I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2014/13

Semiparametric Bayesian frailty model for clustered interval-censored data

CETINYUREK, A. and P. LAMBERT

Semiparametric Bayesian frailty model for clustered interval-censored data

Aysun Çetinyürek Yavuz
Institut des Sciences Humaines et Sociales, Université de Liège
Philippe Lambert
Institut des Sciences Humaines et Sociales, Université de Liège
Institut de Statistique, Université catholique de Louvain

Abstract

The shared frailty model is one of the popular tool to analyze correlated right-censored time-to-event data. In the shared frailty model, the latent frailty is assumed to be shared by the members of a cluster and is assigned a parametric distribution, typically, a gamma distribution due to its conjugacy. However, in case of interval-censored time-to-event data, the inclusion of gamma frailties results in complicated intractable likelihoods, where the conjugacy property does not hold anymore. Here, we propose a semiparametric Bayesian frailty model for analyzing such data. We discuss three parametric specifications for frailty distribution in the analysis of interval-censored data. Afterwards we call particular attention to nonparametric specification of the frailty distribution. The results of the simulation study suggest that the proposed approach is robust to misspecification of the frailty distribution. Moreover, the performance of the proposed methodology is quite good in practical situations where the frailty distribution is multimodal or skewed. The approach is applied to dental data arising from the Signal Tandmobiel® Study.

keywords: flexible; interval-censored; shared frailty; proportional hazards; P-splines, misspecification

1 Introduction

Interval-censored time-to-event data arise frequently in multi-center clinical trials or longitudinal studies, especially in AIDS or cancer research where the patients are followed for a period of time and the event of interest occurs between the visits. Hence, the exact time of the event is not observed, while the times of the last negative test result, L, and of the first positive test result, R, are known. Then the exact time of the event is censored in the interval (L, R) if the subject experiences the event before the end of the study, otherwise it is right censored. This type of data is known as "interval-censored". The approaches proposed for analyzing right-censored data cannot directly be applied to interval-censored data, where the situation is much more complex. A number of parametric, semi-parametric and non-parametric approaches are proposed for analysing interval-censored data in the literature [1, 2, 3, 4, 5, 6, 7, 8, 9]. In these proposed methods, the observations or subjects are assumed to be independent. However, this may not be true in certain situations where the individuals are clustered or subject to multiple measurements. For example, in a dental study, the observations might be obtained from different teeth of the same subject where there may

exist within mouth correlation; in a multi-center trial the subjects from the same hospital could share common features and produce correlated responses.

A popular model accounting for the correlation among time-to-event data is the frailty model. In this approach, it is assumed that there exists a common and unobserved latent variable, named the frailty, that generates the relationship or the dependence of correlated survival times. This group specific parameter is postulated to have a distribution across the population, yielding the shared frailty model. For multivariate right-censored time-to-event data, Vaupel *et al.* [10] was the first to introduce the concept of frailty assuming a gamma distribution. Frailty model has thoroughly been studied since then by many authors to model the heterogeneity between the observations within a cluster or to account for the effect of unobserved covariates [11, 12, 13, 14, 15, 16]. Only a limited number of methods are available for analyzing clustered interval-censored data.

The Cox proportional hazards (PH) model is the most popular regression model to assess the effect of covariates on a time-to-event response [17]. In this model, the subjects or the observations are assumed to be independent. The unknown or unobservable risk factors that cause heterogeneity between the individuals can be accounted for by the introduction of a frailty component that acts multiplicatively on the hazard function. This modified model is known as the shared frailty proportional hazards model,

$$\lambda(t|\boldsymbol{x_{gj}}, \mathfrak{b}_g) = \lambda_0(t)\mathfrak{b}_g \exp(\boldsymbol{x'_{gj}}\boldsymbol{\beta}), \tag{1}$$

where $\lambda(t|\boldsymbol{x}, \mathfrak{b}_g)$ is the hazard function evaluated at the observed time-to-event T given the covariate vector $\boldsymbol{X}$ and the frailty of cluster g, $\mathfrak{b}_g$. Although, the primary interest is usually in the estimation of the regression coefficients in the shared proportional hazards frailty model, ignoring within cluster heterogeneity could result in erroneous conclusions [18]. The inclusion of the frailty term in the PH model provides estimates of standard errors adjusted for the possible effect of within cluster correlation but it complicates the estimation. Furthermore, a good estimate of the survival function might also be of specific interest (see e.g. [19, 20]) .

In the frailty context, a debatable issue is the choice of the frailty distribution since commonly assumed distributions may be too restrictive in practice to represent the correct within-cluster heterogeneity or misspecification of the frailty distribution may disturb the estimation of the regression coefficients as different distributional assumptions for the frailty can lead to considerably different association structures [21]. On the other hand, validating the distribution of the frailty is not trivial and a wrong choice might result in loss of efficiency [22]. Considering these aspects, a nonparametric flexible frailty distribution could be more attractive. This issue has been addressed and discussed in other context by some authors [23, 24, 25, 26, 27], but received very little attention in clustered interval censored data setting [28, 29, 30].

Bellamy *et al.* assumed a parametric Weibull model for the unobserved event times and a log-normal distribution for the frailty [31]. Their method handles right-, left-, and interval-censored data. The proposed Weibull frailty model provides consistent estimates of the population parameters of interest and fairly accurate estimates of variance associated with these estimates. Zuma and Lurie [32] presented two different methods to analyze correlated interval-censored data arising from an HIV study. The authors considered the failure times and the frailty terms as missing in a proportional hazards framework. In 2007, Zuma *et al.* [33] presented another approach by assuming a proportional hazards frailty model with Weibull baseline hazard and gamma frailty by treating the unobserved event times as unknown. Henschel *et al.* [30] proposed a Bayesian model by assuming gamma and lognormal distributed frailty. The authors applied a data augmentation approach by treating the baseline hazard as piecewise constant and reduced the interval-censored data to right-censored setting. Their proposed procedure provides reasonable estimates of the parameters for data sets of sufficient size. Goethals *et al.* [29] proposed a parametric shared gamma frailty model. They showed through a simulation study that their method performs well for estimating all parameters and provides good coverages for the parameter estimates contrary to imputation techniques they consider. Recently, Lam *et al.* [34] proposed a gamma frailty proportional hazards model for interval-censored data where a robust covariance matrix was used for possible misspecification of the

parametric baseline hazard function. The performance of the proposed strategy is reported to be highly satisfactory.

Here, we present a proportional hazards frailty model where the baseline density is estimated following [35, 36, 37] in a Bayesian framework. We model the log density as a linear combination of B-splines associated to a large number of knots and then put a roughness penalty to obtain a smooth density estimate. The composite link model (CLM; [38]) was successfully used by Lambert and Eilers [37] to estimate density from grouped (histogram) data. Then, it was extended by Çetinyürek and Lambert to obtain a smooth density estimate and the regression coefficient estimates from Cox PH for interval-censored data [9]. In this paper an extension of their methodology is proposed for clustered data setting. The rest of the paper is organized as follows. In Section 2, we provide an introduction to the shared proportional hazards frailty model. Afterwards, the baseline density estimation approach from interval-censored data and our strategy for obtaining a smooth flexible estimate for the frailty distribution are explained. The third section introduces the Bayesian counterpart of the proposed models with the description of priors, and the details of parametric and nonparametric modeling of random effects distribution. Section 4 contains the details and the results of an extensive simulation study. The methodology is illustrated with a dataset from Signal Tandmobiel® study in Section 5. Finally, we conclude the paper by a discussion and future research focuses.

2 The shared frailty model

2.1 Model formulation

Let X_{gj} denote the $p\times 1$ covariate vector and $\mathfrak{b}_g$ the latent frailty for unit j in cluster g. It is assumed that the conditional distribution of survival time T_{gj} is related to X_{gj} and $\mathfrak{b}_g$ through the shared proportional hazards frailty model which is an extension of the Cox proportional hazards model to handle correlated time-to-event data. Then the conditional survival function is

$$S(T \mid \boldsymbol{X}=\boldsymbol{x},\mathfrak{b}_g) = Pr(T>t|\boldsymbol{X}=\boldsymbol{x},\mathfrak{b}_g) = S_0(t)^{\mathfrak{b}_g \exp(\boldsymbol{x'\beta})}, \tag{2}$$

with a conditional hazard given by (5). Alternatively one can denote $z_g=\log(\mathfrak{b}_g)$ for the log-frailty and represent the shared proportional hazards frailty model as:

$$S(T|\boldsymbol{X}=\boldsymbol{x},z_g) = S_0(t)^{\exp(z_g+\boldsymbol{x'\beta})}.$$

If $\widetilde{f}_0$ denote a smooth estimate of the baseline density for T (when $\boldsymbol{x}=0,\mathfrak{b}=0$), then one can obtain estimates for the baseline survival and hazard functions using their espressions in terms of f_0:

$$S_0(t) = Pr(T>t|x=0) = 1-F_0(t) = 1-\int_0^t f_0(s)ds, \tag{3}$$

and

$$\lambda_0(t) = \frac{f_0(t)}{S_0(t)}. \tag{4}$$

Given a smooth estimate of the baseline density, the consecutive expressions for baseline survival and hazard, $\widetilde{S_0}(t;\phi,\zeta)$ and $\widetilde{\lambda_0}(t;\phi,\zeta)$, can be derived by substituting $\widetilde{f}_0$ for f_0 in (3) and (4). Finally, the likelihood for the shared frailty model for clustered interval censored data is proportional to

$$L = L(\boldsymbol{\phi},\boldsymbol{\zeta},\boldsymbol{\beta},\mathfrak{b}) = \prod_g\prod_j \widetilde{P}[L_{gj}<T_{gj}<R_{gj}|\boldsymbol{\mathfrak{b}_g},\boldsymbol{x_{gj}}] = \prod_g\prod_j [\widetilde{S}(L_{gj}|\mathfrak{b}_g,\boldsymbol{x_{gj}}) - \widetilde{S}(R_{gj}|\mathfrak{b}_g,\boldsymbol{x_{gj}})]. \tag{5}$$

after the substitution of the spline approximation to $S_0(t)$.

2.2 Density estimation from interval-censored data

To define this more formally, suppose that there are G independent clusters each of which containing n_g sub-units. Let t_{gj} denote the (continuous) time until the event of interest occurs for unit j in cluster g ($j = 1, \ldots, n_g; g = 1, \ldots, G$). We denote the probability density of T as f and the corresponding cumulative distribution function as F. The time, T_{gj} is not observed exactly, instead it is only known to lie within an interval (L_{gj}, R_{gj}) in the support (a, b). The density $f(T)$ will be estimated from interval-censored data $\{(L_{gj}, R_{gj}) : j = 1, \ldots, n_g; g = 1, \ldots, G\}$ on $(a, t_{cens}) \subset (a, b)$ where (a, b) denotes the support of T, by following the approach of Lambert and Eilers [37]. The support is partitioned into I small bins (> 100, say) $I_i = (a_{i-1}, a_i)$ of equal width Δ with midpoints $u_i = a_{i-1} + 0.5\Delta$ ($i = 1, 2, ...I$). The probability to observe T_{gj} in small bin I_i can be written as $\pi_i = \int_{I_i} f_T(t)dt \approx f(u_i)\Delta_i$. The relationship between the intervals and the small bins is provided by a GxIxn_g array $C = [c_{gji}]$ such that $c_{gji} = 1$ if $I_i \subset (L_{gj}, R_{gj})$ and 0 otherwise. Let $d_{gj} = 1$ if $R_{gj} > t_{cens}$ and 0 otherwise. Then, probability γ_{gj} to observe the event-time T_{gj} in (L_{gj}, R_{gj}) for unit j in cluster g is $\gamma_{gj} = \sum_{i=1}^{I} c_{gji} \times \pi_i + d_{gj} \times (1 - \zeta)$ where ζ denotes the probability that $R_{gj} < t_{cens}$.

Therefore, the full likelihood can be written as proportional to:

$$L = \prod_{g=1}^{G} \prod_{j=1}^{n_g} P[L_{gj} < T_{gj} < R_{gj}] = \prod_{g=1}^{G} \prod_{j=1}^{n_g} \gamma_{gj}.$$

We assume here that the π_i's change smoothly over time and model π_i using penalized B-splines following [35]. One can refer to Eilers and Marx [35] for a more detailed information on B-splines. Consider the B-spline basis $\{b_k(\cdot; q)\}_{k=1}^{K}$ of degree q associated to a rich grid (say 20) of equidistant knots and evaluated at the midpoints u_i of the small bins. Let ζ denote the probability to observe the event before t_{cens}. Then π_i is modeled as:

$$\pi_i = \zeta \times \frac{e^{\eta_i}}{e^{\eta_1} + e^{\eta_2} + ... + e^{\eta_I}}, \tag{6}$$

where $\eta_i = \sum_k \phi_k b_k(u_i)$ ensuring that $\sum_i \pi_i = \zeta$. An identifiability constraint is imposed on the spline coefficients, ϕ_k, such that $\sum_i \exp(\eta_i) = 1$ since $\pi_i(\phi) = \pi_i(\phi + c)$ for any constant c.

The model for $\boldsymbol{\gamma}$ can be seen as a composite link model. Originally proposed by Thompson and Baker [38], the composite link model was also used in Bayesian density estimation from grouped (histogram) data by Lambert and Eilers [37]. As we take a generous number of B-splines, it is important to counterbalance that large flexibility by applying a roughness penalty [35]. The roughness penalty is based on squared finite (r^{th} order) differences of the coefficients of adjacent B-splines: it penalizes changes in the r^{th} order differences of the spline coefficients. For example, the second-order ($r = 2$) difference penalty takes the form $\sum_k (\phi_k - 2\phi_{k-1} + \phi_{k-2})^2 = \boldsymbol{\phi' D' D \phi}$. The described density estimation procedure from interval-censored data was successfully used in Çetinyürek and Lambert and was extended to Cox proportional hazards model [9]. After having covered all the details of the baseline density estimation of time-to-event, the focus will move to the frailty part.

2.3 Choice of the frailty distribution

The frailty term, $\mathfrak{b}_g$, is typically assumed to be distributed across clusters according to a gamma distribution (with mean 1 and variance α^2). Other frailty distributions such as log-normal, uniform, Weibull are studied by [39]. The positive stable or inverse Gaussian distributions are also studied for modeling the frailty [40, 44]. The reason for the popularity of a gamma frailty is not of biological but of mathematical relevance as it yields closed forms for the marginal survival and hazard functions. The simple and explicit available form for the Laplace transform of the gamma allows one to use traditional maximum likelihood procedures. However, the conjugacy property does not hold anymore for interval-censored data where the estimation turns out to be more difficult. Most of the software

limits the choice of the frailty distribution to log-normal and gamma distributions [42], see for example *coxph* in package *survival* in R [43].

In frailty models, it is necessary to make sure that the model is identifiable, for example by constraining the mean of the gamma frailty distribution, $E(\mathfrak{b}_g)$, to be 1. The variance of the frailty, $V(\mathfrak{b}_g) = \alpha^2$, that quantifies between-cluster variability must be estimated. In contrast, in a log-normal frailty model, one usually assumes that the log frailty, $z_g = \log(\mathfrak{b}_g)$, has mean 0 and an unknown variance α^2. Log-normal frailty model in fact originates from the link to generalized linear mixed model framework.

There is an ongoing discussion on the selection of the frailty distribution in the literature. Different frailty distributions lead to different correlation structures, for example with a stronger association between larger (resp. smaller) event times with a gamma (resp. positive stable) frailty [45]. It suggests that the frailty distribution should be selected with care as it might imply unrealistic assumptions on the modeled data. It has been shown that the regression parameter estimates are robust to the choice of that random effect distribution in specific contexts [46], but are sensitive to that choice in other frameworks such as in GLMM [22, 47]. Therefore, we will continue with describing a smooth nonparametric density for the frailty term. Still, parametric choices for the frailty distribution remain useful in practice, such as when the number of clusters is small with limited prior information on their dependence structure, or when the interest lies in marginal quantities (such as the survival distribution or hazard function) estimated using robust estimation procedures.

2.4 Flexible frailty density estimation

It is crucial to better characterize the frailty distribution to account for the heterogeneity present in the data since the correlation structure depends on the chosen frailty distribution. As an alternative to parametric specifications, one can choose to work with flexible forms and to use the data to select the association structure. In this spirit, we investigate the use of B-splines to specify the frailty density along the same lines as the density estimation method in Section 2.2.

Recall that $\mathfrak{b}_g$ represents the latent frailty. However, for the sake of computational simplicity, we will use the log-frailty, $z_g = \log(\mathfrak{b}_g)$, and shall refer to it as *log-frailty* from now on. Let α denote the standard deviation and $h(z)$ denote the density of the log-frailty. We start with expressing the log-frailty using a location-scale transformation as

$$z_g^* = \frac{z_g - \nu}{\alpha}$$

such that $E(z^*) = 0$ and $V(z^*) = 1$. By applying this parametrization, one can simply assume a standard range, say $\mathfrak{I} = (-6, 6)$, for the log-frailty. The support of the log-frailty is partitioned into $I = 120$ (say) small bins $I_i = (a_{i-1}, a_i)$ of equal width Δ=0.1 with midpoints $u_i = a_{i-1} + 0.5\Delta$ ($i = 1, 2, ...120$). The probability to observe z_g^* can be written as $\pi^* = \int h^*(z^*)dz^{(*)}$ where π^* is modeled as combination of B-splines as follows:

$$\pi_i^* = \frac{\exp(\eta_i^*)}{\exp(\eta_1^*) + \exp(\eta_2^*) + ... + \exp(\eta_I^*)} \tag{7}$$

where $\eta_i^* = (\sum_\kappa \phi_\kappa^* b_\kappa(u_\mathfrak{i}^*))$ and $\{b_\kappa(\cdot; q)\}_{\kappa=1}^{\kappa^*}$ is the B-spline basis associated to log-frailty with the corresponding spline coefficients ϕ^*. An identifiability constraint is imposed on the spline coefficients, ϕ_κ^*, such that $\sum_\mathfrak{i} \exp(\eta_\mathfrak{i}^*) = 1$ since $\pi_\mathfrak{i}^*(\phi^*) = \pi_\mathfrak{i}^*(\phi^* + c)$ for any constant c.

The density estimation for the log-frailty has some discrepancies from the provided density estimation procedure for time-to-event data,see Section 2.2. Firstly, the log-frailty terms are not subject to censoring, hence the multiplier ζ in (6) equals 1. Initial values for parameters ϕ_k^* can be obtained by selecting them to mimic the shape of a parametric frailty distribution fitted beforehand. Thanks to location-scale transformation, one can use equidistant knots on $\mathfrak{I} = (-6, 6)$ where most of the probability mass lies. This allows one to estimate the standard deviation of the log-frailty. For

identifiability reasons, the mean of z, ν, is fixed to zero. Moreover, z^* is also constrained to have mean 0 and variance 1 using a strong a penalty of the form $\psi \times [(E(z^*)-0)^2 + (var(z^*)-1)^2]$ where ψ is a large penalty (see [48]).

3 Bayesian model

The likelihood based on shared frailty model from interval-censored data is quite complicated to be solved using frequentist methods, thus we adopt Bayesian formulation for the proposed models and explore the joint posterior using a Markov chain Monte Carlo algorithm. In this section, we present the prior distributions for the baseline density estimation. Afterwards, Bayesian model specifications will be presented in two parts: the parametric frailty model formulation and the flexible frailty.

We start by introducing the priors for the model parameters related to the baseline density in the proportional hazards shared frailty model. The smoothness of the models is ensured by penalizing changes in r^{th} order differences of the spline coefficients. In a Bayesian setting, the frequentist roughness penalty is translated into a prior distribution on the finite (r^{th}) order differences of the spline coefficients:

$$(\Delta^r \phi|\tau) \sim N(0, \tau^{-1}).$$

As a result, the joint prior for the B-spline coefficients corresponds to a multivariate normal distribution as:

$$p(\boldsymbol{\phi}|\tau) \propto \tau^{K/2} \exp\left\{-\frac{\tau}{2}\boldsymbol{\phi'P\phi}\right\},$$

with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{P}^{-1}$ where we suggest to take $\boldsymbol{P} = \boldsymbol{D'D} + \epsilon I$. The inverse variance τ plays the role of the penalty parameter in the penalized likelihood of the frequentist setting. A noninformative hyperprior with large variance is usually advocated for τ, e.g. a gamma distribution $\mathcal{G}(\upsilon_1{=}1,\ \upsilon_2{=}0.0001)$ with mean υ_1/υ_2 and variance υ_1/υ_2^2 [49]. Alternative priors are suggested in [50, 51]. This idea was successfully used in many contexts (see e.g. [36, 37, 49, 52]). An improper prior is considered for the regression parameters β. Moreover, a uniform prior on (0,1) is taken for $\zeta = P(T \leq t_{cens}|X = 0)$. Simply, we choose to work with $\xi = \log[\zeta/(1-\zeta)]$. The consequent prior for ξ is thus proportional to

$$\zeta(1-\zeta) = \frac{\exp(\xi)}{(1+\exp(\xi))^2}.$$

The Bayesian model treats the log-frailty terms as unknown parameters. In the next section, we will focus on specification of priors for the log-frailty part.

3.1 Parametric frailty specification

Given the proportional hazards shared frailty model and the conditional independence assumption, the joint posterior of the model parameters and latent log-frailty terms conditional on the observed data could be written as

$$p(\boldsymbol{\phi}, \tau, \xi, \boldsymbol{\beta}, z, \alpha \mid \mathfrak{D}) \propto L \times p(\phi|\tau) \times p(\tau) \times p(\xi) \times p(\beta) \times \left\{\prod_{g=1}^{G} p(z_g|\alpha)\right\} \times p(\alpha) \tag{8}$$

where L is given by (5) and $\mathfrak{D}$ generically denotes the available data. The joint posterior will be explored using MCMC (see Section 3.3). In the Bayesian counterpart of the model, the interest is in the estimation of both fixed and random effects. Thus, we also need to specify the priors related to the frailty part. The log-frailty terms could be assumed, for example, to be Gaussian:

$$z|\alpha \sim N(0, \alpha^2).$$

Typically, the variance of the log-frailty, α^2, is assigned a large variance inverse Gamma prior.

3.2 Flexible frailty specification

The joint posterior of the model parameters for the proportional hazards shared frailty model with a flexible frailty distribution specification can be written as

$$p\left(\boldsymbol{\phi}, \tau, \xi, \boldsymbol{\beta}, z^*, \boldsymbol{\phi^*}, \tau^*, \alpha \mid \mathfrak{D}\right) \propto L \times p(\phi|\tau) \times p(\tau) \times p(\xi) \times p(\beta) \times \left\{ \prod_{g=1}^{G} p(z^*|\alpha) \right\} \times p(\alpha) \times p(\phi^*|\tau^*) \times p(\tau^*) \tag{9}$$

where L is a function of $\gamma_{gj} = \sum_{u_i <= t_{cens}} c_{gji} \times \pi_i + d_{gj} \times (1-\zeta)$ given by (5). The joint posterior will be explored using MCMC (see Section 3.3).

Consider now the priors related to the flexible frailty distribution. The frequentist roughness penalty is translated into a prior distribution on the finite (r^{th}) order differences of the frailty spline coefficients in Bayesian setting as follows:

$$(\Delta^r \phi^*|\tau^*) \sim N(0, \tau^{*-1}).$$

Consequently, the joint prior for the B-splines coefficients from the approximation of the frailty density corresponds to a multivariate normal distribution as:

$$p(\boldsymbol{\phi^*}|\tau^*) \propto (\tau^*)^{\kappa/2} \exp\left\{ -\frac{\tau^*}{2} \boldsymbol{\phi^{*\prime} P^* \phi^*} \right\},$$

with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{P^{*-1}}$ where $\boldsymbol{P^* = D^{*\prime} D^*} + \epsilon I^{(z^*)}$ is a full-rank matrix for some small quantity ϵ (say 10^{-6}). For τ^*, a noninformative hyperprior with large variance is assumed [49].

The (log-)frailty terms are assumed to have a nonparametric density, $h^*(z^*)$, which is constructed via (7) where $h^*(z^*)$ is a zero mean and unit variance density. The standard deviation of the frailty, α, is assigned a noninformative inverse Gamma prior. We set the multiplier in the constraint, ψ, to a large number (say 1000) to force the constraint allow the chains to converge.

3.3 Exploring the model via MCMC

For both frailty specifications, the parameter ζ is restricted to have a value on [0,1] and α is defined to be positive, therefore the Metropolis steps for updating those parameters are performed after logit and log transformations, respectively.

3.3.1 Parametric frailty model

Let $\boldsymbol{\vartheta} = (\boldsymbol{\phi}, \tau, \xi, \boldsymbol{\beta}, z, \alpha)$ be the vector of parameters of length H where $H = K + 1 + 1 + p + G + 1$ and p denotes the number of regression parameters. The samples $\left\{\boldsymbol{\vartheta_{(m)}} : m = 1, ..., M\right\}$ will be drawn from the joint posterior via Markov chain Monte Carlo (MCMC) methods. In our Bayesian model, only the conditional posterior distribution for τ is of a familiar type: $\tau|\mathfrak{D} \sim \mathcal{G}\left(K + \upsilon_1, \left(\boldsymbol{\phi}'_{(m)} P \boldsymbol{\phi}_{(m)}\right)/2 + \upsilon_2\right)$. Hence, we shall follow a Metropolis within Gibbs strategy to sample from the joint posterior $p(\boldsymbol{\vartheta}|\mathfrak{D})$ presented in (8) (with Metropolis steps for all parameters, except τ).

3.3.2 Flexible frailty model

Let $\boldsymbol{\vartheta} = (\boldsymbol{\phi}, \tau, \xi, \boldsymbol{\beta}, z^*, \phi^*, \tau^*, \alpha)$ be the vector of parameters of length H where $H = K+1+1+p+G+\kappa+1+1$ and p denotes the number of regression parameters. The samples $\left\{\boldsymbol{\vartheta_{(m)}} : m = 1, ..., M\right\}$ will be drawn from the joint posterior via MCMC methods. The conditional posterior distributions for τ and τ^* are of familiar type: $\tau|\mathfrak{D} \sim \mathcal{G}\left(K + \upsilon_1, \left(\boldsymbol{\phi'}_{(m)} P \boldsymbol{\phi}_{(m)}\right)/2 + \upsilon_2\right)$, $\tau^*|\mathfrak{D} \sim$

$\mathcal{G}\left(\kappa+\upsilon_1, \left(\boldsymbol{\phi}^*_{(m)}{}' P^* \boldsymbol{\phi}^*_{(m)}\right)/2+\upsilon_2\right)$. Thus, we shall follow a Metropolis within Gibbs strategy to sample from the joint posterior $p(\boldsymbol{\vartheta}|\mathfrak{D})$ presented in (9) (with Metropolis steps for all parameters, except τ).

3.3.3 Frequentist density estimation

For Markov chain Monte Carlo, we need good initial values so that the mixing and the convergence of the chains is faster. In the literature, starting values are often obtained from the frequentist models. In this spirit, we shall explain a frequentist density estimation procedure from time-to-event data when the covariates and the possible heterogeneity are ignored ($\boldsymbol{\beta}=0$, $\mathbf{b_g}=1$ (or $\boldsymbol{z}=0$)). As described in Section 2.2, we start by partitioning the support of T into small bins (more than 100 small bins of equal width) for obtaining an accurate estimate of the density for time-to-event data. Then, following an approach similar to Eilers and Marx [35] we calculate the number of observations in each small bin, namely the pseudo-counts. These pseudo-counts, which are calculated from the C matrix of the composite link model defined in Section 2.2, are later used to build the density estimate. In this spirit, for cluster g, each element of a row in the C matrix is divided by the sum of the elements in that row ($C_{gj.}=\sum_i C_{gji}$). The so-obtained numbers, $W_{gji}=C_{gji}/C_{gj.}$, provide the contribution of the concerned observation in cluster g (e.g. a patient for a multicenter clinical trial) for each small bin partitioning (a, t_{cens}). Then, the contributions of each observation for the i^{th} small bin, I_i, are summed over all observations ($W_{..i}=\sum_g\sum_j W_{gji}$) and rounded to the nearest integer value y_i in order to get the pseudo-count for that small bin. Note that π_i denotes the probability to have an event time in I_i, then the likelihood for these pseudo-counts is proportional to $\prod_i^I \pi_i^{y_i}$. Alternatively, the well known link between the Poisson and the multinomial distributions suggests to assume that the pseudo-counts, y_i, have a Poisson distribution with mean $\mu_i=\pi_i y_+$ conditional on the total number of observations $y_+=\sum_i^I y_i$. Using a rich B-spline basis (over the time axis, see Section 2.2) as regressors in a log-linear model for the mean, one obtains the likelihood

$$\mathcal{L}(\boldsymbol{\phi}|y)=\sum_{i=1}^{N} y_i \log(\mu_i)-\sum_{i=1}^{N}\mu_i,$$

where $\log(\mu_i)=\eta_i=\sum_{k=1}^{K}\phi_k b_k(u_i)$. Then, by subtracting the 2^{nd} order penalty (say) and a small ridge penalty from $\mathcal{L}(\boldsymbol{\phi}|y)$, one obtains the penalized log likelihood function

$$\mathcal{L}_p(\boldsymbol{\phi}|y,\tau)=\mathcal{L}(\boldsymbol{\phi}|y)-\frac{\tau}{2}\boldsymbol{\phi'P\phi},$$

where $\boldsymbol{\phi'D'D\phi}=\sum_k\left(\phi_k-2\phi_{k-1}+\phi_{k-2}\right)^2$ and $\boldsymbol{P}=\boldsymbol{D'D}+\epsilon I$. The function $\mathcal{L}_p$ can be optimized by solving the score equations $B^T(y-\mu)=\tau\boldsymbol{P\phi}$, using iteratively reweighted least squares (IRWLS): iteratively solve (for ϕ)

$$(B^T\tilde{W}B+\tau P)\phi=B^T\tilde{W}(y-\tilde{\mu})+B^T\tilde{W}B\tilde{\phi},$$

where $\widetilde{W}$ is a diagonal matrix with elements $\mu_i(\widetilde{\phi})$ and $\tilde{\phi}$ and $\tilde{\mu}$ are current approximations to the solution. The variance-covariance matrix for the estimated spline coefficients $\boldsymbol{\phi}$ is given by (at convergence),

$$\Sigma_0=(B^TWB+\tau\boldsymbol{P})^{-1}. \tag{10}$$

More detailed information can be found in [35]. Information criteria such as AIC or BIC could be used for choosing the initial optimal (plausible) value of the penalty parameter τ. In our experience, BIC is preferable to AIC as AIC tends to undersmooth the target curve, which was also mentioned by other authors [53].

For obtaining an initial estimate for the frailty density, we also applied the same frequentist procedure as described above. However, as frailty is not observed directly, initial values for the

latent frailty terms should be obtained. It was made in two steps. We first fit a parametric (gamma or log-normal) frailty model using midpoint imputation for the interval-censored data. Then, using the estimated frailty terms as if they were observed, we apply the frequentist density estimation approach to those estimated values to obtain reasonable starting values for ϕ^* and τ^*.

3.3.4 Initialization

Although it is possible to start the chains from arbitrary values, starting the chains at good initial values fasten the convergence. Thus, we initialize the chains at the starting point $\boldsymbol{\vartheta_{(0)}}$. For the parametric frailty specification, the initial state of the chain $\boldsymbol{\vartheta_{(0)}} = \left(\phi_{(0)}, \tau_{(0)}, \xi_{(0)}, \beta_{(0)}, z_{(0)}, \alpha_{(0)}\right)^T$ are chosen as follows:

- A value for $\boldsymbol{\phi}_{(0)}$ can be obtained using the frequentist procedure described in Section 3.3.3.
- We define $\tau_{(0)}$ as the value of τ yielding the smallest BIC for different values taken in a grid.
- $\zeta_{(0)}$ is taken as the proportion of pseudo-counts corresponding to small bins located below t_{cens}.
- In accordance with the estimation of the spline coefficients, we start by ignoring possible covariate effects: $\boldsymbol{\beta}_{(0)} = 0$
- We obtain $z_{(0)}$ from a gamma frailty model applied on data resulting from mid-point imputation.
- The estimated variance from the gamma (parametric) frailty model is used for $\alpha_{(0)}$.

For the nonparametric specification of the frailty, the initial values of the chain $\boldsymbol{\vartheta}$ are chosen along the same lines as described for the parametric frailty model, except for $(\phi^*_{(0)}, \tau^*_{(0)})$ where we use the following:

- A value for $\boldsymbol{\phi}^*_{(0)}$ can be obtained from the frequentist procedure described in Section 3.3.3 after obtaining an approximation for the density of $z^*_{(0)}$.
- Similarly, $\tau^*_{(0)}$ is defined as the value of τ evaluated on a grid yielding the smallest BIC.

3.3.5 MCMC steps

Starting from $\boldsymbol{\vartheta^{(0)}} = \left(\vartheta_1^{(0)}, \ldots \vartheta_{H-1}^{(0)}\right)^T$, the initial state of the chain, the update of the h^{th} component ($h \neq K+1$) is done by drawing from a proposal distribution and then either accepting or rejecting the proposed state with some probability. By running the chains for M iterations (long enough to achieve convergence) and ignoring the first few thousand iterations (say n_b) as an appropriate burn-in period, chains of length $(M - n_b)$ are obtained and thinned (by a factor 50 for the spline coefficients due to large auto-correlations). Finally, the resulting $(M - n_b)/50$ iterations could be seen as a random sample from the joint posterior. Point estimates and corresponding credible regions can be calculated based on these random samples.

The mixing of the chain could be improved by using a Metropolis algorithm on a re-parametrized posterior [52]. In this sense, one can use an approximation to the 2^{nd} order dependence structure of the conditional posterior. The variance covariance matrix, $\boldsymbol{\Sigma_0}$, of the penalized maximum likelihood estimator of the spline parameters $\boldsymbol{\phi}$ could be calculated using (10) for a fixed and reasonably chosen value of the roughness penalty parameter τ. Then, the posterior can be re-parametrized using φ with $\phi = \phi_0 + L\varphi$ where L denotes the lower triangular matrix obtained from the Cholesky decomposition of $\boldsymbol{\Sigma_0}$. Then, the univariate Metropolis algorithm described before can be employed on the re-parametrized posterior. This also fastens the convergence.

3.3.6 Automatic tuning of the algorithm

For optimal convergence of Metropolis algorithm, the asymptotic acceptance probability should be tuned to be approximately 0.44 in one dimensional space decreasing to 0.23 in high dimensional spaces [54, 55]. Good acceptance rates can be achieved via a careful choice of the standard deviation δ_h in the generation of proposals in the preceding univariate Metropolis algorithm [56]. Let δ denote the tuning parameter of interest. The value of δ at iteration $m+1$ can be adjusted using the value at iteration m using (with $\overline{\eta} = 0.44$)

$$\sqrt{\delta_{m+1}} = h\left(\sqrt{\delta_m} + \gamma_m \left(\alpha(\boldsymbol{\theta_{(h)}}, \boldsymbol{\theta_{(h-1)}}) - \overline{\eta}\right)\right)$$

$$h(x) = \begin{cases} \epsilon & \text{if } x < \epsilon \\ \text{x} & \text{if } x \in (\epsilon, A) \\ \text{A} & \text{if } x > A \end{cases}$$

where ϵ is a very small number (say 0.0001) and A a large one (say 10000). If the targeted acceptance level is not achieved, these constants should be changed. The series $\{\gamma_m\}$ is a non-increasing sequence of positive real numbers such that $|\gamma_m - \gamma_{m-1}| \leq m^{-1}$. Possible choices for γ_m are $\frac{10}{m}$ or $\frac{1}{m}$. Practically, the MCMC algorithm is run for a few hundred iterations with the δ_m's automatically updated to achieve the targeted acceptance rate. Then, the last value of δ_m in the so-generated chain can be used in a non-adaptive version of the modified Metropolis algorithm to produce the long chain(s) that will be used for inference.

4 Simulation study

In order to assess the performances of our modeling strategies, we have performed a series of simulations with various number of clusters and cluster sizes under varied interval-censoring schemes. Besides aiming to estimate the regression coefficients and the standard deviation of the frailty accurately, we also aim to obtain good estimates of the survival and frailty density functions. Hence we also investigate the capacity of the proposed method to identify the form of the baseline density and frailty density functions. In line with these objectives, we considered three scenarios for the frailty density (see dashed lines for the assumed frailty distributions in Figure 1). The first scenario (column 1 in Fig. 1), referred as "Gaussian", is characterized by a Weibull baseline hazard and a Gaussian frailty density. In the second (column 2) and third (column 3) scenarios, we assume more challenging mixture models for the frailty distribution. The "Bimodal" scenario refers to the case where the frailty density is a balanced mixture of two Normal distributions, whereas the "Skewed" scenario considers the case with an unbalanced mixture of two Normal distributions.

In the simulation study, we aimed to mimic two different types of settings: the first one is a multicenter study with a small number of centers (clusters) involving each a large number of subjects; the second one is a study with a large number of subjects (clusters) providing each a small number of repeated observations. The results are based on 300 replications with four different values for the number of clusters G (i.e. 20, 50, 100 and 200), each of which has different sizes n_g (i.e. 4, 6, 10, 20 and 50). It should be noted here that large (resp. small) values of G are combined with small (resp. large) ones for n_g. Three different values (0.8, 1.2, and 1.5) were considered for the standard deviation of the frailty. A balanced binary treatment covariate ($X_1 = 0$ for control subject; $X_1 = 1$ for a subject in the intervention group) is used. After generating the frailty terms from the specified distribution, given the value of the frailty and the covariate, we assumed a proportional hazards frailty model with a Weibull baseline hazard, $\lambda_0(t) = 5t^4/70^5$. The corresponding regression coefficient is chosen to be $\log(2)(\cong 0.6931)$ such that given the same frailty, the hazard of a control subject is twice that of the treated one.

For interval-censored data, the performance of the methodology are expected to depend on the width of intervals and the amount of right-censoring [9, 57]. Thus, 6 different scenarios corresponding

to different interval widths (0.5σ, 1.0σ and 1.3σ) and two different amounts of right-censoring (10% and 35%) were used in the simulation.

Our data generation and simulation strategy contain the following steps:

1. Firstly, we generate the log-frailty terms z from one of the specified frailty distributions.
2. Then the values of the covariate are generated.
3. Afterwards, given the values of frailty terms and the covariate, the observations t_{gj} ($g = 1, ..., G; j = 1, ..., n_g$) are generated using the selected proportional hazards frailty model.
4. Each observation t_{gj} is converted into an interval of width w_{gj}, where w_{gj} is generated from a Gamma distribution with a mean equal to the targeted mean width (0.5σ, 1.0σ and 1.3σ) and a variance equal to one fifth of the mean. The interval corresponding to t_{gj} was finally defined as (L_{gj}, R_{gj})=$(t_{gj} - u_{gj} \cdot w_{gj}, t_{gj} + u_{gj} \cdot w_{gj})$ where u_{gj} is randomly generated from a uniform distribution on (0,1).
5. For each simulated dataset, initial values for the spline parameters were obtained using the strategy described in Section 3.3.3.
6. We sample the posterior for the parameters of interest using MCMC (see Section 3.3).
7. Steps 1-6 were repeated for all data sets ($S = 300$ times) to obtain the Monte Carlo estimates for the quantities of interest.

We considered the compact interval (0,120) as (an approximation to) the support of the target Weibull distribution. The observed range of the considered distribution, (0, t_{cens}), changing for different amounts of right-censoring, was divided into small bins of width 1 ($\cong 0.7\sigma$). Cubic B-splines associated to 12 equidistant knots on (0, t_{cens}) and a third order penalty were used. A chain of length M=120000 (including a burn-in period of n_b=54000 runs) was constructed to explore the posterior distribution of the model parameters. The posterior of the spline parameters was re-parametrized using Σ_0 (see Section 3.3). Although the posterior samples from Bayesian estimation using MCMC contain lots of information, we need to use point estimators to summarize the samples and to enable frequentist comparisons. For this reason we use the mean/median of several quantities of interest. Point estimates for β, ζ, τ and α were calculated using the mean of the generated MCMC sample; $(1 - \alpha) \times 100\%$ credible interval can be estimated using the $\alpha/2$ and $(1 - \alpha/2)$ sample quantiles of the chain. The proportions of so-defined credible intervals (one for each of the S datasets) containing the true value of the parameter of interest were reported as an estimate of the corresponding coverage. The fitted baseline density $\widehat{f}_s$ for the s^{th} data set corresponds to the MCMC estimate $\frac{1}{M-n_b}\sum_{m=n_b}^{M}\widetilde{f}^{(m)}$ of the posterior mean of the estimated baseline density. The baseline survival is obtained using the estimated baseline density as in (3). These quantities can be used to derive a point estimate for the mean, standard deviation and some selected quantiles. Further we report on 5%, 15%, 25%, 35%, 50%, 60%, 75%, 85% and 88% quantiles of the baseline survival function. It should however be noted that for some amounts of right censoring, we cannot get the estimates for baseline survival functions at some quantiles since the density is not observed beyond t_{cens}. The frequentist properties of these Bayesian estimators are measured in terms of relative bias (Rbias), empirical standard errors (ESE) and root mean squared error (RMSE). However only relative biases are reported here. If β (say) is the parameter of interest, then the relative bias is defined as $100\left(\frac{\bar{\widetilde{\beta}}-\beta}{\beta}\right)$ where $\bar{\widetilde{\beta}}$ is the mean of the estimates for β over the S data sets.

For each scenario the performances of three different frailty models labeled: "Imputation", "Gaussian" and "Semiparametric" are compared. "Imputation" replaces the interval-censored data by their midpoints and fits a Cox PH model with a log-normal or gamma frailty. "Gaussian" assumes a flexible baseline survival function for interval-censored data and a lognormal frailty density. "Semiparametric" estimates both the baseline survival and the frailty density using Bayesian penalized B-splines.

Only part of the results of the simulation study are presented in the Tables 2-8 for the sake of brevity. In Table 2, the results for regression coefficient and standard deviation for frailty are

summarized. When the generated frailty is unimodal, the regression coefficients β are properly estimated (Rbias $< 4\%$) in both of the flexible and Gaussian frailty models with similar point estimates and 90% credible intervals (except when n_{cl}=20 with 10 observations per cluster). The increase in sample size results in smaller credible intervals. Moreover, the coverages of 90% and 80% credible intervals are very close to nominal values. However the performance of both frailty models deteriorates in the estimation of the standard deviation of the frailty, α, for small cluster sizes (n_i=4 and 6) or when the number of clusters is small (n_{cl}=20). In all cases for both models, the frailty standard deviation is underestimated with low to medium relative bias (between -2% and -11%). In Table 3, a similar setting to Table 2 is presented with a higher standard deviation for frailty (α=1.2). The performance of both models are very similar to those of Table 2 except that the estimation of the standard deviation of the frailty improves (Rbias $< 5\%$).

Tables 4 presents the results for the "Bimodal" setting. One can notice that the relative bias in the estimation of the regression coefficients are very small ($< 3\%$) except again when n_{cl}=20 and n_i=10 ($\approx 7\%$). The coverages of 90% and 80% credible intervals are very close to the nominal values. Here we see again that the flexible frailty model and Gaussian frailty model produce very similar regression coefficient estimates. When we consider the estimation of frailty standard deviation, the performances of the flexible and the Gaussian frailty specifications are similar except when the number of observations per clusters is very small: with 4 observations per cluster the estimation of α produces large relative bias estimates. A similar scenario with a larger mean interval width (1.0σ) was also considered (not presented here): the larger interval width did not affect the results. However when we considered a similar scenario with mean interval width (1.0σ=15) and a larger frailty standard deviation α=1.2 (see Table 5), the estimation gets better. All relative bias estimates are below 5% and the coverages are similar for the flexible and the Gaussian frailty specifications. This progress in the estimation of α is probably due to the separability of the two modes of the frailty distribution when α is larger. It is also noticeable that the estimation of α is more performant with a smaller bias (less than 5%). The coverages of 90% and 80% intervals tend to be larger than the nominal values. It is worthwhile to see that the performances of the two frailty models are similar although the flexible frailty model has more parameters to estimate. The performance of the models in a similar setting to Table 5 but with a larger frailty standard deviation (α=1.5) was also explored (but not presented here). The relative bias is less than 5% for β and less than 6% for α. The coverages for α in the semiparametric model is larger than for the Gaussian model. In the bimodal scenario, the value of the frailty standard deviation is an important factor for the separation of two modes of the mixture distribution.

The results from more challenging settings, ("Skewed" setting), are presented in Table 6. The performances in the estimation of β are quite similar for the flexible and Gaussian frailty models with relative biases smaller than 4% in all cases. The relative biases are also very small for α (less than 5%). The coverages of the credible intervals tend to be larger than the nominal values in the flexible frailty specification for small number of clusters. As sample size increases, the width of credible intervals decreases and coverages gets closer to their nominal values. A similar setting with a greater α=1.5 had a very similar performance (not presented here).

Table 7 results from a setting similar to Table 5 but with 35% right-censoring. The relative bias and coverages are very similar to that of Table 5. However a larger percentage of right-censoring results in larger credible intervals. Another simulation setting similar to Table 6 was also run to see the effect of larger amount of right-censoring (35%): the obtained results (not shown) are consistent with these in Table 6.

The performances of midpoint imputation using a gamma or a lognormal frailty were also compared for all simulated datasets. Except when the widths of the intervals are small (0.5σ), midpoint imputation should be used with caution. Indeed, for a mean interval width 1.0σ, midpoint imputation resulted in relative bias around 5-10% for the regression coefficient, β, and the frailty standard deviation, α. When the mean interval width gets larger, the relative biases increase for both model parameters. Assuming a gamma or a lognormal does not make any difference in the estimation of the regression coefficient, while assuming a lognormal frailty seems preferable when it comes to estimate

the standard deviation of the frailty. The results are not presented here for the sake of brevity.

To sum up all results so far, we can conclude that the flexible frailty and the Gaussian frailty models perform similarly in the estimation of β. Their performance in the estimation of α depends on the different factors (sample size, number of clusters and frailty variance). However, the deteriorations in the estimation of α does not impact the estimation of β. Although the model with flexible frailty has more parameters to estimate, the results are similar for many settings.

4.1 Curve fitting performance

Aside from the regression coefficients, it is also of interest to estimate the underlying frailty density. The choice of the frailty density is reported to be a sensitive and important step in the literature, as it is argued that different distributional assumptions can lead to considerably different association structures [21]. Therefore, it is one of our main aims to investigate the ability of the proposed flexible frailty model to estimate the shape of the frailty density.

The quality of curve fits are explored for different number of clusters, cluster sizes, interval widths and amount of right-censoring. For each frailty scenario, we present here only 5 different settings for the sake of brevity, see Figure 1. From each simulation setting, we simulate 50 datasets and estimate the frailty density. The first row in the figure corresponds to a setting with 20 clusters and 10 observations per cluster (n=200). The third row comes from a setting with n=200 and 50 clusters. The second, fourth and fifth rows correspond to settings with a sample size of 1000 with increasing number of clusters (n_{cl}=20 and n_g=50; n_{cl}=50 and n_g=20; n_{cl}=100 and n_g=10, respectively.

Under the Gaussian frailty scenario, the features of the frailty density are properly captured by the flexible frailty model. Under more challenging scenarios (bimodal or skewed), the performance of the flexible model (in capturing the features of the frailty density) depends on the number of clusters, cluster sizes and frailty standard deviation. When the number of clusters is small (n_{cl}=20) or when the number of data per cluster is insufficient (n_g=4), the shape of the frailty density cannot be properly estimated (with a regularization towards normality). It improves with the number of clusters with even satisfactory results in the challenging "bimodal" frailty scenario (when n_{cl}=50). When the number of clusters and cluster sizes are small, the estimated frailty density resembles a unimodal normal distribution. This is a consequence of the 3^{rd} order difference penalty used in the estimation of the frailty density. Even with a large number of clusters, a cluster size less than 6 might yield a standard normal density. For an accurate representation of the frailty density one needs a large number of clusters of medium to large sizes. The bimodal nature of the frailty density has been well captured with large sample sizes, the largest uncertainty occurring around the modes. It is important to note that the regularization of the estimated frailty density to a Gaussian distribution (for small sample sizes) does not distort the estimation of the regression coefficients: the regression coefficient estimates are robust to misspecification of the frailty density.

The baseline survival function is properly approximated in many scenarios. We also found that the estimations of the baseline survival densities, in Cox PH model are similar to [9] for comparable settings (ignoring frailty model parameters). However, the performance of the baseline survival functions are affected by the number of clusters and cluster sizes. When the cluster sizes are small, then the estimates of the survival curve at selected quantiles tend to be more biased. These results are presented partially for the sake of brevity (only for semiparametric). In many simulation settings, both semiparametric and lognormal frailty specifications resulted in very similar estimates for the survival function.

5 Application to a real data set: Signal Tandmobiel® Study

The Signal tandmobiel dataset results from a longitudinal prospective dental study performed in Flanders (North of Belgium) between the years 1996 and 2001, where 4468 children attending the first year of basic school at the beginning of the study were randomly selected. Then annual dental

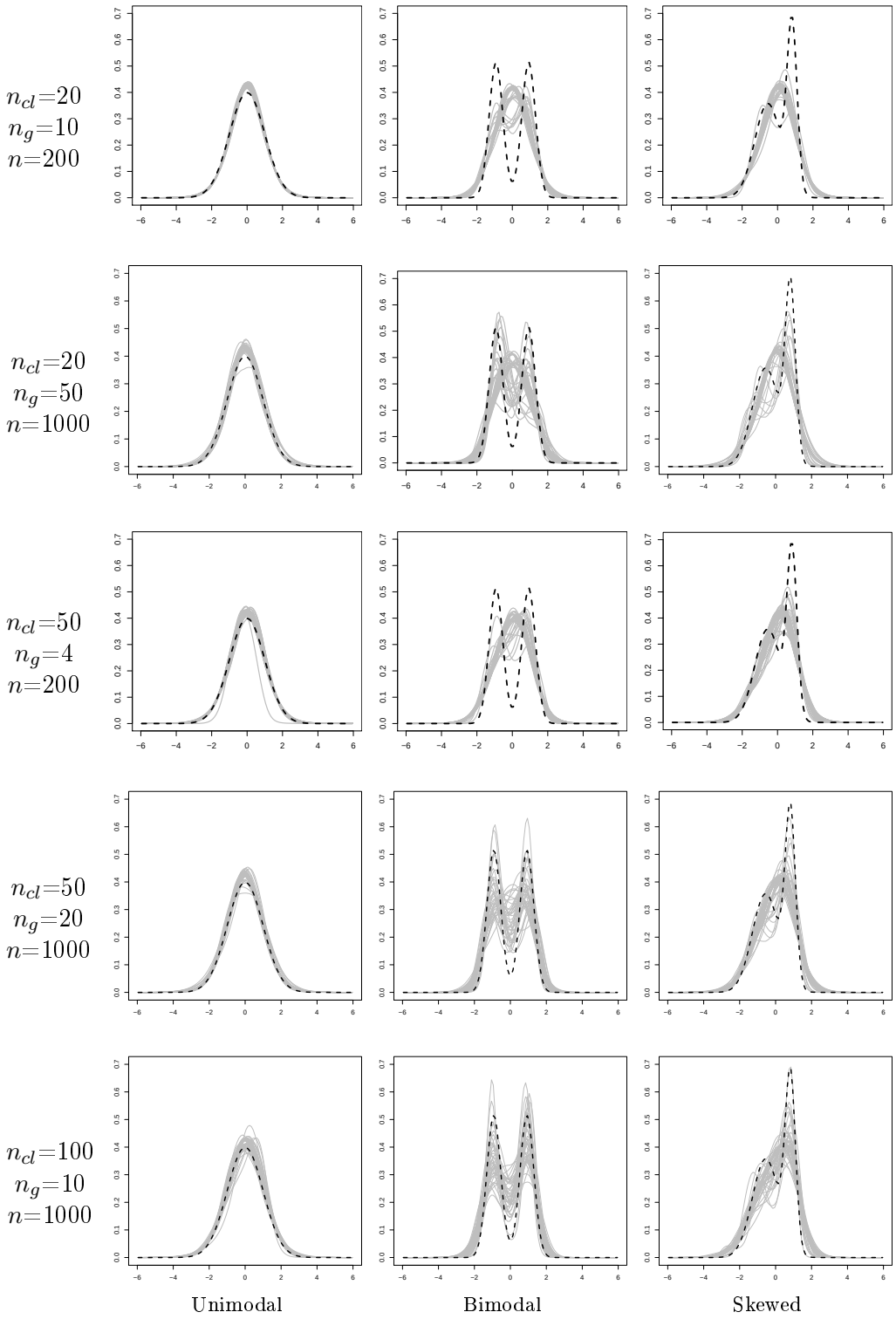


Figure 1: Estimated frailty density aggregated over 50 replications under 5 different settings for 3 scenarios

examinations were performed on the selected cohort by one of 16 trained dentists. The original dataset consists of at most 6 dental observations of dental findings for each child. The data used here are available in the R package *bayesSurv* and contains mainly the information on the emergence timing and caries experience reported as interval-censored observations. Some baseline covariates are also included. For more details on the design of the study, see [58]. Adequate knowledge of timing and patterns of tooth emergence is useful for diagnosis and treatment planning in pediatric dentistry and orthodontics. It is anticipated that boys and girls have different tooth emergence distributions. For this reason, the covariate gender (0=boys, 1=girls) was included in the model. Additionally it was of dental interest to check whether the distribution of emergence time of a permanent tooth changes when its primary predecessor experienced caries or not. Therefore, a binary score, dmf, was associated to each permanent tooth to indicate whether it was preceded by a decayed primary one. The response variable for a particular child, reports for each premolar (teeth 14, 15, 24, 25, 34, 35, 44, 45) his/her age (in years) when it emerged: they will be analyzed jointly and treated as clustered data. It enables not only to quantify the impact of different covariates on the emergence time, but also to study the relationship between the emergence times of different teeth.

A random sample of 150 boys and 150 girls is used for inference. Due to the design of the study with annual planned examinations, the response variable is interval-censored with intervals of approximate length of 1 year, depending on planning of observations. For a better fit, we shifted the time origin to 5 years of age which is clinically the minimal emergence time for the permanent teeth. Namely we replaced T_i by T_i-5 in the model specification. We used 15 knots from -6 to 6 and a third order difference penalty. Five frailty models were fitted to the data: The first one includes only *dmf*, the second *dmf* and *gender*, the third *dmf*, *gender* and their interaction. The fourth model extends the third model by including the *horizontal symmetry* and its interaction with *gender*. Indeed, it was shown by Leroy *et al.* (2003) that there is horizontal symmetry with respect to teeth emergence: the same emergence distributions can be assumed for horizontally symmetric positions. Thus three dummy variables were created: *Man4*, *Max5* and *Man5*, for mandibular first premolars (teeth 34,44), maxillary second premolars (teeth 15,25) and mandibular second premolars (teeth 35,45), respectively. The fourth model includes the main effects and their interactions of *Man4*, *Max5* and *Man5* with *gender*, respectively. For the first four models, the estimated frailty density was bimodal or trimodal (see Fig. 3). It suggested that important covariates were missing. Finally in the fifth model the interactions of *dmf* with *Man4*, *Max5* and *Man5* and three-way interactions of *gender*, *dmf* and *horizontal symmetry* were also included. The estimated frailty density for the fifth model turned to be unimodal. For the sake of brevity, we only present the estimates from Model 1 and 5. Moreover, the estimated survival curves, based on model 5 are shown in Figure 2. The effects of covariates on the time until emergence are given in Table 1. As the main interest in this analysis was the effect of dmf on emergence, we provide the plots of the estimated survival curves for dmf=0 and dmf>0 for boys and girls and the four pairs of horizontally symmetric teeth based on Model 5 (see Figure 2). The figures show that the difference between children with dmf=1 (>0) and dmf=0 is larger for boys than for girls and that the emergence process starts later for boys. Moreover, the subjects who had caries on the primary predecessor significantly tend to have the permanent successor earlier for maxillary tooth. It can also be noticed that for mandibular teeth, there is almost no difference between teeth with and without caries on the primary predecessor. Usually, the emergence process is late for second premolars when compared to first premolars. Our conclusions and figures are in line with Komarek [60] although we use a different modelling approach to analyze the data.

6 Discussion

In this paper, we have presented rather general Bayesian approaches to address an important class of models, the proportional hazards shared frailty model for clustered interval-censored data. Inference ignoring the correlation between observations can be misleading. The main consequence of ignoring

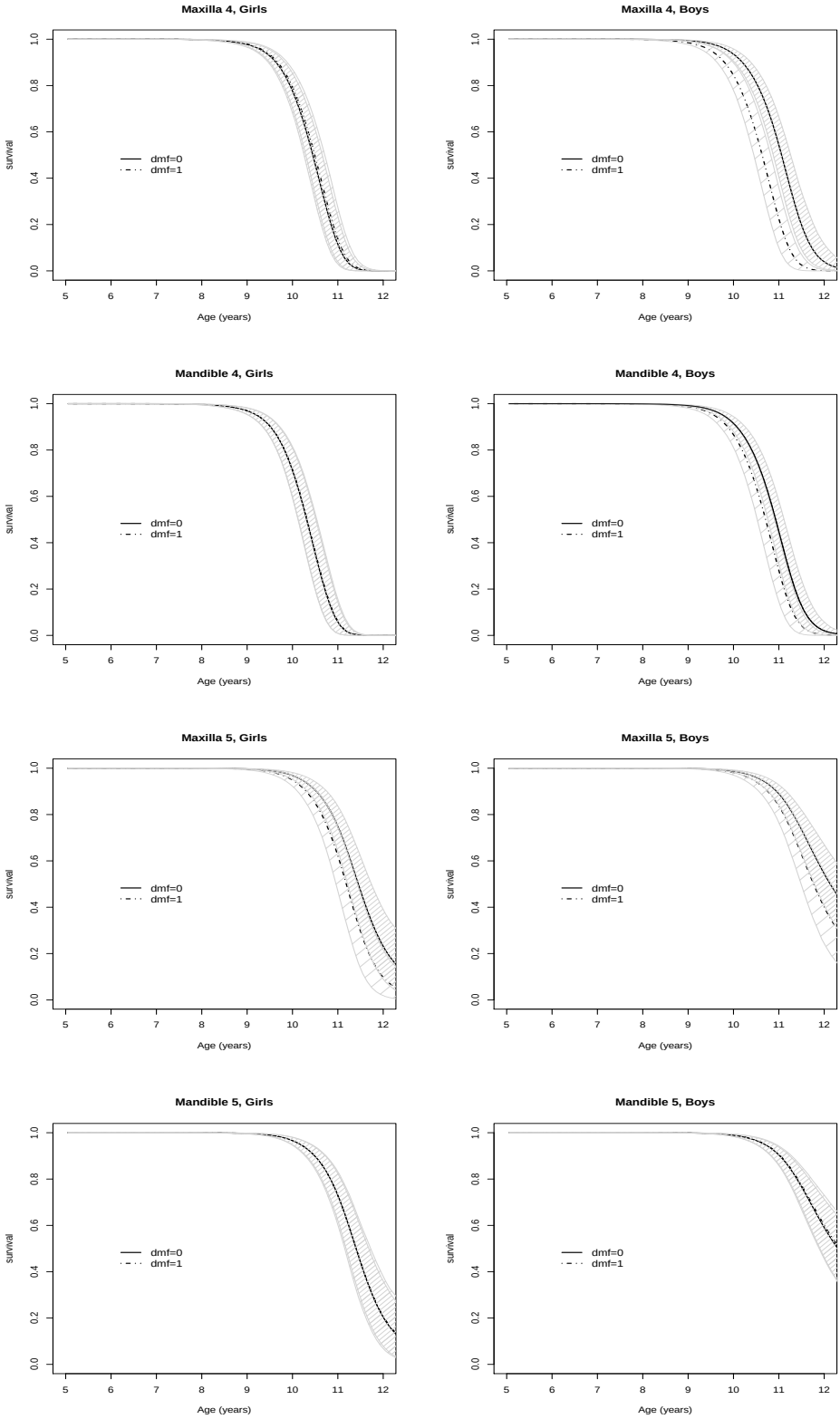
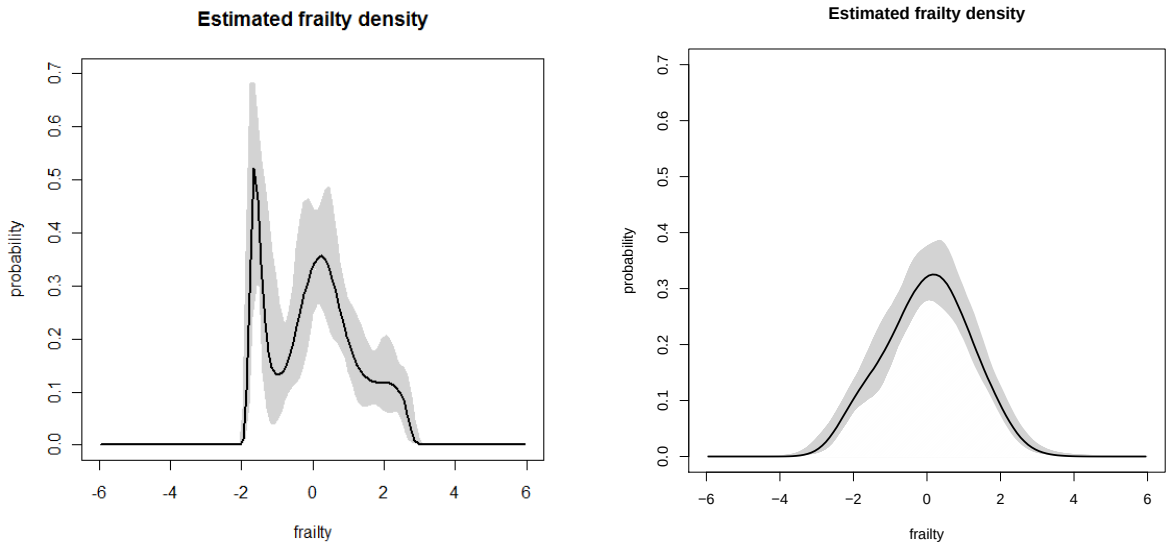


Figure 2: Signa Tandmobiel® Study data: Estimated survival curves based on Model 5 - dmf=0 versus dmf=1



Model 1 Model 5

Figure 3: Signal Tandmobiel® data: Estimated frailty density

the frailty is a reduction in the standard errors that leads to wrongly significant findings. The contribution of this paper revolves around building of flexible Bayesian models for clustered interval-censored data.

On the other hand, it has been indicated in the literature in different contexts that the misspecification of the random effect distribution can influence the estimation of quantities of primary interest, like the fixed effects. To circumvent such misspecification, we have suggested modeling the distribution of the frailty in a flexible way using penalized B-splines. The biggest advantage of using a flexible specification for the density of frailty arise when its shape is of specific interest. If it is considered as a nuisance, assuming a simpler log-normal frailty would be a good enough solution to draw conclusions related to other model parameters, such as regression coefficients and variance of the frailty. Indeed it was shown in the simulation study that the regression parameter estimates in a proportional hazards shared frailty model are robust to the misspecification of the frailty density. However the use of a flexible form for the frailty does not cause any loss in the precision of the estimates when compared to simpler parametric frailty model. Moreover both models provide a possibility of visualizing the baseline density and survival functions.

The authors acknowledges financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'. The Signal Tandmobiel® data set was used in the analyses with a written permission of the authors of Lesaffre, Komarek, and Declerck [59]: they are gratefully acknowledged. Computational resources have been provided by the Consortium des Équipments de Calcul Intensif (CÉCI) funded by the Fonds de la Recherche Scientifique de Belgique (FRS-FNRS) under the Grant No. 2.5020.11.

References

[1] Peto R. Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society (Series C)* 1973, **22**: 86–91.

[2] Turnbull BW. The empirical distribution function with arbitrarily grouped, censored, and truncated data. *Journal of the Royal Statistical Society, Series B (Methodological)* 1976, **38**: 290–295.

[3] Finkelstein DM. A proportional hazards model for interval-censored failure time data. *Biometrics* 1986, **42**: 845–854.

[4] Goetghebeur E, Ryan L. Semiparametric regression analysis of interval-censored data. *Biometrics* 2000, **56**: 1139–1144.

[5] Betensky RA., Lindsey JC, Ryan LM, Wand MP. A local likelihood proportional hazards model for interval-censored data. *Statistics in Medicine* 2002, **21**: 263–275.

[6] Cai T, Betensky RA. Hazard regression for interval-censored data with penalized spline. *Biometrics* 2003, **59**: 570–579.

[7] Komarek A, Lesaffre E, Hilton JF. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics* 2005, **14** (3): 726–745.

[8] Zhang M, Davidian M. Smooth semiparametric regression for arbitrarily censored time-to-event data. *Biometrics* 2008, **64** (2): 567–576.

[9] Cetinyurek A, Lambert P. Smooth estimation of survival functions and hazard ratios from interval-censored data using Bayesian penalized B-splines. *Statistics in Medicine* 2011, **30**: 75–90.

[10] Vaupel JW, Manton KG and Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979, **16**: 439–454.

[11] Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* 1978, **65**: 141–151.

[12] Clayton D , Cuzick J. The semi-parametric Pareto model for regression analysis of survival times. *Proceedings of the Centenary Session of the International Statistical Institute* 1985, **47**: 467–485.

[13] Hougaard P. Analysis of multivariate survival data *Springer, New York* 2000, **38**: 451–455.

[14] Lambert P, Collett D, Kimber A, Johnson R. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine* 2004, **23**: 3177–3192.

[15] Oakes D. A concordence test for independence in the presence of censoring. *Biometrics* 1982, **38**: 451–455.

[16] Aalen OO. Heterogeneity in Survival Analysis. *Statistics in Medicine* 1988, **7**: 1121–1137.

[17] Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society, Series American B* 1972, **34**: 187–220.

[18] Kim M, Xue X. The analysis of multivariate interval-censored survival data. *Statistics in Medicine* 2002, **21**: 3715–3726.

[19] Whittemore AC, Kellner JB. Survival estimation using splines. *Biometrics* 1986, **42**: 495–506.

[20] Rosenberg PS. Hazard function estimation using B-splines. *Biometrics* 1995, **51**: 874–887.

[21] Shih JH, Louis TA. Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis* 1995, **1**: 205–220.

[22] Agresti A, Caffo B, Ohman-Strickland P. Examples in which misspecifications of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics and Data Analysis* 2004, **47**: 639–653.

[23] Klein JP. Semiparametric estimation of random effects Using the Cox model based on the EM algorithm. *Biometrics* 1992, **48 (3)**: 795–806.

[24] Verbeke GP, Lesaffre E. A linear mixed-effects model with heterogeneity in the random-effects population . *Journal of the American Statistical Association* 1996, **91**: 217–221.

[25] Zhang GP, Davidian M. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* 1992, **48 (3)**: 795–806.

[26] Ghidey W, Lesaffre E, Eilers P. Smooth random effects distribution in a linear mixed model. *Biometrics* 2004, **60(4)**: 945–953.

[27] Sheraf E, Strawderman RL, Ruppert D, Cowen M, Halasyamani L. Bayesian adaptive B-spline estimation in proportional hazards frailty models *Electronic Journal of Statistics* 2010, **4**: 606–642.

[28] Komarek A, Lesaffre E. Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions. *Journal of American Statistical Association* 2008, **103** (482): 523–533.

[29] Goethals K, Ampe B, Berkvens D, Laevens H, Janssen P, Duchateau L. Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural Biological and Environmental Statistics* 2009, **26**: 769–781.

[30] Henschel V, Engel J, Hölzel D, Mansmann U. A semiparametric Bayesian proportional hazards model for interval censored data with frailty effects. *BMC Medical Research Methodology* 2009, **9**: 9.

[31] Bellamy HW, Li Y, Ryan LM, Lipsitz S, Canner MJ, Wright R. Analysis of clustered and interval-censored data from a community-based Study in asthma. *Statistics in Medicine* 2004, **23**: 3607–3621.

[32] Zuma K, Lurie M. Application and comparison of methods for analysing correlated interval-censored data from sexual partnerships. *Journal of Data Science* 2005, **3**: 241–256.

[33] Zuma K. A Bayesian analysis of correlated interval-censored data. *Communications in Statistics - Theory and Methods* 2007, **36**: 725–730.

[34] Lam KF, Xu Y, Cheung TL. A multiple imputation approach for clustered interval-censored survival data. *Statistics in Medicine* 2010, **29**: 680–693.

[35] Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties (with comments and rejoinder). *Statistical Science* 1996, **11**: 89–121.

[36] Lambert P, Eilers PHC. Bayesian proportional hazards model with time varying regression coefficients: a penalized Poisson regression approach. *Statistics in Medicine* 2005, **24**: 3977–3989.

[37] Lambert P, Eilers PHC. Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis* 2009, **53**: 1388–1399.

[38] Thompson R, Baker RJ. Composite link functions in generalized linear models. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 1981, **30**: 125–131.

[39] Vaupel JW, Yashin AI. The deviant dynamics of death in heterogeneous populations. *RR-83-1* 1983, International Institute for Applied System Analysis.

[40] Hougaard P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* 1984, **71**: 75–84.

[41] Hougaard P. Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 1986, **73**: 387–396.

[42] Therneau T. survival: Survival Analysis, Including Penalised Likelihood. *R package version 2.36-14*, URL http://CRAN.R-project.org/package=survival. 2012.

[43] R Development Core Team. *R: A Language and Environment for Statistical Computing*. 2012, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[44] Hougaard P. Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 1986, **73**: 387–396.

[45] Hougaard P. Frailty models for survival data. *Lifetime Data Analysis* 1995, **1**: 255–273.

[46] Klein JP, Moeschberger ML, Li YH, Wang ST. Estimating random effects in the Framingham heart study. *Survival Analysis: State of the Art. Kluwer Academic, Boston,MA* 1992, : 99–120.

[47] Chen J, Zhang D, Davidian M. A Monte Carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics* 2002, **3(3)**: 347–360.

[48] Lambert P. Nonparametric additive location-scale models for interval censored data. *Statistics and computing* 2013, **23(1)**: 75–90.

[49] Lang S, Brezger A. Bayesian P-splines. *Journal of Computational and Graphical Statistics* 2004, **13**: 183–212.

[50] Jullion A, Lambert P. Robust specification of roughness penalty prior distribution in spatially adaptive Bayesian P-splines model. *Computational Statistics and Data Analysis* 2007, **51**: 2542–2558.

[51] Scheipl F, Kneib T. Locally adaptive Bayesian P-splines with a normal-exponential-gamma prior. *Computational Statistics and Data Analysis* 2009, **53**: 3533–3552.

[52] Lambert P. Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics and Data Analysis* 2007, **51**: 6307–6320.

[53] Strasak AM, Lang S, Kneib T, Brant LJ, Klenk J, Hilbe W, Oberaigner W, Ruttmann E, Kaltenbacha L, Concin H, Diem G, Pfeiffer KP, Ulmer H, The VHM PP Study Group. Use of penalized splines in extended Cox-type additive hazard regression to flexibly estimate the effect of time-varying serum uric acid on risk of cancer incidence: A prospective, population-based study in 78,850 men. *Annals of Epidemiology* 2009, **19** (1): 15–24.

[54] Gelman A, Roberts GO, Gilks WR. Efficient Metropolis jumping rules. *Bayesian Statistics* 1996, **5**: 599–607.

[55] Roberts GO, Rosenthal JS. Optimal scaling for various Metropolis- Hastings algorithms. *Statistical Science* 2001, **16**: 351–367.

[56] Atchadé YF, Rosenthal JS. On adaptive Markov chain Monte Carlo algorithms. *Bernoulli* 2005, **11**(5): 815–828.

[57] Law G, Brookmeyer R. Effects of mid-point imputation on the analysis of doubly censored data. *Statistics in Medicine* 1992, **11**: 1569–1578.

[58] Vanobbergen J, Martens L, Lesaffre E, Declerck D. The Signal-Tandmobiel project a longitudinal intervention health promotion study in Flanders (Belgium): baseline and first year results. *European Journal of Paediatric Dentistry* 2000, **2**: 87–96.

[59] Lesaffre E, Komárek A, Declerck D. An overview of methods for interval-censored data with an emphasis on applications in dentistry. *Statistical Methods in Medical Research* 2005, **14**: 539–552.

[60] Komarek A, Lesaffre E. Bayesian accelerated failure time model for correlated censored data with a normal mixture as an error distribution. *Statistica Sinica* 2007, **17**: 549–569.

Table 1: Signal Tandmobiel® Study data: Parameter estimates and 90% confidence intervals for the effect of covariates and frailty standard deviation

		Gaussian Frailty		Semiparametric Frailty	
		β	90% Cred. Int.	β	90% Cred. Int.
Model 1	dmf	0.32	(0.17 ; 0.46)	0.34	(0.18 ; 0.48)
	sd(frailty)	1.52	(1.37 ; 1.68)	1.48	(1.35 ; 1.65)
Model 5	Girl	1.47	(0.90 ; 2.08)	1.32	(0.69 ; 1.89)
	dmf	0.92	(0.57 ; 1.28)	0.92	(0.55 ; 1.28)
	Girl - dmf	-1.02	(-1.56 ; -0.57)	-1.01	(-1.52 ; -0.48)
	Mandibular 4	0.29	(0.01 ; 0.56)	0.28	(-0.01 ; 0.57)
	Maxillary 5	-1.80	(-2.12 ; -1.47)	-1.68	(-2.01 ; -1.36)
	Mandibular 5	-1.94	(-2.27 ; -1.63)	-1.83	(-2.16 ; -1.49)
	Girl - mandibular 4	0.05	(-0.37 ; 0.47)	0.03	(-0.39 ; 0.45)
	Girl - maxillary 5	-0.42	(-0.85 ; 0.03)	-0.41	(-0.84 ; 0.04)
	Girl - mandibular 5	-0.18	(-0.61 ; 0.27)	-0.17	(-0.61 ; 0.28)
	dmf - mandibular 4	-0.45	(-0.89 ; 0.01)	-0.44	(-0.91 ; 0.03)
	dmf - maxillary 5	-0.52	(-1.00 ; -0.07)	-0.49	(-0.96 ; 0.00)
	dmf - mandibular 5	-1.00	(-1.50 ; -0.49)	-0.96	(-1.48 ; -0.44)
	Girl - dmf - mandibular 4	0.52	(-0.10 ; 1.18)	0.51	(-0.17 ; 1.20)
	Girl - dmf - maxillary 5	1.15	(0.42 ; 1.80)	1.09	(0.41 ; 1.75)
	Girl - dmf - mandibular 5	1.14	(0.41 ; 1.84)	1.04	(0.30 ; 1.80)
	sd(frailty)	2.80	(2.55 ; 3.07)	2.24	(2.01 ; 2.53)

Table 2: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α with 10% of right-censoring and a mean width of 0.5σ for a varied number of clusters and cluster sizes under a "Unimodal" frailty density with a standard deviation of α=0.8 in S=300 replications

			Semiparametric					Gaussian				
			β=0.693					β=0.693				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.75	(0.47-1.06)	7.7	88	74	0.75	(0.49-1.08)	7.8	88	75
	20	400	0.70	(0.51-0.90)	0.8	89	79	0.70	(0.51-0.90)	0.8	90	79
	50	1000	0.70	(0.58-0.83)	0.7	90	80	0.70	(0.58-0.83)	0.7	89	80
50	4	200	0.72	(0.43-1.02)	3.7	88	80	0.72	(0.43-1.02)	3.8	88	82
	6	300	0.70	(0.48-0.94)	0.8	89	81	0.70	(0.48-0.95)	1.1	90	79
	10	500	0.69	(0.53-0.87)	-0.3	91	82	0.69	(0.54-0.87)	-0.1	90	82
	20	1000	0.70	(0.59-0.81)	0.7	92	82	0.70	(0.60-0.81)	0.7	92	82
100	4	400	0.71	(0.53-0.90)	2.1	92	78	0.71	(0.53-0.90)	2.6	91	77
	6	600	0.70	(0.55-0.85)	0.4	91	81	0.70	(0.55-0.86)	0.6	92	80
	10	1000	0.69	(0.58-0.80)	-0.4	92	84	0.69	(0.58-0.80)	-0.1	93	84
200	4	800	0.70	(0.56-0.84)	1.1	90	77	0.70	(0.56-0.85)	1.6	89	79
	6	1200	0.68	(0.59-0.77)	-2.1	89	77	0.69	(0.61-0.78)	-0.3	91	80
	10	2000	0.68	(0.60-0.78)	-1.4	89	76	0.68	(0.59-0.78)	-1.2	89	79
			α=0.8					α=0.8				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.73	(0.47-1.06)	-8.1	84	77	0.74	(0.49-1.08)	-7.6	83	75
	20	400	0.76	(0.52-1.03)	-4.5	89	82	0.76	(0.53-1.02)	-4.7	87	78
	50	1000	0.76	(0.52-1.00)	-5.1	88	78	0.76	(0.53-0.99)	-5.5	86	76
50	4	200	0.72	(0.20-1.02)	-9.8	88	78	0.72	(0.21-1.02)	-9.7	88	75
	6	300	0.74	(0.55-0.93)	-7.4	87	79	0.74	(0.54-0.93)	-7.3	85	78
	10	500	0.76	(0.57-0.95)	-4.8	88	78	0.77	(0.60-0.95)	-3.4	87	75
	20	1000	0.78	(0.65-0.94)	-2.5	94	82	0.78	(0.64-0.94)	-2.8	90	77
100	4	400	0.76	(0.61-0.93)	-5.0	91	76	0.77	(0.62-0.93)	-3.8	91	76
	6	600	0.77	(0.65-0.91)	-3.3	92	82	0.78	(0.65-0.91)	-2.5	91	82
	10	1000	0.77	(0.65-0.90)	-3.3	89	79	0.79	(0.67-0.90)	-1.6	90	80
200	4	800	0.77	(0.63-0.87)	-4.2	88	80	0.78	(0.64-0.89)	-2.3	90	81
	6	1200	0.77	(0.68-0.87)	-3.6	89	81	0.78	(0.69-0.88)	-1.9	92	82
	10	2000	0.78	(0.70-0.86)	-2.5	91	83	0.79	(0.70-0.87)	-1.6	92	81

Table 3: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α with 10% of right-censoring and a mean width of 0.5σ for a varied number of clusters and cluster sizes under a "Unimodal" frailty density with a standard deviation of α=1.2 in S=300 replications

			Semiparametric					Gaussian				
			β=0.693					β=0.693				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.71	(0.43-0.99)	2.8	90	77	0.71	(0.44-0.99)	3.0	90	78
	20	400	0.70	(0.50-0.89)	1.7	89	78	0.71	(0.50-0.89)	1.8	88	79
	50	1000	0.69	(0.58-0.81)	-0.1	91	83	0.69	(0.58-0.81)	0.1	91	83
50	4	200	0.73	(0.43-1.03)	5.7	88	77	0.74	(0.44-1.04)	6.1	87	76
	6	300	0.71	(0.49-0.93)	1.9	92	84	0.71	(0.49-0.93)	2.1	93	83
	10	500	0.70	(0.54-0.88)	0.8	91	82	0.70	(0.55-0.87)	0.9	92	82
	20	1000	0.70	(0.58-0.83)	1.4	87	81	0.70	(0.58-0.83)	1.5	88	81
100	4	400	0.71	(0.54-0.93)	3.0	90	81	0.72	(0.54-0.94)	3.5	89	81
	6	600	0.71	(0.56-0.84)	2.4	93	85	0.71	(0.56-0.84)	2.8	93	85
	10	1000	0.69	(0.58-0.82)	-0.1	91	82	0.69	(0.58-0.82)	0.1	91	81
200	4	800	0.69	(0.53-0.84)	-0.5	89	82	0.69	(0.54-0.84)	-0.1	89	79
	6	1200	0.70	(0.59-0.85)	1.2	90	80	0.70	(0.59-0.85)	1.4	88	80
	10	2000	0.69	(0.60-0.78)	0.3	88	80	0.70	(0.60-0.78)	0.4	89	80
			α=1.2					α=1.2				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	1.16	(0.77-1.60)	-3.3	88	78	1.15	(0.76-1.56)	-3.9	85	74
	20	400	1.16	(0.79-1.55)	-3.2	90	77	1.15	(0.79-1.52)	-4.0	85	73
	50	1000	1.17	(0.81-1.50)	-2.8	89	80	1.15	(0.80-1.48)	-4.1	88	75
50	4	200	1.16	(0.79-1.50)	-3.7	86	76	1.15	(0.80-1.49)	-3.9	84	73
	6	300	1.15	(0.89-1.40)	-4.0	89	79	1.15	(0.90-1.39)	-4.5	87	77
	10	500	1.18	(0.97-1.41)	-1.5	94	88	1.17	(0.95-1.39)	-2.5	91	82
	20	1000	1.20	(1.01-1.44)	0.1	93	84	1.19	(1.00-1.43)	-1.0	89	78
100	4	400	1.18	(0.93-1.41)	-1.9	89	79	1.19	(0.95-1.41)	-1.2	87	76
	6	600	1.18	(1.00-1.38)	-1.3	92	82	1.18	(1.01-1.39)	-1.3	90	81
	10	1000	1.20	(1.04-1.38)	0.1	90	82	1.19	(1.04-1.37)	-0.4	90	78
200	4	800	1.17	(1.01-1.35)	-2.5	89	76	1.19	(1.03-1.35)	-1.0	86	76
	6	1200	1.19	(1.06-1.34)	-0.9	91	84	1.20	(1.08-1.34)	-0.1	91	77
	10	2000	1.19	(1.08-1.31)	-0.7	91	85	1.19	(1.08-1.33)	-0.5	90	81

Table 4: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α with 10% of right-censoring and a mean width of 0.5σ for a varied number of clusters and cluster sizes under a "Bimodal" frailty density with standard deviation of α=0.8 in S=300 replications

			Semiparametric					Gaussian				
			β=0.693					β=0.693				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.74	(0.47-1.02)	6.6	90	78	0.74	(0.47-1.02)	6.7	88	78
	20	400	0.70	(0.54-0.85)	0.3	94	85	0.70	(0.53-0.85)	0.5	94	85
	50	1000	0.69	(0.57-0.81)	0.1	91	78	0.69	(0.57-0.80)	0.1	90	78
50	4	200	0.71	(0.42-0.99)	1.9	88	79	0.71	(0.44-0.99)	2.6	89	79
	6	300	0.70	(0.46-0.91)	0.8	91	81	0.70	(0.48-0.90)	1.1	91	81
	10	500	0.71	(0.54-0.87)	2.0	91	81	0.71	(0.54-0.86)	1.9	90	81
	20	1000	0.70	(0.57-0.82)	0.4	89	81	0.69	(0.57-0.82)	0.2	88	81
100	4	400	0.69	(0.52-0.91)	0.1	89	83	0.70	(0.53-0.92)	1.2	91	83
	6	600	0.70	(0.54-0.86)	1.0	88	77	0.70	(0.54-0.86)	0.8	88	77
	10	1000	0.70	(0.57-0.83)	1.5	86	78	0.70	(0.57-0.83)	0.9	88	78
200	4	800	0.69	(0.58-0.81)	0.1	89	80	0.69	(0.58-0.80)	-0.5	91	80
	6	1200	0.69	(0.57-0.80)	-0.5	90	80	0.68	(0.56-0.79)	-1.3	89	79
	10	2000	0.69	(0.61-0.78)	0.2	89	79	0.69	(0.61-0.77)	-0.3	91	80
			α=0.80					α=0.80				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.74	(0.33-1.01)	-8.1	89	82	0.75	(0.53-0.98)	-5.8	93	86
	20	400	0.77	(0.60-0.93)	-3.8	97	92	0.77	(0.60-0.92)	-3.9	97	91
	50	1000	0.77	(0.63-0.90)	-3.7	99	96	0.77	(0.65-0.90)	-4.2	99	96
50	4	200	0.59	(0.26-0.95)	-25.8	60	54	0.73	(0.40-0.96)	-8.6	91	77
	6	300	0.73	(0.29-0.93)	-8.2	86	80	0.77	(0.61-0.93)	-3.4	95	86
	10	500	0.79	(0.66-0.92)	-1.8	97	91	0.79	(0.66-0.92)	-1.2	97	91
	20	1000	0.79	(0.68-0.90)	-1.4	95	89	0.80	(0.69-0.90)	-0.6	98	91
100	4	400	0.63	(0.29-0.92)	-21.5	64	60	0.78	(0.63-0.93)	-2.1	92	84
	6	600	0.78	(0.64-0.90)	-2.8	91	85	0.79	(0.66-0.91)	-0.7	93	87
	10	1000	0.79	(0.69-0.89)	-1.2	92	82	0.80	(0.70-0.90)	0.1	95	86
200	4	800	0.63	(0.38-0.88)	-21.0	54	48	0.81	(0.71-0.89)	0.6	93	82
	6	1200	0.79	(0.72-0.86)	-1.1	95	87	0.80	(0.72-0.87)	0.4	96	92
	10	2000	0.79	(0.73-0.86)	-1.3	90	81	0.80	(0.75-0.87)	0.5	97	92

Table 5: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α with 10% of right-censoring and a mean width of 1.0σ for a varied number of clusters and cluster sizes under a "Bimodal" frailty density with standard deviation of α=1.2 in S=300 replications

			Semiparametric					Gaussian				
			β=0.693					β=0.693				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.72	(0.44-1.00)	4.3	93	82	0.72	(0.45-1.01)	4.4	93	82
	20	400	0.70	(0.50-0.90)	0.9	93	82	0.70	(0.50-0.90)	0.8	92	82
	50	1000	0.70	(0.58-0.84)	1.3	91	78	0.70	(0.58-0.84)	1.3	91	78
50	4	200	0.69	(0.41-1.02)	-0.3	90	81	0.69	(0.41-1.02)	-0.3	91	83
	6	300	0.71	(0.48-0.95)	2.0	91	82	0.70	(0.47-0.95)	1.5	90	82
	10	500	0.70	(0.51-0.88)	1.7	89	81	0.70	(0.50-0.88)	1.3	91	81
	20	1000	0.70	(0.58-0.84)	0.8	90	84	0.69	(0.58-0.83)	0.3	91	84
100	4	400	0.70	(0.50-0.91)	1.3	93	81	0.70	(0.50-0.91)	0.7	94	81
	6	600	0.69	(0.50-0.87)	-0.4	87	78	0.68	(0.51-0.85)	-1.6	90	78
	10	1000	0.70	(0.55-0.84)	0.6	93	73	0.69	(0.56-0.82)	-0.2	90	73
200	4	800	0.70	(0.56-0.83)	1.5	87	77	0.69	(0.57-0.81)	-0.2	92	67
	6	1200	0.69	(0.59-0.80)	0.1	88	79	0.68	(0.59-0.80)	-1.2	90	59
	10	2000	0.70	(0.62-0.78)	0.3	90	82	0.69	(0.62-0.76)	-0.4	92	62
			α=1.2					α=1.2				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	1.18	(0.85-1.53)	-1.7	95	90	1.17	(0.88-1.50)	-2.4	95	91
	20	400	1.16	(0.91-1.40)	-3.5	97	92	1.15	(0.91-1.39)	-4.1	97	90
	50	1000	1.16	(0.95-1.36)	-3.4	97	94	1.16	(0.96-1.36)	-3.7	99	94
50	4	200	1.21	(0.93-1.54)	1.1	94	86	1.21	(0.93-1.53)	1.1	94	83
	6	300	1.20	(0.97-1.44)	0.3	95	87	1.20	(0.98-1.42)	0.3	95	89
	10	500	1.19	(1.03-1.38)	-0.8	97	91	1.20	(1.04-1.40)	0.2	98	93
	20	1000	1.16	(1.02-1.32)	-3.1	93	85	1.19	(1.04-1.35)	-0.9	98	92
100	4	400	1.22	(1.01-1.40)	1.3	93	85	1.23	(1.05-1.43)	2.8	93	86
	6	600	1.19	(1.02-1.37)	-0.4	90	80	1.23	(1.06-1.39)	2.3	93	85
	10	1000	1.18	(1.03-1.33)	-2.0	81	74	1.21	(1.10-1.35)	1.1	97	92
200	4	800	1.22	(1.04-1.42)	2.0	86	76	1.26	(1.13-1.39)	4.7	83	72
	6	1200	1.17	(1.04-1.27)	-2.4	93	79	1.23	(1.15-1.33)	3.2	92	81
	10	2000	1.17	(1.07-1.26)	-2.4	87	80	1.23	(1.16-1.30)	2.4	93	85

Table 6: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α with 10% of right-censoring and a mean width of 1.0σ for a varied number of clusters and cluster sizes under a "Skewed" frailty density with standard deviation of α=1.2 in S=300 replications

			Semiparametric					Gaussian				
			β=0.693					β=0.693				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.71	(0.46-1.01)	2.4	92	81	0.71	(0.43-1.01)	2.4	92	81
	20	400	0.70	(0.51-0.87)	1.0	94	85	0.70	(0.51-0.87)	0.9	94	86
	50	1000	0.70	(0.57-0.82)	0.5	91	80	0.70	(0.57-0.82)	0.4	90	81
50	4	200	0.70	(0.41-1.01)	0.9	91	79	0.70	(0.41-1.01)	0.8	91	79
	6	300	0.70	(0.48-0.94)	1.3	91	84	0.70	(0.48-0.94)	1.4	91	84
	10	500	0.70	(0.48-0.88)	0.5	88	76	0.70	(0.49-0.89)	0.5	89	76
	20	1000	0.69	(0.55-0.84)	0.1	88	80	0.69	(0.56-0.83)	0.1	87	80
100	4	400	0.70	(0.46-0.91)	0.3	89	77	0.69	(0.48-0.92)	0.1	89	76
	6	600	0.69	(0.51-0.88)	-0.6	87	80	0.69	(0.51-0.87)	-0.9	90	80
	10	1000	0.69	(0.56-0.81)	-0.5	91	82	0.69	(0.55-0.81)	-0.7	91	81
200	4	800	0.70	(0.57-0.83)	0.4	90	83	0.69	(0.56-0.81)	-0.4	90	81
	6	1200	0.68	(0.56-0.80)	-1.8	88	76	0.68	(0.55-0.80)	-2.4	89	78
	10	2000	0.69	(0.61-0.76)	-0.5	86	77	0.69	(0.61-0.76)	-0.7	90	82
			α=1.2					α=1.2				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	1.20	(0.83-1.63)	-0.2	93	81	1.18	(0.83-1.59)	-2.0	89	79
	20	400	1.16	(0.86-1.46)	-3.0	96	90	1.14	(0.88-1.45)	-4.6	94	89
	50	1000	1.15	(0.87-1.44)	-3.8	95	88	1.14	(0.88-1.42)	-4.9	94	84
50	4	200	1.17	(0.83-1.51)	-2.9	91	79	1.15	(0.84-1.50)	-4.1	87	75
	6	300	1.19	(0.94-1.47)	-0.7	94	83	1.17	(0.93-1.46)	-2.1	93	82
	10	500	1.18	(0.96-1.41)	-1.3	94	85	1.17	(0.96-1.39)	-2.2	92	83
	20	1000	1.18	(1.00-1.36)	-1.9	96	88	1.18	(1.00-1.36)	-1.9	96	86
100	4	400	1.18	(0.97-1.39)	-1.5	91	84	1.18	(0.96-1.39)	-1.7	91	83
	6	600	1.19	(1.01-1.38)	-1.2	92	83	1.18	(1.02-1.39)	-1.2	92	82
	10	1000	1.17	(1.00-1.32)	-2.4	90	83	1.18	(1.03-1.33)	-1.5	94	82
200	4	800	1.17	(1.03-1.32)	-2.7	80	68	1.20	(1.09-1.35)	0.3	91	84
	6	1200	1.16	(1.03-1.31)	-3.0	84	73	1.19	(1.07-1.32)	-0.9	94	82
	10	2000	1.17	(1.07-1.26)	-2.8	88	74	1.19	(1.10-1.28)	-0.9	92	85

Table 7: The mean, relative bias (Rbias in %), 90% credible intervals and corresponding empirical coverages (EC) for β and α with 35% of right-censoring and a mean width of 1.0σ for a varied number of clusters and cluster sizes under a "Bimodal" frailty density with standard deviation of α=1.2 in S=300 replications

			Semiparametric					Gaussian				
			β=0.693					β=0.693				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	0.70	(0.35-1.05)	0.6	90	83	0.70	(0.36-1.05)	0.7	90	81
	20	400	0.70	(0.45-0.92)	-1.0	90	81	0.69	(0.44-0.93)	-0.9	89	81
	50	1000	0.69	(0.54-0.85)	-0.8	88	79	0.69	(0.54-0.85)	-0.7	88	78
50	4	200	0.70	(0.39-1.04)	0.7	90	81	0.70	(0.38-1.05)	1.4	90	80
	6	300	0.70	(0.43-0.98)	1.1	90	80	0.70	(0.43-0.98)	1.3	90	80
	10	500	0.68	(0.48-0.88)	-2.3	90	85	0.68	(0.48-0.88)	-2.3	91	85
	20	1000	0.69	(0.56-0.84)	-0.7	90	81	0.69	(0.56-0.84)	-0.7	91	81
100	4	400	0.70	(0.46-0.95)	1.1	89	78	0.70	(0.45-0.95)	1.0	88	78
	6	600	0.68	(0.50-0.89)	-1.5	88	77	0.68	(0.50-0.89)	-1.6	90	78
	10	1000	0.68	(0.51-0.84)	-1.6	87	77	0.68	(0.52-0.84)	-1.5	89	81
200	4	800	0.68	(0.52-0.83)	-1.7	87	77	0.68	(0.52-0.83)	-2.3	86	76
	6	1200	0.68	(0.50-0.84)	-2.3	89	82	0.68	(0.53-0.82)	-1.8	90	80
	10	2000	0.68	(0.56-0.79)	-2.2	90	78	0.68	(0.57-0.79)	-1.9	89	80
			α=1.2					α=1.2				
n_{cl}	n_g	N	Mean	90% CI	Rbias (%)	EC 90	EC 80	Mean	90% CI	Rbias (%)	EC 90	EC 80
20	10	200	1.24	(0.88-1.67)	3.6	94	87	1.23	(0.88-1.65)	2.2	93	85
	20	400	1.23	(0.98-1.47)	2.2	98	94	1.21	(0.97-1.44)	0.5	98	93
	50	1000	1.19	(0.96-1.38)	-1.2	98	95	1.17	(0.96-1.37)	-2.2	98	94
50	4	200	1.27	(0.95-1.64)	5.9	93	84	1.27	(0.97-1.67)	6.2	92	82
	6	300	1.26	(0.98-1.53)	4.9	93	84	1.25	(0.97-1.54)	4.1	92	81
	10	500	1.23	(1.04-1.42)	2.3	95	90	1.23	(1.04-1.44)	2.8	95	89
	20	1000	1.19	(1.03-1.35)	-0.8	95	87	1.22	(1.05-1.38)	1.5	97	92
100	4	400	1.24	(1.04-1.50)	3.6	92	83	1.26	(1.03-1.53)	4.6	91	81
	6	600	1.23	(1.05-1.42)	2.5	91	82	1.26	(1.08-1.46)	4.8	90	81
	10	1000	1.19	(1.04-1.38)	-0.6	87	78	1.25	(1.11-1.42)	3.8	92	85
200	4	800	1.21	(1.10-1.33)	1.1	92	84	1.27	(1.15-1.39)	5.7	86	71
	6	1200	1.18	(1.02-1.33)	-1.4	92	86	1.27	(1.14-1.38)	5.5	87	76
	10	2000	1.19	(1.03-1.48)	-1.1	91	84	1.25	(1.16-1.36)	4.2	91	81

Table 8: Relative bias (Rbias in %), empirical standard errors (ESE) and root mean squared error (RMSE) for baseline survival at selected quantiles (5%, 15%, 25%, 35%, 50%, 60%, 75%, 85% and 88%) of T for 10% right-censoring, a mean interval width of $1.0\sigma \approx 15$ and frailty standard deviation $\alpha = 1.2$ for a sample of size n=200 and n=1000 in S=300 replications (using the semiparametric frailty model)

		n_{cl}=20,n_g=10			n_{cl}=20,n_g=50			n_{cl}=50,n_g=4			n_{cl}=50,n_g=20		
Frailty		Rbias (%)	ESE	RMSE	Rbias (%)	ESE	RMSE	Rbias (%)	ESE	RMSE	Rbias (%)	ESE	RMSE
	S(39)	0.1	0.014	0.521	-0.1	0.01	0.52	-0.1	0.017	0.521	-0.3	0.008	0.519
	S(49)	0.5	0.034	0.446	0.1	0.025	0.442	0.4	0.037	0.445	-0.6	0.019	0.438
	S(55)	1.1	0.051	0.382	0.3	0.038	0.375	1.0	0.052	0.379	-1.0	0.029	0.367
	S(59)	1.8	0.062	0.329	0.5	0.047	0.324	1.9	0.061	0.333	-1.3	0.035	0.318
Uni.	S(65)	4.6	0.074	0.299	1.0	0.058	0.294	5.7	0.071	0.297	-1.9	0.043	0.295
	S(69)	8.7	0.077	0.312	1.3	0.061	0.315	10.6	0.073	0.303	-2.8	0.044	0.314
	S(75)	22.7	0.073	0.365	5.4	0.057	0.389	28.3	0.070	0.361	-1.2	0.04	0.398
	S(80)	48.1	0.064	0.421	26.4	0.051	0.442	61.0	0.064	0.407	16.3	0.035	0.453
	S(81)	55.4	0.062	0.431	35.0	0.050	0.447	70.7	0.063	0.413	23.8	0.034	0.455
	S(39)	0.1	0.019	0.523	-0.4	0.017	0.519	-0.1	0.016	0.521	-0.4	0.012	0.519
	S(49)	0.2	0.050	0.444	-1.9	0.044	0.434	-0.9	0.038	0.435	-0.8	0.032	0.435
	S(55)	0.3	0.075	0.379	-1.2	0.066	0.375	-1.8	0.053	0.367	-0.8	0.047	0.372
	S(59)	0.8	0.091	0.343	-1.4	0.082	0.332	-2.2	0.063	0.324	-1.1	0.058	0.325
Bimodal	S(65)	4.4	0.108	0.305	-3.8	0.099	0.296	-0.5	0.073	0.292	-3.7	0.069	0.289
	S(69)	11.5	0.112	0.315	-5.4	0.100	0.339	4.9	0.075	0.311	-5.7	0.070	0.330
	S(75)	38.7	0.109	0.363	10.1	0.096	0.395	28.9	0.071	0.365	8.2	0.067	0.389
	S(80)	93.2	0.102	0.384	65.5	0.095	0.402	79.6	0.066	0.385	60.9	0.065	0.403
	S(81)	110.4	0.100	0.389	84.8	0.094	0.415	95.7	0.065	0.393	79.4	0.065	0.411
	S(39)	0.3	0.018	0.523	-0.5	0.015	0.517	-0.2	0.017	0.520	-0.5	0.012	0.517
	S(49)	0.6	0.047	0.447	-1.4	0.040	0.433	-1.1	0.040	0.436	-1.5	0.029	0.433
	S(55)	1.0	0.072	0.385	-2.1	0.061	0.367	-2.4	0.057	0.366	-2.3	0.043	0.362
	S(59)	1.6	0.090	0.336	-2.9	0.075	0.324	-3.4	0.068	0.314	-3.3	0.053	0.317
Skew.	S(65)	5.3	0.109	0.311	-6.2	0.089	0.303	-3.1	0.078	0.300	-7.2	0.063	0.297
	S(69)	12.4	0.114	0.317	-8.5	0.089	0.334	1.2	0.079	0.319	-10.1	0.063	0.329
	S(75)	38.6	0.109	0.356	4.4	0.085	0.399	22.6	0.073	0.366	1.5	0.059	0.398
	S(80)	89.9	0.098	0.389	53.6	0.083	0.417	69.4	0.067	0.400	48.9	0.057	0.453
	S(81)	105.7	0.095	0.390	70.9	0.083	0.418	84.4	0.066	0.403	65.6	0.057	0.455