aptara	BIOM	biom'960	Dispatch: December 5, 2007	CE: SS
	Journal	MSP No.	No. of pages: 9	PE: Adrian

BIOMETRICS DOI: 10.1111/j.1541-0420.2007.00960.x

Fitting Conditional Survival Models to Meta-Analytic Data by Using a Transformation Toward Mixed-Effects Models

Goele Massonnet,* Paul Janssen, and Tomasz Burzykowski

Hasselt University, Center for Statistics, Agoralaan, Building D, B-3590 Diepenbeek, Belgium
**email:* goele.massonnet@uhasselt.be

SUMMARY. 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. We show that the transformation idea also works for multivariate proportional odds models and for multivariate additive risks models. It therefore bridges segregated methodologies as it provides a general way to fit conditional models for multivariate survival data by using mixed models methodology. To study the specific features of the proposed method we focus on frailty models. 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. The proposed method is applied to data from 27 randomized trials in advanced colorectal cancer, which are available through the Meta-Analysis Group in Cancer.

KEY WORDS: Frailty model; Linear mixed model; Model transformation; Random treatment effect.

1. Introduction

Frailty models are widely used to fit clustered survival data. 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. Classical ways to fit frailty models are likelihood based. Likelihood methods to fit shared frailty models include expectation maximization (EM) algorithm (Klein, 1992), penalized partial likelihood (McGilchrist, 1993; Therneau and Grambsch, 2000), and 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), and Bayesian approach (Legrand et al., 2005).

In this article we propose an alternative way to fit frailty models. We start from the following observation: the integral of the weighted (over time) conditional cumulative log hazard 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 pseudodata (the estimated integrals) on which we can apply mixed models methodology. Model transformation also works for multivariate proportional odds models and multivariate additive risks, models. A related reference dealing with model transformation in the classical context of proportional hazards, additive risks, and proportional odds models is Grigoletto and Akritas (1999) and Cao and Gonzalez-Manteiga (2007).

Section 2 introduces the data from 27 randomized trials in advanced colorectal cancer. In Sections 3 and 4 we give, for right censored clustered survival data, the details on how multivariate proportional hazards models (frailty models), multivariate proportional odds models, and multivariate additive risks models can be transformed into mixed-effects models. The finding that parameters of interest of a multivariate survival model become parameters of interest of a related mixed-effects model, provides an interesting link between two seemingly segregated fields. In Sections 5 and 6, we focus on frailty models to study the performance of the proposed method. The simulation study in Section 5 illustrates that, compared to the classical likelihood-based approaches, the transformation 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 6 we discuss the performance of the method for the colorectal cancer data. We finally present some remarks and discuss possible further extensions in Section 7.

2. Advanced Colorectal Cancer Data

The study we consider is described in Burzykowski, Molenberghs, and Buyse (2004) and is typical for data encountered

in this research field. The data come from 27 advanced colorectal cancer trials (Advanced Colorectal Cancer Meta-Analysis Project, 1992, 1994; Meta-Analysis Group in Cancer, 1996, 1998). 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 will be 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, that is, 89.6%). We will investigate the between-trial variation (heterogeneity) in both the baseline risk and the effectiveness of the experimental treatment. For this purpose, we use a frailty model including a fixed treatment effect, a random trial effect, and a random treatment effect.

3. From Frailty Model to Linear Mixed-Effects Model

3.1 *Model Formulation*

Assume we have a total of N patients that come from K different clusters, cluster 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 that 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 cluster effect, and a random treatment effect: the conditional hazard for the jth patient in the ith cluster is then given by

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

where $\lambda_0(t)$ represents the unspecified baseline hazard at time t, β is the fixed overall treatment effect, b_{0i} is the random cluster 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 cluster 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

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

where $\otimes$ is the Kronecker product. The variance components σ_0^2 and σ_1^2 are a measure of the heterogeneity of the hazard due to the random cluster, with respect to random treatment effect; σ_{01} is the covariance between the two random effects within a cluster.

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

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

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

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

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 clusters.

3.2 *The Transformation*

With $\Lambda_{ij}(t) = \int_0^t \lambda_{ij}(s)\, ds$ the cumulative hazard for the jth patient in cluster 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

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

Let $w(\cdot)$ be a weight function ($W(t) = \int_0^t w(s)\, ds$) satisfying $w(s) \geqslant 0, s \in [0, \infty)$, $\int_0^\infty w(s)\, ds = 1$, and assigning zero weight to regions where the logarithm of the cumulative hazard function cannot be estimated due to censoring. The choice of the weight function is discussed in Section 3.3. Integrating both sides in (5) with respect to the weight function we obtain

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

with $\alpha = \int_0^\infty \ln \Lambda_0(t)\, dW(t)$. The patients in cluster i are divided, by the binary covariate x_{ij}, in a control and a treatment group. Let $\Lambda_i^{(0)}(.)$, respectively $\Lambda_i^{(1)}(\cdot)$, be the cumulative hazard function shared by all control, with respect to treated, patients in cluster i. Define, for $k = 0$, 1,

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

Then $\Omega_{i0} = \alpha + b_{0i} - (\beta + b_{1i})$ (control) and $\Omega_{i1} = \alpha + b_{0i} + (\beta + b_{1i})$ (treated). Following the ideas of Grigoletto and Akritas (1999), pseudoobservations for the Ω_{ik}'s can be obtained as

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

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

$$\hat{S}_i^{(0)}(t) = \prod_{j: T_{ij} \leqslant t, x_{ij} = -1} \left\{ \frac{r(T_{ij}) - d(T_{ij})}{r(T_{ij})} \right\},$$

with $r(v)$ the number still at risk at time v and $d(v)$ the number of events at time v and with $\hat{S}_i^{(1)}(t)$ the Kaplan–Meier estimator for the experimental group ($x_{ij} = 1$).

In terms of the pseudo observations we now can propose the model

$$\begin{aligned} \hat{\Omega}_{ik} &= \alpha + b_{0i} + (\beta + b_{1i})x_{ik} + (\hat{\Omega}_{ik} - \Omega_{ik}) \\ &= \alpha + b_{0i} + (\beta + b_{1i})x_{ik} + e_{ik}, \end{aligned} \qquad (6)$$

with $x_{i0} = -1$ and $x_{i1} = 1$. This is a linear mixed model with a random intercept and a random treatment effect. Note that the error terms $e_{ik} = \hat{\Omega}_{ik} - \Omega_{ik}$ correct for the fact that the mixed model is applied to the pseudodata because the

transformed cumulative hazard function cannot directly be observed. 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 3.3 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 cluster. 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 cluster effect is possible. More details on this are given in Section 3.3.

3.3 *The Error Variance*

To apply the methods proposed in Section 3.2, we need estimates for the error variances $\sigma^2_{e,ik}$, with respect to $\sigma^2_{e,i}$, of the random error terms in model (6), with respect to (7). We consider the general model (6). The patients of cluster i are divided in two groups: the control group ($k = 0$) and the treatment group ($k = 1$). Let n_{ik} be the number of patients in group k of cluster i. Note that, with $F_i^{(k)}(t) = 1 - S_i^{(k)}(t)$,

$$\begin{aligned} e_{ik} = \hat{\Omega}_{ik} - \Omega_{ik} &= \int_0^\infty \left\{ \ln \hat{\Lambda}_i^{(k)}(t) - \ln \Lambda_i^{(k)}(t) \right\} dW(t) \\ &\cong \int_0^\infty \frac{1}{\Lambda_i^{(k)}(t)} \frac{1}{S_i^{(k)}(t)} \left\{ \hat{F}_i^{(k)}(t) - F_i^{(k)}(t) \right\} dW(t). \end{aligned} \tag{8}$$

Using the independent and identically distributed (i.i.d.) representation for $\hat{F}_i^{(k)}(t) - F_i^{(k)}(t)$ proposed by Lo and Singh (1986), we easily obtain an i.i.d. representation of e_{ik}. Based on this representation and given an appropriate weight function, an approximation for the (estimated) variance of e_{ik} is obtained through (8). In the sequel we have chosen 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 cluster. The variance of the error term $\hat{\Omega}_{ik} - \Omega_{ik}$ $(i = 1, \ldots, K, k = 0, 1)$ can then be estimated by

$$\begin{aligned} \hat{\sigma}^2_{e,ik} = &\frac{1}{n_{ik}^2} \frac{1}{(B-A)^2} \int_A^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)} \int_A^s \frac{1}{\hat{\Lambda}_i^{(k)}(t)} \\ &\times \sum_{j:\ x_{ij}=k} \frac{I\left(0 \leqslant t_{ij} \leqslant t, \delta_{ij} = 1\right)}{\left\{ 1 - \frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} I\left(T_{ij} < t_{ij}\right) \right\}^2} \, dt \, ds \\ &+ \frac{1}{n_{ik}^2} \frac{1}{(B-A)^2} \int_A^B \frac{1}{\hat{\Lambda}_i^{(k)}(s)} \int_s^B \frac{1}{\hat{\Lambda}_i^{(k)}(t)} \\ &\times \sum_{j:\ x_{ij}=k} \frac{I\left(0 \leqslant t_{ij} \leqslant s, \delta_{ij} = 1\right)}{\left\{ 1 - \frac{1}{n_{ik}} \sum_{j:\ x_{ij}=k} I\left(T_{ij} < t_{ij}\right) \right\}^2} \, dt \, ds. \end{aligned} \tag{9}$$

The technical details of the derivation of (9) are presented in Web Appendix A.

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

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

4. Other Conditional Survival Models

As mentioned in Section 1, model transformation also works for conditional survival models that are different from frailty models. We give two examples.

The multivariate proportional odds model is given by

$$\ln\left\{ \frac{F_{ij}(t)}{1 - F_{ij}(t)} \right\} = h_0(t) + b_{0i} + (\beta + b_{1i}) x_{ij},$$

where $F_{ij}(.)$ is the conditional distribution function for the jth patient in the ith cluster and $h_0(.)$ is the baseline log odds function. Integrating out with respect to the weight function $w(.)$, we obtain

$$\int_0^\infty \ln\left\{ \frac{F_{ij}(t)}{1 - F_{ij}(t)} \right\} dW(t) = \alpha_{PO} + b_{0i} + (\beta + b_{1i}) x_{ij}, \tag{10}$$

with $\alpha_{PO} = \int_0^\infty h_0(t)\, dW(t)$.

Using a marginal likelihood approach, multivariate proportional odds models are studied by Lam, Lee, and Leung (2002) and Lam and Lee (2004).

The multivariate additive risks model is given by

$$\lambda_{ij}(t) = \lambda_0(t) + b_{0i} + (\beta + b_{1i}) x_{ij}.$$

Integrating out the corresponding cumulative hazard function with respect to the weight function $\tilde{W}(t) = W(t) / \int_0^\infty s\, dW(s)$, we obtain

$$\int_0^\infty \Lambda_{ij}(t)\, d\tilde{W}(t) = \alpha_{AR} + b_{0i} + (\beta + b_{1i}) x_{ij}, \tag{11}$$

with $\alpha_{AR} = \int_0^\infty \Lambda_0(t)\, d\tilde{W}(t)$. The additive risks model in the classical survival setting was first studied by Aalen (1980) and is also discussed in Martinussen and Scheike (2006). Yin and Cai (2004) and Martinussen and Scheike (2006) consider marginal additive risks models for multivariate survival data. The study of conditional additive risks models for multivariate survival data seems open.

Starting from (10) and (11), transformations to mixed-effects models are obtained along the lines of the discussion given for the frailty models. The technical details on the estimation of the error variance are given in Web Appendix A.

In the rest of the article we focus on frailty models, as described in Sections 3.1–3.3, to compare the results obtained by the proposed method with the results obtained by a classical likelihood-based approach.

5. Simulations

We study the performance of the proposed method in the context of frailty models by using a simulation study. As simulation model we consider the setting of a multicenter clinical trial. First, we consider the special case of the shared frailty model (4) including 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 Q1 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 misspecification of the frailty density. Next, we consider the general frailty model (1), including a fixed treatment effect, a random center effect, and a random treatment effect. 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.

5.1 *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, λ_0, σ_0^2, σ_{01}, σ_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, pseudodata $\hat{\Omega}_{ik}$ are generated through the model transformation described in Section 3 by using a uniform weight function $w(.)$ on the interval (A, B), chosen so that $0 < \hat{S}_i^{(k)}(t) < 1$ for $t \in (A, B)$. For each cluster i, the estimated variance of $\hat{\Omega}_{ik} - \Omega_{ik}$ is computed Q2 as explained in Section 3.3. To fit model (6), we use the SAS procedure PROC MIXED (see Web Appendix B for details on PROC MIXED). For each data set we obtain estimates for β, σ_0^2, σ_1^2, and σ_{01}.

For the special case of model (4), the data are generated as explained above with $\beta = 0$, $\sigma_1^2 = 0$, and $\sigma_{01} = 0$. Here, we consider moderate censoring (around 30%) and heavy censoring (around 60%). To study the robustness of the proposed method against frailty 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 $E(U_i) = e^{\sigma_0^2/2}$ and variance $\text{var}(U_i) = \theta = e^{\sigma_0^2}(e^{\sigma_0^2} - 1)$. This corresponds to random effects b_{0i} with mean 0 and variance σ_0^2. For each data set, pseudodata $\hat{\Omega}_i$ 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.

5.2 *The Choice of the Parameters*

5.2.1 *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 $\exp\{-\lambda_0 \exp(b_0) T_M\} = 0.5$, with b_0 zero-mean normally distributed, that is, $T_M = \frac{\log 2}{\lambda_0 \exp(b_0)}$. 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: $\text{var}(U_i) = \theta = e^{\sigma_0^2}(e^{\sigma_0^2} - 1)$. The values of σ_0^2 correspond to a frailty variance of $\theta = 0.1$, with respect to 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 $\lambda_0 = 0.1$. The motivation for choosing $\theta = 0.6$ is that the gamma and the lognormal density functions are close for $\theta = 0.1$, whereas for $\theta = 0.6$ these densities are more apart.

5.2.2 *Frailty model with random center and treatment effects.* We consider a situation with 50 centers that have 100 or 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 so that the bladder cancer data considered in Legrand et al. (2005) can serve as a reference. This study investigates heterogeneity in disease-free intervals 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 (σ_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).

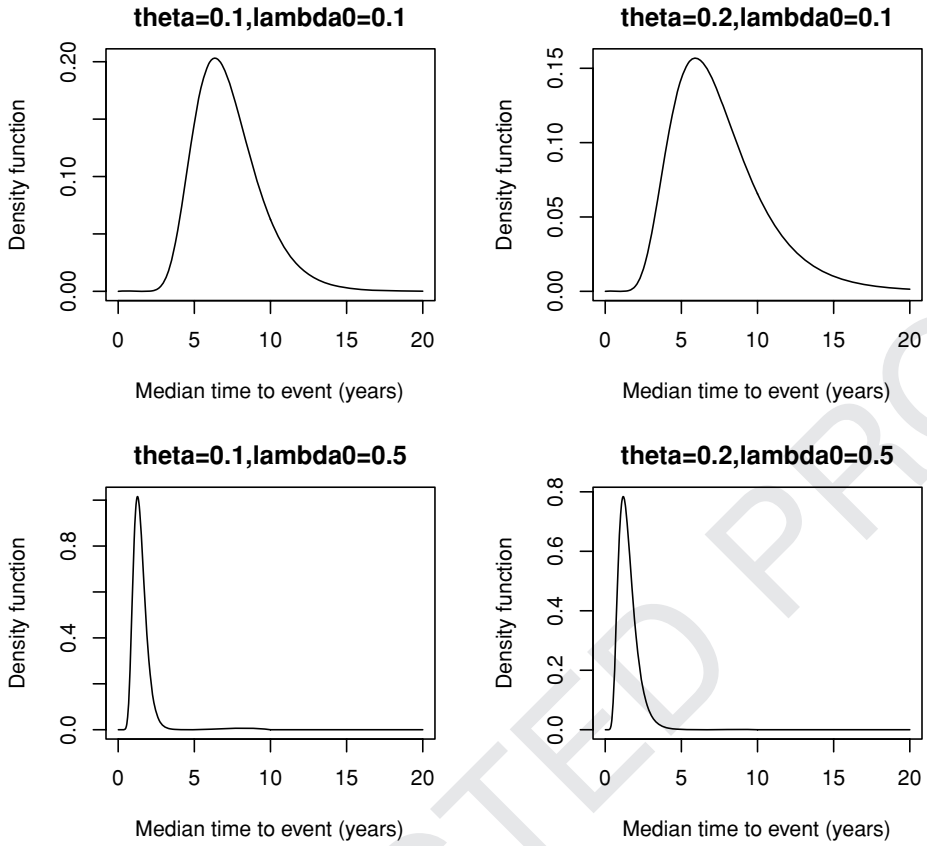

Figure 1. Density function of the median time to event over centers.

Table 1

Relative bias, mean, and empirical standard deviation of the 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)	Relative bias		Empirical standard deviation	Relative bias	Mean	Empirical standard deviation
(100, 100)	−0.0314	0.0849	0.0152	−0.0143	0.0864	0.0169
	−0.0257	0.0854	0.0148	−0.0177	0.0861	0.0162
(100, 50)	−0.0382	0.0843	0.0174	−0.0280	0.0852	0.0198
	−0.0975	0.0791	0.0168	−0.1135	0.0777	0.0197
(50, 100)	−0.0097	0.0868	0.0227	−0.0234	0.0856	0.0223
	0.0029	0.0879	0.0223	−0.0177	0.0861	0.0225
(50, 50)	−0.0188	0.0860	0.0240	−0.0462	0.0836	0.0277
	−0.0667	0.0818	0.0248	−0.1204	0.0771	0.0274
(20, 100)	−0.0439	0.0838	0.0327	−0.0747	0.0811	0.0339
	−0.0154	0.0863	0.0347	−0.0382	0.0843	0.0355
(20, 50)	−0.0690	0.0816	0.0343	−0.0451	0.0837	0.0434
	−0.1010	0.0788	0.0365	−0.1067	0.0783	0.0436

5.3 *Simulation Results*

5.3.1 *Frailty model with a random center effect.* Table 1 presents, for the setting (σ_0^2, λ_0) = (0.08765, 0.1), the relative bias, the mean, and the empirical standard deviation computed for the 500 estimates of the variance of the random center effect. The results for the settings (σ_0^2, λ_0) = (0.08765, 0.5), (0.1577, 0.1), and (0.1577, 0.5) were not substantially different (tables not shown).

The general conclusion for all parameter settings is 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 (coxph in S-Plus 7.0.6)

Table 2
Relative bias, mean, and empirical standard deviation of the 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)	Relative bias	Mean	Emppirical standard deviation
(100, 100)	0.4102	0.4964	0.0862
	0.3226	0.4656	0.0740
(50, 100)	0.3989	0.4925	0.1190
	0.3142	0.4626	0.1047
(20, 100)	0.3563	0.4774	0.1825
	0.3422	0.4725	0.1714

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.

Table 2 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, Burzykowski, and Janssen (2006). It clearly shows the need for lack-of-fit measures for frailty models.

5.3.2 *Frailty model with random center and treatment effects.* In Table 3 we report, for the parameter choice described in Section 5.2.2 and for 50 centers with 200 patients per center, the mean, the empirical standard deviation, and the average of the model-based standard deviations 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. There is no reliable software available to compute the standard errors for the variance–covariance parameters for the penalized partial likelihood approach. 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

Table 3
Mean, empirical standard deviation, and average of the model-based standard deviations of the estimated values over the 500 *simulations;* 50 *centers,* 200 *patients per center* (100 *patients in control and treatment group*); $\lambda_0 = 0.3$, $\beta = -0.2$.

		PROC MIXED			coxme	
	True	Mean	Empirical standard deviation	Model standard	Mean	Empirical standard deviation
β	−0.20	−0.2010	0.0124	0.0133	−0.2001	0.0120
σ_0^2	0	0.0005	0.0009	0.0008	0.0017	0.0011
σ_1^2	0	0.0006	0.0009	0.0008	0.0075	0.0062
σ_{01}	0	0.0002	0.0014	0.0014	0.0027	0.0024
β	−0.20	−0.1998	0.0423	0.0420	−0.2005	0.0422
σ_0^2	0	0.0006	0.0011	0.0009	0.0007	0.0010
σ_1^2	0.08	0.0801	0.0177	0.0182	0.0790	0.0127
σ_{01}	0	−0.0003	0.0047	0.0047	0.0011	0.0016
β	−0.20	−0.1997	0.0407	0.0418	−0.2001	0.0404
σ_0^2	0.04	0.0393	0.0102	0.0100	0.0394	0.0096
σ_1^2	0.08	0.0788	0.0179	0.0180	0.0777	0.0168
σ_{01}	0.0283	0.0277	0.0102	0.0104	0.0276	0.0097
β	−0.20	−0.1993	0.0135	0.0139	−0.1995	0.0123
σ_0^2	0.08	0.0794	0.0181	0.0181	0.0797	0.0124
σ_1^2	0	0.0005	0.0010	0.0008	0.0006	0.0009
σ_{01}	0	0.0000	0.0047	0.0046	0.0016	0.0014
β	−0.20	−0.2009	0.0331	0.0313	−0.2006	0.0321
σ_0^2	0.08	0.0805	0.0182	0.0184	0.0799	0.0179
σ_1^2	0.04	0.0399	0.0096	0.0101	0.0397	0.0092
σ_{01}	0.0283	0.0290	0.0105	0.0107	0.0287	0.0102
β	−0.20	−0.1986	0.0430	0.0421	−0.1996	0.0433
σ_0^2	0.08	0.0776	0.0173	0.0180	0.0776	0.0169
σ_1^2	0.08	0.0794	0.0174	0.0183	0.0790	0.0168
σ_{01}	0.04	0.0393	0.0142	0.0142	0.0394	0.0139

Table 4
Results of the analysis of the survival time of the patients included in the colorectal cancer trials (standard error in parentheses)

Method	β	σ_0^2	σ_1^2	σ_{01}
PROC MIXED	−0.0458 (0.0219)	0.0476 (0.0187)	0.0000 (−)	−0.0084 (0.0056)
coxme	−0.0534 (0.0169)	0.0355	3.34×10^{-10}	1.78×10^{-11}
PROC MIXED	−0.0558 (0.0225)	0.0461 (0.0172)		
coxph	−0.0534 (0.0169)	0.0376		

methods are on average comparable. The estimates produced by coxme have in general the smallest empirical variability. The average of the model-based standard deviations for the fixed effect and for the variance components gives an adequate estimate of the empirical variability for the proposed method.

For situations with 50 centers that have 100 patients per center (50 in the control group and 50 in the treatment group), the proposed method still gives reasonable estimates for β, σ_0^2, σ_1^2, and σ_{01} (table not shown). However, compared to the situation with 50 centers and 200 patients per center, we obtain estimates of a somewhat lower quality. This illustrates that the proposed method needs large enough samples sizes within the control group and the treatment group; this confirms our finding in Section 5.3.1

6. Analysis of Colorectal Cancer Data

To investigate the between-trial variation (heterogeneity) in both the baseline risk and the effectiveness of the therapy, we fit the frailty 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 and the corresponding standard errors, obtained by the proposed method and by the penalized partial likelihood approach (coxme in S-Plus 7.0.6), are presented in Table 4. 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 4.

The estimates obtained by the penalized partial likelihood and the transformation method are a bit different. However, the difference has only a low impact on important medical quantities, for example, on the density of the median time to event in the control group over trials (Figure 2). So both methods provide similar medical conclusions.

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 with the penalized partial likelihood if the cluster sizes are large enough.

The use of frailty models or linear mixed-effects models (for the pseudodata) raises questions on diagnostics. In the context of the method proposed here, we focus on model diagnostics for the linear mixed-effects model for the pseudodata (see, e.g., West, Welch, and Gałecki, 2007). If the diagnostic

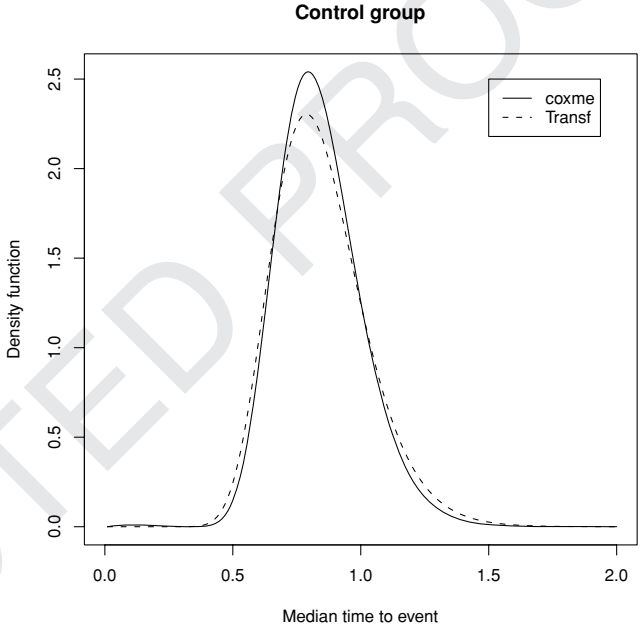

Figure 2. Density function of the median time to event over clusters for the colorectal cancer data.

plots show that the linear mixed-effects model is not appropriate for the pseudodata, it indicates that the corresponding frailty model is not valid for the individual data. However, if the results of the mixed model diagnostics are good, there is no guarantee that the frailty model is the correct model for the individual data. It is indeed possible that other models lead to the same linear mixed-effects model using an appropriate model transformation. To check the fit of the linear mixed-effects model obtained from the shared frailty model using the model transformation, we considered a plot of the studentized conditional residuals, versus the predicted values, a normal QQ plot of the studentized conditional residuals, and a normal QQ plot of the EBLUPs of the random trial effect (figures not shown). These diagnostic plots show that the pseudo-values, obtained from the colorectal cancer data, can be analyzed using the linear mixed-effects model that corresponds to the shared frailty model. Q3

7. Conclusions

In this article, an alternative approach to fit frailty models is proposed. The original problem of "fitting a frailty model" is reformulated into a standard problem of "fitting a

linear mixed-effects model." We show that the integral of the weighted (over time) conditional cumulative log hazard 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 models methodology can be applied. We illustrate that this transformation idea can also be used to fit multivariate proportional odds models and multivariate additive risks models. Most standard statistical packages contain procedures to fit complex linear mixed-effects models but offer only a limited number of procedures to fit conditional (random effects) survival models. The proposed model transformation is therefore a useful practical way to get an insight on heterogeneity in clustered data. The performance of the proposed method was studied by simulation in the context of frailty models. The results indicate a good performance of the proposed method for data sets with a sufficiently large number of clusters (i.e., $K = 20$) and moderate to large sample sizes within covariate-level subgroups in the clusters (i.e., at least $n_{ik} = 50$).

We considered a frailty model with a binary covariate and we therefore could use the Kaplan–Meier estimator for the survival function. It 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). An asymptotic representation for the Beran estimator is proposed by Van Keilegom and Veraverbeke (1997). Such a representation is necessary to estimate the variance of the error terms in the mixed-effects model. 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 readily extends to any censoring scheme for which an i.i.d. representation for a nonparametric estimator for the cumulative hazard or the survival function is available (e.g., for interval-censored data, Lindsey and Ryan (1998), or for left truncated and right censored data, Gijbels and Wang (1993) and Zhou and Yip (1999)).

The performance of the transformation method for multivariate proportional odds models and multivariate additive risks models will be a subject for further study.

8. Supplementary Materials

The R-code to compute the pseudodata, a SAS-macro to fit the linear mixed-effects model to the pseudodata, and the Web Appendices referenced in Sections 3.3 and 5.1 are available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

Acknowledgements

The authors are grateful to the Meta-Analysis Group in Cancer (MAGIC) for the permission to use their data. The authors thank two referees and an associate editor for their valuable comments and suggestions. The authors gratefully acknowledge the financial support from the IAP research network P6/03 of the Belgian Government (Belgian Science Policy).

References

Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. In *Lecture Notes in Statistics 2: Mathematical Statistics and Probability Theory,* W. Klonecki, A. Kozek, and J. Rosinski (eds), 1–25. New York: Springer-Verlag.

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

Burzykowski, T., Molenberghs, G., and Buyse, M. (2004). 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, Series A* **167,** 103–124.

Cao, R. and Gonzalez-Manteiga, W. (2007). Goodness-of-fit tests for conditional models under censoring and truncation. *Journal of Econometrics* (to appear). Q4

Cortiñas Abrahantes, J. and Burzykowski, T. (2005). A version of the EM algorithm for proportional hazards model with random effects. *Biometrical Journal* **47,** 847–862.

Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R., and Sylvester, R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics and Data Analysis* **40,** 603–620. Q5

Ducrocq, V. and Casella, G. (1996). A Bayesian analysis of mixed survival models. *Genetics Selection Evolution* **28,** 505–529.

Gijbels, I. and Wang, J. L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. *Journal of Multivariate Analysis* **47,** 210–229.

Grigoletto, M. and Akritas, M. G. (1999). Analysis of covariance with incomplete data via semiparametric model transformations. *Biometrics* **55,** 1177–1187.

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding*. Guelph, Canada: University of Guelph Press.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48,** 795–806.

Lam, K. F. and Lee, Y. W. (2004). Merits of modelling multivariate survival data using random effects proportional odds model. *Biometrical Journal* **46,** 331–342.

Lam, K. F., Lee, Y. W., and Leung, T. L. (2002). Modeling multivariate survival data by a semiparametric random effects proportional odds model. *Biometrics* **58,** 316–323.

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

Lindsey, J. C. and Ryan, L. M. (1998). Tutorial in biostatistics: Methods for interval-censored data. *Statistics in Medicine* **17,** 219–238.

Lo, S. H. and Singh, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probability Theory and Related Fields* **71,** 455–465.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data.* New York: Springer.

Massonnet, G., Burzykowski, T., and Janssen, P. (2006). Resampling plans for frailty models. *Communications in Statistics—Simulation and Computation* **35,** 497–514.

McGilchrist, C. (1993). REML estimation for survival models with frailty. *Biometrics* **49,** 221–225.

Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56,** 1016–1022.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data, Extending the Cox Model.* New York: Springer.

Vaida, F. and Xu, R. (2000). Proportional hazards model with random effects. *Statistics in Medicine* **19,** 3309–3324.

Van Keilegom, I. and Veraverbeke, N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals of the Institute of Statistical Mathematics* **49,** 467–491.

West, B. T., Welch, K. B., and Gałecki, A. T. (2007). *Linear Mixed Models: A Practical Guide using Statistical Software.* Boca Raton, Florida: Chapman & Hall/CRC.

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika* **91,** 801–818.

Zhou, Y. and Yip, P. S. F. (1999). A strong representation of the product-limit estimator for left truncated and right censored data. *Journal of Multivariate Analysis* **69,** 261–280.

Received February 2007. *Revised October* 2007.
Accepted October 2007.