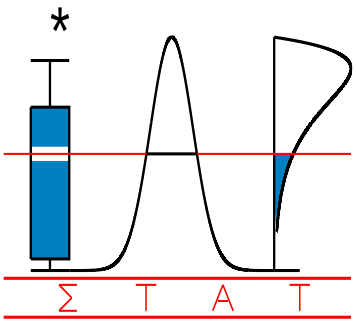
T E C H N I C A L

R E P O R T

0644

FITTING FRAITLY MODELS VIA LINEAR MIXED MODELS USING MODEL TRANSFORMATION

MASSONNET G., JANSSEN P., and T. BURZYKOWSKI



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Fitting Frailty Models via Linear Mixed Models using Model Transformation

Goele Massonnet *, Paul Janssen, Tomasz Burzykowski

Hasselt University, Center for Statistics, Agoralaan, Diepenbeek, Belgium.

Abstract

Frailty models are widely used to model clustered survival data. Classical ways to fit frailty models are likelihood based. We propose an alternative approach in which the original problem of 'fitting a frailty model' is reformulated into the problem of 'fitting a linear mixed model' using model transformation. Based on a simulation study, we show that the proposed method provides a good and simple alternative for fitting frailty models for data sets with a sufficiently large number of clusters and moderate to large sample sizes within covariate level subgroups in the clusters. We illustrate the proposed method using data from 27 randomized trials in advanced colorectal cancer.

Keywords: Frailty model, random treatment effect, model transformation, linear mixed model

1 Introduction

Frailty models are widely used to fit clustered survival data. Classical ways to fit frailty models are likelihood based. Data from multicenter clinical trials are a typical example of clustered data; data within the same center all share the same random cluster effect. The shared frailty model provides an appropriate way to describe the within cluster dependence of outcomes.

*Correspondance: Goele Massonnet, Hasselt University, Center for Statistics, Agoralaan, Building D, B-3590 Diepenbeek, Belgium; Tel: +32-11-268244; Fax:+32-11-268299; E-mail: goele.massonnet@uhasselt.be.

Likelihood methods to fit shared frailty models include: EM-algorithm (Klein, 1992), penalized partial likelihood (Therneau and Grambsch, 2000; McGilchrist, 1993), Bayesian analysis (Ducrocq and Casella, 1996). In recent papers more complex frailty models have been studied. Within the clinical trials context typical examples are frailty models with a random cluster effect and a random treatment effect. To fit such frailty models, the likelihood based methods mentioned above have been adapted to cover this extra complexity in the data: EM algorithm (Vaida and Xu, 2000; Cortiñas and Burzykowski, 2005), penalized partial likelihood (Ripatti and Palmgren, 2000), Bayesian approach (Legrand et al., 2005).

In this paper we propose an alternative way to fit frailty models. We start from the following observation: the integral of the weighted (over time) conditional cumulative loghazard depends in a linear way on the random effects describing the cluster and/or the treatment effect over clusters. Using the data within a cluster we can estimate the integral using nonparametric estimation techniques. Considering the estimated integral as a response we can reformulate the original problem of 'fitting a frailty model' into a standard problem of 'fitting a linear mixed-effects model'. We can summarize the idea as follows: based on the original data we obtain pseudo-data (the estimated integrals) on which we can apply mixed model methodology. Since most standard statistical packages contain procedures to fit complex linear mixed-effects models but do often not contain procedures to fit complex frailty models, the proposed model transformation is a useful practical way to get insight in the heterogeneity in the data. A related reference dealing with model transformation in the context of proportional hazards, additive risks and proportional odds models is Grigoletto and Akritas (1999).

In Section 2.2 we give, for right censored clustered survival data, the details on how frailty models can be transformed into mixed-effects models. The simulation study in Section 3 illustrates that the proposed method provides a good and simple alternative for fitting frailty models for data sets with a sufficiently large number of clusters and moderate to large sample sizes within covariate level subgroups in the clusters. In Section 4 we discuss the performance of the method for a colorectal cancer data set. We finally present some remarks and discuss possible further extensions in Section 5.

2 From frailty model to linear mixed-effects model

2.1 Model formulation

Assume we have a total of N patients that come from K different centers, center i having n_i patients ($N = \sum_{i=1}^{K} n_i$). Each patient is observed from a time zero to a failure time T_{ij}^0 or to a potential right censoring time C_{ij} independent of T_{ij}^0. Let $T_{ij} = \min(T_{ij}^0, C_{ij})$ be the observed time and δ_{ij} be the censoring indicator which is equal to 1 if $T_{ij} = T_{ij}^0$ and 0 otherwise. For each patient, we also have the binary variable x_{ij} representing the treatment to which the patient has been randomized with $x_{ij} = -1$ if the patient is in the control group and $x_{ij} = 1$ if the patient is in the experimental group.

We consider a Cox proportional hazards model including a fixed treatment effect, a random center effect and a random treatment effect: the conditional hazard for the jth patient in the ith center is then given by

$$\lambda_{ij}(t) = \lambda_0(t) \exp\left(b_{0i} + (\beta + b_{1i})x_{ij}\right), \tag{1}$$

where $\lambda_0(t)$ represents the unspecified baseline hazard at time t, β is the fixed overall treatment effect, b_{0i} is the random center effect (contributing the factor $\exp(b_{0i})$ to the hazard) and b_{1i} is the random treatment effect providing information on how the treatment effect within center i deviates from the overall treatment effect captured by the regression coefficient β. The random effects b_{0i} and b_{1i} are assumed to follow zero-mean normal distributions. The variance-covariance matrix of the vector of random effects $\mathbf{b}^T = (b_{01}, b_{11}, \ldots, b_{0i}, b_{1i}, \ldots, b_{0K}, b_{1K})$ takes the form

$$\mathbf{G} = \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \otimes \mathbf{I}_K, \tag{2}$$

where $\otimes$ is the Kronecker product. The variance components σ_0^2 and σ_1^2 can be interpreted as a measure of center and treatment effect over centers heterogeneity of the hazard; σ_{01} is the covariance between the two random effects within a center.

In absence of a random treatment effect model (1) reduces to the shared frailty model

$$\lambda_i(t) = \lambda_0(t) \exp(b_{0i} + \beta x_{ij}) = \lambda_0(t) u_i \exp(\beta x_{ij}), \tag{3}$$

where $u_i = \exp(b_{0i})$ is termed the frailty for center i. In absence of covariates this model further simplifies to

$$\lambda_i(t) = \lambda_0(t)\exp(b_{0i}) = \lambda_0(t)u_i. \tag{4}$$

In (3) and (4) b_{0i}, $i = 1, \ldots, K$, is a sample from a zero-mean normal density with variance σ_0^2, describing the heterogeneity between centers.

2.2 The transformation

With $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)ds$ the cumulative hazard for the jth patient in center i, $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$, and $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, we easily obtain from (1) that

$$\ln \Lambda_{ij}(t) = \ln \Lambda_0(t) + b_{0i} + (\beta + b_{1i})x_{ij}. \tag{5}$$

Let $w(.)$ be a weight function $\left(W(t) = \int_0^t w(s)ds\right)$ satisfying $w(s) \geq 0, s \in [0, \infty)$ and $\int_0^\infty w(s)ds = 1$. Integrating both sides in (5) with respect to the weight function we obtain

$$\Omega_{ij} = \int_0^\infty \ln \Lambda_{ij}(t)dW(t) \quad = \quad \alpha + b_{0i} + (\beta + b_{1i})x_{ij},$$

with $\alpha = \int_0^\infty \ln \Lambda_0(t)dW(t)$. Since the patients in center i are divided, by the binary covariate x_{ij}, in a control and a treatment group we have that $\Omega_{i0} = \alpha + b_{0i} - (\beta + b_{1i})$ (control) and $\Omega_{i1} = \alpha + b_{0i} + (\beta + b_{1i})$ (treated). We also have that, for $k = 0, 1$,

$$\Omega_{ik} = \int_0^\infty \ln \Lambda_{ik}(t)dW(t),$$

with $\Lambda_{i0}(.)$, respectively $\Lambda_{i1}(.)$, the cumulative hazard function shared by all control, resp. treated, patients in group i. Following the ideas of Grigoletto and Akritas (1999), pseudo observations for the Ω_{ik}'s can be obtained as

$$\hat{\Omega}_{ik} = \int_0^\infty \ln \hat{\Lambda}_{ik}(t)dW(t),$$

where $\hat{\Lambda}_{ik}(.)$ is the estimated cumulative hazard based on the observations (T_{ij}, δ_{ij}) for all patients in center i with, for $k = 0$, $x_{ij} = -1$ and, for $k = 1$, $x_{ij} = 1$. As concrete estimator we use $\hat{\Lambda}_{ik}(t) = -\ln \hat{S}_{ik}(t)$ with $\hat{S}_{i0}(t)$ the Kaplan-Meier estimator for the control group ($x_{ij} = -1$):

$$\hat{S}_{i0}(t) = \prod_{j:T_{ij} \leq t, x_{ij} = -1} \left(\frac{r(T_{ij}) - d(T_{ij})}{r(T_{ij})}\right),$$

with $r(v)$ the number still at risk at time v and $d(v)$ the number of events at time v and with $\hat{S}_{i1}(t)$ the Kaplan-Meier estimator for the experimental group ($x_{ij} = 1$).
In terms of the pseudo observations we now can propose the model

$$\hat{\Omega}_{ik} = \alpha + b_{0i} + (\beta + b_{1i})x_{ik} + (\hat{\Omega}_{ik} - \Omega_{ik}) = \alpha + b_{0i} + (\beta + b_{1i})x_{ik} + e_{ik} \tag{6}$$

with $x_{i0} = -1$ and $x_{i1} = 1$. As $e_{ik} = \hat{\Omega}_{ik} - \Omega_{ik}$ it is clear that the random error terms do not satisfy the homogeneity assumption (because different subclusters have different sample sizes). In Section 2.3 we explain how to account for this heterogeneity when mixed models software is used to fit the model. A further remark is that for the special case (4) we obtain the following model after transformation:

$$\hat{\Omega}_i = \alpha + b_{0i} + \left(\hat{\Omega}_i - \Omega_i\right) = \alpha + b_{0i} + e_i. \tag{7}$$

For this one-way random effects model we only have one observation per center. At first glance this leads to identifiability problems. We, however, do have estimators of the variances of the error terms so that estimation of the variance components associated with the random center effect is possible. More details on this is given in Section 2.3.

2.3 The error variance

In this section, we provide estimates for the variances $\sigma^2_{e,ik}$ of the random error terms in model (6) and for the variances $\sigma^2_{e,i}$ of the random error terms in (7) (see the appendix for the technical details). First, we will consider the general model (6). The patients of center i are divided in two groups: the control group ($k = 0$) and the treatment group ($k = 1$). Note that n_{ik} is the number of patients in group k of center i. Define a uniform weight function W on the interval (A, B), where A and B are chosen such that the logarithm of the cumulative hazard can be estimated for $t \in (A, B)$ for the control and the treatment group in each center. The variance of the error term $\hat{\Omega}_{ik} - \Omega_{ik}$ ($i = 1, \ldots, K$, $k = 0, 1$) can then be estimated by

$$\begin{aligned}
\hat{\sigma}^2_{e,ik} &= \hat{\mathrm{Var}}\left(\hat{\Omega}_{ik} - \Omega_{ik}\right) \\
&= \frac{1}{n_{ik}^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_{ik}(s)}\int_A^s \frac{1}{\hat{\Lambda}_{ik}(t)} \sum_{j:\ x_{ij}=k} \frac{I\left(0 \le t_{ij} \le t, \delta_{ij} = 1\right)}{\left(1 - \frac{1}{n_{ik}}\sum_{j:\ x_{ij}=k} I\left(T_{ij} < t_{ij}\right)\right)^2}\, dt\, ds \\
&\quad + \frac{1}{n_{ik}^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_{ik}(s)} \sum_{j:\ x_{ij}=k} \frac{I\left(0 \le t_{ij} \le s, \delta_{ij} = 1\right)}{\left(1 - \frac{1}{n_{ik}}\sum_{j:\ x_{ij}=k} I\left(T_{ij} < t_{ij}\right)\right)^2}\int_s^B \frac{1}{\hat{\Lambda}_{ik}(t)}\, dt\, ds.
\end{aligned}$$

For model (7), we obtain in a similar way the estimated variance of the error term $\hat{\Omega}_i - \Omega_i$ $(i = 1, \ldots, K)$:

$$\begin{aligned}
\hat{\sigma}^2_{e,i} &= \hat{\mathrm{Var}}\left(\hat{\Omega}_i - \Omega_i\right) \\
&= \frac{1}{n_i^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_i(s)}\int_A^s \frac{1}{\hat{\Lambda}_i(t)}\sum_{j=1}^{n_i}\frac{I\left(0 \le t_{ij} \le t, \delta_{ij} = 1\right)}{\left(1 - \frac{1}{n_i}\sum_{j=1}^{n_i} I\left(T_{ij} < t_{ij}\right)\right)^2}\, dt\, ds \\
&\quad + \frac{1}{n_i^2}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_i(s)}\sum_{j=1}^{n_i}\frac{I\left(0 \le t_{ij} \le s, \delta_{ij} = 1\right)}{\left(1 - \frac{1}{n_i}\sum_{j=1}^{n_i} I\left(T_{ij} < t_{ij}\right)\right)^2}\int_s^B \frac{1}{\hat{\Lambda}_i(t)}\, dt\, ds.
\end{aligned}$$

2.4 Fitting the linear mixed-effects model

To fit the transformed models (6) and (7), we use PROC MIXED in SAS. The mixed-effects model is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}, \tag{8}$$

where $\mathbf{y}$ denotes the vector of dependent variable values, $\boldsymbol{\beta}$ is an unknown vector of fixed effects with known model matrix $\mathbf{X}$, $\boldsymbol{\gamma}$ is an unknown vector of random effects with known model matrix $\mathbf{Z}$, and $\mathbf{e}$ is the random error vector. A key assumption is that $\boldsymbol{\gamma}$ and $\mathbf{e}$ are normally distributed with

$$\begin{aligned}
\mathrm{E}\begin{pmatrix}\boldsymbol{\gamma}\\ \mathbf{e}\end{pmatrix} &= \begin{pmatrix}0\\0\end{pmatrix} \\
\mathcal{D}\begin{pmatrix}\boldsymbol{\gamma}\\ \mathbf{e}\end{pmatrix} &= \begin{pmatrix}\mathbf{G} & 0\\ 0 & \mathbf{R}\end{pmatrix}.
\end{aligned}$$

The variance-covariance matrix of $\mathbf{y}$ is therefore $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$. To estimate the variance-covariance components in model (8), PROC MIXED implements two likelihood-based methods: maximum likelihood (ML) and restricted/residual likelihood (REML). We will consider the REML method. The corresponding log likelihood function is:

$$l_R(\mathbf{G}, \mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{r}^T\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\log 2\pi, \tag{9}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of $\mathbf{X}$. PROC MIXED minimizes $-2\, l_R(\mathbf{G}, \mathbf{R})$ over all unknown parameters using a ridge-stabilized Newton-Raphson algorithm. For model (7), $\mathbf{y}^T = \left(\hat{\Omega}_1, \hat{\Omega}_2, \ldots, \hat{\Omega}_K\right)$, $\mathbf{X} = \mathbf{1}_K$, $\boldsymbol{\gamma}^T = (b_{01}, b_{02}, \ldots, b_{0K})$, $\mathbf{G} = \sigma_0^2\mathbf{I}_K$ and $\mathbf{R} = \mathrm{diag}(\sigma^2_{e,1}, \ldots, \sigma^2_{e,K})$. As already mentioned, we only have one observation per level in

model (7). To be able to estimate σ_0^2, we first estimate $\sigma_{e,1}^2, \ldots, \sigma_{e,K}^2$ as explained in Section 2.3. In the PARMS statement of PROC MIXED, initial values for the covariance parameters can be specified. We choose an arbitrary initial value for σ_0^2. The initial values for the error variances are chosen to be $\hat{\sigma}_{e,1}^2, \ldots, \hat{\sigma}_{e,K}^2$. By using the option EQCONS, the initial residual variances will be held constant during the estimation procedure. Maximization of (9) over σ_0^2 gives an estimate for the heterogeneity σ_0^2. The following SAS program fits model (7) to the pseudo data for 20 clusters:

```
proc mixed data=pseudodata;
class cluster;
model omegaihat= ;
random cluster;
repeated/group=cluster;
parms /parmsdata=parmsdataset eqcons= 2 to 21;
run;
```

where *parmsdataset* is a SAS data set that contains the initial values for σ_0^2, $\sigma_{e,1}^2, \ldots, \sigma_{e,K}^2$.
For model (6), $\mathbf{y}^T = \left(\hat{\Omega}_{10}, \hat{\Omega}_{11}, \hat{\Omega}_{20}, \hat{\Omega}_{21}, \ldots, \hat{\Omega}_{K0}, \hat{\Omega}_{K1}\right)$, $\mathbf{X}$ is the model matrix that contains the x_{ik}'s ($i = 1, \ldots, K$ and $k = 0, 1$), $\boldsymbol{\gamma}^{\boldsymbol{T}} = \mathbf{b}^T$ and $\mathbf{G}$ is as defined in (2). Further, $\mathbf{R} = \text{diag}(\sigma_{e,10}^2, \sigma_{e,11}^2, \ldots, \sigma_{e,K0}^2, \sigma_{e,K1}^2)$. The error covariance matrix $\mathbf{R}$ can be estimated as explained in Section 2.3. By maximizing (9) over $\mathbf{G}$ in PROC MIXED while fixing the error variances as described above, we obtain estimates for σ_0^2, σ_1^2 and σ_{01}.
To obtain estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, the mixed model equations are solved (Henderson, 1984). The solutions can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$ and $\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\,\hat{\boldsymbol{\beta}})$.

3 Simulations

We study the performance of the proposed method by using a simulation study. As a simulation model we consider the setting of a multicenter clinical trial. First, we consider the special case in (4) where there is only a random center effect. We compare the results obtained by the proposed method with those obtained by the penalized partial likelihood approach (Therneau and Grambsch, 2000). We use 'coxph' in S-Plus 7.0.6 for the penalized partial likelihood inference. The precision of the parameter estimates is investigated for a varying number of clusters and

number of observations per cluster, the percentage of censored observations, the size of σ_0^2 and the value of the baseline event rate $\lambda_0(t)$ (which we assume constant in time for simplicity). We also discuss the robustness of the proposed method against model misspecification. Next, the general model (1) is considered. For this model, we allow for correlation between b_{0i} and b_{1i}. Also here we compare the results obtained by the proposed method with those based on the penalized partial likelihood approach (Ripatti and Palmgren, 2000) using 'coxme' in S-Plus 7.0.6 for the likelihood inference. We further study the effect of the size of σ_0^2 and σ_1^2 on the precision of the parameter estimates.

3.1 Description of the simulations

We assume a constant sample size per cluster $n_i = n$, for $i = 1, \ldots, K$. For each parameter setting $(K, n, \lambda_0, \sigma_0^2, \sigma_{01}, \sigma_1^2)$, 500 data sets are generated from model (1), assuming a constant baseline hazard. Given a particular parameter setting, observations for a particular data set are generated in the following way. First, K random center effects $b_{01}, b_{02}, \ldots, b_{0K}$ and K random treatment effects $b_{11}, b_{12}, \ldots, b_{1K}$ are generated from a normal distribution with mean 0 and variance-covariance matrix $\mathbf{G}$, as in (2). The time to event for each patient is randomly generated from an exponential distribution with parameter $\lambda_{ij}(t) = \lambda_0 \exp(b_{0i} + (\beta + b_{1i})x_{ij})$, where x_{ij} is generated from a Bernoulli distribution with success probability 0.5. The censoring time for each patient is randomly generated from a uniform distribution, so that approximately 30% censoring is obtained. For each data set, pseudo-data $\hat{\Omega}_{ik}$ are generated through the model transformation described in Section 2.2 by using a uniform weight function $w(.)$ on the interval (A, B), chosen so that $0 < \hat{S}_{ik}(t) < 1$ for $t \in (A, B)$. For each cluster i, the estimated variance of $\hat{\Omega}_{ik} - \Omega_{ik}$ is computed as explained in Section 2.3. To fit model (6), we use the SAS procedure PROC MIXED as explained in Section 2.4. For each data set we obtain an estimate for β, σ_0^2, σ_1^2 and σ_{01}. We consider how the parameter settings compare the precision of $\hat{\beta}$, $\hat{\sigma}_0^2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_{01}$ for the different parameter settings.

For the special case of model (4), the data are generated as explained above with $\beta = 0$, $\sigma_1^2 = 0$ and $\sigma_{01} = 0$. In this case, we consider two censoring settings: moderate censoring (around 30%) and heavy censoring (around 60%).

To study the robustness of the proposed method against model misspecification, the data are generated assuming that the frailties $u_1 = \exp(b_{01}), \ldots, u_K = \exp(b_{0K})$ are gamma distributed

with mean $E(U_i) = e^{\left(\frac{1}{2}\sigma_0^2\right)}$ and variance $Var(U_i) = e^{\sigma_0^2}\left(e^{\sigma_0^2} - 1\right)$. This corresponds to random effects b_{0i} with mean 0 and variance σ_0^2. For each data set, pseudo data $\hat{\Omega}_i$ are generated as explained above. We fit model (4) assuming, incorrectly, that the random effects b_{0i} are normally distributed with mean 0 and variance σ_0^2.

3.2 The choice of the parameters

3.2.1 Frailty model with a random center effect

For the concrete simulation, we take 20, 50, and 100 centers with 50 or 100 patients per center. The parameter values λ_0 and σ_0^2 in both settings are chosen in such a way that a different magnitude of spread in the median time to event from center to center is induced. The median time to event T_M is the solution of $\exp\left(-\lambda_0 \exp(b_0) T_M\right) = 0.5$, with b_0 zero-mean normally distributed, i.e., $T_M = \frac{\log 2}{\lambda_0 \exp(b_0)}$. The magnitude of spread in the median time to event from center to center was determined by computing the density function of T_M (Figure 1). It is easy to show that the density function $f_{T_M}(t)$ is given by

$$f_{T_M}(t) = \frac{1}{t\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{\left(\log\left(\frac{\log 2}{\lambda_0 t}\right)\right)^2}{2\sigma_0^2}\right).$$

As true values for the event rate, we take $\lambda_0 = 0.1$ and 0.5. The heterogeneity parameter is set at $\sigma_0^2 = 0.08765$ and 0.1577. To obtain these values, we use the relation between σ_0^2 and the frailty variance: $Var(U_i) = \theta = e^{\sigma_0^2}\left(e^{\sigma_0^2} - 1\right)$. The values of σ_0^2 correspond to a frailty variance of $\theta = 0.1$, resp. 0.2.

For the settings (σ_0^2, λ_0) = (0.08765, 0.1) and (0.1577, 0.1), there is much spread in the median time to event over the centers. For the settings (σ_0^2, λ_0) = (0.08765, 0.5) and (0.1577, 0.5), there is little spread in the median time to event over the centers, with a bigger spread for $\sigma_0^2 = 0.1577$. To study the robustness of the proposed method, we take $\sigma_0^2 = 0.3520$ ($\theta = 0.6$) and $\lambda_0 = 0.1$. Figure 2 clearly shows our motivation for choosing $\theta = 0.6$. For $\theta = 0.1$ we have a situation where the gamma and the lognormal density functions are close, whereas for $\theta = 0.6$ these densities are more apart.

3.2.2 Frailty model with random center and treatment effects

We consider a situation with 50 centers and 200 patients per center. The baseline hazard is assumed constant and equal to $\lambda_0 = 0.3$. For the treatment effect, we use $\beta = -0.2$. These parameter values are chosen to have a setting that corresponds to the bladder cancer data considered in Legrand et al. (2005). In this case study, Legrand et al. investigated heterogeneity in disease free interval due to center and treatment effect over centers in a large bladder cancer database including data from seven randomized clinical trials. We simulate data using different combinations of values of σ_0^2 and σ_1^2, varying from 0 to 0.08 ($\sigma_0^2, \sigma_1^2 = 0, 0.04$ or 0.08). The covariance parameter σ_{01} is chosen such that the correlation between b_0 and b_1 is equal to 0.5. This value mimics the correlation between the random effects observed in the bladder cancer data (Legrand et al., 2005).

3.3 Simulation results

3.3.1 Frailty model with a random center effect

Tables 1, 2, 3 and 4 present, for each parameter setting, the relative bias, the mean, the empirical standard deviation, the median and the interquartile range computed for the 500 estimates of the variance of the random center effect. One can observe that σ_0^2 is estimated well by the proposed method if the cluster size is large enough (i.e., $n = n_i = 100$). Both for the penalized partial likelihood approach and the proposed method, the absolute relative bias decreases with the increasing cluster size, and is not substantially influenced by the number of clusters. In general, the estimates obtained by the proposed approach are on average closer to the true value σ_0^2 if the cluster size is large enough (i.e., $n = n_i = 100$). For a smaller cluster size ($n = n_i = 50$), the estimates obtained by the penalized partial likelihood are more precise. In general, the absolute relative bias increases if the amount of censoring increases. However, if the cluster size is large enough, σ_0^2 is estimated well by the proposed method. By increasing θ from 0.1 to 0.2, the bias decreases for both methods when there is 30% censoring, except for the penalized partial likelihood when $\theta = 0.2$, $\lambda_0 = 0.5$ and $(K, n) = (20, 50)$.

Table 5 shows the results obtained by the penalized partial likelihood approach and the proposed method if the 'true' frailties are gamma distributed with variance 0.6. The results illustrate

that, for both methods, the point estimates of σ_0^2 are biased if the model is misspecified. This lack of robustness is also discussed in the bootstrap context by Massonnet et al. (2006). It clearly shows the need for lack-of-fit measures for frailty models.

3.3.2 Frailty model with random center and treatment effects

In Table 6 we report, for the setting described in Section 3.2.2, the mean, the empirical standard deviation, the median and the interquartile range computed over the 500 estimates of the fixed treatment effect and the variance-covariance components of the random effects. We compare the results obtained by the proposed method with those obtained by coxme in S-Plus 7.0.6. The parameter β is in general estimated well by both methods. The bias of the fixed effect estimates obtained by coxme is in general a bit smaller than for the proposed method. The empirical variability of estimates of β is similar for both methods. The estimates of σ_0^2, σ_1^2 and σ_{01} for both methods are on average comparable. The estimates produced by coxme have in general the smallest empirical variability.

4 Case study

As an example we will analyze the data from 27 advanced colorectal cancer trials (Advanced Colorectal Cancer Meta-Analysis Project, 1992, 1994; Meta-Analysis Group in Cancer, 1996, 1998). This study is described in Burzykowski et al. (2004). In the four meta-analyses, the comparison was between an experimental treatment and a control treatment. In total there are 4007 patients, 1871 (46.7 %) in the control group and 2136 (53.3 %) in the experimental group. The number of patients per trial varies from 15 to 382 patients (the mean (median) number of patients per trial is 149 (148)). Our analysis is based on the survival time, defined as the time from randomization to death from any cause. Most patients have died (3591 out of 4007 patients, i.e., 89.6 %). First, we fit model (1) including a fixed treatment effect, a random trial effect and a random treatment effect. In this model, we also take into account a possible correlation between the two random effects within a trial. The parameter estimates, obtained by the proposed method and by the penalized partial likelihood approach (coxme in S-Plus 7.0.6), are presented in Table 7. The point estimates for σ_1^2 and σ_{01} are very small (almost

zero). For this reason, we fit the shared frailty model (3) including a fixed treatment effect and a random trial effect. The results are shown in Table 8. The estimates obtained by the penalized partial likelihood and the transformation method are a bit different. However, the plot of the density function of the median time to event in the control group over trials (Figure 3) shows that the shape of this density function is the same. A possible explanation for the difference between the estimates obtained by both methods, is that only 16 out of 27 trials have sample sizes of both the treatment and the control group larger than 50 patients. From the simulations, we know that the accuracy of the transformation method is comparable to the penalized partial likelihood if the cluster sizes are large enough.

We performed simulations to evaluate the performance of the proposed method in a setting similar to the setting of the advanced colorectal cancer data set where the heterogeneity due to treatment effect over centers is larger. To simulate data, values of the baseline hazard $\lambda_0(t)$ (assumed to be constant over time $\lambda_0(t) = \lambda_0$), of the treatment effect β and of the variance component σ_0^2 are chosen to resemble the colorectal cancer data set. We also consider the same number of trials and the same number of patients in the control and the treatment group in each trial. We use a constant baseline hazard $\lambda_0 = 0.84$, a treatment effect of $\beta = -0.0533$ and $\sigma_0^2 = 0.04$. The estimate for the cumulative baseline hazard in the conditional model, estimated by the Breslow estimator given in Duchateau et al. (2002), supports the assumption of a constant baseline hazard (figure not shown). Further we choose $\sigma_1^2 = 0.08$ and $\sigma_{01} = 0.0283$ (which means a correlation for b_0 and b_1 of 0.5 as in the simulations in section 3), so that the values of σ_1^2 and σ_{01} are different from zero. The censoring time for each patient is generated from a uniform distribution so that approximately 11% censoring is obtained, as in the colorectal cancer data. For this parameter setting, 500 data sets are generated from model (1), assuming a constant baseline hazard. For each simulated data set, model (1) is fitted using the proposed method and the penalized partial likelihood. In Table 9, we report the mean, the empirical standard deviation, the median and the interquartile range computed over the 500 estimates of the treatment effect and the variance components. The bias of the fixed effect estimates obtained by the proposed method is on average smaller than for the penalized partial likelihood approach. The empirical variability of the estimates of β is similar for both methods. The estimates of the variances of the random effects obtained by the proposed method are on average only slightly different from those obtained by the penalized partial likelihood approach. The empirical variability of the estimates obtained by coxme is a bit smaller.

5 Conclusions

In this paper, an alternative approach to fit frailty models is proposed. In this approach the original problem of 'fitting a frailty model' is reformulated into a standard problem of 'fitting a linear mixed-effects model'. For this purpose, the integral of the weighted (over time) conditional cumulative loghazard is considered. This integral depends in a linear way on the random effects describing the cluster and/or the treatment effect over clusters. Using the data within a cluster, the integral can be estimated using nonparametric estimation techniques. Considering the estimated integrals as a response, linear mixed-effects model methodology can be applied. Since most standard statistical packages contain procedures to fit complex linear mixed-effects models but do often not contain procedures to fit complex frailty models, the proposed model transformation is a useful practical way to get insight in the heterogeneity in clustered data. The performance of the proposed method was studied by simulation. The results obtained by the transformation method were compared with those obtained by an algorithm for the penalized partial likelihood that is available in S-Plus 7.0.6. The results indicate a good performance of the proposed method for data sets with a sufficiently large number of clusters and moderate to large sample sizes within covariate level subgroups in the clusters.

The transformation method was illustrated using data from advanced colorectal cancer trials. The analysis shows that the heterogeneity due to treatment effect over centers and the covariance between both random effects is very small (almost zero). We also performed simulations to evaluate the performance of the proposed method in a setting similar to the colorectal cancer data where the heterogeneity due to treatment effect over centers is larger. A general conclusion is that the estimates based on the transformation idea and the estimates obtained from the penalized partial likelihood approach are similar.

We considered a frailty model with a binary covariate and we therefore could use the Kaplan-Meier estimator for the survival function. A possible alternative is to use the Nelson-Aalen estimator to estimate the cumulative hazard in a direct way. It also would be of interest to extend the transformation idea to frailty models with a continuous covariate. We then need the Beran estimator to estimate the survival function (Beran, 1981). From the above discussion it is also clear that the transformation method is useful for censoring schemes that are different from the right censoring scheme discussed so far. Indeed, the transformation idea can be used for any censoring scheme for which a nonparametric estimator for the cumulative hazard or the

survival function is available (e.g., for interval-censored data, Lindsey and Ryan (1998)). These topics will be subjects for further research.

Acknowledgment

The authors are grateful to the Meta-Analysis Group in Cancer (MAGIC) for the permission to use their data. The authors gratefully acknowledge the financial support from the IAP research network nr P5/24 of the Belgian Government (Belgian Science Policy).

Appendix

To apply the method proposed in Section 2.2 we need estimated values for the error variance. To obtain estimates we can rely on an asymptotic representation, proposed by Lo and Singh (1986), decomposing $\hat{F}_{ik}(t) - F_{ik}(t)$ as an average of i.i.d. terms and a lower order remainder term $r_{ik}(t)$, where F_{ik} is the continuous failure time distribution function for subjects in center i with $x_{ij} = k$.

Let G be the censoring distribution, $1 - H_{ik}(s) = (1 - F_{ik}(s))(1 - G(s))$ and $H^u_{ik}(s) = P\left(T_{ij} \leq s, \delta_{ij} = 1 | x_{ij} = k\right) = \int_0^s (1 - G(y^-))dF_{ik}(y)$. It follows from Lo and Singh (1986) that

$$\hat{F}_{ik}(t) - F_{ik}(t) = \frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} \xi_{ik}(T_{ij}, \delta_{ij}, t) + r_{ik}(t),$$

where

$$\xi_{ik}(T_{ij}, \delta_{ij}, t) = \frac{I\left(T_{ij} \leq t, \delta_{ij} = 1\right)}{1 - H_{ik}(T_{ij})} - \int_0^t \frac{I\left(T_{ij} > s\right)}{\left(1 - H_{ik}(s)\right)^2}\, dH^u_{ik}(s),$$

for a subject in cluster i with observed information (T_{ij}, δ_{ij}) and $x_{ij} = k$.

By using the relationship $\Lambda_{ik}(t) = -\ln\left(1 - F_{ik}(t)\right)$ and first order Taylor expansions, we obtain

$$\begin{aligned} & \ln \hat{\Lambda}_{ik}(t) - \ln \Lambda_{ik}(t) \\ \cong\ & \frac{1}{\Lambda_{ik}(t)} \frac{1}{\left(1 - F_{ik}\left(t\right)\right)} \left(\hat{F}_{ik}(t) - F_{ik}(t)\right) \\ \cong\ & \frac{1}{\Lambda_{ik}(t)} \frac{1}{S_{ik}(t)} \frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} \xi_{ik}(T_{ij}, \delta_{ij}, t) \end{aligned}$$

Integrating both sides with respect to $w(.)$ gives

$$\begin{aligned}
& \hat{\Omega}_{ik} - \Omega_{ik} \\
= & \int_0^\infty \ln \hat{\Lambda}_{ik}(t) dW(t) - \int_0^\infty \ln \Lambda_{ik}(t) dW(t) \\
\cong & \frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} \int_0^\infty \frac{\xi_{ik}(T_{ij}, \delta_{ij}, t)}{\Lambda_{ik}(t) S_{ik}(t)} dW(t) \\
= & \frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} \eta_{ik}(T_{ij}, \delta_{ij}),
\end{aligned}$$

with $\eta_{ik}(T_{ij}, \delta_{ij}) = \int_0^\infty \frac{\xi_{ik}(T_{ij}, \delta_{ij}, t)}{\Lambda_{ik}(t) S_{ik}(t)} dW(t)$.

Noting that the function $\xi_{ik}(T_{ij}, \delta_{ij}, t)$ is a conditional version (conditioned on the center i and the subgroup with $x_{ij} = k$) of the function ξ in Lo and Singh (1986), it follows that for a subject with observed information (T_{ij}, δ_{ij}) in the subgroup with $x_{ij} = k$:

$$\begin{aligned}
E\left[\xi_{ik}(T_{ij}, \delta_{ij}, t)\right] &= 0 \\
\mathrm{Cov}\left[\xi_{ik}(T_{ij}, \delta_{ij}, t), \xi_{ik}(T_{ij}, \delta_{ij}, s)\right] &= (1 - F_{ik}(t))(1 - F_{ik}(s)) \int_0^{t \wedge s} \frac{dH^u_{ik}(y)}{(1 - H_{ik}(y))^2}.
\end{aligned}$$

Assume that subject l in cluster i is in the subgroup with $x_{il} = k$. The asymptotic variance of the error terms $\hat{\Omega}_{ik} - \Omega_{ik}$ is given by

$$\begin{aligned}
\sigma^2_{e,ik} &= \mathrm{Var}(\hat{\Omega}_{ik} - \Omega_{ik}) \\
&= \mathrm{Var}(\frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} \eta_{ik}(T_{ij}, \delta_{ij})) \\
&= \frac{1}{n_{ik}} \mathrm{Var}(\eta(T_{il}, \delta_{il})) \\
&= \frac{1}{n_{ik}} \mathrm{E}\left\{ \int_0^\infty \frac{\xi_{ik}(T_{il}, \delta_{il}, t)}{\Lambda_{ik}(t) S_{ik}(t)} dW(t) \int_0^\infty \frac{\xi_{ik}(T_{il}, \delta_{il}, s)}{\Lambda_{ik}(s) S_{ik}(s)} dW(s) \right\} \\
&= \frac{1}{n_{ik}} \int_0^\infty \int_0^\infty \frac{1}{\Lambda_{ik}(t) S_{ik}(t) \Lambda_{ik}(s) S_{ik}(s)} \mathrm{Cov}\left(\xi_{ik}(T_{il}, \delta_{il}, t), \xi(T_{ij}, \delta_{ij}, s)\right) dW(t) dW(s) \\
&= \frac{1}{n_{ik}} \int_0^\infty \int_0^\infty \frac{1}{\Lambda_{ik}(t) \Lambda_{ik}(s)} \int_0^{s \wedge t} \frac{dH^u_{ik}(y)}{(1 - H_{ik}(y))^2} dW(t) dW(s).
\end{aligned}$$

Let $w(.)$ be a uniform weight function on the interval (A, B), where A and B are chosen such that for each center the logarithm of the cumulative hazard can be estimated well for $t \in (A, B)$

for the control and the treatment group in each center. Then

$$\begin{aligned}\sigma^2_{e,ik} &= \mathrm{Var}\left(\hat{\Omega}_{ik} - \Omega_{ik}\right) \\ &= \frac{1}{n_{ik}}\frac{1}{(B-A)^2}\int_A^B\int_A^s \frac{1}{\Lambda_{ik}(t)\Lambda_{ik}(s)}\int_0^t \frac{dH^u_{ik}(y)}{\left(1-H_{ik}(y^-)\right)^2}dtds \\ &\quad + \frac{1}{n_{ik}}\frac{1}{(B-A)^2}\int_A^B\int_s^B \frac{1}{\Lambda_{ik}(t)\Lambda_{ik}(s)}\int_0^s \frac{dH^u_{ik}(y)}{\left(1-H_{ik}(y^-)\right)^2}dtds.\end{aligned}$$

To obtain an estimate of the asymptotic error variance, we replace $H_{ik}(y^-)$ and $H^u_{ik}(y)$ by the following empirical estimators:

$$\begin{aligned}H^u_{ik}(y) &= \frac{1}{n_{ik}}\sum_{j:\ x_{ij}=k} I\left(T_{ij}\le y, \delta_{ij}=1\right) \\ H_{ik}(y) &= \frac{1}{n_{ik}}\sum_{j:\ x_{ij}=k} I\left(T_{ij}< y\right).\end{aligned}$$

This gives the following estimated variances of the error terms:

$$\begin{aligned}\hat{\sigma}^2_{e,ik} &= \hat{\mathrm{Var}}\left(\hat{\Omega}_{ik} - \Omega_{ik}\right) \\ &= \frac{1}{n^2_{ik}}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_{ik}(s)}\int_A^s \frac{1}{\hat{\Lambda}_{ik}(t)}\sum_{j:\ x_{ij}=k}\frac{I\left(0\le t_{ij}\le t, \delta_{ij}=1\right)}{\left(1-\frac{1}{n_{ik}}\sum_{j:\ x_{ij}=k} I\left(T_{ij}<t_{ij}\right)\right)^2}\,dt\,ds \\ &\quad + \frac{1}{n^2_{ik}}\frac{1}{(B-A)^2}\int_A^B \frac{1}{\hat{\Lambda}_{ik}(s)}\sum_{j:\ x_{ij}=k}\frac{I\left(0\le t_{ij}\le s, \delta_{ij}=1\right)}{\left(1-\frac{1}{n_{ik}}\sum_{j:\ x_{ij}=k} I\left(T_{ij}<t_{ij}\right)\right)^2}\int_s^B \frac{1}{\hat{\Lambda}_{ik}(t)}\,dt\,ds.\end{aligned}$$

References

R. Beran, "Nonparametric regression with randomly censored survival data." Technical Report, University of California, Berkeley, 1981.

T. Burzykowski, G. Molenberghs and M. Buyse, "The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer," Journal of the Royal Statistical Society A vol. 167 pp. 103-124, 2004.

J. Cortiñas Abrahantes and T. Burzykowski, "A version of the EM algorithm for proportional hazards model with random effects," Biometrical Journal vol. 47 pp. 847-862, 2005.

L. Duchateau, P. Janssen, P. Lindsey, C. Legrand, R. Nguti and R. Sylvester, "The shared frailty model and the power for heterogeneity tests in multicenter trials," Computational Statistics and Data Analysis vol. 40 pp.603-620, 2002.

V. Ducrocq and G. Casella, "A Bayesian analysis of mixed survival models," Genetics Selection Evolution vol. 28 pp. 505-529, 1996.

M. Grigoletto and M.G. Akritas, "Analysis of covariance with incomplete data via semiparametric model transformations," Biometrics vol. 55 pp. 1177-1187, 1999.

C.R. Henderson, Applications of linear models in animal breeding, University of Guelph Press: Guelph, Canada, 1984.

J.P. Klein, "Semiparametric estimation of random effects using the Cox model based on the EM algorithm," Biometrics vol. 48 pp. 795-806.

C. Legrand, V. Ducrocq, P. Janssen, R. Sylvester and L. Duchateau, "A Bayesian approach to jointly estimate center and treatment by center heterogeneity in a proportional hazards model," Statistics in Medicine vol. 24 pp. 3789-3804, 2005.

J.C. Lindsey and L.M. Ryan, "Tutorial in Biostatistics: methods for interval-censored data," Statistics in Medicine vol. 17 pp. 219-238, 1998.

S.H. Lo and K. Singh, "The product-limit estimator and the bootstrap: Some asymptotic representations," Probability Theory and Related Fields vol. 71 pp. 455-465, 1986.

G. Massonnet, T. Burzykowski and P. Janssen, "Resampling plans for frailty models," Communications in Statistics - Simulation and Computation vol. 35 pp. 497-514, 2006.

C. McGilchrist, "REML estimation for survival models with frailty," Biometrics vol. 49 pp. 221-225, 1993.

S. Ripatti and J. Palmgren, "Estimation of multivariate frailty models using penalized partial likelihood," Biometrics vol. 56 pp. 1016-1022, 2000.

T.M. Therneau and P.M. Grambsch, Modeling Survival Data, Extending the Cox model, Springer: New York, 2000.

F. Vaida and R. Xu, "Proportional hazards model with random effects," Statistics in Medicine vol. 19, pp. 3309-3324, 2000.

Figure 1: Density function of the median time to event over centers .

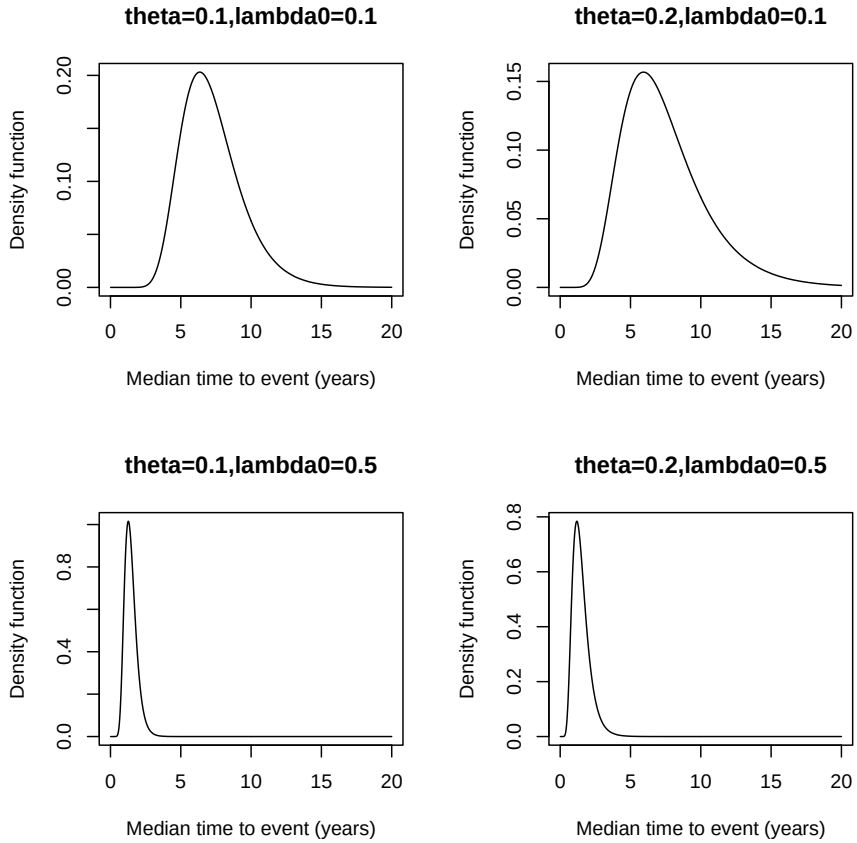


Figure 2: Density function for the lognormal and the gamma distribution .

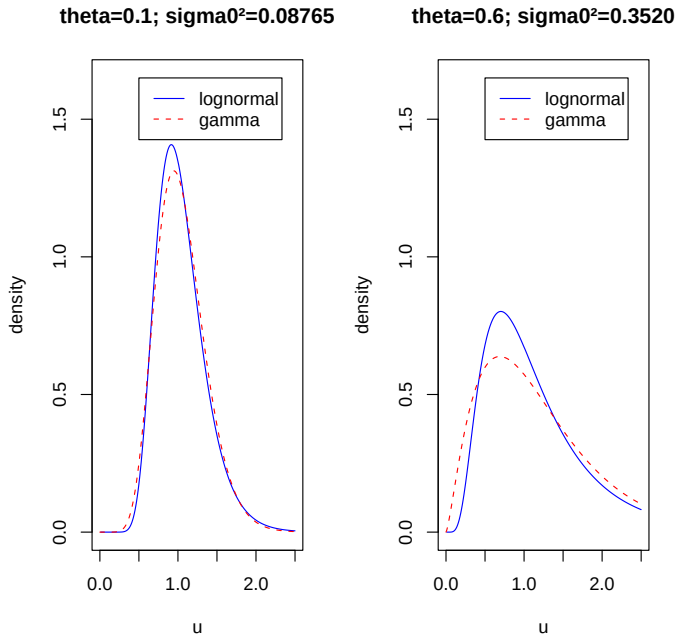


Table 1: Relative bias, mean, empirical standard deviation, median and interquartile range of estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.08765$ ($\theta = 0.1$), $\lambda_0 = 0.5$; first line for coxph, second line for PROC MIXED; left part 30% censoring, right part 60 % censoring.

(K, n)	Rel. bias	Mean	Emp. std	Median	IQR	Rel. bias	Mean	Emp. std	Median	IQR
(100,100)	-0.0314	0.0849	0.0152	0.0850	0.0202	-0.0143	0.0864	0.0169	0.0867	0.0225
	-0.0257	0.0854	0.0148	0.0848	0.0207	-0.0177	0.0861	0.0162	0.0860	0.022
(100,50)	-0.0382	0.0843	0.0174	0.0841	0.0228	-0.0280	0.0852	0.0198	0.0846	0.025
	-0.0975	0.0791	0.0168	0.0780	0.0214	-0.1135	0.0777	0.0197	0.0772	0.0271
(50,100)	-0.0097	0.0868	0.0227	0.0847	0.0297	-0.0234	0.0856	0.0223	0.0852	0.0286
	0.0029	0.0879	0.0223	0.0865	0.0285	-0.0177	0.0861	0.0225	0.0851	0.0300
(50,50)	-0.0188	0.0860	0.0240	0.0838	0.0337	-0.0462	0.0836	0.0277	0.0818	0.0354
	-0.0667	0.0818	0.0248	0.0788	0.0328	-0.1204	0.0771	0.0274	0.0756	0.0341
(20,100)	-0.0439	0.0838	0.0327	0.0795	0.0463	-0.0747	0.0811	0.0339	0.0775	0.0432
	-0.0154	0.0863	0.0347	0.0819	0.0483	-0.0382	0.0843	0.0355	0.0796	0.0443
(20,50)	-0.0690	0.0816	0.0343	0.0781	0.0452	-0.0451	0.0837	0.0434	0.0787	0.0555
	-0.1010	0.0788	0.0365	0.0745	0.0493	-0.1067	0.0783	0.0436	0.0701	0.0559

Table 2: Relative bias, mean, empirical standard deviation, median and interquartile range of estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.1577$ ($\theta = 0.2$), $\lambda_0 = 0.5$; first line for coxph, second line for PROC MIXED; left part 30% censoring, right part 60 % censoring.

(K, n)	Rel. bias	Mean	Emp. std	Median	IQR	Rel. bias	Mean	Emp. std	Median	IQR
(100,100	-0.0120	0.1558	0.0264	0.1558	0.0335	-0.0292	0.1531	0.0253	0.1533	0.0329
	-0.0120	0.1558	0.0261	0.1553	0.0348	-0.0323	0.1526	0.0253	0.1520	0.0330
(100,50)	-0.0355	0.1521	0.0272	0.1511	0.0368	-0.0406	0.1513	0.0282	0.1501	0.0356
	-0.0938	0.1429	0.0263	0.1416	0.0337	-0.1046	0.1412	0.0281	0.1399	0.0377
(50,100)	-0.0184	0.1548	0.0348	0.1528	0.0470	-0.0374	0.1518	0.0375	0.1511	0.0490
	-0.0101	0.1561	0.0351	0.1537	0.0474	-0.0317	0.1527	0.0382	0.1524	0.0491
(50,50)	-0.0146	0.1554	0.0388	0.1526	0.0515	-0.0609	0.1481	0.0400	0.1450	0.0497
	-0.0590	0.1484	0.0392	0.1455	0.0518	-0.1141	0.1397	0.0396	0.1365	0.0521
(20,100)	-0.0140	0.1555	0.0536	0.1478	0.0705	-0.0615	0.1480	0.0561	0.1441	0.0776
	0.0184	0.1606	0.0578	0.1528	0.0788	-0.0247	0.1538	0.0593	0.1503	0.0839
(20,50)	-0.0704	0.1466	0.0576	0.1427	0.0735	-0.0653	0.1474	0.0668	0.1384	0.0816
	-0.0900	0.1435	0.0610	0.1363	0.0834	-0.0964	0.1425	0.0672	0.1338	0.0850

Table 3: Relative bias, mean, empirical standard deviation, median and interquartile range of estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.08765$ ($\theta = 0.1$), $\lambda_0 = 0.1$; first line for coxph, second line for PROC MIXED; left part 30% censoring, right part 60 % censoring.

(K, n)	Rel. bias	Mean	Emp. std	Median	IQR	Rel. bias	Mean	Emp. std	Median	IQR
(100,100)	-0.0222	0.0857	0.0146	0.0850	0.0194	-0.0200	0.0859	0.0170	0.0850	0.0226
	-0.0177	0.0861	0.0146	0.0854	0.0204	-0.0234	0.0856	0.0167	0.0844	0.0234
(100,50)	-0.0131	0.0865	0.0163	0.0856	0.0215	-0.0188	0.0860	0.0190	0.0859	0.0272
	-0.0793	0.0807	0.0164	0.0792	0.0215	-0.1010	0.0788	0.0189	0.0781	0.0263
(50,100)	-0.0051	0.0872	0.0204	0.0865	0.0276	-0.0280	0.0852	0.0232	0.0844	0.0322
	0.0017	0.0878	0.0210	0.0873	0.0294	-0.0234	0.0856	0.0238	0.0843	0.0321
(50,50)	-0.0200	0.0859	0.0245	0.0847	0.0323	-0.0211	0.0858	0.0281	0.0860	0.0386
	-0.0793	0.0807	0.0238	0.0807	0.0294	-0.0975	0.0791	0.0286	0.0785	0.0393
(20,100)	-0.0416	0.0840	0.0303	0.0824	0.0438	-0.0941	0.0794	0.0339	0.0753	0.0447
	-0.0086	0.0869	0.0319	0.0855	0.0461	-0.0645	0.0820	0.0354	0.0782	0.0468
(20,50)	-0.0690	0.0816	0.0355	0.0795	0.0490	-0.0747	0.0811	0.0397	0.0775	0.0517
	-0.0998	0.0789	0.0373	0.0747	0.0498	-0.1409	0.0753	0.0405	0.0706	0.0536

Table 4: Relative bias, mean, empirical standard deviation, median and interquartile range of estimated values $\hat{\sigma}_0^2$ over the 500 simulations; true $\sigma_0^2 = 0.1577$ ($\theta = 0.2$), $\lambda_0 = 0.1$; first line for coxph, second line for PROC MIXED; left part 30% censoring, right part 60 % censoring.

(K, n)	Rel. bias	Mean	Emp. std	Median	IQR	Rel. bias	Mean	Emp. std	Median	IQR
(100,100)	-0.0127	0.1557	0.0265	0.1553	0.0338	-0.0120	0.1558	0.0276	0.1545	0.0343
	-0.0127	0.1557	0.0260	0.1553	0.0346	-0.0127	0.1557	0.0272	0.1537	0.0350
(100,50)	-0.0082	0.1564	0.0265	0.1547	0.0347	-0.0431	0.1509	0.0290	0.1500	0.0381
	-0.0634	0.1477	0.0267	0.1442	0.0379	-0.1053	0.1411	0.0296	0.1405	0.0403
(50,100)	-0.0152	0.1553	0.0345	0.1556	0.0461	-0.0520	0.1495	0.0357	0.1429	0.0450
	-0.0108	0.1560	0.0354	0.1551	0.0443	-0.0476	0.1502	0.0370	0.1441	0.0475
(50,50)	-0.0342	0.1523	0.0390	0.1470	0.0522	-0.0311	0.1528	0.0429	0.1487	0.0557
	-0.0755	0.1458	0.0398	0.1422	0.0520	-0.0881	0.1438	0.0436	0.1419	0.0565
(20,100)	-0.0317	0.1527	0.0538	0.1476	0.0672	-0.0653	0.1474	0.0571	0.1407	0.0834
	0.0038	0.1583	0.0573	0.1514	0.0734	-0.0317	0.1527	0.0599	0.1479	0.0880
(20,50)	-0.0330	0.1525	0.0568	0.1445	0.0820	-0.0977	0.1423	0.0635	0.1338	0.0845
	-0.0476	0.1502	0.0595	0.1406	0.0841	-0.1344	0.1365	0.0654	0.1287	0.0809

Table 5: Relative bias, mean, empirical standard deviation, median and interquartile range of estimated values $\hat{\sigma}_0^2$ over the 500 simulations; 'True' gamma frailties, $\sigma_0^2 = 0.3520$ ($\theta = 0.6$), $\lambda_0 = 0.1$; first line for coxph, second line for PROC MIXED; 30 % censoring.

(K, n)	Rel. bias	Mean	Emp. std	Median	IQR
(100,100)	0.4102	0.4964	0.0862	0.4913	0.1128
	0.3226	0.4656	0.0740	0.4588	0.0945
(50,100)	0.3989	0.4925	0.1190	0.4846	0.1683
	0.3142	0.4626	0.1047	0.4563	0.1413
(20,100)	0.3563	0.4774	0.1825	0.4450	0.2240
	0.3422	0.4725	0.1714	0.4495	0.2297

Table 6: Mean, empirical standard deviation, median, interquartile range of the estimated values over the 500 simulations; 50 centers, 200 patients per center (100 patients in control and treatment group); $\lambda_0 = 0.3$, $\beta = -0.2$; left part for PROC MIXED, right part for coxme.

	True	Mean	Emp. std.	Median	IQR	Mean	Emp. std.	Median	IQR
β	-0.20	-0.2010	0.0124	-0.2010	0.0170	-0.2001	0.0120	-0.2001	0.0156
σ_0^2	0	0.0005	0.0009	0.0000	0.008	0.0017	0.0011	0.0016	0.0017
σ_1^2	0	0.0006	0.0009	0.0000	0.0009	0.0075	0.0062	0.0100	0.0128
σ_{01}	0	0.0002	0.0014	0.0002	0.0020	0.0027	0.0024	0.0024	0.0050
β	-0.20	-0.1996	0.0423	-0.1983	0.0567	-0.2005	0.0422	-0.1998	0.0545
σ_0^2	0	0.0006	0.0011	0.0000	0.0009	0.0007	0.0010	0.0000	0.0010
σ_1^2	0.08	0.0801	0.0177	0.0781	0.0244	0.0790	0.0127	0.0800	0.0036
σ_{01}	0	-0.0003	0.0048	-0.0003	0.0061	0.0011	0.0016	0.0008	0.0012
β	-0.20	-0.1997	0.0407	-0.2014	0.0540	-0.2001	0.0404	-0.2011	0.0513
σ_0^2	0.04	0.0393	0.0102	0.0387	0.0138	0.0394	0.0096	0.0388	0.0132
σ_1^2	0.08	0.0788	0.0179	0.0773	0.0240	0.0777	0.0168	0.0766	0.0218
σ_{01}	0.0283	0.0277	0.0102	0.0272	0.0130	0.0276	0.0097	0.0267	0.0123
β	-0.20	-0.1993	0.0135	-0.1993	0.0194	-0.1995	0.0123	-0.1996	0.0180
σ_0^2	0.08	0.0795	0.0181	0.0792	0.0233	0.0797	0.0124	0.0800	0.0010
σ_1^2	0	0.0005	0.0010	0.0000	0.0006	0.0006	0.0009	0.0001	0.0008
σ_{01}	0	0.0000	0.0048	0.0002	0.0066	0.0016	0.0014	0.0017	0.0021
β	-0.20	-0.2009	0.0331	-0.1997	0.0446	-0.2006	0.0321	-0.1987	0.0418
σ_0^2	0.08	0.0805	0.0182	0.0796	0.0253	0.0799	0.0179	0.0788	0.0248
σ_1^2	0.04	0.0399	0.0096	0.0395	0.0133	0.0397	0.0092	0.0387	0.0130
σ_{01}	0.0283	0.0290	0.0105	0.0284	0.0135	0.0287	0.0102	0.0280	0.0126
β	-0.20	-0.1986	0.0430	-0.1999	0.0587	-0.1996	0.0433	-0.2018	0.0581
σ_0^2	0.08	0.0776	0.0173	0.0760	0.0218	0.0776	0.0169	0.0757	0.0225
σ_1^2	0.08	0.0794	0.0174	0.0791	0.0220	0.0790	0.0168	0.0790	0.0230
σ_{01}	0.04	0.0393	0.0142	0.0386	0.0186	0.0394	0.0139	0.0394	0.0174

Table 7: Results of the analysis of the survival time of the patients included in the colorectal cancer trials.

Method	β	σ_0^2	σ_1^2	σ_{01}
PROC MIXED	-0.0458	0.0476	0.0000	-0.0084
coxme	-0.0533	0.0301	3.34×10^{-10}	1.78×10^{-11}

Table 8: Results of the analysis of the survival time of the patients included in the colorectal cancer trials.

Method	β	σ_0^2
PROC MIXED	-0.0558	0.0438
coxph	-0.0534	0.0376

Figure 3: Density function of the median time to event over centers for the colorectal cancer data.

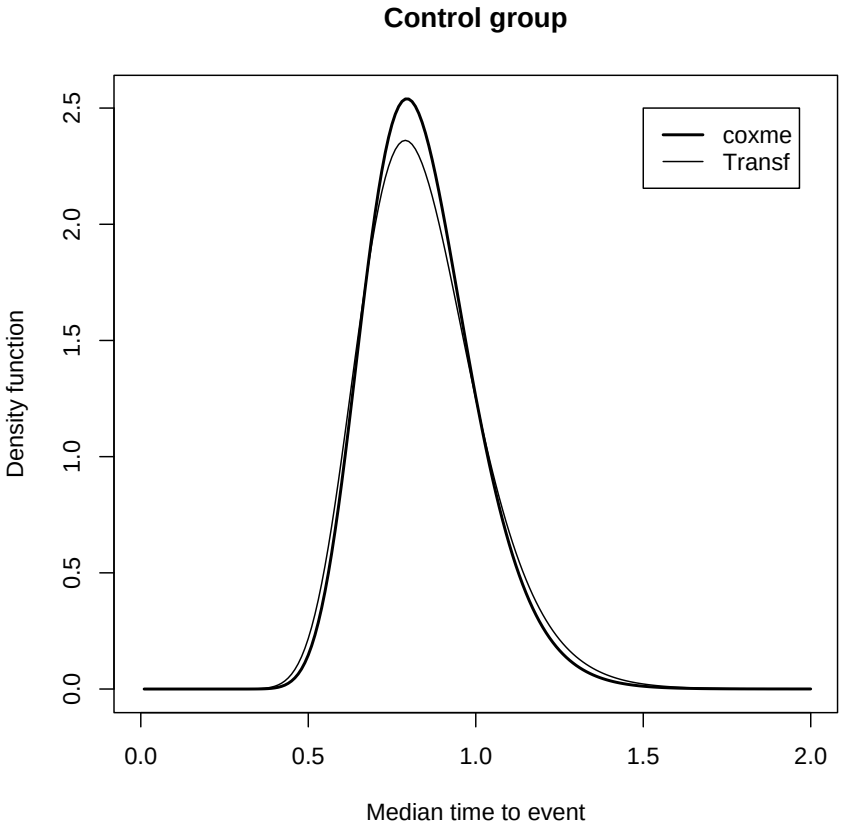


Table 9: Mean, empirical standard deviation, median, interquartile range of the estimated values over the 500 simulations; the same design as the colorectal cancer data; $\lambda_0 = 0.84$; left part for PROC MIXED, right part for coxme.

	True	Mean	Emp. std.	Median	IQR	Mean	Emp. std.	Median	IQR
β	-0.0533	-0.0537	0.0608	-0.0562	0.0877	-0.0588	0.0607	-0.0613	0.0863
σ_0^2	0.04	0.0378	0.0151	0.0357	0.0221	0.0400	0.0146	0.0387	0.0208
σ_1^2	0.08	0.0743	0.0269	0.0717	0.0338	0.0775	0.0255	0.0743	0.0332
σ_{01}	0.0283	0.0263	0.0150	0.0256	0.0201	0.0276	0.0147	0.0259	0.0199