



Simulated maximum likelihood estimation in joint models for multiple longitudinal markers and recurrent events of multiple types, in the presence of a terminal event

M. H. Hof, J. Z. Musoro, R. B. Geskus, G. H. Struijk, I. J. M. ten Berge & A. H. Zwinderman

To cite this article: M. H. Hof, J. Z. Musoro, R. B. Geskus, G. H. Struijk, I. J. M. ten Berge & A. H. Zwinderman (2016): Simulated maximum likelihood estimation in joint models for multiple longitudinal markers and recurrent events of multiple types, in the presence of a terminal event, Journal of Applied Statistics, DOI: [10.1080/02664763.2016.1262336](https://doi.org/10.1080/02664763.2016.1262336)

To link to this article: <http://dx.doi.org/10.1080/02664763.2016.1262336>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 01 Dec 2016.



[Submit your article to this journal](#)



Article views: 147



[View related articles](#)



[View Crossmark data](#)



Simulated maximum likelihood estimation in joint models for multiple longitudinal markers and recurrent events of multiple types, in the presence of a terminal event

M. H. Hof^a, J. Z. Musoro^a, R. B. Geskus^a, G. H. Struijk^b, I. J. M. ten Berge^b and A. H. Zwinderman^a

^aDepartment of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands; ^bDepartment of Nephrology, Academic Medical Center, Amsterdam, The Netherlands

ABSTRACT

In medical studies we are often confronted with complex longitudinal data. During the follow-up period, which can be ended prematurely by a terminal event (e.g. death), a subject can experience recurrent events of multiple types. In addition, we collect repeated measurements from multiple markers. An adverse health status, represented by 'bad' marker values and an abnormal number of recurrent events, is often associated with the risk of experiencing the terminal event. In this situation, the missingness of the data is not at random and, to avoid bias, it is necessary to model all data simultaneously using a joint model. The correlations between the repeated observations of a marker or an event type within an individual are captured by normally distributed random effects. Because the joint likelihood contains an analytically intractable integral, Bayesian approaches or quadrature approximation techniques are necessary to evaluate the likelihood. However, when the number of recurrent event types and markers is large, the dimensionality of the integral is high and these methods are too computationally expensive. As an alternative, we propose a simulated maximum-likelihood approach based on quasi-Monte Carlo integration to evaluate the likelihood of joint models with multiple recurrent event types and markers.

ARTICLE HISTORY

Received 24 September 2015
Accepted 9 November 2016

KEYWORDS

Joint model; numerical integration; quasi-Monte Carlo integration; shared frailties; simulated maximum likelihood

CLASSIFICATION CODES

62N02; 65D30

1. Introduction

Joint modeling of longitudinal and time-to-event data is increasingly being used in medical research. Several studies monitor markers of patient's state of health over time in relation to the occurrence of event(s). In the situation in which we only have a small number of marker measurements and the markers are associated with events that end the follow-up period (e.g. death), it is necessary to estimate the marker trajectories and the risk of experiencing this final event simultaneously to obtain unbiased estimates of the parameters in the models of the markers and of the events [14]. Some key references on joint modeling techniques are [12,24,46], and comprehensive overviews can be found in [41,45,48].

CONTACT M. H. Hof m.h.hof@amc.uva.nl

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Recent years have seen the extension of the application of joint models from the traditional single longitudinal outcome and single terminal event type to settings with; (i) multiple longitudinal outcomes and a single terminal event type [39,44,47], (ii) a single longitudinal outcome and multiple terminal event types (i.e. a competing risks situation) [10,20,28], (iii) multiple longitudinal outcomes and multiple terminal event types [6], and (iv) multiple longitudinal outcomes, a single recurrent event type, and a single terminal event type [30].

In this paper, we consider joint models for data consisting of multiple markers, multiple recurrent event types, and a terminal event. Our model consists of (i) a multivariate mixed effects submodel where the markers are allowed to develop simultaneously through time, thus accounting for the interrelation between the marker measurements from the same individual, as well as the intra-individual dependency over time. (ii) A survival submodel that shares latent terms with the longitudinal submodel, and contained recurrent event-type specific frailties which were assumed to be interrelated to each other. The longitudinal markers were assumed to influence the risk of each recurrent event type via their true values. (iii) A survival model for the terminal event which was associated with both the longitudinal and the recurrent event submodels via different shared latent terms.

The main problem of joint models for multiple markers and multiple (recurrent) events is the dimensionality of the (unobserved) random effects and frailties in, respectively, the submodels in (i) and (ii). Bayesian approaches [3,32], likelihood approaches using quadrature methods for evaluating the integrals [30,39], and EM-type of approaches [31] have been used to estimate the parameters of joint models. However, even for a small number of markers and recurrent event types (e.g. 3 markers and 3 recurrent event types), the number of random effects and frailty terms may become large. Moreover, the number of random effects and frailties may be very large when nonlinear time trends are assumed. Consequently, these traditional methods to maximize the log likelihood of joint models sometimes fail to converge or are very computationally expensive.

To overcome the computational problems of joint models with high dimensional random effects and frailties, we propose to use simulated maximum likelihood (SML) techniques [19]. With SML estimation, Monte Carlo (MC) techniques are used to integrate the random effects and frailties out of the likelihood which is maximized to estimate the structural parameters of the joint model. Vectors of random effects and frailties are drawn from their multivariate distributions, and the average (log) likelihood over all drawn vectors is the approximated log likelihood. The number of vectors that is necessary is usually a trade-off between computational speed and desired accuracy; when too few vectors are drawn, simulation bias may be introduced in the estimation procedure. Generally, random draws from the random effects and frailties distributions are inefficient and a large number of draws is necessary to obtain accurate results. Quasi-Monte Carlo (QMC) techniques have been proposed to improve the performance of the MC integration method [26]. As opposed to MC techniques, in which a set of random vectors of random effects and frailties is used, QMC techniques use deterministic points that approximate the distributions of the random effects and frailties as closely as possible.

The rest of this paper is organized as follows. Firstly, we introduce the post kidney transplantation data, which is used as example throughout this paper. Secondly, we introduce the submodels of the joint model, along with the joint model likelihood. Thirdly, the SML estimator is introduced using the quasi-random sequences to approximate the likelihood

Table 1. Number of recurrences within patients for all 4 recurrent infections.

Infection type	Frequency										
	0	1	2	3	4	5	6	7	8	9	10
Cytomegalovirus	244	87	21	4	0	0	1				
Upper respiratory	224	83	31	13	4	1	1				
Wound	260	71	17	4	3	1	0	0	0	1	
Urinary tract	199	61	44	13	13	11	6	5	3	1	1

of the joint model and we describe how we can use the QMC technique for prediction. Finally, we illustrate the properties of the SML estimator using simulations and we fit our joint model to the post kidney transplantation data.

2. Post kidney transplant data

This study was motivated by the follow-up data from 357 patients who had received a kidney transplant between 2000 and 2009 at the Academic Medical Center in Amsterdam [42]. At the time of the transplant, the baseline covariates age (in years), gender (female), type of immunosuppressive treatment (type *A*/ type *B*) were registered. The maximum follow-up time for all patients was seven years, after which all patients were assumed to be right-censored non-informatively. From all patients, 44 (22%) reached the maximum follow-up time.

The health status of each patient was measured by four longitudinal markers of the patient’s immune system, that is, CD4+ T cells, CD8+ T cells, B cells, and natural killer cells, and the occurrence of four recurrent infection types. All four markers were measured at the same time points. The number of repeated marker measurements per patient varied between 0 and 18, with a median of 2. There were 21 (6%) patients without any marker measurement. The recurrent infection types and their number of recurrences are given in Table 1.

Our objectives were to estimate the associations between the marker trajectories and the risk of experiencing a recurrent infection type, the associations between the recurrent event types, and the effect of the marker trajectories and the occurrence of the recurrent event types on the risk of dropping out before the maximum follow-up time. Patients who were more frail might have experienced more recurrent infections. In addition, more frail patients might have dropped-out from the study prematurely, causing informative right-censoring. To obtain unbiased estimates of all associations, a joint model was fitted to the data. We simultaneously estimated the submodels regarding the evolution of the markers through time, the risks of the recurrent event types, and the risk of dropping out before the maximum follow-up time.

3. Methods

3.1. Submodels specification

Consider that we have longitudinal data of n subjects from M measured markers, and R recurrent event types. For subject i ($i = 1, \dots, n$), let T_i^* and C_i be the terminal event (e.g. death) and censoring time. Write $T_i = \min(T_i^*, C_i)$ as the follow-up time and $\delta_{i0} =$

$I(T_i^* < C_i)$ as the indicator whether the follow-up period of subject i was ended by a terminal event, where $I()$ is the indicator function. The censoring time C_i is assumed to be independent of the markers and events processes. The time T_i^* at which subject i experiences a terminal event is allowed to depend on the markers and events processes. During the follow-up period, subject i can experience recurrent events of different types and have repeated measurements of multiple markers.

Let $y_{im}(t)$ be the value of marker $m = 1, \dots, M$ at time t for the i th subject. The m th marker is observed J_{mi} times in the interval $[0, T_i)$, leading to the vector of observed measurements $\mathbf{y}_{im} = \{y_{im}(t_{im,j}); 0 \leq t_{im,j} < T_i \text{ and } j = 1, \dots, J_{mi}\}$. Note that the number of measurements may differ between individuals and that all markers could be measured at different time-points. The times at which we observe recurrent event types $r = 1 \dots, R$ are described by the vector $\mathbf{t}_{ir} = \{t_{ir,k}; 0 < t_{ir,k} \leq T_i \text{ and } k = 1, \dots, K_{ri}\}$, where K_{ri} is the number of times recurrent event type r occurs in subject i .

We assume that, given the random effects of the M markers $\mathbf{v}_i = \{v_{im}; m = 1, \dots, M\}$, all marker values are uncorrelated. Each marker has its own submodel, where the parametrization of each submodel depends on the marker type (e.g. continuous [23] or binary [38]). For the post kidney transplant data, we assumed that the m th marker for subject i at time t could be described by the linear mixed effects model

$$\begin{aligned} y_{im}(t) &= y_{im}^*(t) + \epsilon_{im}(t), \\ y_{im}^*(t) &= \boldsymbol{\phi}_m' \mathbf{z}_{im}(t) + \mathbf{v}_{im}' \mathbf{q}_{im}(t), \end{aligned} \quad (1)$$

where $y_{im}^*(t)$ is the true marker value. In addition, $\mathbf{z}_{im}(t)$ and $\mathbf{q}_{im}(t)$ denote the design vectors of, respectively, fixed and random effects at time t , $\boldsymbol{\phi}_m$ are fixed effects, and \mathbf{v}_{im} is the vector of random effects. The vectors $\mathbf{z}_{im}(t)$ and $\mathbf{q}_{im}(t)$ may contain functions of time t to capture any (non)linear effect of time on the marker trajectories. Finally, $\epsilon_{im}(t)$ is a normally distributed measurement error with variance α_m^2 , which is assumed to be white noise.

For subject i , $h_{ir}(t)$ and $\lambda_i(t)$ are the hazards of the r th recurrent event type and the terminal event at time t . For the recurrent events, we assume that, given the frailties of the R recurrent event types $\mathbf{w}_i = \{w_{ir}; r = 1, \dots, R\}$, all recurrent event processes are uncorrelated. The hazard for the r th recurrent event type at time t is

$$h_{ir}(t) = h_{0r}(t) \exp \left\{ \boldsymbol{\beta}_r' \mathbf{x}_{ir}(t) + w_{ir} + \sum_{m=1}^M \gamma_{rm} y_{im}^*(t) \right\}, \quad (2)$$

where $h_{0r}(t)$ is the baseline hazard for event type r at time t , $\mathbf{x}_{ir}(t)$ is a vector of (possibly time-varying) characteristics of subject i at time t , $\boldsymbol{\beta}_r$ are the fixed effects of $\mathbf{x}_{ir}(t)$, w_{ir} is a frailty term capturing the correlation between the recurrent events of type r , and the last term of Equation (2) denotes the effects of the true marker values $y_{im}^*(t)$ of the M markers on the risk of recurrent event type r at time t , where the effect of the m th marker value $y_{im}^*(t)$ on recurrent event type r is given by γ_{rm} . For example, $\gamma_{rm} = 0$ implies that there is no effect of the m th marker value on the risk of experiencing the r th recurrent event type.

The hazard of the terminal event at time t depends on the R frailty terms of the recurrent event types \mathbf{w}_i and on the M true marker values $y_{im}^*(t)$ at time t . Therefore, the hazard for

the terminal event at time t can be written as

$$\lambda_i(t) = \lambda_0(t) \exp \left\{ \boldsymbol{\beta}'_0 \mathbf{x}_{i0}(t) + \sum_{m=1}^M \gamma_{0m} y_{im}^*(t) + \sum_{r=1}^R \chi_r w_{ir} \right\}, \quad (3)$$

where $\lambda_0(t)$ is the baseline hazard of the terminal event at time t , $\mathbf{x}_{i0}(t)$ the vector of (possibly time-varying) covariates related at time t to the terminal event, and $\boldsymbol{\beta}_0$ the regression weights related to $\mathbf{x}_{i0}(t)$. The last two terms denote, respectively, the effect of the M markers and the effect of the R frailties of the recurrent event types on the hazard of experiencing the terminal event at time t . Because the event and marker submodels share the random effects of the markers, joint models that are based on these assumptions are also known as shared parameter models [29–31,39].

3.2. The joint likelihood function

We assume that \mathbf{v}_i and \mathbf{w}_i follow independent multivariate normal distributions centered at zero with covariance matrices $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_w$, respectively, that is, $\mathbf{v}_i \sim N(0, \boldsymbol{\Sigma}_v)$ and $\mathbf{w}_i \sim N(0, \boldsymbol{\Sigma}_w)$. In the post kidney transplant data, we assumed independence between the random effects of *different* markers, that is,

$$\boldsymbol{\Sigma}_v = \begin{pmatrix} \boldsymbol{\Sigma}_{v1} & 0 & \dots & 0 \\ 0 & \boldsymbol{\Sigma}_{v2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{\Sigma}_{vM} \end{pmatrix}.$$

This choice reduced the number of parameters in our joint model substantially. A disadvantage of this choice is that we could lose some efficiency, because the dependencies *between* markers were ignored.

Let $\boldsymbol{\theta} = \{\mathbf{h}_0, \lambda_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\chi}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_w\}$ be the collection of all joint model parameters, where $\mathbf{h}_0 = \{h_{0r}; r = 0, \dots, R\}$ and λ_0 are, respectively, the baseline hazards of the recurrent and the terminal event types. In addition, the vector $\boldsymbol{\gamma} = \{\gamma_{rm}; r = 0, \dots, R \text{ and } m = 1, \dots, M\}$ contains the associations between all markers and event types. The parameter vectors $\boldsymbol{\chi} = \{\chi_r; r = 1, \dots, R\}$, $\boldsymbol{\phi} = \{\phi_m; m = 1, \dots, M\}$ and $\boldsymbol{\alpha} = \{\alpha_m; m = 1, \dots, M\}$ are used in the marker submodels from Equation (1). The *observed* data from subject i is $\mathcal{D}_i(T_i) = \{\mathcal{Y}_i, \mathcal{Z}_i, T_i\}$, where $\mathcal{Y}_i = \{y_{im}(t_{im,j}); 0 \leq t_{im,j} \leq T_i \text{ and } m = 1, \dots, M \text{ and } j = 1, \dots, J_{m_i}\}$ and $\mathcal{Z}_i = \{t_{ir,k}; 0 \leq t_{ir,k} \leq T_i \text{ and } r = 0, \dots, R \text{ and } k = 1, \dots, K_{r_i}\}$ are, respectively, all observed marker measurements and all observed event times until time T_i . Assuming that the data from the different subjects are mutually independent, the log joint likelihood of the data is given by $\sum_{i=1}^n \log[\mathcal{L}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta})]$, where $\mathcal{L}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta})$ is the subject-specific likelihood contribution, that is,

$$\begin{aligned} \mathcal{L}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta}) &= \int_{(\mathbf{v}_i, \mathbf{w}_i)} L_{i0}(T_i | \lambda_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\phi}, \boldsymbol{\chi}, \mathbf{v}_i, \mathbf{w}_i) \times \prod_{r=1}^R L_{ir}(\mathbf{t}_{ir} | h_{0r}, \boldsymbol{\beta}_r, \boldsymbol{\gamma}_r, \boldsymbol{\phi}, \mathbf{v}_i, \mathbf{w}_{ir}) \\ &\times \prod_{m=1}^M L_{im}(\mathbf{y}_{im} | \boldsymbol{\phi}_m, \boldsymbol{\alpha}_m, \mathbf{v}_{im}) \times p(\mathbf{v}_i | \boldsymbol{\Sigma}_v) p(\mathbf{w}_i | \boldsymbol{\Sigma}_w) d(\mathbf{v}_i, \mathbf{w}_i). \end{aligned} \quad (4)$$

The integral in this $\mathcal{L}_i(\mathcal{D}_i(T_i) | \theta)$ is of dimension $d = d_v + d_w$, which are the sizes of, respectively, the vectors \mathbf{v}_i and \mathbf{w}_i . Both d_v and d_w depend on the number of markers M , the number of recurrent event types R , and the flexibility of the event and marker submodels. For instance, when we have enough data, we might want to estimate multiple random effects per marker m to capture nonlinear trends over time that differ per individual.

The contribution to the likelihood of the measurements of the m th marker from individual i depends on the submodel specification. With the specification from Equation (1), the contribution is

$$L_{im}(\mathbf{y}_{im} | \phi_m, \alpha_m, \mathbf{v}_{im}) = \prod_{j=1}^{J_{m_i}} \frac{1}{\alpha_m \sqrt{2\pi}} \exp \left\{ \frac{-[y_{im}(t_{im,j}) - y_{im}^*(t_{im,j})]^2}{2\alpha_m^2} \right\},$$

In addition, the likelihood contribution of the r th recurrent event type is [7]

$$L_{ir}(\mathbf{t}_{ir} | h_{0r}, \beta_r, \gamma_r, \phi, \mathbf{v}_i, \mathbf{w}_{ir}) = \left[\prod_{k=1}^{K_{r_i}} h_{ir}(t_{ir,k}) \right]^{\delta_{ir}} \times \exp \left\{ - \int_0^{T_i} h_{ir}(u) du \right\}, \quad (5)$$

where δ_{ir} is the indicator that describes whether individual i experienced at least one recurrent event of type r , that is, $\delta_{ir} = I(K_{r_i} > 0)$. Furthermore, the likelihood contribution of the terminal event is

$$L_{i0}(T_i | \lambda_0, \beta_0, \gamma_0, \phi, \chi, \mathbf{v}_i, \mathbf{w}_i) = \lambda_i(T_i)^{\delta_{i0}} \exp \left\{ - \int_0^{T_i} \lambda_i(u) du \right\}. \quad (6)$$

The baseline hazards $h_{0r}(t)$ and $\lambda_0(t)$ in, respectively, (2) and (3) are modeled with parametric functions that may be different for the various recurrent event types and for the terminal event. For instance, we can assume that the baseline hazards can be modeled with spline functions or piecewise linear constant functions.

Calculating the survival functions from Equations (5) and (6) involves the approximation of uni-dimensional integrals. Depending on the shape of the baseline hazard functions, different methods are necessary to accurately estimate the survival functions. In this paper, we only consider smooth functions in the marker submodels and the (recurrent) event submodels. Because these smooth functions are well-approximated by a polynomial functions, we used Gauss–Kronrod quadrature methods to approximate the integrals in the survival functions [36,39].

4. SML estimator

The maximum likelihood (ML) estimator is defined as

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} \sum_{i=1}^n \log[\mathcal{L}_i(\mathcal{D}_i(T_i) | \theta)], \quad (7)$$

where Θ is the parameter space. To obtain the ML estimate, the subject-specific likelihood $\mathcal{L}_i(\mathcal{D}_i(T_i) | \theta)$ needs to be calculated. Note that this likelihood contains an analytically intractable and (potentially) high dimensional integral. Numerical approximation techniques are needed for the evaluation of $\mathcal{L}_i(\mathcal{D}_i(T_i) | \theta)$.

Because we use a fully parametric model, we could use Gauss–Hermite quadrature (GHQ) integration to approximate the integral in Equation (4). For low-dimensional integration problems, observed in joint models with a low number of random effects and frailty terms (e.g. fitting a joint model with only one marker, one recurrent event type, and one terminal event), GHQ has been frequently used [10,24,27,29]. However, since the number of necessary evaluations of the likelihood increases exponentially with the number of random effects and frailty terms, the GHQ method quickly becomes too computationally expensive for high-dimensional integration problems.

An alternative to GHQ integration is to use a Bayesian approach based on Markov chain Monte Carlo (MCMC) algorithms. A problem of MCMC algorithms is that they are inherently serial, that is, each Markov chain iteration depends on the results of the previous iteration. Therefore, it is not possible to decrease the computational costs with, for instance, parallel computing. Musoro *et al.* used a Bayesian approach to fit a joint model to the post kidney transplant data [32]. However, because of the computational costs, they could only include two recurrent event types in their joint model.

A different method, which has not been used before in joint modeling, is to evaluate the integrals with simulation techniques leading to an SML estimator of θ [19,21]. The SML estimator relies on the specification of an unbiased simulator $\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \theta)$ of the density $\mathcal{L}_i(\mathcal{D}_i(T_i) | \theta)$ in such a way that

$$\mathcal{L}_i(\mathcal{D}_i(T_i) | \theta) = \tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \theta) = E_{v_i, w_i} \{ \mathcal{L}_i(\mathcal{D}_i(T_i) | v_i, w_i, \theta) p(v_i | \Sigma_v) p(w_i | \Sigma_w) \},$$

where

$$\begin{aligned} \mathcal{L}_i(\mathcal{D}_i(T_i) | v_i, w_i, \theta) &= L_{i0}(T_i | \lambda_0, \beta_0, \gamma_0, \phi, \chi, v_i, w_i) \times \prod_{r=1}^R L_{ir}(t_{ir} | h_{0r}, \beta_r, \gamma_r, \phi, v_i, w_{ir}) \\ &\times \prod_{m=1}^M L_{im}(y_{im} | \phi_m, \alpha_m, v_{im}). \end{aligned} \quad (8)$$

An unbiased estimator can be obtained with MC integration techniques. We can draw Z independent vectors $(v_i^{(z)}, w_i^{(z)})$, $z = 1, \dots, Z$, from their respective multivariate normal distributions $p(v_i | \Sigma_v)$ and $p(w_i | \Sigma_w)$. With these Z vectors, we calculate

$$\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \theta) = \frac{1}{Z} \sum_{z=1}^Z \mathcal{L}_i(\mathcal{D}_i(T_i) | v_i^{(z)}, w_i^{(z)}, \theta). \quad (9)$$

Following the strong law of large numbers, $\lim_{Z \rightarrow \infty} \tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \theta) = \mathcal{L}_i(\mathcal{D}_i(T_i) | \theta)$. The corresponding SML estimator is

$$\hat{\theta}_{\text{SML}} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^n \log \left[\frac{1}{Z} \sum_{z=1}^Z \mathcal{L}_i(\mathcal{D}_i(T_i) | v_i^{(z)}, w_i^{(z)}, \theta) \right]. \quad (10)$$

An advantage of the SML estimator over MCMC algorithms is that we can use parallel computing for the evaluation of Equation (9); the likelihood given each vector $(v_i^{(z)}, w_i^{(z)})$ can be calculated independently of each other. In addition, compared to GHQ methods

integration, the SML estimator can use an arbitrary number of vectors Z to evaluate the likelihood.

However, careful consideration of the number of vectors Z is necessary. Gouriéroux showed that the SML estimator is only consistent and asymptotically equivalent to the ML estimator if $\sqrt{n}/Z \rightarrow 0$ as $Z \rightarrow \infty$ (see Proposition 3.2, page 44 of [18,19]). Gouriéroux and Montfort also showed that the bias, caused by the simulation procedure, is of order \sqrt{n}/Z when one draws independent vectors $(\mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)})$ from their corresponding densities (see property 3, page 78 of [18]). Therefore, if Z is chosen to be large enough, the simulation bias is negligible [19,21,43].

4.1. Corrected SML estimation

The bias introduced by the simulation procedure should be negligible to obtain valid parameter estimates, which requires a large Z . However, locating $\hat{\boldsymbol{\theta}}_{\text{SML}}$ requires the use of optimization methods (e.g. Newton–Raphson optimization), which generally need to evaluate the log likelihood a substantial number of times. To limit the computational costs, the number of vectors Z might be chosen too small, leading to biased parameter estimates. To (partially) remove the bias in the SML estimator, we propose therefore to perform one additional Gauss–Newton step leading to the corrected SML (CSML) estimator in such a way that

$$\hat{\boldsymbol{\theta}}_{\text{CSML}} = \hat{\boldsymbol{\theta}}_{\text{SML}} + H_n(\hat{\boldsymbol{\theta}}_{\text{SML}})^{-1} E_n[s(\hat{\boldsymbol{\theta}}_{\text{SML}})], \quad (11)$$

where $H_n(\hat{\boldsymbol{\theta}}_{\text{QMC}})$ and $E_n[s(\hat{\boldsymbol{\theta}}_{\text{QMC}})]$ are, respectively, the information matrix and the expected score equations at $\hat{\boldsymbol{\theta}}_{\text{QMC}}$. The information matrix and the expected score equations are defined as

$$H_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \log[\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right)' \left(\frac{\partial \log[\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right), \quad (12a)$$

$$E_n[s(\boldsymbol{\theta})] = \frac{1}{n} \sum_{i=1}^n \frac{\partial \log[\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}. \quad (12b)$$

For the CSML estimator, the information matrix and the expected score equations only have to be obtained once. Therefore, the likelihood has to be evaluated only a few number of times and we can use a huge number of vectors \tilde{Z} , where $\tilde{Z} \gg Z$, to approximate the likelihood. Assuming that the information matrix and the expected score equations are estimated with negligible simulation bias, we can expect that the CSML parameter estimate $\hat{\boldsymbol{\theta}}_{\text{CSML}}$ lies closer to the ML estimate than the initial, less accurate SML estimate $\hat{\boldsymbol{\theta}}_{\text{SML}}$.

5. QMC method

Because MC integration contributes variance to the (C)SML estimates, variance reduction is desirable. One way to reduce the variance in the MC integration approximation is to replace the underlying random sequence with a deterministic sequence, referred

to in the literature as QMC integration [9,16]. Generally, QMC integration gives better approximations of integrals than MC approximation [2,11].

To implement the QMC integration approach, we first re-write the densities regarding \mathbf{v}_i and \mathbf{w}_i into a d -dimensional multivariate normal distribution with covariance matrix $\mathbf{\Omega}$, that is, $p(\mathbf{v}_i | \mathbf{\Sigma}_v)p(\mathbf{w}_i | \mathbf{\Sigma}_w) = p(\mathbf{v}_i, \mathbf{w}_i | \mathbf{\Omega}) \sim N(0, \mathbf{\Omega})$ with

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Sigma}_v & 0 \\ 0 & \mathbf{\Sigma}_w \end{pmatrix}. \quad (13)$$

The deterministic point set, used in QMC integration, is obtained by transforming a low-discrepancy point set, which is generated in the d -dimensional unit hypercube $[0, 1]^d$. Discrepancy is a measure of equidistance between the vectors in the point set; the more even the vectors are spread over the unit hypercube, the lower the discrepancy [33,34]. Examples of sequences that can be used to create low-discrepancy points sets are Halton, Faure, or Sobol sequences [26]. With QMC integration, the simulator $\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta})$ is calculated as follows:

- (1) *Generate the deterministic QMC point set:* Use a low-discrepancy sequence to generate a QMC point set \mathbf{B}_i of size $Z \times d$.
- (2) *Change the integration region:* \mathbf{B}_i contains values $b_{i,zj}$ ($z = 1, \dots, Z$ and $j = 1, \dots, d$) in the d -dimensional unit hypercube $[0, 1]^d$. We transform the values of \mathbf{B}_i to reflect a d -dimensional normal distribution with means zero, variances one, and covariances zero. Let $\tilde{\mathbf{B}}_i$ be the matrix of transformed values with the elements $\tilde{b}_{i,zj} = \Phi^{-1}(b_{i,zj})$, for $j = 1, \dots, d$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function [35].
- (3) *Add the correlation structure:* The point set $\mathbf{U}_i = \{\mathbf{u}_{iz}; z = 1, \dots, Z\}$ is obtained via $\mathbf{U}_i = \tilde{\mathbf{B}}_i \mathbf{Q}$, where \mathbf{Q} is the upper triangular matrix from the Cholesky decomposition $\mathbf{\Omega} = \mathbf{Q}^T \mathbf{Q}$ [17].
- (4) *Obtain the approximated likelihood:* The point set \mathbf{U}_i is used to obtain the subject specific likelihood contribution

$$\tilde{\mathcal{L}}_i(\mathcal{D}_i(T_i) | \boldsymbol{\theta}) = \frac{1}{Z} \sum_{z=1}^Z \mathcal{L}_i(\mathcal{D}_i(T_i) | \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}, \boldsymbol{\theta}), \quad \text{where } \mathbf{u}_{ih} = (\mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}).$$

Throughout this paper, we use Sobol sequences to obtain low-discrepancy point sets. A detailed description and an efficient algorithm to obtain Sobol sequences can be found in [1,4,13,25].

6. Dynamic prediction

In this section, we use the fitted joint model and the QMC integration method to predict the probability of experiencing events for a new subject i . More specifically, consider that subject i has been followed until time t_{obs} and has not experienced a terminal event yet, that is, $T_i^* > t_{\text{obs}}$. Given the follow-up history of subject i until t_{obs} , denoted as $\mathcal{D}_i(t_{\text{obs}})$, we calculate the probability of event-free survival (pEFS) until time $t > t_{\text{obs}}$. The pEFS can

be written as

$$\begin{aligned} \tilde{S}_i(t | \mathcal{D}_i(t_{\text{obs}}), t > t_{\text{obs}}, \boldsymbol{\theta}) &= \int_{(\mathbf{v}_i, \mathbf{w}_i)} S_{i0}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i) \times \prod_{r=1}^R S_{ir}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i) \\ &\times p(\mathbf{v}_i, \mathbf{w}_i | \mathcal{D}_i(t_{\text{obs}}), \boldsymbol{\theta}) d(\mathbf{v}_i, \mathbf{w}_i), \end{aligned} \quad (14)$$

where

$$S_{ir}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i) = \exp \left\{ - \int_{t_{\text{obs}}}^t h_{ir}(u) du \right\}$$

is the probability of not experiencing the r th recurrent event type before t . In addition, the probability of not experiencing the terminal event before t is

$$S_{i0}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i) = \exp \left\{ - \int_{t_{\text{obs}}}^t \lambda_i(u) du \right\}.$$

Using Bayes' theorem, $p(\mathbf{v}_i, \mathbf{w}_i | \mathcal{D}_i(t_{\text{obs}}), \boldsymbol{\theta})$ can be written as

$$p(\mathbf{v}_i, \mathbf{w}_i | \mathcal{D}_i(t_{\text{obs}}), \boldsymbol{\theta}) = \frac{\mathcal{L}_i(\mathcal{D}(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i) p(\mathbf{v}_i, \mathbf{w}_i | \boldsymbol{\Omega})}{\int_{(\mathbf{v}_i, \mathbf{w}_i)} \mathcal{L}_i(\mathcal{D}_i(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i) p(\mathbf{v}_i, \mathbf{w}_i | \boldsymbol{\Omega}) d(\mathbf{v}_i, \mathbf{w}_i)},$$

where $\mathcal{L}_i(\mathcal{D}_i(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i, \mathbf{w}_i)$ is the conditional joint likelihood of the observed history $\mathcal{D}_i(t_{\text{obs}})$ calculated with (8). Because the integrals in Equation (14) are analytically intractable and potentially high dimensional, we use the QMC method to estimate the conditional survival function $\tilde{S}_i(t | \mathcal{D}_i(t_{\text{obs}}), t > t_{\text{obs}}, \boldsymbol{\theta})$. Note that a large point set can be used to approximate $\tilde{S}_i(t | \mathcal{D}_i(t_{\text{obs}}), t > t_{\text{obs}}, \boldsymbol{\theta})$ since the integrals only need to be evaluated once.

Thus, $\tilde{S}_i(t | \mathcal{D}_i(t_{\text{obs}}), t > t_{\text{obs}}, \boldsymbol{\theta})$ is approximated as

$$\tilde{S}_i(t | \mathcal{D}_i(t_{\text{obs}}), t > t_{\text{obs}}, \boldsymbol{\theta}) = \frac{\frac{1}{Z} \sum_{z=1}^Z f_{iz}}{\frac{1}{Z} \sum_{z=1}^Z \mathcal{L}_i(\mathcal{D}_i(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)})},$$

where

$$\begin{aligned} f_{iz} &= \mathcal{L}_i(\mathcal{D}_i(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}) \times S_{i0}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}) \\ &\times \prod_{r=1}^R S_{ir}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}). \end{aligned}$$

Note that other survival functions, such as the probability of experiencing a recurrent event of type r given that the individual is still alive, can be estimated in a similar way.

7. Simulation

7.1. Setup

We simulated longitudinal follow-up data of $M=2$ markers and $R=2$ recurrent event types from $n=350$ subjects. For all subjects the maximum follow-up time was $C_i = 10$, after which each subject was assumed to be censored non-informatively. Moreover, the follow-up period of subjects could be ended before $t=10$ by a terminal event.

For subject i , the data was generated as follows. To generate the times for subject i at which we observed both markers, 8 time points t_{ij} were drawn from a uniform distribution over $[0, 10)$. Consequently, all subjects had up to 8 time points at which we observed the markers. Marker observations at time t_{ij} were obtained from the relations

$$\begin{aligned} y_{i1}(t_{ij}) &= y_{i1}^*(t_{ij}) + \epsilon_{i1}(t_{ij}), \quad \text{where } y_{i1}^*(t_{ij}) = (\phi_{10} + v_{i1,0}) + (\phi_{11} + v_{i1,1})t_{ij}, \\ y_{i2}(t_{ij}) &= y_{i2}^*(t_{ij}) + \epsilon_{i2}(t_{ij}), \quad \text{where } y_{i2}^*(t_{ij}) = (\phi_{20} + v_{i2,0}) + (\phi_{21} + v_{i2,1})t_{ij}. \end{aligned}$$

Average marker trajectories were described by a fixed intercept and slope, where $(\phi_{10}, \phi_{11}, \phi_{20}, \phi_{21}) = (0.5, 0.05, 1, 0)$. In addition, both markers had a random intercept and slope with $(v_{i1,0}, v_{i1,1}) \sim N(0, \mathbf{A}_{v1}^T \mathbf{A}_{v1})$ and $(v_{i2,0}, v_{i2,1}) \sim N(0, \mathbf{A}_{v2}^T \mathbf{A}_{v2})$ for, respectively, the first and second marker, with upper triangular matrices

$$\mathbf{A}_{v1} = \begin{pmatrix} \exp(-0.45) & 0 \\ 0 & \exp(-1.75) \end{pmatrix}, \quad \mathbf{A}_{v2} = \begin{pmatrix} \exp(-0.6) & 0.02 \\ 0 & \exp(-1.6) \end{pmatrix}.$$

There was no correlation *between* the random effects of both markers and both markers were observed at the same times. The measurement errors $\epsilon_{i1}(t)$ and $\epsilon_{i2}(t)$ were drawn independently from mean zero normal distributions with variances of, respectively, $\alpha_1^2 = \exp(-0.9)$ and $\alpha_2^2 = \exp(-0.9)$.

For each subject, a binary covariate X_i was sampled from a binomial distribution with $p(X_i = 1) = 0.5$. The hazards for both recurrent events were

$$\begin{aligned} h_{i1}(t) &= h_{01} \exp\{\beta_1 x_i + w_{i1} + \gamma_{11} y_{i1}^*(t) + \gamma_{12} y_{i2}^*(t)\}, \\ h_{i2}(t) &= h_{02} \exp\{\beta_2 x_i + w_{i2} + \gamma_{21} y_{i1}^*(t) + \gamma_{22} y_{i2}^*(t)\}, \end{aligned}$$

where the true parameter values were $(h_{01}, h_{02}) = \exp(-1.2, -1.4)$, $(\beta_1, \beta_2) = (-0.15, 0.2)$, and $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}) = (0.1, 0.08, 0.09, 0.12)$ and frailties $(w_{i1}, w_{i2}) \sim N(0, \mathbf{A}_w^T \mathbf{A}_w)$ with upper triangular matrix

$$\mathbf{A}_w = \begin{pmatrix} \exp(-0.45) & 0.07 \\ 0 & \exp(-0.6) \end{pmatrix}.$$

Using the hazards, we simulated recurrent event times for subject i until time $T_i = \min(T_i^*, C_i = 10)$. To generate the terminal event time T_i^* , we used the hazard

$$\lambda_i(t) = \lambda_0 \exp\{\beta_0 x_i + \gamma_{01} y_{i1}^*(t) + \gamma_{02} y_{i2}^*(t) + \chi_1 w_{i1} + \chi_2 w_{i2}\}, \quad (15)$$

with constant hazard $\lambda_0 = \exp(-3)$, and $(\beta_0, \gamma_{01}, \gamma_{02}) = (0.35, 0.15, 0.1)$. The effects of the frailty terms from the recurrent events were $(\chi_1, \chi_2) = (0.3, 0.2)$.

7.2. Estimation methods

The following three estimators were fitted to the data:

SML estimator

Obtain $\hat{\theta}_{\text{SML}}$ using Equation (10). The variance of $\hat{\theta}_{\text{SML}}$ was derived using the information matrix from Equation (12a) evaluated at $\hat{\theta}_{\text{SML}}$. The deterministic point set

was generated with Sobol sequences. To evaluate the behavior of the SML estimator, we used a relatively small ($Z = 10,000$) and relatively large ($Z = 60,000$) Sobol point set. The simulation bias was expected to be smaller with a larger point set, leading to more accurate estimates. The information matrix was approximated using a Sobol point set with $\tilde{Z} = 200,000$ vectors.

CSML estimator

Update the estimate $\hat{\theta}_{\text{SML}}$ using Equation (11) to obtain $\hat{\theta}_{\text{CSML}}$. Both SML estimates, obtained with a small ($Z = 10,000$) and large ($Z = 60,000$) Sobol point set, were updated by approximating the information matrix with Sobol point set containing $\tilde{Z} = 200,000$ vectors. The variances of $\hat{\theta}_{\text{CSML}}$ were derived using the information matrix from Equation (12a) evaluated at $\hat{\theta}_{\text{CSML}}$. The information matrix was approximated using a Sobol point set with $\tilde{Z} = 200,000$ vectors.

GHQ estimator

To compare the SML and CSML estimators with existing methods in terms of bias and computational time, we used a multidimensional GHQ to approximate the integral from Equation (4). We generated a multidimensional grid of evaluation points by combining uni-dimensional quadratures with the product rule [8]. This grid of evaluation points was rescaled to fit the integration region using a Cholesky decomposition of the covariance matrix Ω from Equation (13). Let $(v_i^{(z)}, w_i^{(z)})$ be a vector from the rescaled grid of evaluation points with corresponding weight w_z . The GHQ estimator was defined as

$$\hat{\theta}_{\text{GHQ}} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^n \log \left[\sum_{z=1}^Z w_z \mathcal{L}_i(\mathcal{D}_i(T_i) | v_i^{(z)}, w_i^{(z)}, \theta) \right]. \quad (16)$$

To have a GHQ estimator that was comparable in computational time to the SML estimator with a large Sobol point set, we used six evaluation points per dimension leading to a quadrature with $Z = 6^6 = 46,656$ vectors. The variances of $\hat{\theta}_{\text{GHQ}}$ were derived from the information matrix evaluated at $\hat{\theta}_{\text{GHQ}}$. We used the same multidimensional GHQ with six evaluation points per dimension for this approximation.

The simulations were performed in the statistical software package R [37], using the `parallel` package to implement the parallel computing to speed up the calculations of Equation (9). The simulated likelihoods from Equations (16) and (11) and the approximated likelihood from Equation (10) were maximized with trust region optimization, implemented by the `trustOptim` package [5]. The expected scores were obtained using the `numDeriv` package [15]. In addition, the `numDeriv` package was also used to calculate the standard errors of both estimators.

The simulation was repeated 500 times. To investigate the performance of all three estimators, we reported the mean ($E[\hat{\theta}]$), mean absolute bias ($E[|\theta - \hat{\theta}|]$), mean squared error (MSE) ($E[(\theta - \hat{\theta})^2]$), mean estimated variance ($E[\operatorname{var}(\hat{\theta})]$), and coverage of the 95% confidence interval for all model parameter estimates.

If the approximation of the integral in Equation (4) is accurate enough, all average parameter estimates should be similar to the true parameter values with low mean absolute biases. In addition, because the MSE consists of a variance and bias component, the MSEs

of all estimated model parameters should be approximately similar to their mean estimated variances. When the MSEs are substantially larger than the mean estimated variances, this means that the parameter estimates are contaminated with bias.

7.3. Results

The results for the SML and CSML estimator using a small Sobol point set ($Z = 10,000$) are given in Table 2. With the SML estimator, all average parameter estimates were similar to the true parameter values. For the parameters of both marker submodels, substantial differences were found between the MSEs and the mean estimated variances of the parameters regarding the random effects, the fixed intercepts, and the fixed slopes. This meant that simulation bias was present in the parameter estimates of the marker submodels. Moreover, because we assumed that the simulation bias was negligible in the calculation of the standard errors of the model parameters, the coverages of the 95% confidence interval were too low.

The parameters of the event submodels did not suffer from simulation bias and the MSEs were approximately similar to the mean variance of the estimated parameters. Consequently, the event submodel parameters had excellent coverages of the 95% confidence interval. Interestingly, although the marker submodel parameters were affected by the simulation error, the association parameters between the true values of the markers and the event processes were not.

An improvement of the parameter estimates was found with the CSML estimator. The extra Gauss–Newton step reduced the simulation error, which was reflected by lower mean absolute biases and lower MSEs for all marker submodels parameters. However, some simulation error still remained. Although the coverages of the 95% confidence interval of all marker submodels parameters improved, they were too low.

In Table 3, the results for the SML and CSML estimator using a large Sobol point set ($Z = 60,000$) are given. As expected, using more vectors Z to simulate the likelihood, the parameters estimates from the SML and CSML estimator were more accurate. Less simulation bias was observed for the parameters regarding the random effects, the fixed intercepts, and the fixed slopes of both markers.

With the GHQ estimator, the estimates of the parameters regarding the random effects and the residual variance of both markers submodels were biased, leading to high mean absolute errors, MSEs, and poor coverages of the 95% confidence interval (see Table 4). In addition, compared to the SML estimator, the mean absolute biases and MSEs of the fixed intercepts and slopes were substantially higher. The parameters of the event submodels were accurately estimated with the GHQ estimator. With respect to the MSEs and mean absolute errors of these parameters, the GHQ estimator was comparable to the SML estimator using a large Sobol point set.

The biased parameter estimates of the GHQ estimator were caused by the low number of evaluation points per dimension, leading to a very crude approximation of the likelihood from (4). Although we could have increased the accuracy of the approximation by adding more evaluation points to the uni-dimensional quadratures, the size of the multivariate GHQ would have increased exponentially. For instance, using uni-dimensional quadratures rule with seven evaluation points would require the evaluation of $Z = 7^6 = 117,649$

Table 2. Simulation results for the SML estimator and CSML estimator using a small ($Z = 10,000$) Sobol point set.

	Parameter	θ	SML estimator (θ_{SML})					CSML estimator (θ_{CSML})				
			$E[\hat{\theta}]$	$E[\theta - \hat{\theta}] \times 1000$	$E[(\theta - \hat{\theta})^2] \times 1000$	$E[\text{var}(\hat{\theta})] \times 1000$	Coverage 95% CI	$E[\hat{\theta}]$	$E[\theta - \hat{\theta}] \times 1000$	$E[(\theta - \hat{\theta})^2] \times 1000$	$E[\text{var}(\hat{\theta})] \times 1000$	Coverage 95% CI
Random effects markers and recurrent events	$\log(\alpha_{1,1}^{v_1})$	-0.450	-0.444	48.932	3.755	3.144	0.92	-0.465	50.798	3.880	3.144	0.91
	$\alpha_{1,2}^{v_1}$	0.000	0.004	17.213	0.443	0.145	0.69	0.003	14.699	0.330	0.145	0.78
	$\log(\alpha_{2,2}^{v_1})$	-1.750	-1.726	64.693	5.914	2.674	0.77	-1.775	58.908	5.053	2.674	0.84
	$\log(\alpha_{1,1}^{v_2})$	-0.600	-0.600	52.482	4.555	3.760	0.92	-0.618	53.271	4.292	3.760	0.93
	$\alpha_{1,2}^{v_2}$	0.020	0.025	20.355	0.664	0.176	0.71	0.022	15.155	0.371	0.176	0.81
	$\log(\alpha_{2,2}^{v_2})$	-1.600	-1.577	54.291	4.405	2.339	0.81	-1.618	57.767	5.429	2.339	0.81
	$\log(\alpha_{1,1}^w)$	-0.450	-0.475	78.031	8.884	8.172	0.95	-0.481	73.983	8.138	8.172	0.93
	$\alpha_{1,2}^w$	0.070	0.073	67.724	7.044	5.232	0.90	0.066	60.925	5.788	5.232	0.93
Marker submodels	$\log(\alpha_{2,2}^w)$	-0.600	-0.646	97.386	14.332	11.571	0.93	-0.637	89.675	12.091	11.571	0.95
	ϕ_{10}	0.500	0.497	40.373	2.461	1.832	0.91	0.500	38.476	2.225	1.832	0.93
	ϕ_{11}	0.050	0.049	14.220	0.324	0.143	0.79	0.050	11.622	0.209	0.143	0.91
	$\log(\alpha_1)$	-0.900	-0.892	15.761	0.353	0.418	0.94	-0.901	15.370	0.381	0.418	0.94
	ϕ_{20}	1.000	1.000	33.813	1.768	1.468	0.91	1.002	32.250	1.600	1.468	0.94
	ϕ_{21}	0.000	-0.001	17.757	0.473	0.171	0.75	0.000	12.996	0.274	0.171	0.88
	$\log(\alpha_2)$	-0.900	-0.892	16.460	0.353	0.415	0.96	-0.901	15.578	0.379	0.415	0.96
Recurrent event submodels	$\log(h_{01})$	-1.200	-1.190	77.156	9.195	8.753	0.95	-1.200	76.475	8.923	8.753	0.95
	β_1	-0.150	-0.150	89.106	12.297	11.407	0.95	-0.148	82.861	11.301	11.407	0.95
	γ_{11}	0.100	0.100	37.417	2.260	2.204	0.95	0.099	35.840	2.136	2.204	0.94
	γ_{12}	0.080	0.079	33.524	1.854	1.792	0.94	0.081	32.905	1.728	1.792	0.95
	$\log(h_{02})$	-1.400	-1.386	73.057	8.651	8.408	0.93	-1.398	74.007	8.774	8.408	0.94
	β_2	0.200	0.198	81.102	10.197	9.852	0.94	0.201	81.185	10.191	9.852	0.95
	γ_{21}	0.090	0.090	38.374	2.279	1.999	0.93	0.090	37.831	2.202	1.999	0.94
	γ_{22}	0.120	0.117	32.922	1.693	1.601	0.94	0.117	34.373	1.796	1.601	0.94
Terminal event submodel	$\log(\lambda_0)$	-3.000	-3.023	119.121	22.664	24.998	0.96	-2.995	123.289	23.306	24.998	0.96
	β_0	0.350	0.358	125.003	25.217	25.406	0.95	0.359	129.600	26.793	25.406	0.94
	γ_{01}	0.150	0.154	58.680	5.572	6.080	0.97	0.153	61.984	6.095	6.080	0.95
	γ_{02}	0.100	0.106	52.092	4.228	4.906	0.96	0.108	53.328	4.593	4.906	0.95
	χ_1	0.300	0.308	177.805	54.036	52.435	0.95	0.308	182.523	56.144	52.435	0.94
	χ_2	0.200	0.231	214.066	71.351	83.475	0.97	0.228	217.644	74.118	83.475	0.96

Table 3. Simulation results for the SML estimator and CSML estimator using a large ($Z = 60,000$) Sobol point set.

	Parameter	θ	SML estimator (θ_{SML})					CSML estimator (θ_{CSML})				
			$E[\hat{\theta}]$	$E[\theta - \hat{\theta}]$ $\times 1000$	$E[(\theta - \hat{\theta})^2]$ $\times 1000$	$E[\text{var}(\hat{\theta})]$ $\times 1000$	Coverage 95% CI	$E[\hat{\theta}]$	$E[\theta - \hat{\theta}]$ $\times 1000$	$E[(\theta - \hat{\theta})^2]$ $\times 1000$	$E[\text{var}(\hat{\theta})]$ $\times 1000$	Coverage 95% CI
Random effects markers and recurrent events	$\log(\alpha_{1,1}^{v_1})$	-0.450	-0.448	46.195	3.344	3.089	0.93	-0.457	46.838	3.348	3.089	0.92
	$\alpha_{1,2}^{v_1}$	0.000	0.002	14.769	0.346	0.137	0.77	0.002	13.661	0.282	0.137	0.83
	$\log(\alpha_{2,2}^{v_1})$	-1.750	-1.749	48.546	3.704	2.421	0.86	-1.762	49.165	3.632	2.421	0.88
	$\log(\alpha_{1,1}^{v_2})$	-0.600	-0.599	44.929	3.277	3.650	0.96	-0.606	45.104	3.296	3.650	0.95
	$\alpha_{1,2}^{v_2}$	0.020	0.023	16.828	0.438	0.166	0.76	0.023	15.636	0.370	0.166	0.80
	$\log(\alpha_{2,2}^{v_2})$	-1.600	-1.596	48.511	3.710	2.154	0.86	-1.608	48.121	3.632	2.154	0.86
	$\log(\alpha_{1,1}^w)$	-0.450	-0.452	71.970	8.117	7.946	0.95	-0.462	69.453	7.568	7.946	0.95
	$\alpha_{1,2}^w$	0.070	0.066	56.076	5.112	5.247	0.94	0.066	56.001	4.928	5.247	0.95
Marker submodels	$\log(\alpha_{2,2}^w)$	-0.600	-0.623	82.866	10.742	11.515	0.96	-0.630	82.519	10.728	11.515	0.94
	ϕ_{10}	0.500	0.500	33.786	1.862	1.780	0.95	0.500	32.806	1.751	1.780	0.95
	ϕ_{11}	0.050	0.051	11.276	0.207	0.131	0.87	0.051	10.559	0.186	0.131	0.90
	$\log(\alpha_1)$	-0.900	-0.898	15.314	0.363	0.405	0.97	-0.898	15.440	0.368	0.405	0.97
	ϕ_{20}	1.000	1.000	30.694	1.446	1.431	0.95	1.000	31.023	1.454	1.431	0.96
	ϕ_{21}	0.000	-0.000	13.936	0.309	0.157	0.85	-0.001	12.518	0.261	0.157	0.87
Recurrent event submodels	$\log(\alpha_2)$	-0.900	-0.898	15.523	0.375	0.401	0.95	-0.898	15.599	0.378	0.401	0.95
	$\log(h_{01})$	-1.200	-1.199	67.432	7.276	9.083	0.98	-1.202	68.101	7.397	9.083	0.97
	β_1	-0.150	-0.156	84.043	11.095	11.800	0.96	-0.152	83.149	11.174	11.800	0.94
	γ_{11}	0.100	0.101	36.762	2.089	2.263	0.95	0.101	38.116	2.208	2.263	0.94
	γ_{12}	0.080	0.080	31.466	1.573	1.855	0.97	0.079	31.790	1.599	1.855	0.97
	$\log(h_{02})$	-1.400	-1.401	71.542	8.276	8.621	0.95	-1.402	71.496	8.258	8.621	0.94
	β_2	0.200	0.202	78.657	9.586	10.107	0.95	0.202	78.441	9.585	10.107	0.96
	γ_{21}	0.090	0.090	36.839	2.163	2.074	0.94	0.091	36.417	2.108	2.074	0.95
Terminal event submodel	γ_{22}	0.120	0.120	32.058	1.613	1.654	0.94	0.120	32.547	1.624	1.654	0.95
	$\log(\lambda_0)$	-3.000	-3.023	121.437	22.222	25.026	0.96	-3.012	120.290	22.151	25.026	0.96
	β_0	0.350	0.365	121.682	23.075	25.420	0.95	0.365	123.127	23.515	25.420	0.95
	γ_{01}	0.150	0.152	60.637	5.587	6.155	0.96	0.152	60.798	5.740	6.155	0.96
	γ_{02}	0.100	0.103	51.556	4.211	4.912	0.96	0.103	52.131	4.312	4.912	0.96
	χ_1	0.300	0.313	173.087	49.005	48.407	0.96	0.315	175.603	49.292	48.407	0.96
	χ_2	0.200	0.211	219.150	78.011	77.404	0.95	0.214	216.717	77.137	77.404	0.95

Table 4. Simulation results for the GHQ estimator.

	Parameter	θ	One-step estimator (θ_{SML})				Coverage 95% CI
			$E[\hat{\theta}]$	$E[\theta - \hat{\theta}]$ $\times 1000$	$E[(\theta - \hat{\theta})^2]$ $\times 1000$	$E[\text{var}(\hat{\theta})]$ $\times 1000$	
Random effects markers and recurrent events	$\log(\alpha_{1,1}^{v_1})$	-0.450	-0.691	240.688	3.890	1.420	0.01
	$\sigma_{1,2}^{v_1}$	0.000	0.006	9.311	0.097	0.014	0.48
	$\log(\alpha_{2,2}^{v_1})$	-1.750	-1.872	126.505	7.123	0.226	0.12
	$\log(\alpha_{1,1}^{v_2})$	-0.600	-0.871	270.994	5.345	1.976	0.00
	$\sigma_{1,2}^{v_2}$	0.020	0.027	10.090	0.128	0.013	0.45
	$\log(\alpha_{2,2}^{v_2})$	-1.600	-1.712	117.162	6.560	0.148	0.09
	$\log(\alpha_{1,1}^w)$	-0.450	-0.471	68.567	6.735	7.640	0.97
	$\sigma_{1,2}^w$	0.070	0.074	48.803	3.668	5.341	0.98
	$\log(\alpha_{2,2}^w)$	-0.600	-0.623	76.353	8.869	11.012	0.97
Marker submodels	ϕ_{10}	0.500	0.500	44.758	3.139	0.513	0.56
	ϕ_{11}	0.050	0.051	20.260	0.614	0.019	0.25
	α_1	-0.900	-0.822	77.831	0.365	0.387	0.02
	ϕ_{20}	1.000	1.000	34.632	1.828	0.478	0.68
	ϕ_{21}	0.000	-0.002	22.020	0.727	0.017	0.21
	α_2	-0.900	-0.818	81.864	0.369	0.384	0.01
Recurrent event submodels	h_{01}	-1.200	-1.198	71.763	7.943	8.878	0.96
	β_1	-0.150	-0.148	84.129	10.845	11.295	0.97
	γ_{11}	0.100	0.100	36.403	2.102	2.243	0.94
	γ_{12}	0.080	0.079	32.851	1.766	1.848	0.95
	h_{02}	-1.400	-1.399	68.683	7.596	8.694	0.96
	β_2	0.200	0.202	77.269	9.199	10.081	0.96
	γ_{21}	0.090	0.091	34.764	1.873	2.113	0.97
	γ_{22}	0.120	0.119	30.242	1.422	1.707	0.98
Terminal event submodel	λ_0	-3.000	-3.014	119.664	21.680	24.895	0.97
	β_0	0.350	0.352	126.241	24.282	25.434	0.96
	γ_{01}	0.150	0.151	63.075	6.474	6.227	0.92
	γ_{02}	0.100	0.102	55.525	4.785	4.987	0.96
	χ_1	0.300	0.313	171.530	45.619	49.391	0.96
	χ_2	0.200	0.200	206.844	71.781	76.290	0.95

vectors. In terms of computational costs, this would mean that the GHQ estimator would be two times slower than the SML estimator with a large Sobol point set ($Z = 60,000$).

A possible solution to decrease the computational costs of the GHQ estimator is to combine the uni-dimensional quadratures in a different way, leading to sparse GHQ grids. Unfortunately, a problem of these sparse GHQ grids is that some of the weights are negative [22]. In our simulation, this led to negative approximations of the joint likelihood from Equation (4). Only if we increased the number of vectors in the sparse GHQ grid substantially (i.e. $Z > 1,000,000$), this problem disappeared.

8. Analysis of the post kidney transplant data

8.1. Estimation

The post kidney transplant data, introduced in Section 2, was analyzed with the following joint model. The marker submodels followed the specification of Equation (1), where the m th marker submodel was parametrized by a fixed intercept and slope and a random intercept and slope. For the random effects, we assumed independence between the random effects of *different* markers. The four infection types with frailty terms followed the specification of Equation (2); that is, the subject specific frailty w_{ir} and all four true marker

values at time t influenced the hazard at t of the r th recurrent infection. The drop out sub-model was parametrized as Equation (3); the frailties of four recurrent infections and all four true marker values could influence the risk of dropping out.

The baseline hazards in all the recurrent infections and dropout submodels were described by B-spline functions. Our joint model had eight random effects in the marker submodels and four frailty terms. Therefore the evaluation of the likelihood function from the joint model involved a 12-dimensional integral. We used our SML estimator to obtain the parameter estimates $\hat{\theta}_{\text{SML}}$. To simulate the likelihood, we used a Sobol point set containing $Z = 120,000$ vectors. For the approximation of the Hessian matrix, we used a point set with $\tilde{Z} = 1,000,000$ vectors.

8.2. Results

The parameter estimates have been summarized in Table 5. Women had an increased risk of experiencing an urinary tract infection (log-hazard: 0.799; standard error: 0.265). In addition, adding a year to the age at baseline increased the risk of dropping out of the study (0.022; 0.008). In addition, negative associations were found between the CD4+ cell count and both the cytomegalovirus (−3.694; 1.232) and the upper respiratory infection type (−4.414; 0.997); an increased CD4+ cell count was associated with a decreased risk of experiencing these two infection types. A similar association was found between CD4+ cell count and the risk of dropping out of the study (−7.050; 1.843).

Positive associations were found between CD8+ cell count and upper respiratory infection (4.386; 1.183); an increased CD8+ cell count was associated with an increased risk of experiencing an upper respiratory infection. The risk of dropping out of the study was also positively associated (6.392; 1.933) with CD8+ cell count.

The variance of the frailty from the cytomegalovirus infection type was really small ($\sigma_{\text{Cyto}} < 0.001$). The variances of the frailties for, respectively, the upper respiratory infection type, wound infection type, and urinary tract infection type were $\sigma_{\text{Resp}} = 0.093$, $\sigma_{\text{Wound}} = 0.833$, and $\sigma_{\text{Urinary}} = 1.489$. With exception of the cytomegalovirus infection type, the frailties of all infection types were positively correlated with each other. No associations were found between the frailties of all four infection types and the risk of dropping out of the study.

8.3. Dynamic prediction

To illustrate the use of QMC integration for dynamic prediction, we have predicted the probability that patient i is still at risk (i.e. has not dropped out of the study) at time $t > t_{\text{obs}}$ and did not experience infection type r . This probability depends on the history, denoted as $\mathcal{D}_i(t_{\text{obs}})$ until time t_{obs} and can be calculated as

$$\frac{\frac{1}{Z} \sum_{z=1}^Z f_{iz}}{\frac{1}{Z} \sum_{z=1}^Z \mathcal{L}_i(\mathcal{D}_i(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)})},$$

where

$$\begin{aligned} f_{iz} &= \mathcal{L}_i(\mathcal{D}_i(t_{\text{obs}}) | \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}) \times S_{i0}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}) \\ &\times S_{ir}(t | t > t_{\text{obs}}, \boldsymbol{\theta}, \mathbf{v}_i^{(z)}, \mathbf{w}_i^{(z)}). \end{aligned}$$

Table 5. Joint model fixed parameter estimates (Est.) and standard errors (Std. err.) for the post kidney transplant data.

Marker submodels		Fixed effects									
		CD8+ T Cells (cd8)		CD4+ T Cells (cd4)		B Cells (bcl)		Natural killer cells (nk)			
		Est.	Std.err.	Est.	Std.err.	Est.	Std.err.	Est.	Std.err.		
	ϕ_0	0.680	0.028	0.827	0.030	0.520	0.019	0.586	0.019		
	ϕ_1	0.045	0.010	0.000	0.009	−0.007	0.005	0.018	0.006		
	$\log(\alpha)$	−2.017	0.029	−1.954	0.025	−2.565	0.022	−2.304	0.024		
		Random effects									
		$\sigma_{0,cd8}$	$\sigma_{1,cd8}$	$\sigma_{0,cd4}$	$\sigma_{1,cd4}$	$\sigma_{0,bcl}$	$\sigma_{1,bcl}$	$\sigma_{0,nk}$	$\sigma_{1,nk}$		
	$\sigma_{0,cd8}$	0.016	−0.003								
	$\sigma_{1,cd8}$	−0.003	0.002								
	$\sigma_{0,cd4}$			0.036	−0.007						
	$\sigma_{1,cd4}$			−0.007	0.002						
	$\sigma_{0,bcl}$					0.015	−0.001				
	$\sigma_{1,bcl}$					−0.001	< 0.001				
	$\sigma_{0,nk}$							0.017	−0.003		
	$\sigma_{1,nk}$							−0.003	0.001		
		Fixed effects									
		Cytomegalovirus (Cyto)		Upper respiratory (Resp)		Wound		Urinary tract (urinary)		Terminal event	
		Est.	Std.err.	Est.	Std.err.	Est.	Std.err.	Est.	Std.err.	Est.	Std.err.
	β_{female}	−0.054	0.214	−0.179	0.183	−0.174	0.293	0.799	0.265	−0.176	0.196
	β_{age}	0.006	0.009	−0.012	0.007	0.008	0.011	0.005	0.009	0.022	0.008
	$\beta_{\text{treatment } B}$	0.258	0.428	−0.034	0.298	0.791	0.492	0.084	0.528	0.046	0.445
	γ_{cd8}	2.356	1.650	4.386	1.183	2.168	2.034	0.109	1.385	6.392	1.933
	γ_{cd4}	−3.694	1.232	−4.414	0.997	0.384	1.674	−1.898	1.372	−7.050	1.843
	γ_{bcl}	2.788	1.696	−0.426	1.663	−1.344	2.045	−0.297	1.796	4.011	2.226
	γ_{nk}	−1.831	1.599	0.201	1.423	−1.968	2.123	1.960	1.537	1.597	1.908
	χ_{Cyto}									0.016	> 10
	χ_{Resp}									0.375	> 10
	χ_{Wound}									−0.054	> 10
	χ_{Urinary}									0.085	5.347
		Frailties									
		σ_{Cyto}	σ_{Resp}	σ_{Wound}	σ_{Urinary}						
	σ_{Cyto}	< 0.001	−0.002	−0.005	−0.005						
	σ_{Resp}	−0.002	0.093	0.266	0.234						
	σ_{Wound}	−0.005	0.266	0.833	0.426						
	σ_{Urinary}	−0.005	0.234	0.426	1.489						

Notes: The baseline hazard estimates of the (recurrent) event submodels are not shown. The intercept was based on a 48 year old man who was given treatment A at baseline and had cell counts of zero for all four longitudinal markers.

For example, let $t_{\text{obs}} = 1$ years. We used the estimated model parameters $\hat{\theta}$ and $Z = 500,000$ to obtain the predictions. In Figure 1, we illustrate the prediction of the various event type risks, as well as the impact of history on prediction using two patients (patient A and patient B) who have exactly the same baseline characteristics, but differ in their history information in the past 1 year.

Patient A did not experience any infections during the first year of follow-up. Moreover, no marker measurements were performed. Since this individual did not experience any infections before $t = 1$, the model predicted higher than average CD4+ cell count and lower

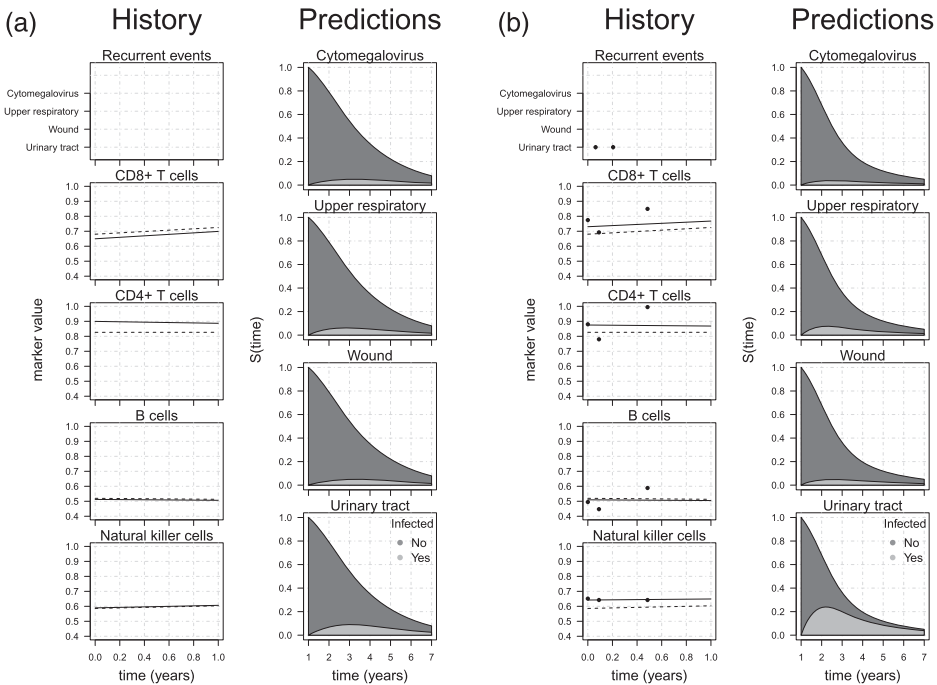


Figure 1. History and predicted probabilities of experiencing events for two patients. The dotted lines in the marker figures denote the fixed marker trajectories, that is, zero random effects. The continuous lines are the predicted marker trajectories. For the predictions, the probability of experiencing a particular infection is given. In addition, the sum of the infected no/yes probabilities is the probability that the patient has not dropped out of the study. (a) Patient without events and marker measurements until $t = 1$ and (b) Patient with events and marker measurements until $t = 1$.

than average CD8+ cell count. Following Table 5, these predictions could be expected since a higher CD4+ cell count and a lower CD8+ cell count were associated with decreased risks of all event types (except for wound infections).

Patient *B* on the other hand experienced two urinary tract infections in the first year of follow-up, and also had three measurements of all marker types. Compared to the mean, patient *B* had increased CD4+, natural killer, and CD8+ cell count. We also observed that patient *B* had a higher probability of dropping out of the study compared to patient *A*. For instance, at $t = 3$, the probability of that patient *A* has not dropped out of the study was about 55% and for patient *B* about 35%. Moreover, the probability of experiencing an urinary tract infection was higher for patient *B* than for patient *A*. For the other infections, the probabilities were similar for both patients.

9. Discussion

With joint modeling of multiple longitudinal markers, multiple recurrent events, and a terminal event we are confronted with a high dimensional, analytically intractable integral. To approximate the integral we have proposed to use an SML approach based on QMC integration techniques. Moreover, these integration techniques can be used when

we want to use the joint model for individualized prediction. The strong law of large numbers guarantees that, when we draw an infinite number of vectors from the random effects distribution, the SML estimator is equivalent to an ML estimator. However, with a finite number of vectors, non-ignorable simulation error could influence the parameter estimates. With simulations, we have investigated the performance of the SML estimator. The results from this simulation showed that, even with a relatively small number of vectors, the SML estimator could give accurate parameter estimates of the recurrent and terminal event processes. Moreover, we compared the SML estimator to an estimator based on GHQ approximation. Using approximations with the same computational costs (i.e. approximately the same number of evaluation points), the SML estimator was more accurate.

An important challenge is to have ignorable simulation error while keeping the computational costs of evaluating the likelihood to a minimum. Further research is necessary to investigate the possibility to test whether the integration error is small enough. A starting point might be the statistical test proposed by Hajivassiliou for MC integration [21].

During the development of our joint model, we assumed that all the dependencies between the marker trajectories and the events processes were captured by using the true marker value in the event submodels. However, there are other ways to capture the dependencies between marker trajectories and the recurrent events, which we did not consider in this paper [41]. For instance, we could assume that all dependencies between the markers and the recurrent event types are captured by the covariances between \mathbf{v}_i and \mathbf{w}_i [24]. Following this assumption, we would then parameterize the full covariance matrix $\mathbf{\Omega}$ from Equation (13) and omit the true marker values from Equation (2). The SML approach can easily be adapted to these joint models, since the joint model structure from Equation (4) remains the same.

To maximize the simulated likelihood, we used numerical approximation techniques to estimate the first and second derivatives of the joint likelihood function. Although this approach is flexible and can be used for many different parametrizations of the joint model (e.g. mixtures of longitudinal marker types), it is a computationally expensive procedure. As an alternative to numerical approximation techniques, we could use automatic differentiation techniques to obtain exact evaluations of first and second derivatives [40]. Further research is necessary to implement automatic differentiation in the SML estimator.

In conclusion, we have showed that SML estimation could be useful when we jointly model highly dimensional longitudinal data, consisting of multiple longitudinal markers, multiple recurrent events, and a terminal event.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- [1] I.A. Antonov and V.M. Saleev, *An economic method of computing $LP\tau$ -sequences*, USSR Comput. Math. Math. Phys. 19 (1979), pp. 252–256.
- [2] S. Asmussen and P.W. Glynn, *Chapter IX: Numerical integration*, in *Stochastic Simulation: Algorithms and Analysis: Algorithms and Analysis*, Stochastic Modelling and Applied Probability, Springer, New York, 2007, pp. 260–273.

- [3] T. Baghfalaki, M. Ganjali, and R. Hashemi, *Bayesian joint modeling of longitudinal measurements and time-to-event data using robust distributions*, J. Biopharm. Stat. 24 (2014), pp. 834–855.
- [4] P. Bratley and B.L. Fox, *Algorithm 659: Implementing Sobol's quasirandom sequence generator*, ACM Trans. Math. Softw. 14 (1988), pp. 88–100.
- [5] M. Braun, *trustOptim: Trust region nonlinear optimization, efficient for sparse Hessians*, 2013. Available at <http://CRAN.R-project.org/package=trustOptim>, r package version 0.8.3.
- [6] Y.-Y. Chi and J.G. Ibrahim, *Joint models for multivariate longitudinal and multivariate survival data*, Biometrics 62 (2006), pp. 432–445.
- [7] R.J. Cook and J.F. Lawless, *The Statistical Analysis of Recurrent Events*, Statistics for Biology and Health, Springer, New York, NY, 2007.
- [8] J. Dick, F.Y. Kuo, and I.H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer. 22 (2013), pp. 133–288.
- [9] J. Dick and F. Pillichshammer, *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, New York, NY, 2010.
- [10] R.M. Elashoff, G. Li, and N. Li, *A joint model for longitudinal measurements and survival data in the presence of multiple failure types*, Biometrics 64 (2008), pp. 762–771.
- [11] K.T. Fang, Y. Wang, and P.M. Bentler, *Some applications of number-theoretic methods in statistics*, Statist. Sci. 9 (1994), pp. 416–428.
- [12] C.L. Faucett and D.C. Thomas, *Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach*, Stat. Med. 15 (1996), pp. 1663–1685.
- [13] J.E. Gentle, *Chapter 3. Quasirandom numbers*, in *Random Number Generation and Monte Carlo Methods*, 2nd ed., Springer, New York, 2005, p. 387.
- [14] R.B. Geskus, *Which individuals make dropout informative?* Stat. Methods Med. Res. 23 (2014), pp. 91–106.
- [15] P.I. Gilbert and R.I. Varadhan, *numDeriv: Accurate Numerical Derivatives*, 2012. Available at <http://CRAN.R-project.org/package=numDeriv>, r package version 2014.2-1.
- [16] P. Glasserman, *Chapter 5. Quasi-Monte Carlo*, in *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004, pp. 281–339.
- [17] J. González, F. Tuerlinckx, P. De Boeck, and R. Cools, *Numerical integration in logistic-normal models*, Comput. Statist. Data Anal. 51 (2006), pp. 1535–1548.
- [18] C. Gouriéroux and A. Monfort, *Simulation based inference in models with heterogeneity*, Annales d'Économie et de Statistique 20 (1990), pp. 69–107.
- [19] C. Gouriéroux and A. Monfort, *Simulation-Based Econometric Methods*, Oxford University Press, New York, 1997.
- [20] R. Gueorguieva, R. Rosenheck, and H. Lin, *Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial*, J. Roy. Statist. Soc. Ser. A 175 (2012), pp. 417–433.
- [21] V. Hajivassiliou, *Some practical issues in maximum simulated likelihood*, in *Simulation-Based Inference in Econometrics: Methods and Applications*, R. Mariano, T. Schuermann, and M.J. Weeks, eds., Cambridge University Press, Cambridge, 2000, pp. 71–99.
- [22] F. Heiss and V. Winschel, *Likelihood approximation by numerical integration on sparse grids*, J. Econometrics 144 (2008), pp. 62–80.
- [23] C.R. Henderson, *Analysis of covariance in the mixed model: Higher-level, nonhomogeneous, and random regressions*, Biometrics 38 (1982), pp. 623–640.
- [24] R. Henderson, P. Diggle, and A. Dobson, *Joint modelling of longitudinal measurements and event time data*, Biostatistics 1 (2000), pp. 465–480.
- [25] S. Joe and F.Y. Kuo, *Remark on Algorithm 659: Implementing Sobol's quasirandom sequence generator*, ACM Trans. Math. Softw. 29 (2003), pp. 49–57.
- [26] C. Lemieux, *Chapter 5. Quasi-Monte Carlo constructions*, in *Monte Carlo and Quasi-Monte Carlo sampling*, Springer, New York, 2009, pp. 139–200.
- [27] N. Li, R.M. Elashoff, and G. Li, *Robust joint modeling of longitudinal measurements and competing risks failure time data*, Biom. J. 51 (2009), pp. 19–30.

- [28] N. Li, R.M. Elashoff, G. Li, and J. Saver, *Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial*, Stat. Med. 29 (2010), pp. 546–557.
- [29] L. Liu and X. Huang, *The use of Gaussian quadrature for estimation in frailty proportional hazards models*, Stat. Med. 27 (2008), pp. 2665–2683.
- [30] L. Liu and X. Huang, *Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome*, J. Roy. Statist. Soc. Ser. C 58 (2009), pp. 65–81.
- [31] L. Liu, R.A. Wolfe, and X. Huang, *Shared frailty models for recurrent events and a terminal event*, Biometrics 60 (2004), pp. 747–756.
- [32] J.Z. Musoro, R.B. Geskus, and A.H. Zwinderman, *A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant*, Biom. J. 57 (2014), pp. 185–200.
- [33] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, in CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.
- [34] H. Niederreiter, *Error bounds for quasi-Monte Carlo integration with uniform point sets*, J. Comput. Appl. Math. 150 (2003), pp. 283–292.
- [35] T. Pillards and R. Cools, *Using Box–Muller with low discrepancy points*, in *Computational Science and its Applications – ICCSA 2006: Lecture Notes in Computer Science*, M.L. Gavrilova, O. Gervasi, V. Kumar, C.K. Tan, and D. Taniar, eds., Vol. 3984, Berlin, Heidelberg, Springer, 2006, pp. 780–788.
- [36] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, New York, NY, 2007.
- [37] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. Available at <http://www.R-project.org/>.
- [38] M. Rezaul Karim and S.L. Zeger, *Generalized linear models with random effects; salamander mating revisited*, Biometrics 48 (1992), pp. 631–644.
- [39] D. Rizopoulos, *Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data*, Biometrics 67 (2011), pp. 819–829.
- [40] H.J. Skaug, *Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models*, J. Comput. Graph. Statist. 11 (2002), pp. 458–470.
- [41] I. Sousa, *A review on joint modelling of longitudinal measurements and time-to-event*, REVSTAT 9 (2011), pp. 57–81.
- [42] G.H. Struijk, R.C. Minnee, S.D. Koch, A.H. Zwinderman, K.A.M.I. van Donselaar-van der Pant, M.M. Idu, I.J.M. Bergeten, and F.J. Bemelman, *Maintenance immunosuppressive therapy with everolimus preserves humoral immune responses*, Kidney Int. 78 (2010), pp. 934–940.
- [43] Y.J. Sung and C.J. Geyer, *Monte Carlo likelihood inference for missing data models*, Ann. Statist. 35 (2007), pp. 990–1011.
- [44] M.J. Sweeting and S.G. Thompson, *Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture*, Biom. J. 53 (2011), pp. 750–763.
- [45] A.A. Tsiatis and M. Davidian, *Joint modeling of longitudinal and time-to-event data: An overview*, Statist. Sinica 14 (2004), pp. 809–834.
- [46] S.M. Wulfsohn and A.A. Tsiatis, *A joint model for survival and longitudinal data measured with error*, Biometrics 53 (1997), pp. 330–339.
- [47] J. Xu and S.L. Zeger, *Joint analysis of longitudinal data comprising repeated measures and times to events*, Society 50 (2001), pp. 375–387.
- [48] M. Yu, N. Law, J. Taylor, and H. Sandler, *Joint longitudinal-survival-cure models and their application to prostate cancer*, Statist. Sinica 14 (2004), pp. 832–835.