, \dots, V_J < v_J) = C(v_1, \dots, v_J)$  the joint cumulative distribution function. A flexible parametric joint model for  $\mathbf{V}$  can be obtained by specifying the Gaussian copula (Nelsen, 2007) given by

$$C(v_1, \dots, v_J; \boldsymbol{\rho}) = \Phi(\Phi^{-1}(v_1) + \dots, \Phi^{-1}(v_J)),$$

and the Gaussian copula density is written as

$$c(v_1, \dots, v_J; \boldsymbol{\rho}) = \frac{1}{\sqrt{\det R}} \exp \left( -\frac{1}{2} \begin{pmatrix} \Phi^{-1}(v_1) \\ \vdots \\ \Phi^{-1}(v_J) \end{pmatrix}' (R^{-1} - I) \begin{pmatrix} \Phi^{-1}(v_1) \\ \vdots \\ \Phi^{-1}(v_J) \end{pmatrix} \right) \quad (3)$$

where  $\Phi(\cdot)$  is the cumulative distribution function of a standard normal random variable,  $R$  is a correlation matrix with  $(j, k)$  component  $\rho_{jk} = \text{corr}(V_j, V_k | x_i)$ , and  $\boldsymbol{\rho} = (\rho_{12}, \dots, \rho_{(J-1)J})$ .

The joint density of  $U_i$  is obtained using the probability integral transform and setting  $V_j = G(u_{ij}; \sigma_j)$  and substituting it into (3). Then differentiating gives the joint density which can be written

$$dG(\mathbf{u}_i; \boldsymbol{\phi}) = c(G_1(u_{i1}; \sigma_1), \dots, G_J(u_{iJ}; \sigma_J); \boldsymbol{\rho}) \cdot \prod_{j=1}^J dG_j(u_{ij}; \sigma_j),$$

where the parameters of the multivariate random effect distribution is denoted by  $\boldsymbol{\phi} = (\boldsymbol{\sigma}', \boldsymbol{\rho}')$ . We let  $\boldsymbol{\psi}_j = (\boldsymbol{\theta}'_j, \sigma_j)'$  be the marginal parameters for process  $j$ ,  $\boldsymbol{\psi} = (\boldsymbol{\psi}'_1, \dots, \boldsymbol{\psi}'_J)'$  and  $\boldsymbol{\Omega} = (\boldsymbol{\psi}', \boldsymbol{\rho}')$  contain all parameters. The marginal likelihood for  $n$  independent multivariate counting processes is then

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^n \int_0^\infty \dots \int_0^\infty \left\{ \prod_{j=1}^J L_{ij}(\boldsymbol{\theta}_j | u_{ij}) \right\} dG(\mathbf{u}_i; \boldsymbol{\phi}) \quad (4)$$

where the conditional likelihood  $L_{ij}(\boldsymbol{\theta}_j|u_{ij})$  is given by (Cook and Lawless, 2007)

$$L_{ij}(\boldsymbol{\theta}_j|u_{ij}) = \prod_{k=1}^{N_{ij}(C_i)} u_{ij} \lambda_{j0}(t_{ijk}; \boldsymbol{\alpha}_j) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j) \exp\left(-\int_0^\infty u_{ij} Y_i(v) \lambda_{j0}(v; \boldsymbol{\alpha}_j) \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j) dv\right). \quad (5)$$

In the development here we are both describing the formation of the model and likelihood construction for the parametric setting. Semiparametric models can be fitted where  $\lambda_{j0}(t)$  is not assumed to have any parametric form, in this case we let  $d\Lambda_{j0}$  denote the corresponding infinite dimensional parameter; see Section 3.2.

### 3 ESTIMATION BASED ON COMPOSITE LIKELIHOOD

#### 3.1 COMPOSITE LIKELIHOOD CONSTRUCTION

When the dimension of  $J$  is large, model fitting, construction of the full likelihood, and inference is computationally challenging. We adopt composite likelihood methods to resolve the computational difficulty. Thus instead of maximizing the full log-likelihood (4), a pairwise log-likelihood is used for inference. In this case, we consider each pair of events to form a composite pairwise likelihood as

$$CL^2(\boldsymbol{\Omega}) = \prod_{i=1}^n \prod_{(j,k) \in \mathcal{M}} L_{ijk}(\boldsymbol{\eta}_{jk})^{w_{jk}} \quad (6)$$

where

$$L_{ijk}(\boldsymbol{\eta}_{jk}) = \int_0^\infty \int_0^\infty L_{ij}(\boldsymbol{\theta}_j|u_{ij}) L_{ik}(\boldsymbol{\theta}_k|u_{ik}) dG_{jk}(u_{ij}, u_{ik}; \boldsymbol{\phi}_{jk}),$$

$$dG_{jk}(u_{ij}, u_{ik}; \boldsymbol{\phi}_{jk}) = dG_j(u_{ij}; \sigma_j) dG_k(u_{ik}; \sigma_k) c(G_j(u_{ij}; \sigma_j), G_k(u_{ik}; \sigma_k); \rho_{jk}), \quad (7)$$

$\boldsymbol{\eta}_{jk} = (\boldsymbol{\psi}_j, \boldsymbol{\psi}'_k, \rho_{jk})'$ ,  $\boldsymbol{\phi}_{jk} = (\sigma_j, \sigma_k, \rho_{jk})'$  and  $\mathcal{M}$  is the collection of  $J(J-1)$  pairs of  $(j, k)$  of event types. We note that  $w_{jk} = 1/(J-1)$  is chosen to make a single effective contribution to the composite likelihood for data for each type of event from each individual.

For parametric analysis we optimize (6) and make use of the sandwich variance formula in (2) to ensure valid inference. In this case the expectations are approximated empirically after substitution of the maximum composite likelihood estimates.

#### 3.2 A SEMIPARAMETRIC EM ALGORITHM FOR ESTIMATION WITH PAIRWISE LIKELIHOOD

When margins are specified semiparametrically even solving (6) directly is difficult due to the high dimension of the parameters. Here we adopt an expectation-maximization algorithm in which we treat random effects as missing data and the data on the event process as observed (Dempster et al., 1977). We specify, in this case,  $\lambda_{j0} = d\Lambda_{j0}$  as an unspecified function and let  $\boldsymbol{\theta}_j = (d\Lambda'_{j0}, \boldsymbol{\beta}'_j)'$ . Given random effects we decompose the complete pairwise composite log-likelihood in (6) into the following parts as

$$\log CL^2(\boldsymbol{\Omega}) = \sum_{i=1}^n \sum_{(j,k) \in \mathcal{M}} w_{jk} (\log L_{ij}(\boldsymbol{\theta}_j|u_{ij}) + \log L_{ik}(\boldsymbol{\theta}_k|u_{ik}) + \log dG_{jk}(u_{ij}, u_{ik}; \boldsymbol{\phi}_{jk})) \quad (8)$$

where the first two terms are given by the log of (5) and the last term by the log of (7). In the E-step, we take the conditional expectation of the complete pairwise composite log-likelihood given the data

on the corresponding pairs of processes. Here we define  $H_{ijk}(t) = \{H_{ij}(s), H_{ik}(s), 0 < s < t\}'$  as the history of  $j$  and  $k$  types of events for an individual  $i$ . Then

$$\begin{aligned} E[\log CL^2(\Omega)|H_{ijk}(C_i); \hat{\Omega}^{(r-1)}] &= \sum_{i=1}^n \sum_{(j,k) \in \mathcal{M}} w_{jk} \{E[\log L_{ij}(\boldsymbol{\theta}_j|u_{ij})|H_{ijk}(C_i); \hat{\Omega}^{(r-1)}] \\ &+ E[\log L_{ik}(\boldsymbol{\theta}_k|u_{ik})|H_{ijk}(C_i); \hat{\Omega}^{(r-1)}] + E[\log dG_{jk}(u_{ij}, u_{ik}; \boldsymbol{\phi}_{jk})|H_{ijk}(C_i); \hat{\Omega}^{(r-1)}]\} \end{aligned} \quad (9)$$

where  $\hat{\Omega}^{(r-1)}$  is an estimate of  $\Omega$  at the  $(r-1)$ st iteration. The estimating equation in the M-step at the  $r$ th iteration (Klein, 1992) is

$$U_j^{(r-1)}(\boldsymbol{\beta}_j) = \sum_{i=1}^n \int_0^\infty \sum_{k=1, j \neq k}^J w_{jk} \bar{Y}_i(s) W_{ij}^{(r-1)}(s; \boldsymbol{\beta}_j) dN_{ij}(s) \quad (10)$$

where

$$W_{ij}^{(r-1)}(s; \boldsymbol{\beta}_j) = \left( \mathbf{x}_{ij} - \frac{R_j^{(1,r-1)}(s; \boldsymbol{\beta}_j)}{R_j^{(0,r-1)}(s; \boldsymbol{\beta}_j)} \right),$$

and we set

$$R_j^{(h,r)}(s; \boldsymbol{\beta}_j) = \sum_{i=1}^n \bar{Y}_i(s) \left[ \sum_{k=1, j \neq k}^J w_{jk} E[U_{ij}|H_{ijk}(C_i), \hat{\Omega}^{(r)}] \right] \exp(\mathbf{x}'_{ij} \boldsymbol{\beta}_j) x_{ij}^h. \quad (11)$$

The calculation of  $E[U_{ij}|H_{ijk}(C_i), \hat{\Omega}^{(r)}]$  is given as

$$E[U_{ij}|H_{ijk}(C_i); \hat{\Omega}^{(r)}] = \frac{\int_0^\infty \int_0^\infty u_{ij} P(H_{ijk}(C_i)|u_{ij}, u_{ik}, \mathbf{x}_{ij}, \mathbf{x}_{ik}; \hat{\boldsymbol{\theta}}_j^{(r)}, \hat{\boldsymbol{\theta}}_k^{(r)}) dG_{jk}(u_{ij}, u_{ik}; \hat{\boldsymbol{\phi}}_{jk}^{(r)})}{\int_0^\infty \int_0^\infty P(H_{ijk}(C_i)|u_{ij}, u_{ik}, \mathbf{x}_{ij}, \mathbf{x}_{ik}; \hat{\boldsymbol{\theta}}_j^{(r)}, \hat{\boldsymbol{\theta}}_k^{(r)}) dG_{jk}(u_{ij}, u_{ik}; \hat{\boldsymbol{\phi}}_{jk}^{(r)})} \quad (12)$$

where  $P(H_{ijk}(C_i)|u_{ij}, u_{ik}, \mathbf{x}_{ij}, \mathbf{x}_{ik}; \boldsymbol{\theta}_j, \boldsymbol{\theta}_k) = L_{ij}(\boldsymbol{\theta}_j|u_{ij})L_{ik}(\boldsymbol{\theta}_k|u_{ik})$  and  $\hat{\boldsymbol{\theta}}_j^{(r)}, \hat{\boldsymbol{\theta}}_k^{(r)}$  and  $\hat{\boldsymbol{\phi}}_{jk}^{(r)}$  are the estimates of  $\boldsymbol{\theta}_j, \boldsymbol{\theta}_k$ , and  $\boldsymbol{\phi}_{jk}$  at the  $r$ -th iteration, respectively. Let  $\hat{\boldsymbol{\beta}}_j^{(r)}$  denote the solution to  $U_j^{(r-1)}(\boldsymbol{\beta}_j) = 0$  in (10). The cumulative baseline rates are then estimated using the Breslow formula as

$$\hat{\Lambda}_{j0}^{(r)}(s) = \sum_{i=1}^n \int_0^\infty \sum_{k=1, j \neq k}^J w_{jk} \bar{Y}_i(s) dN_{ij}(s) / R_j^{(0,r-1)}(s; \hat{\boldsymbol{\beta}}_j^{(r)}). \quad (13)$$

The maximization of (9) in semi-parametric setting can be easily carried out using the `coxph` function in R with  $\log((E[U_{ij}|H_{ijk}(C_i)] + E[U_{ik}|H_{ijk}(C_i)])/2)$  treated as an offset term. The variance of the random effects and the dependence parameter are estimated by maximizing

$$\sum_{i=1}^n \sum_{(j,k) \in \mathcal{M}} w_{jk} E[\log dG_{jk}(u_{ij}, u_{ik}; \boldsymbol{\phi}_{jk})|H_{ijk}(C_i), \hat{\Omega}^{(r-1)}] \quad (14)$$

using the standard optimization software such as the `optim` function in R. The E-step and M-step are repeated iteratively until the following stopping rule is satisfied;

$$\max \left( \frac{|\log CL(\hat{\Omega}^{(r+1)}) - \log CL(\hat{\Omega}^{(r)})|}{|\log CL(\hat{\Omega}^{(r)})|} \right) \leq 10^{-10}, \text{ or } \max(|\hat{\Omega}^{(r+1)} - \hat{\Omega}^{(r)}|) \leq 10^{-4}.$$

Under the regularity conditions (Godambe, 1960), we have

$$\sqrt{n}(\hat{\Omega} - \Omega) \xrightarrow{d} N(0, \mathcal{A}^{-1}(\Omega)\mathcal{B}(\Omega)[\mathcal{A}^{-1}(\Omega)]') \quad (15)$$

as  $n \rightarrow \infty$  and the asymptotic covariacne matrix is estimated by

$$\hat{A}^{-1}(\hat{\Omega})\hat{B}(\hat{\Omega})[\hat{A}^{-1}(\hat{\Omega})]'$$

In a semiparametric setting, the calculation of variance of infinite dimensional parameters leads to the inverse of infinite dimensional information matrix. To overcome the difficulty of calculation, we adopt the profile likelihood approach (Murphy and Van der Vaart, 2000) to obtain the variance estimates. We first define the profile likelihood function

$$pl_n(\gamma) = \max_{\Lambda} \frac{1}{n} \sum_{i=1}^n q_i(\gamma, \Lambda),$$

where  $\gamma = (\beta', \phi')'$  indicates the finite dimensional parameters,  $\Lambda = (\Lambda'_{10}, \dots, \Lambda'_{J0})'$ , and  $q_i(\gamma, \Lambda) = \sum_{(j,k) \in \mathcal{M}} w_{jk} \log CL_{ijk}(\Omega)$  is the log of the observed pairwise likelihood,  $i = 1, \dots, n$ . Then, we can calculate  $\hat{A}$  and  $\hat{B}$  at  $\hat{\gamma}$  by

$$\hat{A}[k, l] = - \frac{pl_n(\hat{\gamma} + h\mathbf{e}_k + h\mathbf{e}_l) - pl_n(\hat{\gamma} + h\mathbf{e}_k - h\mathbf{e}_l) - pl_n(\hat{\gamma} - h\mathbf{e}_k + h\mathbf{e}_l) + pl_n(\hat{\gamma} - h\mathbf{e}_k - h\mathbf{e}_l)}{4h^2}, \quad (16)$$

and

$$\hat{B}[k, l] = - \sum_{i=1}^n \frac{1}{4nh^2} [q_i(\hat{\gamma} + h\mathbf{e}_k, \hat{\Lambda}_{\hat{\gamma} + h\mathbf{e}_k}) - q_i(\hat{\gamma} - h\mathbf{e}_k, \hat{\Lambda}_{\hat{\gamma} - h\mathbf{e}_k})] [q_i(\hat{\gamma} + h\mathbf{e}_l, \hat{\Lambda}_{\hat{\gamma} + h\mathbf{e}_l}) - q_i(\hat{\gamma} - h\mathbf{e}_l, \hat{\Lambda}_{\hat{\gamma} - h\mathbf{e}_l})], \quad (17)$$

where  $\mathbf{e}_k$  is the  $k$ th coordinate vector and  $h$  is the increment in numerical differentiation. Here  $\Lambda_{\gamma}$  is the cumulative baseline hazard function which maximizes the observed pairwise likelihood given  $\gamma$ . The details of the variance estimation are presented in the Appendix A. The calculation of conditional expectation in (12), (14), (16), and (17) requires to use numerical integration. To facilitate shared-memory multi-processor, we implement the OpenMP (Open Multi-Processing) interface in C++ to carry out numerical integration by Gaussian-Quadrature with 20 nodes for each dimension.

### 3.3 TWO-STAGE SEMIPARAMETRIC ESTIMATION WITH PAIRWISE LIKELIHOOD

The implementation of a two-stage estimation procedure can ease computation (Andersen, 2004; Zhao and Joe, 2005). In the first stage, the parameters for each type of event,  $\psi_j = (\boldsymbol{\theta}_j, \sigma_j)'$ , are estimated under a working independence assumption. We also use the expectation-maximization algorithm treating a random effect for each type as missing data to obtain  $\hat{\psi}_j$ . Given a random effect  $u_{ij}$ , the complete likelihood function for each type is  $L_{ij}(\boldsymbol{\theta}_j | u_{ij})$ . In the E-step, we obtain the conditional expectation given the observed data for each type

$$E[\log L_{ij}(\boldsymbol{\theta}_j | u_{ij}) | H_{ij}(C_i), \hat{\boldsymbol{\psi}}^{(r)}]$$

at the  $r$ -th iteration. In the M-step, the estimating equation (10) changes to

$$U_j^{(r-1)}(\boldsymbol{\beta}_j) = \sum_{i=1}^n \int_0^{\infty} \bar{Y}_i(s) W_{ij}^{(r-1)}(s; \boldsymbol{\beta}_j) dN_{ij}(s)$$

where  $R_j^{(h,r)}(s; \beta_j)$  in (11) becomes

$$R_j^{(h,r)}(s; \beta_j) = \sum_{i=1}^n \bar{Y}_i(s) E[U_{ij} | H_{ij}(C_i), \hat{\psi}_j^{(r)}] \exp(\mathbf{x}'_{ij} \beta_j) \mathbf{x}_{ij}^h.$$

We calculate  $E[U_{ij} | H_{ij}(C_i); \hat{\psi}_j^{(r)}]$  as

$$E[U_{ij} | H_{ij}(C_i); \hat{\psi}_j^{(r)}] = \frac{\int_0^\infty u_{ij} P(H_{ij}(C_i) | u_{ij}, \mathbf{x}_{ij}; \hat{\theta}_j^{(r)}) dG_j(u_{ij}; \hat{\sigma}_j^{(r)})}{\int_0^\infty P(H_{ij}(C_i) | u_{ij}, \mathbf{x}_{ij}; \hat{\theta}_j^{(r)}) dG_j(u_{ij}; \hat{\sigma}_j^{(r)})}$$

where  $P(H_{ij}(C_i) | u_{ij}, \mathbf{x}_{ij}; \theta) = L_{ij}(\theta_j | u_{ij})$ . The estimated cumulative baseline rates are then obtained as

$$\hat{\Lambda}_{j0}^{(r)}(s) = \sum_{i=1}^n \int_0^\infty \bar{Y}_i(s) dN_{ij}(s) / R_j^{(0,r-1)}(s; \hat{\beta}_j^{(r)}).$$

The procedure is iterated until convergence. In the second stage, we solve the composite score function from the pairwise likelihood with respect to  $\rho$  plugging the estimates  $\hat{\psi}$  from the stage 1. Again, we implement the expectation-maximization algorithm with random effects treated as missing data in which we obtain the dependence parameter by maximizing the following estimating equation given the the estimates of marginal parameters

$$\sum_{i=1}^n \sum_{(j,k) \in \mathcal{M}} w_{jk} E[\log dG_{jk}(u_{ij}, u_{ik}; \phi_{jk}) | H_{ijk}(C_i), \hat{\rho}^{(r-1)}, \hat{\psi}_j, \hat{\psi}_k].$$

Zhao and Joe (2005) commented that two-stage estimation in composite likelihood is recommended with a weak dependence. In a strong dependence case, the simultaneous estimation method gives better estimates. The variance estimates in two-stage estimation are present in Appendix. The code can be downloaded from <https://github.com/joolee0918/Mfrailty>.

## 4 SIMULATION STUDIES

Simulation studies are conducted to evaluate the performance of estimators from the joint models introduced in Section 2. We consider three different types of infections ( $J=3$ ). For a randomized treatment, we let  $X_{i1}$  the indicator of treatment where  $X_{i1} = 1$  having treatment otherwise  $X_i = 0$  where  $P(X_{i1} = 1) = 0.5$ . We also consider a continuous variable  $X_{i2}$  which was generated from the normal distribution,  $N(5, 1)$ . We generate the data over the interval  $(0, 1]$  with an independent random censoring  $C_i$ . We assume that  $C_i$  follows an exponential distribution with rate  $-\log(0.9)$  and  $-\log(0.7)$  indicating 10% and 30% censoring, respectively. Each event times follow Weibull distribution with the survival function  $\mathcal{F}(t; \lambda, \alpha) = \exp(-(t/\lambda)^\alpha)$ . We set  $(\lambda_1, \alpha_1) = (0.667, 1.00)$ ,  $(\lambda_2, \alpha_2) = (0.639, 1.25)$ ,  $(\lambda_3, \alpha_3) = (0.630, 1.50)$  and we set the coefficients  $\beta_{11} = \beta_{21} = \beta_{31} = \log(0.8)$ ,  $\beta_{12} = \beta_{22} = \beta_{32} = \log(1.1)$ . We consider the Gaussian copula with log-normal margins for random effects where  $E(U_{ij}) = 1$  and  $\text{Var}(\log(U_{ij})) = \sigma_j^2$ . We set the frailty parameters as  $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0.16$ , and the association parameters as  $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, 0.25, 0.25)$ , and  $(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.30, -0.50, 0.30)$ . We generate 1000 samples with 300 and 500 individuals. For each data set, parametric and semi-parametric analyses are carried out based on the pairwise likelihood, and two-stage procedure based on the pairwise likelihood in Section 3. The empirical bias (EBIAS), average asymptotic (large sample) standard error (ASE), empirical standard error (ESE) and empirical coverage probability (ECP) are evaluated for all parameter estimates and reported in reported in Table 1, Table 2, and Table 3. The average of number of each type of events (NAVE) is also reported.



Table 1: Frequency properties of estimators obtained by fitting a Weibull-model using the pairwise likelihood with the censoring rate 10% and 30%, and  $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, 0.25, 0.25)$  and  $(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.30, 0.50, 0.30)$ ; the sample size is 300 and 500, and  $nsim = 1000$

n	cen%	TYPE	PARAM	$(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, 0.25, 0.25)$				$(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.30, -0.50, 0.30)$			
				EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP
300	10%	1	$\log(\lambda_1)$	0.027	0.260	0.245	0.939	0.023	0.256	0.245	0.932
			$\log(\alpha_1)$	0.001	0.040	0.039	0.951	0.002	0.038	0.039	0.953
			$\beta_{11}$	0.005	0.096	0.094	0.949	0.005	0.099	0.094	0.930
			$\beta_{12}$	0.004	0.049	0.046	0.939	0.004	0.048	0.046	0.941
			$\sigma_1^2$	-0.005	0.047	0.050	0.968	-0.003	0.044	0.050	0.970
		2	$\log(\lambda_2)$	0.011	0.202	0.189	0.925	0.007	0.200	0.187	0.921
			$\log(\alpha_2)$	0.001	0.039	0.037	0.922	-0.001	0.036	0.037	0.949
			$\beta_{21}$	-0.001	0.089	0.089	0.950	0.002	0.088	0.089	0.951
			$\beta_{22}$	0.002	0.048	0.044	0.921	0.001	0.048	0.044	0.925
			$\sigma_2^2$	-0.001	0.043	0.046	0.971	-0.003	0.041	0.045	0.975
		3	$\log(\lambda_3)$	0.000	0.150	0.150	0.946	0.002	0.152	0.150	0.945
			$\log(\alpha_3)$	-0.001	0.034	0.034	0.950	0.000	0.035	0.034	0.944
			$\beta_{31}$	0.002	0.083	0.085	0.947	-0.004	0.085	0.085	0.958
			$\beta_{32}$	0.000	0.043	0.043	0.951	0.001	0.043	0.043	0.953
			$\sigma_3^2$	-0.002	0.042	0.042	0.958	-0.001	0.039	0.042	0.972
		Copula	$\rho_{12}$	0.016	0.214	0.235	0.969	0.000	0.212	0.243	0.962
$\rho_{13}$	0.021		0.215	0.224	0.954	-0.004	0.196	0.223	0.958		
$\rho_{23}$	0.012		0.204	0.208	0.956	0.017	0.189	0.208	0.964		
300	30%	1	$\log(\lambda_1)$	0.020	0.278	0.260	0.933	0.014	0.270	0.259	0.945
			$\log(\alpha_1)$	0.000	0.039	0.041	0.959	0.000	0.039	0.041	0.960
			$\beta_{11}$	0.003	0.104	0.099	0.947	0.008	0.104	0.099	0.943
			$\beta_{12}$	0.003	0.053	0.049	0.929	0.002	0.051	0.049	0.945
			$\sigma_1^2$	-0.004	0.050	0.054	0.968	-0.002	0.050	0.054	0.966
		2	$\log(\lambda_2)$	0.016	0.207	0.199	0.944	0.021	0.210	0.199	0.932
			$\log(\alpha_2)$	0.000	0.038	0.038	0.951	0.000	0.038	0.038	0.956
			$\beta_{21}$	-0.001	0.095	0.095	0.947	0.000	0.100	0.095	0.936
			$\beta_{22}$	0.004	0.049	0.047	0.941	0.005	0.050	0.047	0.925
			$\sigma_2^2$	-0.003	0.046	0.049	0.968	-0.001	0.046	0.055	0.970
		3	$\log(\lambda_3)$	-0.004	0.171	0.160	0.921	-0.006	0.160	0.160	0.951
			$\log(\alpha_3)$	0.001	0.037	0.036	0.939	-0.001	0.036	0.036	0.945
			$\beta_{31}$	-0.005	0.092	0.091	0.951	-0.004	0.090	0.091	0.941
			$\beta_{32}$	-0.001	0.049	0.045	0.921	-0.002	0.046	0.045	0.944
			$\sigma_3^2$	-0.003	0.044	0.046	0.967	0.000	0.044	0.049	0.966
		Copula	$\rho_{12}$	0.020	0.239	0.256	0.954	0.010	0.231	0.268	0.967
$\rho_{13}$	0.028		0.238	0.246	0.950	0.015	0.215	0.246	0.957		
$\rho_{23}$	0.020		0.210	0.229	0.966	0.017	0.214	0.228	0.966		

(Continues)



Table 2: Frequency properties of estimators obtained by fitting a semiparametric model using the pairwise likelihood and two-stage pairwise estimation method based on the pairwise likelihood with the censoring rate 10% and 30%, and  $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, 0.25, 0.25)$ ; the sample size is 300 and 500, and  $n_{sim} = 1000$

n	cen%	TYPE (NAVE <sup>†</sup> )	PARAM	Pairwise Likelihood				Two-stage Pairwise Likelihood			
				EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP
300	10%	1 (623)	$\beta_{11}$	0.002	0.095	0.094	0.948	0.002	0.095	0.094	0.947
			$\beta_{12}$	-0.001	0.048	0.046	0.943	-0.001	0.048	0.046	0.942
			$\sigma_1^2$	-0.009	0.051	0.050	0.963	-0.009	0.051	0.049	0.964
		2 (722)	$\beta_{21}$	-0.003	0.089	0.089	0.948	-0.003	0.089	0.089	0.949
			$\beta_{22}$	-0.001	0.047	0.044	0.940	-0.001	0.047	0.044	0.942
			$\sigma_2^2$	-0.005	0.047	0.045	0.969	-0.005	0.047	0.045	0.973
		3 (821)	$\beta_{31}$	-0.003	0.083	0.085	0.946	-0.003	0.083	0.085	0.948
			$\beta_{32}$	0.000	0.042	0.043	0.952	0.000	0.042	0.043	0.951
			$\sigma_3^2$	-0.005	0.045	0.042	0.956	-0.005	0.045	0.041	0.956
		Copula	$\rho_{12}$	0.009	0.242	0.241	0.952	0.008	0.240	0.241	0.954
			$\rho_{13}$	0.015	0.238	0.246	0.942	0.014	0.237	0.231	0.943
			$\rho_{23}$	0.004	0.223	0.212	0.942	0.004	0.221	0.212	0.944
300	30%	1 (552)	$\beta_{11}$	0.000	0.103	0.099	0.945	0.000	0.103	0.099	0.948
			$\beta_{12}$	-0.001	0.051	0.049	0.941	-0.001	0.051	0.049	0.940
			$\sigma_1^2$	-0.010	0.056	0.053	0.961	-0.010	0.056	0.053	0.967
		2 (631)	$\beta_{21}$	-0.004	0.094	0.095	0.953	-0.004	0.094	0.095	0.954
			$\beta_{22}$	0.000	0.048	0.047	0.939	0.000	0.048	0.047	0.938
			$\sigma_2^2$	-0.006	0.049	0.049	0.973	-0.007	0.049	0.049	0.973
		3 (709)	$\beta_{31}$	-0.006	0.092	0.091	0.948	-0.006	0.092	0.091	0.949
			$\beta_{32}$	-0.001	0.047	0.045	0.931	-0.001	0.047	0.045	0.929
			$\sigma_3^2$	-0.006	0.047	0.045	0.964	-0.006	0.047	0.045	0.967
		Copula	$\rho_{12}$	0.011	0.264	0.264	0.940	0.010	0.260	0.266	0.943
			$\rho_{13}$	0.020	0.261	0.262	0.934	0.019	0.259	0.256	0.937
			$\rho_{23}$	0.012	0.231	0.237	0.955	0.011	0.230	0.238	0.958
500	10%	1 (1035)	$\beta_{11}$	0.000	0.076	0.073	0.942	0.000	0.076	0.073	0.943
			$\beta_{12}$	0.002	0.037	0.036	0.948	0.002	0.037	0.036	0.947
			$\sigma_1$	-0.005	0.040	0.039	0.967	-0.005	0.040	0.039	0.969
		2 (1202)	$\beta_{21}$	-0.001	0.069	0.069	0.950	0.000	0.069	0.069	0.951
			$\beta_{22}$	0.000	0.035	0.035	0.943	0.000	0.035	0.035	0.943
			$\sigma_2$	-0.002	0.037	0.035	0.947	-0.002	0.036	0.035	0.950
		3 (1364)	$\beta_{31}$	-0.001	0.069	0.066	0.938	-0.001	0.069	0.066	0.939
			$\beta_{32}$	0.001	0.035	0.033	0.936	0.001	0.035	0.033	0.953
			$\sigma_3^2$	-0.004	0.031	0.032	0.970	-0.004	0.031	0.032	0.969
		Copula	$\rho_{12}$	0.008	0.184	0.180	0.941	0.007	0.182	0.180	0.969
			$\rho_{13}$	0.006	0.172	0.172	0.947	0.005	0.171	0.172	0.950
			$\rho_{23}$	0.013	0.152	0.159	0.962	0.013	0.152	0.159	0.969
500	30%	1 (917)	$\beta_{11}$	0.000	0.082	0.077	0.934	0.000	0.082	0.077	0.934
			$\beta_{12}$	0.000	0.039	0.038	0.946	0.000	0.039	0.038	0.946
			$\sigma_1$	-0.003	0.043	0.042	0.970	-0.003	0.043	0.042	0.969
		2 (1050)	$\beta_{21}$	0.000	0.074	0.073	0.949	0.000	0.074	0.074	0.949
			$\beta_{22}$	0.001	0.038	0.037	0.945	0.001	0.038	0.037	0.944
			$\sigma_2^2$	-0.002	0.039	0.038	0.955	-0.002	0.039	0.038	0.957
		3 (1179)	$\beta_{31}$	0.000	0.073	0.071	0.946	0.000	0.073	0.071	0.947
			$\beta_{32}$	0.001	0.035	0.035	0.955	0.001	0.035	0.035	0.954
			$\sigma_3^2$	-0.004	0.036	0.035	0.963	-0.004	0.036	0.035	0.963
		Copula	$\rho_{12}$	-0.001	0.202	0.195	0.942	-0.002	0.200	0.195	0.969
			$\rho_{13}$	0.009	0.188	0.186	0.951	0.008	0.187	0.186	0.957
			$\rho_{23}$	0.007	0.176	0.175	0.955	0.006	0.175	0.175	0.963

† NAVE is the average number of events per individual over the course of observation.

Table 3: Frequency properties of estimators obtained by fitting a semiparametric model using the pairwise likelihood and two-stage pairwise estimation method based on the pairwise likelihood with the censoring rate 10% and 30%, and  $(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.30, -0.50, 0.30)$ ; the sample size is 300 and 500, and  $nsim = 1000$

n	cen%	TYPE (NAVE <sup>†</sup> )	PARAM	Pairwise Likelihood				Two-stage Pairwise Likelihood			
				EBIAS	ESE	ASE	ECP	EBIAS	ESE	ASE	ECP
300	10%	1 (623)	$\beta_{11}$	0.004	0.096	0.094	0.938	0.004	0.097	0.094	0.939
			$\beta_{12}$	-0.002	0.047	0.047	0.942	-0.002	0.047	0.047	0.943
			$\sigma_1^2$	-0.006	0.050	0.050	0.962	-0.006	0.050	0.050	0.960
		2 (722)	$\beta_{21}$	-0.005	0.090	0.089	0.948	-0.005	0.090	0.089	0.947
			$\beta_{22}$	-0.004	0.045	0.044	0.938	-0.004	0.045	0.044	0.938
			$\sigma_2^2$	-0.006	0.046	0.045	0.963	-0.006	0.046	0.045	0.963
		3 (819)	$\beta_{31}$	-0.002	0.084	0.085	0.956	-0.002	0.084	0.085	0.955
			$\beta_{32}$	0.001	0.042	0.043	0.954	0.001	0.042	0.043	0.954
			$\sigma_3^2$	-0.004	0.043	0.042	0.959	-0.004	0.043	0.041	0.958
		Copula	$\rho_{12}$	-0.026	0.231	0.243	0.945	-0.025	0.230	0.243	0.944
			$\rho_{13}$	-0.018	0.212	0.228	0.963	-0.017	0.210	0.227	0.942
			$\rho_{23}$	0.012	0.207	0.213	0.959	0.011	0.206	0.214	0.959
300	30%	1 (553)	$\beta_{11}$	0.006	0.103	0.099	0.949	0.006	0.103	0.099	0.949
			$\beta_{12}$	-0.002	0.051	0.049	0.954	-0.002	0.051	0.049	0.955
			$\sigma_1^2$	-0.008	0.056	0.054	0.954	-0.008	0.056	0.053	0.954
		2 (631)	$\beta_{21}$	-0.001	0.099	0.095	0.937	-0.001	0.099	0.095	0.938
			$\beta_{22}$	0.001	0.049	0.047	0.931	0.001	0.049	0.047	0.933
			$\sigma_2^2$	-0.006	0.050	0.049	0.970	-0.006	0.050	0.049	0.969
		3 (709)	$\beta_{31}$	-0.004	0.090	0.091	0.941	-0.004	0.090	0.091	0.941
			$\beta_{32}$	-0.001	0.045	0.045	0.951	-0.001	0.045	0.045	0.951
			$\sigma_3^2$	-0.003	0.048	0.046	0.964	-0.003	0.048	0.045	0.964
		Copula	$\rho_{12}$	-0.005	0.255	0.270	0.945	-0.004	0.253	0.267	0.947
			$\rho_{13}$	0.005	0.232	0.253	0.935	0.006	0.230	0.262	0.969
			$\rho_{23}$	0.016	0.236	0.234	0.944	0.015	0.234	0.233	0.964
500	10%	1 (1037)	$\beta_{11}$	-0.001	0.077	0.073	0.927	-0.001	0.077	0.073	0.925
			$\beta_{12}$	-0.001	0.037	0.036	0.948	-0.001	0.037	0.036	0.949
			$\sigma_1$	-0.004	0.039	0.039	0.963	-0.004	0.039	0.039	0.966
		2 (1202)	$\beta_{21}$	-0.001	0.070	0.069	0.943	0.000	0.070	0.069	0.942
			$\beta_{22}$	0.001	0.036	0.034	0.937	0.001	0.036	0.034	0.935
			$\sigma_2$	-0.005	0.036	0.035	0.964	-0.005	0.036	0.035	0.967
		3 (1365)	$\beta_{31}$	0.000	0.067	0.066	0.945	0.000	0.067	0.066	0.945
			$\beta_{32}$	0.002	0.033	0.033	0.951	0.002	0.033	0.033	0.949
			$\sigma_3^2$	-0.003	0.031	0.033	0.961	-0.003	0.031	0.033	0.961
		Copula	$\rho_{12}$	-0.015	0.185	0.184	0.951	-0.015	0.183	0.184	0.952
			$\rho_{13}$	-0.019	0.166	0.173	0.958	-0.018	0.164	0.173	0.958
			$\rho_{23}$	0.008	0.160	0.160	0.945	0.007	0.160	0.160	0.945
500	30%	1 (919)	$\beta_{11}$	-0.002	0.079	0.077	0.944	-0.002	0.079	0.077	0.945
			$\beta_{12}$	0.002	0.039	0.038	0.943	0.002	0.039	0.038	0.944
			$\sigma_1$	-0.004	0.042	0.042	0.968	-0.005	0.042	0.042	0.968
		2 (1049)	$\beta_{21}$	-0.002	0.072	0.074	0.951	-0.002	0.072	0.074	0.949
			$\beta_{22}$	0.000	0.035	0.037	0.962	0.000	0.035	0.037	0.962
			$\sigma_2^2$	-0.001	0.040	0.038	0.954	-0.002	0.040	0.038	0.953
		3 (1180)	$\beta_{31}$	-0.001	0.076	0.071	0.931	-0.001	0.076	0.071	0.932
			$\beta_{32}$	0.000	0.036	0.037	0.948	0.000	0.036	0.035	0.947
			$\sigma_3^2$	-0.004	0.035	0.035	0.968	-0.004	0.035	0.035	0.967
		Copula	$\rho_{12}$	-0.024	0.199	0.200	0.958	-0.024	0.196	0.199	0.961
			$\rho_{13}$	-0.021	0.183	0.193	0.947	-0.019	0.181	0.190	0.948
			$\rho_{23}$	0.019	0.181	0.173	0.938	0.019	0.180	0.173	0.938

† NAVE is the average number of events per individual over the course of observation.

The empirical biases are small for all estimates of parameters and empirical standard errors and average estimated standard errors are in good agreement. The empirical coverage probability are close to the nominal confidence level of 95%. Comparing parametric and semi-parametric model, we find that all estimates and standard errors are very close between two models. However, to protect from the model misspecification, a semi-parametric model may be recommended despite the more intensive computational burden. The standard errors of estimators of marginal parameters are very close between the pairwise simultaneous and two-stage model. When the sample size increases, the variance of the estimators decrease. We note that the variances of the estimators increase as the censoring rate is higher. We report the average of computation time in seconds in Table 4 with the sample size 300 and 500. The two-stage pairwise likelihood estimation method is less computationally intensive. In this simulation studies, we only consider three types of events in which pair-wise likelihood approach is feasible. However, if more than three events are analyzed two-stage methods are easy to implement.

Table 4: The average of computation time (in seconds) for fitting a semiparametric model using the pairwise likelihood and two-stage pairwise estimation method based on the pairwise likelihood with the censoring rate 10% and 30%, and  $(\rho_{12}, \rho_{13}, \rho_{23}) = (0.25, 0.25, 0.25)$   $(\rho_{12}, \rho_{13}, \rho_{23}) = (-0.30, 0.50, -0.30)$ ; the sample size is 300 and 500, and  $n_{sim} = 1000$

n	$\rho$	cen%	Pairwise Likelihood	Two-stage Pairwise Likelihood
			time (sec)	time (sec)
300	(0.25, 0.25, 0.25)	10%	409.659	273.671
		30%	434.067	290.511
	(-0.30, 0.50, -0.30)	10%	437.675	295.806
		30%	487.316	328.956
500	(0.25, 0.25, 0.25)	10%	727.007	485.660
		30%	766.560	518.123
	(-0.30, 0.50, -0.30)	10%	713.243	483.049
		30%	724.884	491.380

## 5 RECURRENT INFECTIONS IN A PEDIATRIC TRIAL OF IRON SUPPLEMENTATION

Malnutrition in children in low-income countries has been identified as a cause of immune deficiency and susceptibility to infectious diseases since activation of the immune system in response to infection requires additional energy. Examples infectious disease due to malnutrition are opportunistic pathogens and fungus, noma, respiratory, intestinal infections, tuberculosis, measles and other chronic infections (Ambrus, 2004; Schaible and Stefan, 2007). Iron deficiency, which is also prevalent in developing countries, causes anemia and the deficiency of red blood cells. Lemaire et al. (2011) conducted a study to examine the efficacy of iron-containing micro-nutrient powder (iron MNP) on the risk of infectious diseases to malnourished children. In a randomized clinical trial, 268 Bangladeshi children, aged 12-24 month, and moderate-to-severe malnourished with a hemoglobin concentration

between 70 and 110 g/L, were recruited for two-phase, 12/2007-06/2008 and 07/2008-01/2009, respectively. Iron MNP were provided to 136 children daily for 2 months and the remaining children were provided placebo powder. The primary endpoint was the occurrence of infectious diseases such as diarrhea, dysentery, vomiting, lower respiratory tract infections (LRTIs), coughing, sneezing and fever. During the 2 months intervention period, the incidence of infectious disease was assessed every other day, whereas after the intervention period, it was assessed weekly. Given this data, interest lies on how to examine the effect of treatment on the incidence of multiple diseases which are caused by malnutrition and iron deficiency. An analysis based on a composite score is feasible (Lemaire et al., 2011), however it leads to losing information on distinct episodes. Often interest lies in a treatment effect on specific disease, however, since the lack of nutrition directly or indirectly affects the immune system, the different types of infections in consequence of malnutrition may be associated. Therefore somewhat related diseases should be considered together. In the analysis of this data, we consider three types of events including digestive disorders including diarrhea, dysentery, and vomiting, LRTI, and other infections. The number of events of digestive disorders is 323, that of LRTI 513, and other infections has a total number of 100 events.

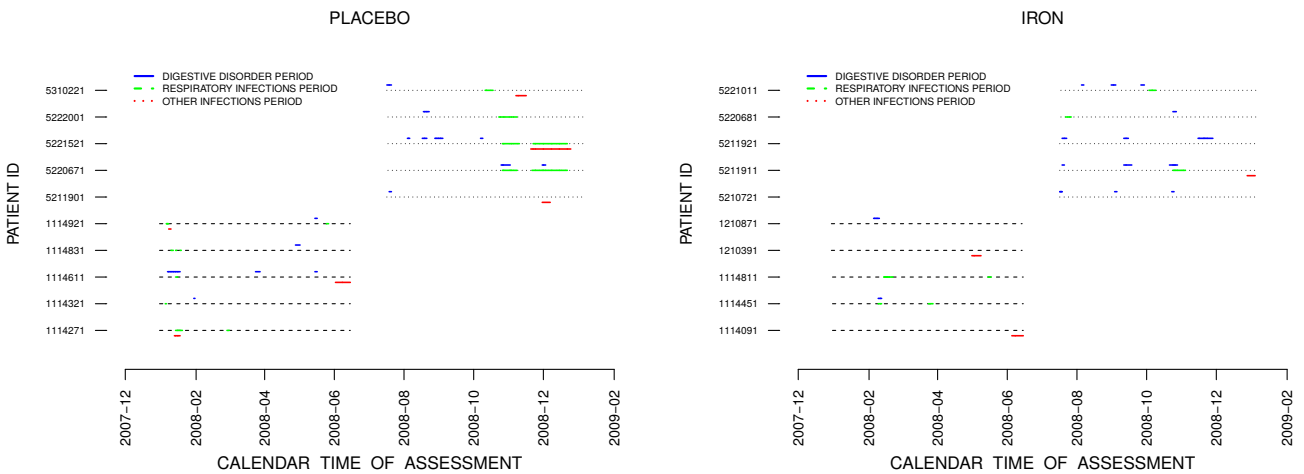


Figure 2: Digestive disorders, lower respiratory tract infections, and other infections event plots for phase 1 and phase 2 showing the onset and the duration of episodes

Figure 2 displays digestive disorders, respiratory infections, and other infections data for each individual with two-phase where lines represent a period of diseases. Due to the possible correlations between these events, the joint analysis should be performed to take into account for the dependence. We include three covariates; iron supplement as a treatment, phase, and a hemoglobin level which is a binary covariate having 1 if the hemoglobin level is less than 9.1 g/L. We conduct pairwise likelihood and two-stage pairwise likelihood analysis in application.

The results are summarized in Table 5. We observe the iron supplements do not have significant effects on the occurrence of all three types of events. The iron supplements reduce the occurrence of digestive disorders ( $RR : 0.82$ ;  $95\% CI : (-0.44, 0.04)$ ,  $p = 0.101$  with pairwise likelihood) and LRTI ( $RR : 0.85$ ;  $95\% CI : (-0.35, 0.01)$ ,  $p = 0.062$  with pairwise likelihood) whereas it increases the onset of other infections ( $RR : 1.41$ ;  $95\% CI : (-0.04, 0.73)$ ,  $p = 0.083$  with pairwise likelihood). We note that the change of phase from 1 to 2 significantly increases the occurrence of all type of events; phase 2 represents the winter period so that viral gastroenteritis increases onset of digestive disorders ( $RR : 1.45$ ;  $95\% CI : (0.13, 0.61)$ ,  $p = 0.003$  with pairwise likelihood). Also the onset of LRTI may increase due to seasonal factors ( $RR : 1.95$ ;  $95\% CI : (0.47, 0.87)$ ,  $p < 0.001$  with pairwise likelihood). The other infections have also significantly higher onset for the winter period ( $RR : 2.66$ ;  $95\% CI : (0.56, 1.39)$ ,  $p < 0.001$  with pairwise likelihood). The level of hemoglobin does not affect the onset of all three types of events. Digestive disorders and other infections show het-

Table 5: Joint analysis of three types of infections based on a semiparametric model; digestive disorder, lower respiratory infections (LRTI) and other infections including an iron supplement, phase, and hemoglobin level (Hb) as covariates with lognormal marginal distribution for random effects and Gaussian copula functions.

<i>Semiparametric</i>		Pairwise Likelihood			Two-stage PW Likelihood		
TYPE	Covariate	EST	SE	p	EST	SE	p
COEFFICIENTS							
Digestive disorder	Iron	-0.202	0.123	0.101	-0.204	0.123	0.100
	Phase	0.368	0.123	0.003	0.369	0.123	0.003
	Hb	0.020	0.132	0.877	0.020	0.133	0.883
LRTI	Iron	-0.169	0.091	0.062	-0.170	0.091	0.062
	Phase	0.669	0.101	< 0.001	0.669	0.101	< 0.001
	Hb	-0.052	0.104	0.618	-0.051	0.104	0.623
Other infections	Iron	0.341	0.197	0.083	0.340	0.198	0.086
	Phase	0.979	0.212	< 0.001	0.980	0.212	< 0.001
	Hb	-0.106	0.240	0.660	-0.103	0.241	0.669
RANDOM EFFECTS							
	$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	0.166, (0.100)	0.045, (0.001)	0.119 (0.105)	0.158 (0.078)	0.046 (0.001)	0.109 (0.069)
DEPENDENCE PARAMETERS							
	$(\rho_{12}, \rho_{13}, \rho_{23})$	-0.329 (0.535)	-0.738 (0.148)	0.411 (0.532)	-0.324 (0.532)	-0.715 (0.150)	0.384 (0.418)

erogeneity where  $\hat{\sigma}_1^2 = 0.17$ ,  $\hat{\sigma}_3^2 = 0.12$  with pairwise likelihood whereas LRTI has small variance of random effects ( $\hat{\sigma}_2^2 = 0.05$ ). From the estimates of dependence parameters, digestive disorders have a strong negative association with other infections ( $\hat{\rho}_{13} = -0.74$  with pairwise likelihood). There is a negative association between digestive infections and LRTI ( $\hat{\rho}_{12} = -0.33$ ), and a positive association between LRTI and other infections ( $\hat{\rho}_{23} = 0.41$  with pairwise likelihood).

Figure 3 shows the estimated cumulative baseline rates of each episode under the pairwise likelihood. The disorders of digestion for the placebo group having hemoglobin level less than 9.1 g/L have approximately 1.15 events occurrence for the period 12/2007-06/2008 (phase 1). The expected number of events of LRTI for the same group is approximately 1.5 which is higher incidence compared to the onset of digestive disorders. The other infections have less than 0.2 events for these group.

## 6 DISCUSSION

We propose the use of a multivariate random effects distribution to model heterogeneity in the risk of several types of recurrent events based on a mixed Poisson model formulation. The joint distribution of the random effects is constructed via a Gaussian copula model which means that the measures of overdispersion for each type of event are functionally independent and that the dependence structure is quite general. Semiparametric estimation is carried out using a composite likelihood expectation-

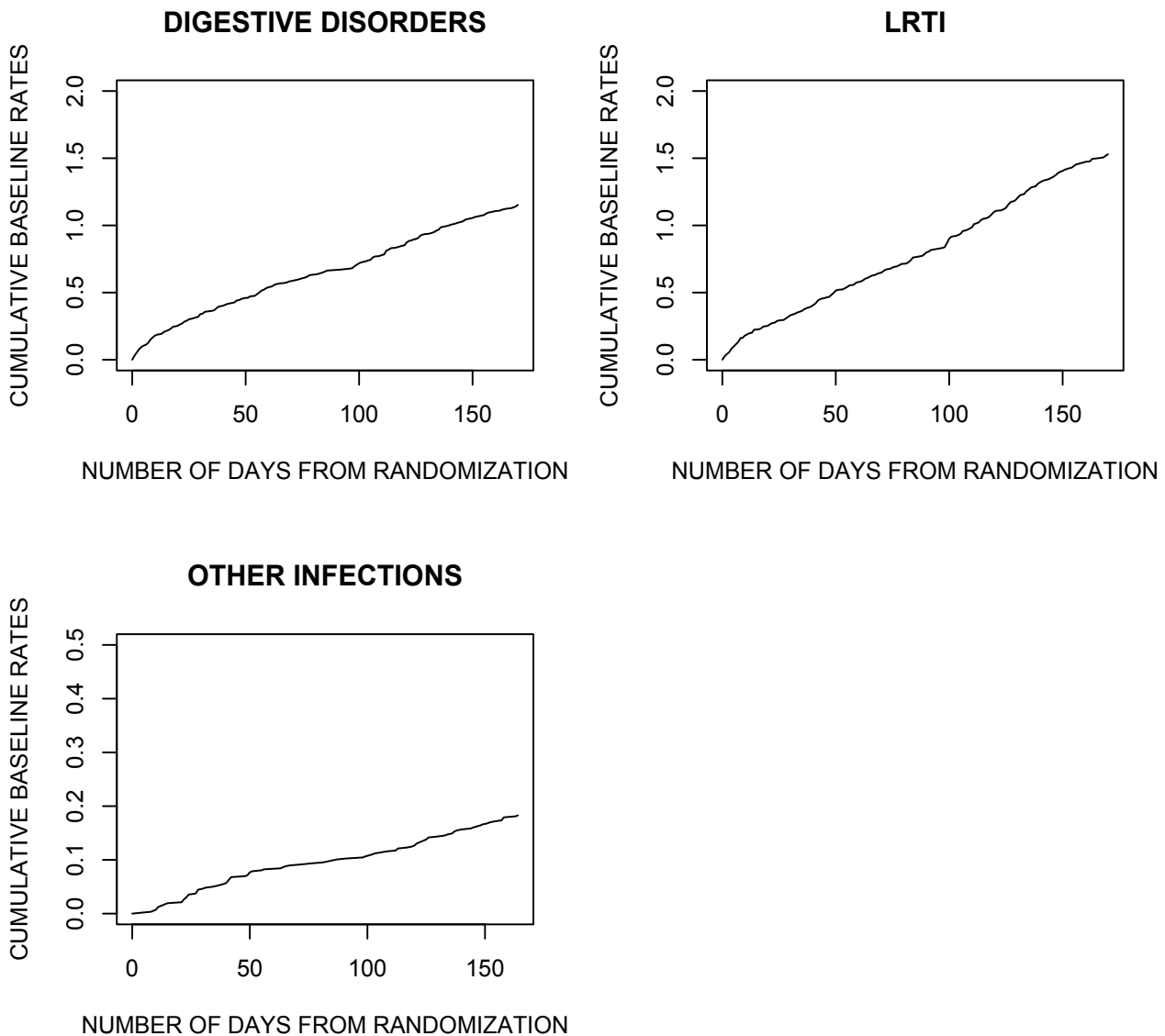


Figure 3: Plots of estimated cumulative baseline rates of digestive disorders, LRTI, and other infections for the pairwise likelihood analysis using the joint model

maximization algorithm; an even more computationally efficient two-stage estimation procedure is also developed which simply uses a working independence assumption at the first stage. Large sample variance estimates are derived for both approaches and are shown to be valid in finite samples in empirical studies. The approach is particularly appealing for use in settings with many different types of events. We note that a marginal distribution of  $U_{ij}$  can be a log-normal distribution, a gamma distribution, and etc. However, the provided code includes a log-normal distribution since it has a closed form in E-step in EM algorithm as shown in Appendix B.

This general approach can be naturally extended to accommodate multi-type interval-censored recurrent event data of the sort studied by Chen et al. (2005) where the exact event times are unavailable but the interval of event times of each type is known. The fact that this method is implemented using an expectation-maximization algorithm means that it can also be generalized to accommodate settings where the event types are partially missing; see Chen and Cook (2009).

An alternative framework for analyzing multi-type recurrent events is via marginal methods and estimating functions (e.g. Cai and Schaubel, 2004). While this can be appealing because it is based on



partially specified models it does not lend itself naturally to prediction. Fully specified models, even when fitted under composite likelihood, can ensure prediction of future events of any type or all types, exploits the history of the joint processes which should yield more accurate and precise predictions of features of the recurrent event processes (Fredette and Lawless, 2007).

## ACKNOWLEDGEMENTS

The authors thank Dr. Stanley Zlotkin and Dr. Mathieu Lemaire for collaboration on the iron supplementation study and permission to use the data here and Jerry Lawless for comments on an earlier version of this manuscript. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (RGPIN 155849) and the Canadian Institutes for Health Research (FRN 13887). Richard Cook is a Canada Research Chair in Statistical Methods for Health Research.

## CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

## REFERENCES

- Ambrus, J. (2004). Nutrition and infectious diseases in developing countries and problems of acquired immunodeficiency syndrome. *Experimental Biology and Medicine*, 229(6):464–472.
- Andersen, E. (2004). Composite likelihood and two-stage estimation in family studies. *Biostatistics*, 5(1):15–30.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B*, pages 192–236.
- Cai, J. and Schaubel, D. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis*, 10(2):121–138.
- Chen, B. and Cook, R. (2009). The analysis of multivariate recurrent events with partially missing event types. *Lifetime Data Analysis*, 15(1):41.
- Chen, B., Cook, R., Lawless, J., and Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine*, 24(5):671–691.
- Chen, X., Wang, Q., Cai, J., and Shankar, V. (2012). Semiparametric additive marginal regression models for multiple type recurrent events. *Lifetime Data Analysis*, 18(4):504–527.
- Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media.
- Cook, R., Lawless, J., and Lee, K. (2010). A copula-based mixed poisson model for bivariate recurrent events under event-dependent censoring. *Statistics in Medicine*, 29(6):694–707.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B*, pages 1–38.
- Fredette, M. and Lawless, J. (2007). Finite-horizon prediction of recurrent events, with application to forecasts of warranty claims. *Technometrics*, 49(1):66–80.

- Gladman, D., Ibañez, D., and Urowitz, M. (2002). Systemic lupus erythematosus disease activity index 2000. *Journal of Rheumatology*, 29(2):288–291.
- Godambe, V. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211.
- Jayaram, L., Pizzichini, M., Cook, R., Boulet, L., Lemiere, C., Pizzichini, E., Cartier, A., Hussack, P., Goldsmith, C., Laviolette, M., et al. (2006). Determining asthma treatment by monitoring sputum cell counts: effect on exacerbations. *European Respiratory Journal*, 27(3):483–494.
- Klein, J. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, pages 795–806.
- Lawless, J. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association*, 82(399):808–815.
- Lemaire, M., Islam, Q., Shen, H., Khan, M., Parveen, M., Abedin, F., Haseen, F., Hyder, Z., Cook, R., and Zlotkin, S. (2011). Iron-containing micronutrient powder provided to children with moderate-to-severe malnutrition increases hemoglobin concentrations but not the risk of infectious morbidity: a randomized, double-blind, placebo-controlled, noninferiority safety trial. *The American Journal of Clinical Nutrition*, 94(2):585–593.
- Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80(1):221–39.
- Mazroui, Y., Mathoulin-Pélissier, S., MacGrogan, G., Brouste, V., and Rondeau, V. (2013). Multivariate frailty models for two types of recurrent events with a dependent terminal event: application to breast cancer data. *Biometrical Journal*, 55(6):866–884.
- Murphy, S. A. and Van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–465.
- Nadeau, C. and Lawless, J. (1998). Inference for means and covariances of point processes through estimating functions. *Biometrika*, 85(4):893–906.
- Nelsen, R. (2007). *An Introduction to Copulas*. Springer Science & Business Media.
- Ng, E. and Cook, R. (1999). Robust inference for bivariate point processes. *Canadian Journal of Statistics*, 27(3):509–524.
- Schaible, U. and Stefan, H. (2007). Malnutrition and infection: complex mechanisms and global impacts. *PLoS Med*, 4(5):e115.
- Silberstein, S. D., Olesen, J., Bousser, M., Diener, H., Dodick, D., First, M., Goadsby, P., Göbel, H., Lainez, M., Lance, J., et al. (2005). The international classification of headache disorders, (ichd-ii)—revision of criteria for 8.2 medication-overuse headache. *Cephalalgia*, 25(6):460–465.
- Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92(1):1–28.
- Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.
- Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canadian Journal of Statistics*, 33(3):335–356.

## APPENDIX A: CALCULATION OF THE VARIANCE ESTIMATES IN TWO-STAGE ESTIMATION

In (15),  $A(\Omega)$  and  $B(\Omega)$  are as follows in two-stage estimation. We first let

$$mpl_n^j(\zeta) = \sum_{i=1}^n \frac{1}{n} mq_i^j(\zeta, \Lambda)$$

where  $\zeta = (\beta_j', \sigma_j)'$ , and  $mq_i^j(\zeta, \Lambda) = \log \int_0^\infty L_{ij}(\theta_j | u_{ij}) dG_j(u_{ij}; \sigma_j)$ . We let  $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$ , where

$$A_1 = \begin{pmatrix} A_{1,1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{1,2} & \mathbf{0} & \vdots & \vdots \\ \vdots & \mathbf{0} & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \cdots & A_{1,J} & \mathbf{0} \end{pmatrix} \text{ with}$$

$$A_{1,j}[k, l] = \max_{\Lambda_{j0}} \frac{1}{n} \frac{mpl_n^j(\zeta + h\mathbf{e}_k + h\mathbf{e}_l) - mpl_n^j(\zeta + h\mathbf{e}_k - h\mathbf{e}_l) - mpl_n^j(\zeta - h\mathbf{e}_k + h\mathbf{e}_l) + mpl_n^j(\zeta - h\mathbf{e}_k - h\mathbf{e}_l)}{4h^2},$$

$j = 1, \dots, J$ , and

$$A_{2,j}[k, l] = \max_{\mathbf{\Lambda}} \frac{1}{n} \frac{pl_n^M(\gamma + h\mathbf{e}_k + h\mathbf{e}_l) - pl_n^M(\gamma + h\mathbf{e}_k - h\mathbf{e}_l) - pl_n^M(\gamma - h\mathbf{e}_k + h\mathbf{e}_l) + pl_n^M(\gamma - h\mathbf{e}_k - h\mathbf{e}_l)}{4h^2},$$

$j = 1, \dots, J$ .

Likewise, we can write  $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$ , where  $B_1 = \begin{pmatrix} B_{1,1} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B_{1,2} & \mathbf{0} & \vdots & \vdots \\ \vdots & \mathbf{0} & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \cdots & B_{1,J} & \mathbf{0} \end{pmatrix}$  with

$$B_{1,j}(k, l) = \frac{1}{4nh^2} \sum_{i=1}^n [mq_i^j(\zeta + h\mathbf{e}_k, \Lambda_{\zeta+h\mathbf{e}_k}) - mq_i^j(\zeta - h\mathbf{e}_k, \Lambda_{\zeta-h\mathbf{e}_k})] \\ \times [mq_i^j(\zeta + h\mathbf{e}_l, \Lambda_{\zeta+h\mathbf{e}_l}) - mq_i^j(\zeta - h\mathbf{e}_l, \Lambda_{\zeta-h\mathbf{e}_l})],$$

and

$$B_{2,j}(k, l) = \frac{1}{4nh^2} \sum_{i=1}^n [q_i(\gamma + h\mathbf{e}_k, \Lambda_{\gamma+h\mathbf{e}_k}) - q_i(\gamma - h\mathbf{e}_k, \Lambda_{\gamma-h\mathbf{e}_k})] \\ \times [q_i(\gamma + h\mathbf{e}_l, \Lambda_{\gamma+h\mathbf{e}_l}) - q_i(\gamma - h\mathbf{e}_l, \Lambda_{\gamma-h\mathbf{e}_l})].$$

## APPENDIX B: THE CONDITIONAL SCORE VECTOR

The conditional score vector  $\partial L_{ij}(\theta_j | u_{ij}) / \partial \theta_j$  for  $\beta_j$ ,  $d\Lambda_{j0}(\cdot)$  from the complete data pairwise likelihood (8) is given as

$$\frac{\partial L_{ij}(\theta_j | u_{ij})}{\partial \beta_j} = \sum_{i=1}^n \left\{ N_{ij}(C_i) x_{ij} - u_{ij} \int_0^\infty \bar{Y}_i(v) x_{ij} \exp(x'_{ij} \beta_j) d\Lambda_{j0}(v) \right\},$$

$$\frac{\partial L_{ij}(\theta_j | u_{ij})}{\partial d\Lambda_{j0}(t_{jk})} = \frac{1}{d\Lambda_{j0}(t_{jk})} - \sum_{i=1}^n u_{ij} \bar{Y}_i(t_{jk}) \exp(x'_{ij} \beta_j),$$

where  $t_{jk}$  is the  $k^{\text{th}}$  time of type  $j$  event occurrence.

In (8),  $\log dG(u_{ij}, u_{ik}; \phi_{jk})$  is given as

$$\begin{aligned} \log dG(u_{ij}, u_{ik}; \phi_{jk}) = & \sum_{i=1}^n \left[ \frac{3}{2} \log(u_{ij}) - \frac{\log^2(u_{ij})}{2\sigma_j^2} - \log(\sigma_j) - \frac{\sigma_j^2}{8} + \frac{3}{2} \log(u_{ik}) - \frac{\log^2(u_{ik})}{2\sigma_k^2} - \log(\sigma_k) - \frac{\sigma_k^2}{8} \right. \\ & - \frac{\log(1 - \rho_{jk}^2)}{2} - \frac{\rho_{jk}^2(\sigma_j^2 + \sigma_k^2)}{8(1 - \rho_{jk}^2)} + \frac{\rho_{jk}\sigma_j\sigma_k}{4(1 - \rho_{jk}^2)} \\ & \left. - \frac{\rho_{jk}^2}{2(1 - \rho_{jk}^2)} \left\{ \frac{\log^2(u_{ij})}{\sigma_j^2} + \log(u_{ij}) + \frac{\log^2(u_{ik})}{\sigma_k^2} + \log(u_{ik}) \right\} \right. \\ & \left. + \frac{\rho_{jk}}{1 - \rho_{jk}^2} \left\{ \frac{\log(u_{ij})\log(u_{ik})}{\sigma_j\sigma_k} + \frac{\sigma_k \log(u_{ij})}{2\sigma_j} + \frac{\sigma_j \log(u_{ik})}{2\sigma_k} \right\} \right]. \end{aligned}$$

The components of the conditional score vector  $\partial \log dG(u_{ij}, u_{ik}; \phi_{jk}) / \partial \phi_{jk}$  are given as follows.

$$\begin{aligned} \frac{\partial \log dG(u_{ij}, u_{ik}; \phi_{jk})}{\partial \sigma_j} &= -\frac{n}{\sigma_j} - \frac{n\sigma_j}{4(1 - \rho_{jk}^2)} + \frac{n\rho_{jk}\sigma_k}{4(1 - \rho_{jk}^2)} \\ &+ \sum_{i=1}^n \frac{\log^2(u_{ij})}{(1 - \rho_{jk}^2)\sigma_j^3} - \sum_{i=1}^n \frac{\rho_{jk} \log(u_{ij}) \log(u_{ik})}{(1 - \rho_{jk}^2)\sigma_j^2\sigma_k} + \frac{\rho_{jk}}{1 - \rho_{jk}^2} \sum_{i=1}^n \left\{ \frac{-\sigma_k \log(u_{ij})}{2\sigma_j^2} + \frac{\log(u_{ik})}{2\sigma_k} \right\}, \\ \frac{\partial \log dG(u_{ij}, u_{ik}; \phi_{jk})}{\partial \sigma_k} &= -\frac{n}{\sigma_k} - \frac{n\sigma_k}{4(1 - \rho_{jk}^2)} + \frac{n\rho_{jk}\sigma_j}{4(1 - \rho_{jk}^2)} \\ &+ \sum_{i=1}^n \frac{\log^2(u_{ik})}{(1 - \rho_{jk}^2)\sigma_k^3} - \sum_{i=1}^n \frac{\rho_{jk} \log(u_{ij}) \log(u_{ik})}{(1 - \rho_{jk}^2)\sigma_j\sigma_k^2} + \frac{\rho_{jk}}{1 - \rho_{jk}^2} \sum_{i=1}^n \left\{ \frac{-\sigma_j \log(u_{ik})}{2\sigma_k^2} + \frac{\log(u_{ij})}{2\sigma_j} \right\}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \log dG(u_{ij}, u_{ik}; \phi_{jk})}{\partial \rho_{jk}} &= \frac{n\rho_{jk}}{1 - \rho_{jk}^2} - \frac{n\rho_{jk}(\sigma_j^2 + \sigma_k^2)}{4(1 - \rho_{jk}^2)^2} + \frac{n\sigma_j\sigma_k(1 + \rho_{jk}^2)}{4(1 - \rho_{jk}^2)^2} \\ &- \frac{\rho_{jk}}{(1 - \rho_{jk}^2)^2} \sum_{i=1}^n \left\{ \frac{\log^2(u_{ij})}{\sigma_j^2} + \frac{\log^2(u_{ik})}{\sigma_k^2} + \log(u_{ij}) + \log(u_{ik}) \right\} \\ &+ \frac{1 + \rho_{jk}^2}{(1 - \rho_{jk}^2)^2} \sum_{i=1}^n \left\{ \frac{\log(u_{ij})\log(u_{ik})}{\sigma_j\sigma_k} + \frac{\sigma_k \log(u_{ij})}{2\sigma_j} + \frac{\sigma_j \log(u_{ik})}{2\sigma_k} \right\}. \end{aligned}$$

In two-stage estimation,  $\log dG(u_{ij}; \sigma_j)$  is given as

$$\log dG(u_{ij}; \sigma_j) = \sum_{i=1}^n \left[ \frac{3}{2} \log(u_{ij}) - \frac{\log^2(u_{ij})}{2\sigma_j^2} - \log(\sigma_j) - \frac{\sigma_j^2}{8} \right].$$

The conditional score function for  $\sigma_j$  from  $\log dG(u_{ij}; \sigma_j)$  is given as

$$\frac{\partial \log dG(u_{ij}; \sigma_j)}{\partial \sigma_j} = -\frac{n}{\sigma_j} + \frac{\log^2(u_{ij})}{\sigma_j^3} - \frac{\sigma_j^3}{4},$$

for  $j = 1, 2, \dots, J$ .