Statistical challenges in observational cohort studies

Hof, M.H.P.

Publication date
2015
Document Version
Final published version

5 Quasi Monte Carlo approximation in joint models for multiple longitudinal markers and recurrent events of multiple types, in the presence of a terminal event

In medical studies we are often confronted with complex longitudinal data consisting of multiple time to event data and multiple markers. During the follow up period, subject can experience recurrent events (e.g. infections, strokes) of multiple types. In addition, the follow up period can be ended by a terminal event (e.g. death). After the terminal event, data from the recurrent events and the markers are missing. An adverse health status, represented by 'bad' marker values and an abnormal number of recurrent events, is often correlated with the terminal event. In this situation, the missingness of the data from the recurrent events and the markers is not at random and all available data should be modeled jointly to obtain unbiased parameter estimates of all submodels. In our joint model, the correlations between the repeated observations of a specific marker or certain event type are represented by normally distributed random effects. We have an analytically intractable integral in our joint model, which can be evaluated using Bayesian approaches or quadrature approximation techniques. However, when we have a large number of recurrent event types and markers, the dimensionality of the integral is high and these methods are too computationally expensive. In this chapter, we propose a quasi-Monte Carlo (QMC) approach to evaluate the likelihood. With the QMC approach the integrals are evaluated using a deterministic point set. The accuracy of the approximation depends on the size of the QMC point set. Our simulations showed that the QMC approximation method gives accurate parameter estimates of all event submodels. In addition, the associations between the markers and the event submodels were also accurately estimated. However, the standard errors of the parameter estimates of the marker submodels were underestimated which was observed by too low coverages of the 95% confidence intervals.

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 limited 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 [98]. For instance, consider data from patients after kidney transplant [99]. In this study, subjects were monitored longitudinally using five immunological markers, and could experience up to ten infection types, all at multiple times. Subjects also dropped out of the study before the follow-up period ended. Drop-out could be caused by death; subjects who were more frail and thus experienced more recurrent infections, might have consequently died at some point due to impaired worsening of their immune status. This could be reflected as well in their marker trajectories (for instance by declining marker values). Joint models have been used for this type of data; some key references on joint modeling techniques are [100, 101, 102], and a comprehensive overview can be found in [32, 103, 104].

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 [33, 34, 35], (*ii*) a single longitudinal outcome and multiple terminal event types (i.e. a competing risks situation) [36, 37, 38], (*iii*) multiple longitudinal outcomes and multiple terminal event types [39], and (*iv*) multiple longitudinal outcomes, a single recurrent event type, and a single terminal event type [40].

In this paper, we consider joint models for data of consisting of multiple markers, multiple recurrent event types, and a terminal event. Our model comprised of (*i*) a multivariate mixed effects submodel where the markers could develop simultaneously through time, thus accounting for the interrelation between the markers, as well as the intra-individual dependency over time. (*ii*) A survival submodel that shared 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 [105, 106], likelihood approaches using quadrature methods for evaluating the integrals [40, 33], and EM-type of approaches [107] have been used to estimate the parameters of joint models. However, even for a small number of markers and recurrent event types (e.g. > 2 markers and > 2 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 time varying random effects or frailties are necessary. Consequently, these traditional methods to maximize the log likelihoods 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 Monte Carlo (MC) techniques to integrate out the random effects and frailties [108]. 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 error 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 [109]. Instead of using a set of random vectors of random effects and frailties, a deterministic point is used which approximates the distributions of the random effects and frailties as closely as possible. To achieve this property, we use quasi-random sequences (e.g. Faure, Halton, or Sobol sequences [109]) which are formed by points in a hypercube and are designed to achieve a high level of uniformity [110, 111]. With transformations, these point sets can be used to approximate a large number of distributions, e.g. the multivariate normal with a given covariance matrix [112].

The rest of this chapter is organized as follows. First, the submodels of the joint model are introduced, along with the joint model likelihood. Secondly, the QMC estimator is introduced using the quasi-random sequences to approximate the likelihood of the joint model. Thirdly, we describe how we can use the QMC

estimator for prediction. Finally, we illustrate properties of the QMC estimator using simulations and we fit our joint model on kidney transplantation data.

Methods

Submodels specification

Consider that we have longitudinal data of n subjects from M measured markers, and R recurrent event types. During the follow-up period, subject $i = 1, \ldots, n$ could experience recurrent events of different types and had repeated measurements of multiple markers. Both the markers and events are observed until time T_i, which is the minimum of a censoring time C_i and a terminal event time T_i^*, i.e. $T_i = \min(T_i^*, C_i)$. 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 depends on the markers and events processes. In addition, for individual i we have the terminal event indicator δ_{i0} defined as $\delta_{i0} = I(T_i^* < C_i)$, where $I()$ is the indicator function.

Let $y_{im}(t)$ be the value of marker $m = 1, \ldots, M$ at time t for the i^{th} subject. The m^{th} marker is observed J_{m_i} times in the interval $[0, T_i)$, leading to the vector of observed measurements $\mathbf{y}_{im} = \{y_{im}(t_{im,j}); 0 \leq t_{im,j} < T_i \text{ and } j = 1, \ldots, J_{m_i}\}$. 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 $r = 1 \ldots, R$ are described by the vector $\mathbf{t}_{ir} = \{t_{ir,k}; 0 < t_{ir,k} < T_i \text{ and } k = 1, \ldots, K_{r_i}\}$, where K_{r_i} is the number of times recurrent event type r occurs in subject i. Our aim is to estimate the associations between the marker trajectories and the risk of experiencing a recurrent event type or a terminal event. We propose a joint model approach in which we simultaneously estimate the submodels regarding the evolution of the markers through time, the risks of the recurrent event types, and the risk of the terminal event. The dependencies between all submodels are based on the following assumptions [107, 113, 40, 33], where for subject i

1. given the random effects of the M markers $\boldsymbol{v}_i = \{\boldsymbol{v}_{im}; m = 1, \ldots, M\}$, the repeated measures of the markers are uncorrelated,

2. given the frailties of the R recurrent event types $\boldsymbol{w}_i = \{w_{ir}; r = 1, \ldots, R\}$, the occurrences of the recurrent event types are uncorrelated with each other and uncorrelated with the terminal event,
3. the hazards of the recurrent event types at time t depend on the vector of true marker values at t, denoted as $\{y^*_{im}(t); m = 1, \ldots, M\}$,
4. the hazard of the terminal event at time t depends on the R frailty terms of the recurrent event types $\boldsymbol{w}_i$ and on the M true marker values 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 [104]. We have the following submodel for the m^{th} marker for subject i

$$\begin{aligned} y_{im}(t) &= y^*_{im}(t) + \epsilon_{im}(t), \qquad \epsilon_{im}(t) \sim N(0, \alpha^2_m) \\ y^*_{im}(t) &= \boldsymbol{\phi}'_m \mathbf{z}_{im}(t) + \boldsymbol{v}'_{im} \mathbf{q}_{im}(t), \end{aligned} \tag{5.1}$$

where $y_{im}(t)$ is the observed marker value of marker m at time t. 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 $\boldsymbol{v}_{im}$ is the vector of random effects. The vectors $\mathbf{z}_{im}(t)$ and $\mathbf{q}_{im}(t)$ may contain (transformed) values of time t to capture any (non-)linear effect of time on the marker value. Finally, $\epsilon_{im}(t)$ is a normally distributed measurement error with variance α^2_m which is assumed to be white noise; i.e. $\epsilon_{im}(t)$ is independent of $\boldsymbol{v}_{im}$, the R frailty terms w_{ir}, and the measurement error of the other markers.

For subject i, $h_{ir}(t)$ is the hazard of the r^{th} recurrent event type at time t, and $\lambda_i(t)$ is the hazard of the terminal event at time t. The hazard for the r^{th} recurrent event type is written as

$$h_{ir}(t) = h_{0r}(t) \exp\left\{ \boldsymbol{\beta}'_r \mathbf{x}_{ir}(t) + w_{ir} + \sum_{m=1}^{M} \gamma_{rm} y^*_{im}(t) \right\}, \tag{5.2}$$

where $h_{0r}(t)$ is the baseline hazard at time t, $\mathbf{x}_{ir}(t)$ is a vector of (possibly time-varying) characteristics of subject i at time t associated with the risk of experiencing event r, $\boldsymbol{\beta}_r$ 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 (5.2) denotes the weighted sum of the effects of the true 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 for the terminal event is modeled as

$$\lambda_i(t) = \lambda_0(t) \exp\left\{ \boldsymbol{\beta}_0' \mathbf{x}_{i0}(t) + \sum_{m=1}^{M} \gamma_{0m} y^*_{im}(t) + \sum_{r=1}^{R} \chi_r w_{ir} \right\}, \tag{5.3}$$

where $\lambda_0(t)$ is the baseline hazard 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. The effect of the m^{th} marker value is given by γ_{m0} and the effect of the r^{th} frailty term by χ_r.

The joint likelihood function

In this chapter, we assume that $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$ follow independent multivariate normal distributions centered at zero with covariance matrices $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_w$ respectively, i.e. $\boldsymbol{v}_i \sim N(0, \boldsymbol{\Sigma}_v)$ and $\boldsymbol{w}_i \sim N(0, \boldsymbol{\Sigma}_w)$. Dependencies between the markers are captured by the covariances in $\boldsymbol{\Sigma}_v$. Similarly, the dependencies between the recurrent event types are described by the covariances in $\boldsymbol{\Sigma}_w$. Furthermore, all the dependencies between the markers and events submodels are captured by the last terms of (5.2) and (5.3). The covariances between $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$ are assumed to be zero.

Let $\boldsymbol{\theta} = \{\boldsymbol{h}_0(t), \lambda_0(t), \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\chi}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\Sigma}_v, \boldsymbol{\Sigma}_w\}$ be the collection of all joint model parameters. The *observed* data from subject i is $\mathcal{D}_i(T_i) = \{\mathcal{Y}_i(T_i), \mathcal{Z}_i(T_i)\}$, where

$$\begin{aligned}
\mathcal{Y}_i(T_i) =& \{y_{im}(t_{im,j}); 0 \leq t_{im,j} \leq T_i \text{ and } m = 1, \ldots, M \text{ and } j = 1, \ldots, J_{m_i}\}, \\
\mathcal{Z}_i(T_i) =& \{t_{ir,k}; 0 \leq t_{ir,k} \leq T_i \text{ and } r = 0, \ldots, R \text{ and } k = 1, \ldots, K_{r_i}\},
\end{aligned}$$

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

$$\sum_{i=1}^{n} \log[\mathcal{L}_i(\mathcal{D}_i(T_i)|\boldsymbol{\theta})], \tag{5.4}$$

where $\mathcal{L}_i(\mathcal{D}_i(T_i)|\boldsymbol{\theta})$ is the subject-specific likelihood contribution

$$\begin{aligned}\mathcal{L}_i(\mathcal{D}_i(T_i)|\boldsymbol{\theta}) = \\ \int_{(\boldsymbol{v}_i,\boldsymbol{w}_i)} L_{i0}(T_i|\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0,\boldsymbol{\phi},\boldsymbol{\chi},\boldsymbol{v}_i,\boldsymbol{w}_i) \times \prod_{r=1}^{R} L_{ir}(\mathbf{t}_{ir}|\boldsymbol{\beta}_r,\boldsymbol{\gamma}_r,\boldsymbol{\phi},\boldsymbol{v}_i,w_{ir}) \times \\ \prod_{m=1}^{M} L_{im}(\mathbf{y}_{im}|\boldsymbol{\phi}_m,\alpha_m,\boldsymbol{v}_{im}) \times p(\boldsymbol{v}_i,\boldsymbol{w}_i|\boldsymbol{\Omega}) d(\boldsymbol{v}_i,\boldsymbol{w}_i),\end{aligned} \tag{5.5}$$

where $p(\boldsymbol{v}_i,\boldsymbol{w}_i|\boldsymbol{\Omega}) \sim N(0,\boldsymbol{\Omega})$, with covariance matrix $\boldsymbol{\Omega}$, which can be written as

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{v}} & 0 \\ 0 & \boldsymbol{\Sigma}_w \end{pmatrix}. \tag{5.6}$$

The contribution to the likelihood of the measurements of the m^{th} marker is

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

where $y^*_{im}(t_{im,j})$ denotes the true marker value at time $t_{im,j}$ from (5.1). In addition, the likelihood contribution of the r^{th} recurrent event type is [114]

$$L_{ir}(\mathbf{t}_{ir}|\boldsymbol{\beta}_r,\boldsymbol{\gamma}_r,\boldsymbol{\phi},\boldsymbol{v}_i,w_{ir}) = \left[\prod_{k=1}^{K_{r_i}} h_{ir}(t_{ir,k})\right] \times \exp\left\{-\int_0^{T_i} h_{ir}(u)\mathrm{d}u\right\}, \tag{5.7}$$

where $h_{ir}(t)$ is the subject specific hazard at time t from (5.2). Furthermore, the likelihood contribution of the terminal event is

$$L_{i0}(T_i|\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\phi}, \boldsymbol{\chi}, \boldsymbol{v}_i, \boldsymbol{w}_i) = \lambda_i(T_i)^{\delta_{i0}} \exp\left\{-\int_0^{T_i} \lambda_i(u)\mathrm{d}u\right\}, \tag{5.8}$$

where $\lambda_i(t)$ is the subject specific hazard for a terminal event at time t from (5.3). In this chapter, the baseline hazards $h_{0r}(t)$ and $\lambda_0(t)$ in respectively (5.2) and (5.3) are modeled with piecewise functions that may be different for the various recurrent event types and for the terminal event. For the r^{th} recurrent event type, the follow-up time is divided into B_r intervals with cut-off points $\tau_{r0} < \tau_{r1} < \ldots < \tau_{rB_r}$, with $\tau_{r0} = 0$ and $\tau_{rB_r} = \infty$. The baseline hazard $h_{0r}(t)$ is then calculated as

$$h_{0r}(t) = \sum_{l=1}^{B_r} I(\tau_{rl-1} \leq t < \tau_{rl})\tilde{h}_{rl}$$

where $\tilde{h}_{rl}$ denotes the constant baseline hazard for the interval $[t_{rl-1}, t_{rl})$. We can perform the same procedure for the terminal event baseline hazard, where we use the cut-off points $\tau_{00} < \tau_{01} < \ldots < \tau_{0B_r}$, with $\tau_{00} = 0$ and $\tau_{0B_0} = \infty$.

Judicious choice of the cut-off points allows to approximate almost any type of baseline hazard, using smaller intervals where the baseline hazards rapidly change and larger intervals where the baseline hazards remain almost constant. Feng et al. showed that, for a frailty survival model, a piecewise constant hazard with 8-10 intervals often yields excellent estimates of the other model parameters [115]. The integrals in (5.7) and (5.8) are approximated with uni dimensional quadrature methods [33].

Evaluating the joint model likelihood

The joint log likelihood of (5.4) is maximized with respect to all joint model parameters $\boldsymbol{\theta}$ to obtain the MLE $\widehat{\boldsymbol{\theta}}$. The integral in (5.5) is of dimension $d = d_v + d_w$, which are the sizes of respectively the vectors $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$. 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 non-linear trends over time.

The problem of evaluating (5.5) is that the integral in $\mathcal{L}_i(\mathcal{D}_i(T_i)|\boldsymbol{\theta})$ is analytically intractable and (potentially) high dimensional. Therefore, we require numerical approximation techniques for the evaluation of $\mathcal{L}_i(\mathcal{D}_i(T_i)|\boldsymbol{\theta})$. Because we use a fully parametric model, we could use well-known numerical techniques to approximate the integral, such as Gauss-Hermite quadrature (GHQ) integration, and Monte-Carlo (MC) integration [116]. With GHQ, the locations of the evaluation points and their corresponding weights (usually) depend on the number of evaluation points chosen per dimension [117], and often each dimension has the same number of points. To create multidimensional quadratures, we combine unidimensional rules by taking their Kronecker product. Consequently, the size of the point set increases exponentially with the number of dimensions. Consider that we have a d dimensional integral and we decide to have seven evaluation points per dimension. The point set contains $Z = 7^d$ vectors of dimension d, which is substantial even when we have a moderate number of dimensions (e.g. $Z = 7^5 = 16807$, $Z = 7^6 = 117649$, $Z = 7^7 = 823543$ vectors) and the GHQ method quickly becomes too computationally expensive. However, for low-dimensional integration problems, GHQ is usually considered to be the preferred method. Therefore, GHQ has been frequently used in the estimation of joint models in which the dimensionality of the random effects were relatively small (e.g. fitting a joint model with only one marker, one recurrent event type, and one terminal event [113]).

When we have moderate to high dimensional random effects (i.e. $d > 5$), we propose to use MC integration to approximate (5.5). For a given set of random vectors $\left(\boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)}\right)$ drawn from $p(\boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\Omega})$, where $(z = 1, \ldots, Z)$, the approximated likelihood contribution of subject i is

$$\mathcal{L}^*_{iZ}(\mathcal{D}_i(T_i)|\boldsymbol{\theta}) = \frac{1}{Z}\sum_{z=1}^{Z}\Big\{L_{i0}(T_i|\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\phi}, \boldsymbol{\chi}, \boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)}) \times \prod_{r=1}^{R} L_{ir}(\mathbf{t}_{ir}|\boldsymbol{\beta}_r, \boldsymbol{\gamma}_r, \boldsymbol{\phi}, \boldsymbol{v}_i^{(z)}, w_{ir}^{(z)}) \times \prod_{m=1}^{M} L_{im}(\mathbf{y}_{im}|\boldsymbol{\phi}_m, \alpha_m, \boldsymbol{v}_{im}^{(z)})\Big\}, \tag{5.9}$$

Following the strong law of large numbers, we have $\lim_{Z\to\infty} \mathcal{L}^*_{iZ}(\mathcal{D}_i(T_i)|\boldsymbol{\theta}) = \mathcal{L}_i(\mathcal{D}_i(T_i)|\boldsymbol{\theta})$ [108]. The integration error-bound of the MC approximation is of the order $O(1/\sqrt{Z})$, regardless of the dimensionality of the integral. Unfortunately, the strong law of large numbers does not guarantee that the MC method will always behave well for point sets of finite size [112].

When we randomly drawn vectors from $p(\boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\Omega})$, some parts of the density will have more points than necessary (clusters of vectors) and other parts less than necessary (large gaps between vectors). To improve the convergence rate of MC approximation, quasi-Monte Carlo (QMC) methods have been developed [118, 110]. Instead of using a random sample from $p(\boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\Omega})$, a deterministic point set is used which is chosen to approximate the distribution $p(\boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\Omega})$ as closely as possible. Consequently, the integration error of the QMC method *could* come close to the order $O(1/Z)$. Generally, QMC approximation gives better approximations of integrals than MC approximation [119, 120, 121]

The deterministic point set for our joint model 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[122, 111]. Examples of sequences that can be used to create low-discrepancy points sets are Halton, Faure, or Sobol sequences [109]. The steps in approximating the likelihood are:

1. *Generate the deterministic QMC point set*: QMC point sets can be generated with the R package randtoolbox [123]. Similarly to [124], we assign different QMC point sets to each subject in order to avoid correlation in the integration errors across subjects. Therefore each subject i has its own QMC point set $\mathbf{B}_i$ of size $Z \times d$. All n matrices $\mathbf{B}_i$ are obtained by generating a $Zn \times d$ dimensional QMC point set $\mathbf{B}$, from which the first Z rows are assigned to $\mathbf{B}_1$, the second Z rows to $\mathbf{B}_2$, and so on.

2. *Change the integration region*: The QMC point set $\mathbf{B}_i$ contains values $b_{i,zj}$ ($z = 1, \ldots, Z$ and $j = 1, \ldots, d$), 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 $\widetilde{\mathbf{B}}_i$ be the matrix of transformed values with the elements $\tilde{b}_{i,zj} = \Phi^{-1}(b_{i,zj})$, for $j = 1, \ldots, d$, where $\Phi^{-1}(.)$ is the inverse of the standard normal cumulative distribution function [125].

3. *Add the correlation structure*: The point set $\mathbf{U}_i = \{\mathbf{u}_{iz}; z = 1, \ldots, Z\}$ is obtained via $\mathbf{U}_i = \widetilde{\mathbf{B}}_i\mathbf{Q}$, where $\mathbf{Q}$ is the upper triangular matrix from the Cholesky decomposition $\boldsymbol{\Omega} = \mathbf{Q}^\intercal\mathbf{Q}$ [126].
4. *Obtain the approximated likelihood*: The point set $\mathbf{U}_i$ is used to obtain the subject specific likelihood contribution $\mathcal{L}^*_{iZ}(\mathcal{D}_i(T_i)|\boldsymbol{\theta})$ from (5.9), where $\mathbf{u}_{ih} = \left(\boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)}\right)$. The log likelihood of the data from all subjects is $\sum_{i=1}^n \log[\mathcal{L}^*_{iZ}(\mathcal{D}_i(T_i)|\boldsymbol{\theta})]$.

Throughout this chapter, 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 [127, 128, 129, 130].

Quasi Monte Carlo maximum likelihood estimation

To obtain the parameter estimates, we maximize the approximated log likelihood of the data to get the QMC maximum likelihood estimate

$$\hat{\boldsymbol{\theta}}_{QMC} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \log[\mathcal{L}^*_{iZ}(\mathcal{D}_i(T_i)|\boldsymbol{\theta})].$$

where $\boldsymbol{\Theta}$ is the parameter space. The QMC estimator is consistent, if $Z \to \infty$ and $n \to \infty$ and the likelihood is correctly specified. Moreover, when $n, Z \to \infty$ and $\sqrt{n}/Z \to 0$ the QMC estimator is asymptotically equivalent to an ML estimator (proposition 3.2, page 44 of [108]).

Finding $\hat{\boldsymbol{\theta}}_{QMC}$ requires the use of optimization methods (e.g. Newton-Raphson optimization), which generally need to evaluate the approximated log likelihood a substantial number of times. To limit the computational costs of locating $\hat{\boldsymbol{\theta}}_{QMC}$, the number of vectors Z used to approximate the integral in (5.5) might be chosen too small to obtain the required accuracy. Consequently, $\hat{\boldsymbol{\theta}}_{QMC}$ is contaminated with simulation error and the QMC estimator does not necessarily share the same properties as the ML estimator [131].

Hajivassiliou suggested a two-step estimator based on linearized maximum likelihood (LML) estimation to (partially) remove the simulation error [132]. The main principle behind the LML estimator is that, when we have to evaluate the

log likelihood a relatively small number of times, we can approximate the log likelihood with a very large QMC point set and the simulation error is then negligible. Importantly, this means that the LML estimator has all the standard asymptotic properties of estimators not based on simulation [132].

The two-step approach works as follows. First, we obtain the QMC estimate $\hat{\boldsymbol{\theta}}_{QMC}$ by maximizing the approximated log likelihood with a large enough Z. Given the initial QMC estimate $\hat{\boldsymbol{\theta}}_{QMC}$, the second step is

$$\hat{\boldsymbol{\theta}}_{LML} = \hat{\boldsymbol{\theta}}_{QMC} + H_n(\hat{\boldsymbol{\theta}}_{QMC})^{-1} E_n[s(\hat{\boldsymbol{\theta}}_{QMC})],$$

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

$$\begin{aligned} H_n(\boldsymbol{\theta}) =& \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \log[\mathcal{L}^*_{iZ}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right)' \left(\frac{\partial \log[\mathcal{L}^*_{iZ}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \right), \\ E_n[s(\boldsymbol{\theta})] =& \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log[\mathcal{L}^*_{iZ}(\boldsymbol{\theta})]}{\partial \boldsymbol{\theta}}, \end{aligned} \tag{5.10}$$

using a very large Z. The variance of the LML estimate is derived from the approximated information matrix $H_n(\hat{\boldsymbol{\theta}}_{LML})$.

Dynamic prediction

In this section, we use the fitted joint model and the QMC approximation 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. Given the follow-up history of subject i until t_{obs}, we predict the probability of experiencing no new recurrent event type r (or terminal event) until time $T > t_{obs}$. Let $\mathcal{D}(t_{obs})$ be the observed history of subject i until t_{obs}. The conditional survival function for event type r is

$$S_{ir}(T|\mathcal{D}(t_{obs}), T > t_{obs}, \boldsymbol{\theta}, \boldsymbol{\Omega}) = \int_{(\boldsymbol{v}_i, \boldsymbol{w}_i)} S_{ir}(T|T > t_{obs}, \boldsymbol{\theta}, \boldsymbol{v}_i, \boldsymbol{w}_i) p(\boldsymbol{v}_i, \boldsymbol{w}_i|\mathcal{D}(t_{obs}), \boldsymbol{\Omega}) d(\boldsymbol{v}_i, \boldsymbol{w}_i). \tag{5.11}$$

For the recurrent events types $r = 1, \ldots, R$, we have

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

and for the terminal event

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

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

$$p(\boldsymbol{v}_i, \boldsymbol{w}_i|\mathcal{D}(t_{obs}), \boldsymbol{\Omega}) = \frac{\mathcal{L}_i(\mathcal{D}(t_{obs})|\boldsymbol{\theta}, \boldsymbol{v}_i, \boldsymbol{w}_i) p(\boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\Omega})}{\int_{(\boldsymbol{v}_i, \boldsymbol{w}_i)} \mathcal{L}_i(\mathcal{D}(t_{obs})|\boldsymbol{\theta}, \boldsymbol{v}_i, \boldsymbol{w}_i) p(\boldsymbol{v}_i, \boldsymbol{w}_i|\boldsymbol{\Omega}) d(\boldsymbol{v}_i, \boldsymbol{w}_i)}, \tag{5.12}$$

where $\mathcal{L}_i(\mathcal{D}(t_{obs})|\boldsymbol{\theta}, \boldsymbol{v}_i, \boldsymbol{w}_i)$ is the conditional joint likelihood of the observed history $\mathcal{D}(t_{obs})$ calculated as

$$\mathcal{L}_i(\mathcal{D}(t_{obs})|\boldsymbol{\theta}, \boldsymbol{v}_i, \boldsymbol{w}_i) = L_{i0}(T_i|\beta_0, \boldsymbol{\gamma}_0, \boldsymbol{\phi}, \boldsymbol{\chi}, \boldsymbol{v}_i, \boldsymbol{w}_i) \times \prod_{r=1}^{R} L_{ir}(\mathbf{t}_{ir}|\beta_r, \boldsymbol{\gamma}_r, \boldsymbol{\phi}, \boldsymbol{v}_i, w_{ir}) \times \prod_{m=1}^{M} L_{im}(\mathbf{y}_{im}|\boldsymbol{\phi}_m, \alpha_m, \boldsymbol{v}_{im}).$$

Because the integrals in (5.11) are analytically intractable and potentially high dimensional, we use the QMC method to estimate the conditional survival function $S_{ir}(T|\mathcal{D}(t_{obs}), T > t_{obs}, \boldsymbol{\theta})$. We could estimate the integrals in (5.11) and (5.12) with the same accuracy by using $\mathbf{U}_i = \left\{ \left(\boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)} \right); z = 1, \ldots, Z \right\}$ of size Z to approximate both integrals. Note that a relatively large point set can be used to approximate $S_{ir}(t|\mathcal{D}(t_{obs}), t > t_{obs}, \boldsymbol{\theta})$ since the integrals only need to be evaluated once. Thus, $S_{ir}(T|\mathcal{D}(t_{obs}), T > t_{obs}, \boldsymbol{\theta})$ is approximated as

$$\hat{S}_{ir}(T|\mathcal{D}(t_{obs}), T > t_{obs}, \boldsymbol{\theta}, \boldsymbol{\Omega}) = \frac{\sum_{z=1}^{Z} S_{ir}(T|T > t_{obs}, \boldsymbol{\theta}, \boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)}) \mathcal{L}_{iZ}^{*}(\mathcal{D}(t_{obs})|\boldsymbol{\theta}, \boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)})}{\sum_{z=1}^{Z} \mathcal{L}_{iZ}^{*}(\mathcal{D}(t_{obs})|\boldsymbol{\theta}, \boldsymbol{v}_i^{(z)}, \boldsymbol{w}_i^{(z)})}.$$

Simulation

Setup

To illustrate the properties of our approach, 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. Marker observations at time t were obtained from the relations

$$\begin{aligned} y_{i1}(t) &= (\phi_{10} + v_{i1,0}) + (\phi_{11} + v_{i1,1})t + \epsilon_{i1}(t), \\ y_{i2}(t) &= (\phi_{20} + v_{i2,0}) + (\phi_{21} + v_{i2,1})t + \epsilon_{i2}(t). \end{aligned}$$

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 $(v_{i1,0}, v_{i1,1}) \sim N(0, \boldsymbol{\Sigma}_{v1})$ and $(v_{i2,0}, v_{i2,1}) \sim N(0, \boldsymbol{\Sigma}_{v2})$ for respectively the first and second marker, with covariance matrices

$$\boldsymbol{\Sigma}_{v_1} = \begin{pmatrix} 0.40 & 0.00 \\ 0.00 & 0.03 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{v_2} = \begin{pmatrix} 0.30 & 0.01 \\ 0.01 & 0.04 \end{pmatrix}.$$

There was no correlation *between* the random effects of both markers and both markers were observed at the same times. The measurement errors $(\epsilon_{i1}(t), \epsilon_{i2}(t))$ were drawn independently from mean zero normal distributions with variances of respectively $\alpha_1^2 = 0.16$ and $\alpha_2^2 = 0.16$.

To generate the times for subject i at which we observed both markers, ten time points were drawn from a uniform distribution over $[0, C_i)$. Consequently, all subjects had up to 10 time points at which we observed the markers.

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

$$h_{i1}(t) = h_{01} \exp\left\{\beta_1 x_i + w_{i1} + \gamma_{11} f_{i1}(t) + \gamma_{12} f_{i2}(t)\right\},$$
$$h_{i2}(t) = h_{02} \exp\left\{\beta_2 x_i + w_{i2} + \gamma_{21} f_{i1}(t) + \gamma_{22} f_{i2}(t)\right\},$$

where the true parameter values are $(h_{01}, h_{01}) = (0.3, 0.2)$, $(\beta_1, \beta_2) = (-0.15, 0.2)$, $(\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}) = (0.1, 0.08, 0.09, 0.12)$, and frailties $(w_{i1}, w_{i2}) \sim N(0, \boldsymbol{\Sigma}_w)$ with

$$\boldsymbol{\Sigma}_w = \begin{pmatrix} 0.40 & 0.05 \\ 0.05 & 0.30 \end{pmatrix}.$$

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

$$\lambda_i(t) = \lambda_0 \exp\left\{\beta_0 x_i + \gamma_{01} f_{i1}(t) + \gamma_{02} f_{i2}(t) + \chi_1 w_{i1} + \chi_1 w_{i2}\right\}, \tag{5.13}$$

with constant baseline hazard $\lambda_0 = 0.05$, and $(\beta_0, \gamma_{01}, \gamma_{01}) = (0.35, 0.15, 0.1)$. The effects of the frailty terms from the recurrent events were $(\chi_1, \chi_2) = (0.3, 0.2)$.

For the simulation, we used Sobol sequences to generate the deterministic point set which was used to approximate the integral in (5.5). To investigate whether the LML estimator improved the estimates of the QMC estimator, we fitted the following models to the data

One-step estimator
: Obtain $\hat{\boldsymbol{\theta}}_{QMC}$ by maximizing the approximated log likelihood. The variances of the parameter estimates $\hat{\boldsymbol{\theta}}_{QMC}$ were derived from a approximated hessian matrix evaluated at $\hat{\boldsymbol{\theta}}_{QMC}$ using (5.10).

Two-step estimator
: Update $\hat{\boldsymbol{\theta}}_{QMC}$ using the approximated scores and hessian matrix to obtain $\hat{\boldsymbol{\theta}}_{LML}$. The variances of the parameter estimates $\hat{\boldsymbol{\theta}}_{LML}$ were derived from a approximated hessian matrix evaluated at $\hat{\boldsymbol{\theta}}_{LML}$ using (5.10).

Two scenarios were considered, in which the number of vectors Z in the Sobol point set differed in the first step of the estimation procedure (i.e. to obtain $\hat{\boldsymbol{\theta}}_{QMC}$). The number of vectors in the Sobol point set for each step of the estimation procedure were:

	One-step estimator		Two-step estimator	
	$\hat{\boldsymbol{\theta}}_{QMC}$	$\widehat{\text{var}}(\hat{\boldsymbol{\theta}}_{QMC})$	$\hat{\boldsymbol{\theta}}_{LML}$	$\widehat{\text{var}}(\hat{\boldsymbol{\theta}}_{LML})$
scenario 1	2500	50000	50000	50000
scenario 2	10000	50000	50000	50000

For all parameters of the one-step estimator and the two-step estimator in both runs, we reported the mean estimate, mean squared error (MSE), and coverage of the 95% confidence interval based on repeating the simulation 400 times. The simulation was performed with the statistical software R. Sobol sequences, generated with the randtoolbox package [123], were used to create the QMC point set that was used to approximate the likelihood. For the maximization of the approximated likelihood, we used a trust region optimization routine [133], with a numerically approximated gradient function of the likelihood. In addition, the LML estimation was based on numerically approximated scores, using the numDeriv package [134].

Results

The simulation results are summarized in tables 5.1 and 5.2. On average, in both scenarios of the simulation, both the one-step estimator and the two-step estimator gave unbiased estimates of all joint model parameters. When we used a relatively small number of vectors to obtain $\hat{\boldsymbol{\theta}}_{QMC}$, the coverages of the 95% confidence interval were underestimated. This meant that we had non-ignorable simulation bias in the analysis, which was not dealt with in the variance estimates of the parameters.

For scenario 1 of the simulation, the one-step estimator gave too low coverages of the 95% confidence intervals for the parameters of the random effect covariance matrix, the marker submodels, and the recurrent event submodels. However, the coverages of the 95% confidence intervals of the terminal event submodel were good. Regarding the differences between the one-step and two-step estimator, a clear improvement in accuracy of the parameter estimates was found with the two-step estimator. Not only was the two-step estimator more accurate, which

was reflected in better or equally good coverages of the 95% confidence interval for all parameters, it also gave a lower MSE for all the random effect parameters and the marker submodel parameters. However, a small increase in MSE was observed for the terminal event submodel parameters.

The joint model fitted in scenario 2 of the simulation gave more accurate results because more vectors were used to integrate out the random effects in (5.5); for all parameters in the terminal and recurrent event submodels both estimators had good coverages of the 95% confidence intervals. The coverages of the 95% confidence intervals of the parameters in the marker submodel parameters were however still underestimated. A similar trend was observed in the random effect and frailty parameters; while the frailty variances and covariances of the recurrent event submodels had accurate coverages of the 95% confidence interval, the variances and covariances of the random effects of the markers had too small coverages of the 95% confidence intervals.

			One-step estimator ($\boldsymbol{\theta}_{QMC}$)			Two-step estimator ($\boldsymbol{\theta}_{LML}$)		
	Parameter	True	Estimate	MSE×1000	Coverage 95% C.I.	Estimate	MSE×1000	Coverage 95% C.I.
Random effects markers and recurrent events	$\sigma^{v_1}_{1,1}$	0.400	0.411	3.98	0.812	0.398	2.77	0.865
	$\sigma^{v_1}_{1,2}$	0.000	0.008	0.31	0.460	0.006	0.20	0.465
	$\sigma^{v_1}_{2,2}$	0.030	0.034	0.03	0.578	0.031	0.02	0.610
	$\sigma^{v_2}_{1,1}$	0.300	0.303	1.94	0.873	0.300	1.66	0.885
	$\sigma^{v_2}_{1,2}$	0.010	0.016	0.32	0.512	0.014	0.22	0.550
	$\sigma^{v_2}_{2,2}$	0.040	0.045	0.05	0.575	0.042	0.04	0.608
	$\sigma^{w}_{1,1}$	0.400	0.348	9.65	0.752	0.377	7.21	0.880
	$\sigma^{w}_{1,2}$	0.050	0.048	4.51	0.802	0.049	3.26	0.912
	$\sigma^{w}_{2,2}$	0.300	0.227	5.40	0.777	0.274	5.06	0.925
Marker submodels	ϕ_{10}	0.500	0.499	3.57	0.784	0.503	2.51	0.867
	ϕ_{11}	0.050	0.051	0.58	0.560	0.050	0.39	0.608
	α_1	0.160	0.171	0.04	0.618	0.160	0.04	0.952
	ϕ_{20}	1.000	0.995	2.65	0.829	0.995	1.98	0.864
	ϕ_{21}	0.000	-0.000	0.66	0.548	-0.000	0.37	0.648
	α_2	0.160	0.171	0.04	0.651	0.160	0.04	0.942
Recurrent event submodels	h_{01}	0.300	0.312	0.96	0.907	0.304	0.95	0.927
	β_1	-0.150	-0.156	15.02	0.882	-0.146	13.51	0.932
	γ_{11}	0.100	0.099	2.82	0.907	0.099	3.22	0.912
	γ_{12}	0.080	0.081	2.33	0.912	0.078	2.48	0.912
	h_{02}	0.250	0.262	0.74	0.862	0.255	0.73	0.905
	β_2	0.200	0.191	13.02	0.892	0.196	12.18	0.920
	γ_{21}	0.090	0.093	2.52	0.887	0.091	2.39	0.922
	γ_{22}	0.120	0.120	2.03	0.897	0.121	2.26	0.905
Terminal event submodel	λ_0	0.050	0.049	0.07	0.957	0.053	0.09	0.932
	β_0	0.350	0.356	28.34	0.955	0.355	30.63	0.930
	γ_{01}	0.150	0.158	6.67	0.950	0.155	7.67	0.932
	γ_{02}	0.100	0.100	5.52	0.952	0.101	6.27	0.930
	χ_1	0.300	0.351	86.32	0.947	0.350	85.95	0.912
	χ_2	0.200	0.250	180.35	0.970	0.245	170.96	0.905

Table 5.1: Simulation results for scenario 1 (small Sobol point set).

			One-step estimator ($\boldsymbol{\theta}_{QMC}$)			Two-step estimator ($\boldsymbol{\theta}_{LML}$)		
	Parameter	True	Estimate	MSE×1000	Coverage 95% C.I.	Estimate	MSE×1000	Coverage 95% C.I.
Random effects markers and recurrent events	$\sigma^{v_1}_{1,1}$	0.400	0.410	2.70	0.882	0.393	2.37	0.885
	$\sigma^{v_1}_{1,2}$	0.000	0.003	0.21	0.535	0.003	0.16	0.590
	$\sigma^{v_1}_{2,2}$	0.030	0.033	0.02	0.610	0.031	0.02	0.665
	$\sigma^{v_2}_{1,1}$	0.300	0.309	1.64	0.895	0.299	1.40	0.900
	$\sigma^{v_2}_{1,2}$	0.010	0.012	0.18	0.570	0.012	0.14	0.588
	$\sigma^{v_2}_{2,2}$	0.040	0.043	0.03	0.647	0.041	0.03	0.677
	$\sigma^{w}_{1,1}$	0.400	0.383	7.60	0.895	0.378	6.32	0.917
	$\sigma^{w}_{1,2}$	0.050	0.049	3.42	0.902	0.049	2.77	0.920
	$\sigma^{w}_{2,2}$	0.300	0.272	4.64	0.948	0.272	3.81	0.968
Marker submodels	ϕ_{10}	0.500	0.501	2.69	0.868	0.501	2.19	0.882
	ϕ_{11}	0.050	0.050	0.41	0.603	0.050	0.29	0.672
	α_1	0.160	0.163	0.03	0.948	0.161	0.04	0.950
	ϕ_{20}	1.000	1.000	1.88	0.887	1.000	1.82	0.875
	ϕ_{21}	0.000	0.001	0.43	0.618	0.001	0.34	0.682
	α_2	0.160	0.164	0.03	0.925	0.161	0.03	0.948
Recurrent event submodels	h_{01}	0.300	0.307	0.90	0.948	0.303	0.91	0.945
	β_1	-0.150	-0.150	14.36	0.930	-0.147	13.44	0.942
	γ_{11}	0.100	0.095	2.68	0.935	0.096	3.09	0.925
	γ_{12}	0.080	0.078	2.23	0.927	0.081	2.32	0.927
	h_{02}	0.250	0.256	0.72	0.917	0.254	0.71	0.922
	β_2	0.200	0.200	12.81	0.938	0.196	11.87	0.915
	γ_{21}	0.090	0.093	2.47	0.917	0.093	2.46	0.935
	γ_{22}	0.120	0.117	1.93	0.922	0.118	2.17	0.922
Terminal event submodel	λ_0	0.050	0.050	0.07	0.955	0.051	0.08	0.942
	β_0	0.350	0.360	28.66	0.950	0.355	29.74	0.955
	γ_{01}	0.150	0.158	6.67	0.955	0.159	7.47	0.930
	γ_{02}	0.100	0.097	5.47	0.955	0.100	5.97	0.950
	χ_1	0.300	0.338	60.20	0.948	0.349	65.23	0.940
	χ_2	0.200	0.230	98.90	0.955	0.259	107.81	0.950

Table 5.2: Simulation results for scenario 2 (large Sobol point set).

Post kidney transplant data

Data description

To illustrate the use of our joint model in more complex situations, we analyzed the follow-up data from 357 patients who had a kidney transplant between 2000 and 2009 at the Academic Medical Center in Amsterdam [99]. During the followup, patients could experience up to ten different recurrent infection types. The types of infections and the number of recurrences are given in table 5.3. We excluded the parasitic infection type from the analysis because it was only observed in 8 patients. In addition, five infection types were included in the model without frailty term since the number of recurrences within patients was low.

Included in the joint model

	Infection type	Frequency										
		0	1	2	3	4	5	6	7	8	9	10
With frailty	Viral	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
Without frailty	Cytomegalovirus	158	194	5								
	Fungal	306	40	8	2	0	1					
	Bacterial	325	24	3	3	2						
	Gastroenteritis	333	21	2	1							
	Lower respiratory	316	34	4	3							

Excluded from the joint model

	Infection type	Frequency										
		0	1	2	3	4	5	6	7	8	9	10
	Parasitic	349	8									

Table 5.3: Number of recurrences within patients, and the choice of submodel (in/out of the model, with/without frailty term) for all ten infections.

The patients were followed for up to seven years. During this period 313 (88%) patients dropped out of the study. In our joint model, we treated dropout as a terminal event. At baseline, the covariates age, gender, type of immunosuppressive treatment (type A/ type B; for more information see [106]) and duration of dialysis prior to the transplant were registered. During the follow-up period five longitudinal markers of the patient's immune system were measured; CD3+

T cells, CD4+ T cells, CD8+ T cells, natural killer cells and B cells. All five markers were measured at the same time points. Since almost all variation in CD3+ T cells could be explained by CD4+ T cells and CD8+ T cells, this marker was excluded from the analysis.

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 marker submodels followed the specification of (5.1), 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, which reduced the number of parameters in our joint model substantially. A disadvantage of this choice is that we might lose some detail; the dependencies *between* markers were ignored. For each marker m, the two random effects were assumed to have a mean zero bivariate normal distribution.

The four infection types with frailty terms followed the specification of (5.2); i.e. 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 submodel was parametrized as (5.3); 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 modeled with piecewise constant functions. The choice of cut-off points for the piecewise constant functions depended on the distribution of the event times. In figure 5.1 the event times and the cut-off points for all piecewise constant functions are given.

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 twelve dimensional integral. We used our QMC maximum likelihood estimator to obtain the parameter estimates $\boldsymbol{\theta}_{QMC}$. To approximate the integral, we used a Sobol point set containing $Z = 30000$ vectors. The variance of the parameter estimates was estimated via the approximated information matrix in (5.10). For this approximation, we used a point set with $Z = 250000$ vectors.

The focus of this analysis was to investigate (i) the associations between the markers and recurrent event types, (ii) the associations between the recurrent event types (i.e. the covariance matrix of the frailties), and (iii) the effect of the markers and the four frailty terms on the terminal event.

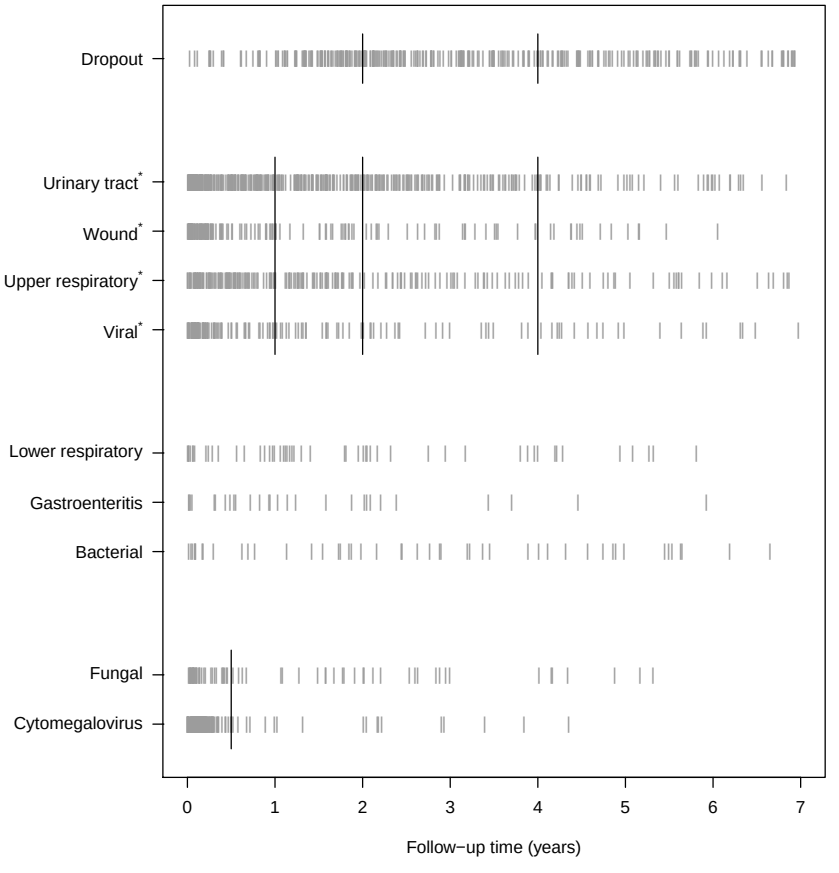

Figure 5.1: Distribution of all nine infection types included in the joint model. Each vertical gray line denotes an event, the black lines represent the cut-off points for the piecewise constant baseline hazards, and the infections with a $*$ are the infection types with frailty terms in their submodels.

Results

The associations between the markers and the events have been summarized in figure 5.2. All estimated associations were non-significant (all p-values > 0.49),

which might be caused by the fact that we only had relatively few marker measurements per patient. With exception of the effect of B cells on the probability of experiencing a cytomegalovirus infection, all associations between the markers and the infections had negative parameter estimates; i.e. higher marker value lowered the probability of experiencing an infection. The estimated associations between the marker values and drop out were close to zero.

The covariance matrix of the frailties and the associations of the frailties with the drop out probability are given in table 5.4. For the viral infection type, the variance of the frailty was really small (0.001). The variances of the frailties for respectively the upper respiratory infection type, wound infection type, and urinary tract infection type were 0.640, 0.925, and 1.568. The frailties of these three infection types were positively correlated.

Although all associations between the frailties and the probability of dropping out of the study were not significant, both the frailty of the upper respiratory infection type and the frailty of the urinary tract infection type had a positive impact on the probability of dropping out of the study. The estimated association between the risk of dropping out of the study and the frailty of the wound infection type was approximately zero.

	Viral	Upper respiratory	Wound Wound	Urinary tract	$\hat{\chi}$ (s.e.)
Viral	0.001	0.008	0.003	0.004	0.006 (0.125)
Upper respiratory	0.008	0.640	0.255	0.309	0.217 (0.112)
Wound	0.003	0.255	0.925	0.474	-0.016 (0.095)
Urinary tract	0.004	0.309	0.474	1.568	0.167 (0.105)

Table 5.4: Estimated variances and covariances of the four frailty terms in the joint model. In addition, the estimated impacts of the frailty terms on the drop out probability $\hat{\chi}$ and their standard errors are given.

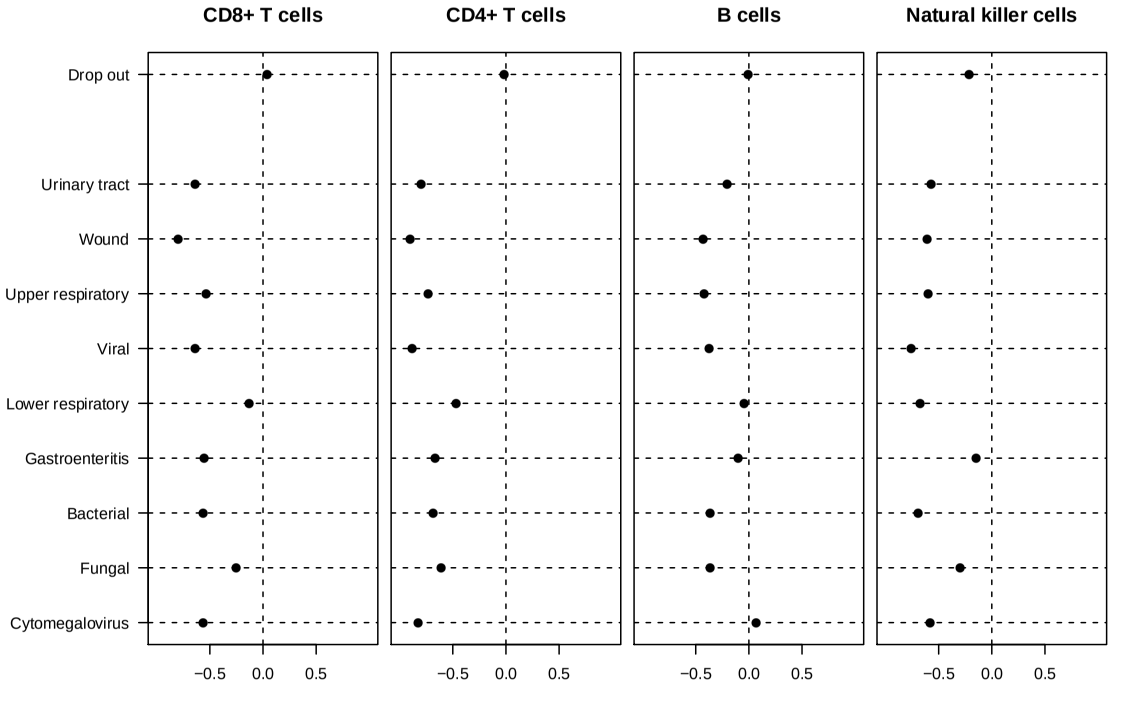

Figure 5.2: Estimated associations between the markers and the events. All the associations were non-significant. We did not include the 95% confidence interval in the figures since they were too large compared to the estimated associations.

Dynamic prediction

To illustrate the use of the QMC method for dynamic prediction, we have predicted the probability of experiencing events before time $T > t_{obs}$ given the history until time t_{obs} for a particular patient. For the example, let $t_{obs} = 0.5$ years. In figure 5.3 the history of the patient is given. The patient had experienced two viral infections, one upper respiratory infection, one urinary tract infection, and a cytomegalovirus infection in the first 0.5 years of follow-up. In addition, all four markers had been measured four times.

We used a Sobol point set of $Z = 500000$ vectors to obtain the probability of experiencing a new infection or dropping out of the study with the QMC method from page 109. The predicted probabilities are given in figure 5.4. Compared to an individual who experienced no infections until 0.5 years of follow-up and had average marker trajectories, the probability of experiencing a urinary tract infection, a wound infection, and a urinary tract infection increased. The probability of experiencing a viral infection was only slightly increased. In addition, the probability of dropping out was also increased. On the other hand, the probabilities of experiencing one of the five infections without a frailty term did not change.

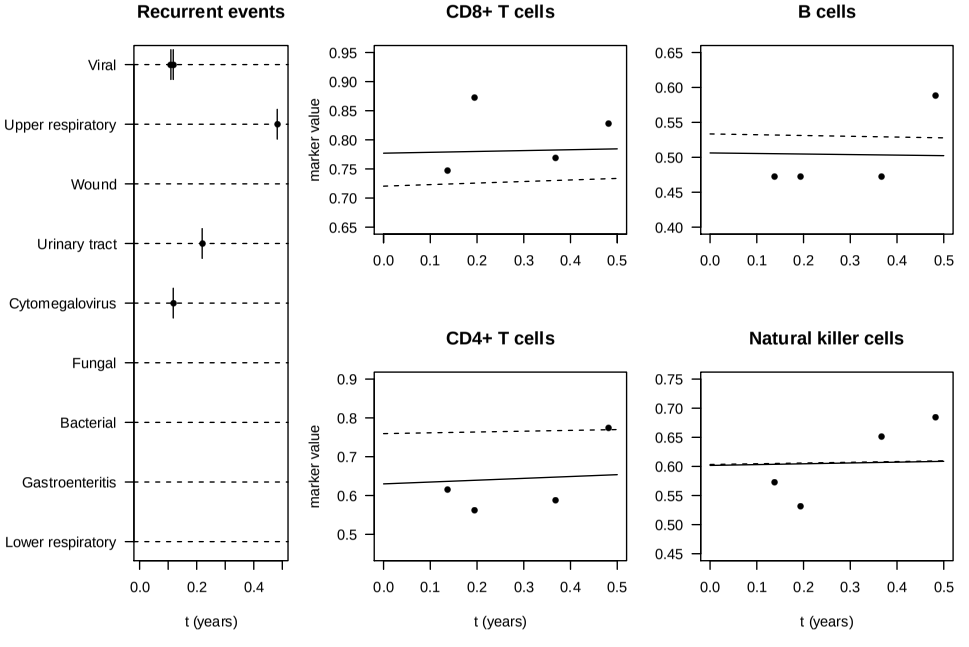

Figure 5.3: Observed history for the example patient until $t_{obs} = 0.5$. For all four markers, the dotted lines represent the average marker trajectories (i.e. random slope and intercept are zero). The solid lines are the estimated patient specific marker trajectories.

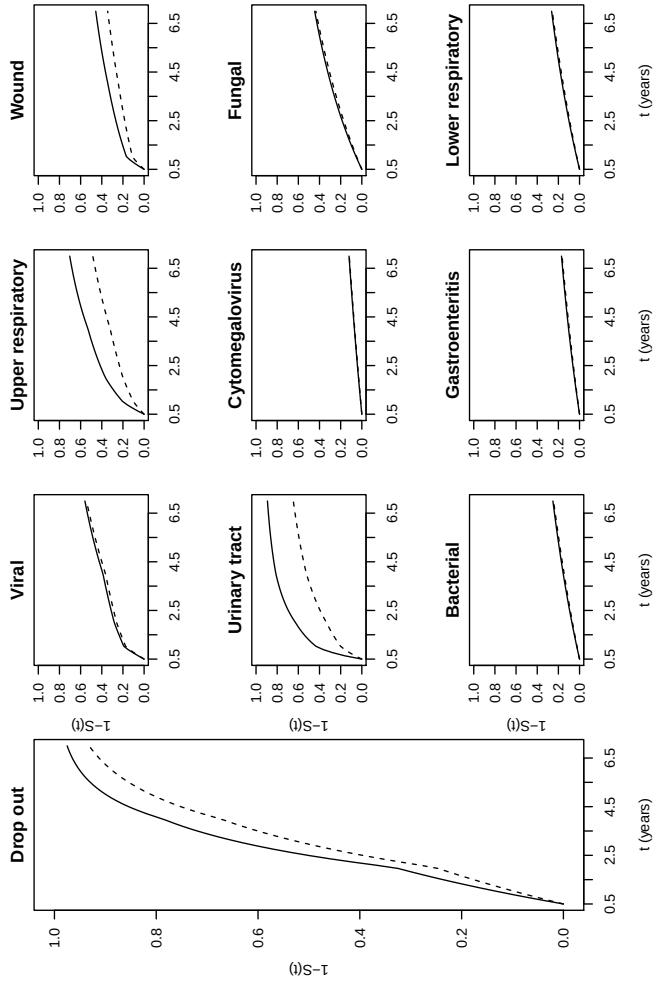

Figure 5.4: Predicted probabilities of experiencing a particular event before $T > t_{obs}$ based on the observed history from figure 5.3. The dotted lines represent the probability of experiencing a particular event given for a patient who did not suffer from any infections until t_{obs} and had average marker trajectories. The solid lines are the predictions for the patient with the observed history from figure 5.3.

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 a quasi-Monte Carlo (QMC) approach. Moreover, the QMC approach can also be used when we want to use the joint model for prediction. With simulations we have showed that our QMC approach gives valid parameter estimates of the recurrent and terminal event processes.

The simulation showed that, with a relatively small set of vectors to approximate the integral in the joint model likelihood, the standard errors of the estimated parameters in the marker submodels were too small. We performed additional analyses to investigate this problem by fitting the model to simulated data from just one marker (with random slope and intercept) and one recurrent event (with one frailty term). First we used an extremely large number of vectors to approximate the integral, i.e. $Z = 50000$ to obtain $\boldsymbol{\theta}_{QMC}$ and $Z = 500000$ to approximate the Hessian matrix. As expected, the standard errors of all parameters of the joint model were large enough to obtain good coverages of the 95% confidence interval. When we lowered Z, the standard errors of the parameters in the marker submodel were affected quickly. Especially decreasing the number of vectors used to approximate the Hessian matrix from (5.10) had a strong effect. The standard errors of the parameters from the event submodels were less affected by a smaller Z (data not shown). Therefore, when we are primarily interested in the event submodels and the effects of the markers on the hazards of experiencing particular events, we could use a relatively small Z.

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 whether it is possible to develop methods to test whether the integration error is small enough. Starting points might be the statistical test proposed by Hajivassiliou for Monte Carlo integration [132] and randomized QMC methods [135].

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 chapter [104]. For instance, we could assume that all dependencies between the markers and the recurrent event types are captured

by the covariances between $\boldsymbol{v}_i$ and $\boldsymbol{w}_i$ [102, 85]. Following this assumption, we would then parameterize the full covariance matrix $\boldsymbol{\Omega}$ from (5.6) and omit the true marker values from (5.2). The QMC approach can easily be adapted to these joint models, since the joint model structure from (5.5) remains the same.

For the kidney transplant data analysis, we made the assumption that the occurrences of infections *without* a frailty term (e.g. fungal infection or gastroenteritis) had no effect on the risk of experiencing another infection type or dropping out. When we suspect that the number of previous infections has an effect on these risks, we could include this observed number as a covariate in the event submodels, which allows the hazard of experiencing particular events to change depending on the infection history. In addition, we assumed that the frailties had no effect on the risk of experiencing an infection type without frailty term. When we suspect that the frailties have impact, we add the frailty terms to the hazard function of the infection without frailty term in a similar way as in the terminal event submodel from (5.3). However, these alternative parameterizations are outside the scope of this chapter and are interesting subjects for further investigation.

In addition, more research is needed to investigate whether the use of more sophisticated point sets could improve the accuracy of the QMC estimator for joint models. A problem of the deterministic point sets used in QMC integration is that points tend to cluster in higher dimensions [109]. To remove this characteristic, hybrid generators have been proposed which scramble deterministic quasi-Monte Carlo point sets using some random process [129]. Further research is needed to evaluate whether this improves the accuracy and efficiency of parameter estimation and at what computational cost.

In conclusion, we have showed that QMC integration could be useful when we jointly model multiple longitudinal markers, multiple recurrent events, and a terminal event.