Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution

Arnošt Komárek[1,‡], Emmanuel Lesaffre[1,*,†] and Catherine Legrand[2,§]

[1] *Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium*
[2] *European Organisation for Research and Treatment of Cancer, E. Mounierlaan 83/11, 1200 Brussels, Belgium*

SUMMARY

Nowadays, most clinical trials are conducted in different centers and even in different countries. In most multi-center studies, the primary analysis assumes that the treatment effect is constant over centers. However, it is also recommended to perform an exploratory analysis to highlight possible center by treatment interaction, especially when several countries are involved. We propose in this paper an exploratory Bayesian approach to quantify this interaction in the context of survival data. To this end we used and generalized a random effects accelerated failure time model. The generalization consists in using a penalized Gaussian mixture as an error distribution on top of multivariate random effects which are assumed to follow a normal distribution. For computational convenience, the computations are based on Markov chain Monte Carlo techniques. The proposed method is illustrated on the disease free survival times of early breast cancer patients collected in the EORTC trial 10854. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: multi-center study; penalized Gaussian mixture; regression; survival analysis

[‡] Current address: Department of Probability and Mathematical Statistics, Charles University, Sokolovská 83, 186 75 Praha 8-Karlín, the Czech Republic

[†] E-mail: emmanuel.lesaffre@med.kuleuven.be

[§] Current address: Institute of Statistics, Université Catholique de Louvain, 20 voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium

[*] Correspondence to: Emmanuel Lesaffre, Biostatistical Centre, Katholieke Universiteit Leuven, Kapucijnenvoer 35, 3000 Leuven, Belgium

Contract/grant sponsor: Research Funds Katholieke Universiteit Leuven; contract/grant number: PDM/06/242

Contract/grant sponsor: Belgian Federal Science Policy Office; contract/grant number: P6/03

Contract/grant sponsor: National Cancer Institute, U.S.A.; contract/grant number: 5U10 CA11488-35

1. INTRODUCTION

1.1. Motivation: EORTC trial 10854

The EORTC trial 10854 (Clahsen et al. [1]; van der Hage et al. [2]) is a large multi-center study ($n = 2\,793$ patients in $N = 14$ centers) aiming to compare perioperative polychemotherapy (POP FAC arm) with no further treatment (control arm) on the disease free survival (DFS) time in early breast cancer patients who underwent potentially curative surgery. The centers are located in 5 geographical regions: the Netherlands, Poland, France, Southern Europe, and South Africa. To improve the efficiency with which the treatment effect is evaluated, we wish to account for known sources of variability – known patient- and center-specific characteristics (covariates) and use an appropriate regression model. For many patients, the observed DFS time is right-censored.

1.2. Heterogeneity

In multi-center, multinational studies, like the EORTC trial 10854, there are often unknown sources of heterogeneity between centers, despite the use of a common protocol. This can happen for many reasons: geographical differences, different working habits of the staff in different centers, different patient populations attracted by different centers, etc. This applies even more when several countries are involved, see, e.g., Anello, O'Neill and Dubey [3]. We need to distinguish between two types of heterogeneity with respect to: (a) baseline characteristics and (b) treatment efficacy. In the latter case one speaks of a treatment by center interaction. If this interaction is large, the interpretation of the effect of treatment needs to be done with caution, especially when the treatment by center interaction is qualitative (reverses in direction from one center to another). Figure 1 shows Kaplan-Meier estimates of the DFS distribution for the POP FAC arm and the control arm, separately for each center. From these curves, there seems to be some heterogeneity among the centers. Not only the overall proportion of DFS patients differs at each time point and in each treatment arm from center to center (*baseline heterogeneity*) but also the effect of treatment on DFS, expressed by the relative position of the two curves in the control and treatment arm seems to vary across centers both quantitatively and qualitatively (*treatment effect heterogeneity*). A possible approach to take into account the heterogeneity between centers is a model with the center indicator and the treatment by center interaction as a part of the covariates (fixed effects model). However, in this paper, we have opted for a random effects model, see Section 5 for a more detailed discussion of this option.

<Figure 1 about here.>

1.3. Aim and outline of the paper

Random effects regression models constitute a widely used approach for regression analysis when heterogeneity resulting from clustering of the data cannot be ruled out. Further, since it is difficult to assess the distributional assumptions with censored data, it is preferred to leave the distribution of survival times unspecified as, e.g., in Cox's proportional hazards (PH) model (Cox [4]) or, alternatively, to specify it in a flexible way.

The main objective of this paper is to present a random effects regression model in which the distribution of survival times is not specified in a conventional parametric way like, e.g., log-normal, log-logistic, or Weibull. The PH model is certainly the most popular survival regression model. This has probably to do with the elegant concept of partial likelihood (Cox [5]). However, this does not imply that the PH assumption is taken to be granted, as is often done in practice. In this paper, we concentrate on the AFT model (e.g., Kalbfleisch and Prentice [6], Chap. 7), in which the covariates directly accelerate or decelerate the expected survival time. It has been pointed out by David Cox (in Reid [7]) that the AFT models are "in many ways more appealing because of their quite direct physical interpretation". However, it is not the purpose of this paper to balance the two approaches or to give statements preferring one approach over the other.

To model a survival distribution flexibly, we use a smoothing technique based on a penalized Gaussian mixture (PGM, see Section 2) exploited in Komárek, Lesaffre and Hilton [8] and Komárek and Lesaffre [9]. Specifically, in Ref. [8] an AFT model for independent observations is proposed. However, their model does not allow for the inclusion of random effects and hence cannot be used to take heterogeneity into account. The extension allowing a random intercept and hence modeling baseline heterogeneity is presented in Ref. [9]. Here, our main intention is to modify the model from Ref. [9] by inclusion of multivariate random effects such that heterogeneity of an arbitrary type can be considered. Further, we wish to illustrate the use of this modification on modeling the baseline and treatment effect heterogeneity in the analysis of the EORTC trial 10854.

The paper is organized as follows. Section 2 reviews the PGM approach of Ref. [8] and [9], and explains its use in the AFT model with multivariate random effects. In Section 3, we describe the inferential procedure for the suggested model based on Markov chain Monte Carlo methodology and explain its motivation by the penalized maximum-likelihood method. The analysis of the DFS time in early breast cancer patients is presented in Section 4. We finalize the paper by a discussion in Section 5.

2. RANDOM EFFECTS AFT MODEL WITH PENALIZED GAUSSIAN MIXTURE AS AN ERROR DISTRIBUTION

2.1. Random effects AFT model

Let $T_{i,l}$ $(i = 1, \dots, N,\ l = 1, \dots, n_i)$ denote the event time for the lth patient in the ith center. Our approach not only allows for right-censored data but also for left- or interval-censored data. Therefore, assume that $T_{i,l}$ occurred within an interval of time $\lfloor t^L_{i,l},\, t^U_{i,l}\rfloor$, where the symbols $\lfloor$ and $\rfloor$ are used for the lower and upper limits of the interval which can be open, closed or half-closed according to the context. For instance, for an exactly observed event time, $\lfloor t^L_{i,l},\, t^U_{i,l}\rfloor = [t_{i,l},\, t_{i,l}]$, for a right-censored observation, $\lfloor t^L_{i,l},\, t^U_{i,l}\rfloor = (t_{i,l},\, \infty)$. Further, assume that the observed intervals are a result of an independent censoring process.

The random effects AFT model is in fact a classical linear mixed model of Laird and Ware [10] with the logarithmic link function, i.e.

$$\log(T_{i,l}) = \boldsymbol{b}_i'\boldsymbol{z}_{i,l} + \boldsymbol{\beta}'\boldsymbol{x}_{i,l} + \varepsilon_{i,l}, \qquad (i = 1, \dots, N,\ l = 1, \dots, n_i), \tag{1}$$

where $\varepsilon_{i,l}$ are i.i.d. error terms having the distribution of the baseline log-event time with a density g_ε, $\boldsymbol{x}_{i,l} = (x_{i,l,1}, \dots, x_{i,l,s})'$ are vectors of patient- and center-specific covariates assumed to have a *homogeneous* effect across centers and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_s)'$ is a vector of corresponding regression coefficients – fixed effects. Further, $\boldsymbol{z}_{i,l} = (z_{i,l,1}, \ldots, z_{i,l,q})'$ are vectors of factors with possibly varying (*heterogeneous*) effect across centers. For example, to model the baseline and treatment effect heterogeneity between centers we take $\boldsymbol{z}_{i,l} = (1,\ \mathsf{treat}_{i,l})'$, where $\mathsf{treat}_{i,l}$ is the treatment indicator for the (i,l)th patient. Finally, $\boldsymbol{b}_i = (b_{i,1}, \ldots, b_{i,q})'$ are i.i.d. vectors of center-specific random effects with some (parametric) density g_b having a mean $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ which expresses an overall effect of the covariates included in $\boldsymbol{z}_{i,l}$.

2.2. *Baseline survival distribution*

Any parametric assumption concerning the baseline survival distribution in the AFT model (1) represented by the density g_ε is difficult to check with censored data. For this reason, it is our intention to leave g_ε either unspecified or to specify it in a flexible way. This goal can be achieved in different ways. For example, Pan and Louis [11] and Pan and Connett [12] consider the random effects AFT model and estimate the distribution of the error term by inclusion of a non-parametric Kaplan-Meier estimation step in their estimation procedure. Komárek and Lesaffre [13] specify g_ε as a Gaussian mixture with unknown number of components, unknown means and variances.

An alternative route, namely the use of smoothing techniques, was taken by Ref. [8] and [9] and will be followed also here. In both papers, the error density g_ε is expressed as a shifted and scaled penalized Gaussian mixture (PGM), which is specified as

$$g_\varepsilon(\varepsilon) = \tau^{-1} \sum_{j=-K}^{K} w_j(\boldsymbol{a})\, \varphi\bigl\{\tau^{-1}(\varepsilon - \alpha) \,|\, \mu_j,\, \sigma^2\bigr\}, \tag{2}$$

where α and τ are the (unknown) intercept and scale parameter, respectively, and

$$w_j(\boldsymbol{a}) = \frac{\exp(a_j)}{\sum_{k=-K}^{K} \exp(a_k)}, \quad j = -K, \ldots, K \tag{3}$$

are (unknown) mixture weights. The weights in (3) are reparametrized to ensure that g_ε is a density for which we need $0 < w_j < 1$, $j = -K, \ldots, K$ and $\sum_j w_j = 1$. Therefore, we will work with the parameter vector $\boldsymbol{a} = (a_{-K}, \ldots, a_K)'$ instead of the vector $\boldsymbol{w} = (w_{-K}, \ldots, w_K)'$. Further, $\boldsymbol{\mu} = \{\mu_{-K}, \ldots, \mu_K\}$ is a *fine* grid of equidistant knots centered around zero ($\mu_0 = 0$) and σ^2 is a fixed basis variance, common for all mixture components. In fact, Gaussian densities $\varphi(\cdot|\mu_{-K}, \sigma^2)$, $\ldots$, $\varphi(\cdot|\mu_K, \sigma^2)$ form a set of basis functions which are used, through the estimation of weights $\boldsymbol{w}$, to smooth the unknown density g_ε. The following choice has been used in the analysis of Section 4, i.e.: $K = 15$, $\mu_{-K} = -4.5$, $\mu_K = 4.5$, $\sigma = 0.2$ and $\mu_{j+1} - \mu_j = 0.3$, see Ref. [8] for a motivation.

Identification problems stemming from a high number of unknown parameters (e.g., in the analysis in Section 4, we have 31 unknown mixture weights) are prevented by putting a roughness penalty on the weights, see Section 3. In survival analysis, the baseline survival distribution is more often specified by modeling the hazard function. For example, similarly to our PGM approach, Lambert and Eilers [14], Kneib [15], or Kneib and Fahrmeir [16] use penalized B-splines to express the baseline hazard function. In the context of an AFT the density of the baseline log-event time (error term) seems to be more advantageous since this density enters directly the likelihood (see further in Section 3).

2.3. Random effects distribution

We will assume that g_b, the distribution of random effects $\boldsymbol{b}_i$ $(i = 1, \dots, N)$, is multivariate normal with unknown mean $\boldsymbol{\gamma}$ and unknown covariance matrix $\mathbb{D}$. For example, when modeling the baseline and treatment effect heterogeneity in Section 4, we have $\boldsymbol{b}_i = (b_{i,1},\, b_{i,2})'$, $\boldsymbol{z}_{i,l} = (1,\, \mathsf{treat}_{i,l})$, and

$$\mathrm{E}(\boldsymbol{b}_i) = \boldsymbol{\gamma} = (0,\, \gamma_2)', \qquad \mathrm{var}(\boldsymbol{b}_i) = \mathbb{D} = \begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{1,2} & d_{2,2} \end{pmatrix}, \tag{4}$$

where γ_2 is the mean treatment effect and $d_{1,1}$, $d_{2,2}$, $d_{1,2}$ variance components of the random effects distribution. Note that γ_1 (the mean of the random effect expressing the baseline heterogeneity) cannot be distinguished from the PGM parameter α (intercept of the error term). Hence, for identifiability reasons, γ_1 is constantly equal to zero.

The reasons for choosing a normal distribution for the random effects and do not a smoothing approach as for the error term are: (i) The number of centers in our application is quite low (14) providing only a low number of (moreover latent) "observations" to estimate the shape of the distribution; (ii) It has been shown in the literature (Keiding, Andersen and Klein [17], Lambert et al. [18]) that the regression parameters, which are usually of the primary interest are robust against misspecification of the random effects distribution; (iii) When interest lies in the marginal characteristics like the hazard or survival functions, a possible misspecification of the random effects distribution is at least partly corrected by the estimation of the error distribution.

3. ESTIMATION AND INFERENCE

3.1. Penalized maximum likelihood

In the original work, Ref. [8], penalized maximum likelihood (PML) was used for estimation. In the current context, we would have to maximize the penalized likelihood

$$L^{penal}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) \times \exp\Bigl\{-\frac{\lambda}{2}\sum_{j=-K+s}^{K}(\Delta^d a_j)^2\Bigr\} \tag{5}$$

with respect to $\boldsymbol{\theta} = \bigl(\boldsymbol{\beta}', \boldsymbol{\gamma}', \mathrm{vec}(\mathbb{D}),\, \alpha,\, \tau,\, \boldsymbol{a}',\, \lambda\bigr)'$. In expression (5), $L(\boldsymbol{\theta})$ denotes the likelihood which is equal to

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N}\Biggl[\int_{\mathbb{R}^q}\Biggl\{\prod_{l=1}^{n_i}\int_{t_{i,l}^L}^{t_{i,l}^U} p_{i,l}(t\,|\,\boldsymbol{a},\, \alpha,\, \tau,\, \boldsymbol{\beta},\, \boldsymbol{b})\,dt\Biggr\}\varphi_q(\boldsymbol{b}\,|\,\boldsymbol{\gamma},\, \mathbb{D})\,d\boldsymbol{b}\Biggr], \tag{6}$$

where

$$p_{i,l}(t\,|\,\boldsymbol{a},\, \alpha,\, \tau,\, \boldsymbol{\beta},\, \boldsymbol{b}) = (t\tau)^{-1}\sum_{j=-K}^{K} w_j(\boldsymbol{a})\varphi\biggl(\frac{\log(t) - \alpha - \boldsymbol{b}'\boldsymbol{z}_{i,l} - \boldsymbol{\beta}'\boldsymbol{x}_{i,l}}{\tau}\,\bigg|\,\mu_j,\, \sigma^2\biggr). \tag{7}$$

Further, Δ^d denotes the dth-order difference operator ($d = 3$ was used in the analysis presented in Section 4). The roughness penalty, $-\frac{\lambda}{2}\sum_{j=-K+s}^{K}(\Delta^d a_j)^2$, which can also be written as

$-\frac{\lambda}{2}\,\boldsymbol{a}'\mathbb{P}_d'\mathbb{P}_d\boldsymbol{a}$ for an appropriate difference operator matrix $\mathbb{P}_d$, avoids identifiability problems or overfitting the data, see Eilers and Marx [19]. A trade-off between the smoothness of the density g_ε and fitting the data is driven by the smoothing parameter λ, which has to be estimated as well.

3.2. Bayesian specification

Maximization of the penalized likelihood (5) is quite difficult. However, as pointed out by Wahba [20], the penalized likelihood is proportional to the posterior density in an appropriately specified Bayesian model. Estimates and inference can then be used on the sample from this posterior distribution obtained using the Markov chain Monte Carlo (MCMC) method (see, e.g., Robert and Casella [21]). This approach was followed in Ref. [9] and will be used also here, with some modifications stemming from the use of multivariate random effects combined with covariates.

In agreement with Ref. [9], we specify the prior distribution of the model parameters, $p(\boldsymbol{\theta})$, as a product of vague, but proper distributions and a Gaussian Markov random field (GMRF, see, e.g., Rue and Held [22]) prior for transformed mixture weights $\boldsymbol{a}$, that is

$$p(\boldsymbol{\theta}) = \underbrace{\prod_{j=1}^{s} p(\beta_j) \times \prod_{j=2}^{q} p(\gamma_j) \times p(\mathbb{D}) \times p(\alpha) \times p(\tau^{-2}) \times p(\lambda)}_{p(\boldsymbol{\theta}_{-\boldsymbol{a}})} \times p(\boldsymbol{a}\,|\,\lambda), \tag{8}$$

where $p(\beta_j)$ $(j = 1, \dots, s)$, $p(\gamma_j)$ $(j = 2, \dots, q)$, $p(\alpha)$ are densities of the normal distribution with (zero) mean and large variance, e.g., $\mathcal{N}(0,\, 10^2)$ was used in the analysis of Section 4. Further, $p(\mathbb{D})$ is the inverse Wishart distribution with a small number of degrees of freedom df_b and a diagonal scale matrix $\mathbb{S}_b$ with small values on the diagonal. In Section 4, we used $df_b = q = 2$ and $\mathbb{S}_b = \mathrm{diag}(0.002)$. Furthermore, $p(\tau^{-2})$ and $p(\lambda)$, prior densities of the parameters that can be interpreted as inverse variances, are densities of a dispersed gamma distribution, e.g., Gamma(1, 0.005) distributions were used in Section 4. Finally, the GMRF prior of the vector $\boldsymbol{a}$ is such that $p(\boldsymbol{a}\,|\,\lambda) \propto \exp\bigl(-\frac{\lambda}{2}\,\boldsymbol{a}'\mathbb{P}_d'\mathbb{P}_d\boldsymbol{a}\bigr)$, and corresponds to the penalty part of the penalized likelihood (5). Using the Bayes' rule, the posterior distribution is

$$p(\boldsymbol{\theta}\,|\,\mathrm{data}) \;\propto\; L(\boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \;=\; L(\boldsymbol{\theta}) \times p(\boldsymbol{\theta}_{-\boldsymbol{a}}) \times p(\boldsymbol{a}\,|\,\lambda) \tag{9}$$

and it is seen that provided $p(\boldsymbol{\theta}_{-\boldsymbol{a}}) \overset{\text{approx.}}{\propto} 1$, the posterior distribution (9) is approximately proportional to the penalized likelihood (5).

3.3. Bayesian data augmentation

When using a Bayesian approach, it is advantageous to consider latent quantities, further denoted as $\boldsymbol{\psi}$, which are explicitly or implicitly integrated out from the likelihood (6), as additional model parameters (Bayesian data augmentation, see Tanner and Wong [23]). For convenience in notation we explain it in the case when all event times $t_{i,l}$ are censored, i.e. $[\mathrm{data}] = (t_{1,1}^L,\, t_{1,1}^U, \dots, t_{N,n_N}^L,\, t_{N,n_N}^U)'$ and $t_{i,l}^L < t_{i,l}^U$ for all i and l.

Let $\boldsymbol{\psi} = (\boldsymbol{t}',\, \boldsymbol{r}',\, \boldsymbol{B}')'$, with $\boldsymbol{t} = (t_{1,1}, \dots, t_{N,n_N})'$, $\boldsymbol{r} = (r_{1,1}, \dots, r_{N,n_N})'$, and $\boldsymbol{B} = (\boldsymbol{b}_1', \dots, \boldsymbol{b}_N')'$, be the vector of latent exact event times, component labels (explained below), and random effects, respectively. The prior distribution of the full parameter vector $(\boldsymbol{\psi}', \boldsymbol{\theta}')'$ is specified as

$$p(\boldsymbol{\psi}, \boldsymbol{\theta}) = \underbrace{p(\boldsymbol{t}\,|\,\boldsymbol{r}, \boldsymbol{B}, \boldsymbol{\theta}) \times p(\boldsymbol{r}\,|\,\boldsymbol{B}, \boldsymbol{\theta}) \times p(\boldsymbol{B}\,|\,\boldsymbol{\theta})}_{p(\boldsymbol{\psi}\,|\,\boldsymbol{\theta})} \times p(\boldsymbol{\theta}), \tag{10}$$

where

$$p(\boldsymbol{t}\,|\,\boldsymbol{r}, \boldsymbol{B}, \boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{l=1}^{n_i}\biggl\{(t_{i,l}\tau)^{-1}\varphi\biggl(\frac{\log(t_{i,l}) - \alpha - \boldsymbol{b}_i'\boldsymbol{z}_{i,l} - \boldsymbol{\beta}'\boldsymbol{x}_{i,l}}{\tau}\,\bigg|\,\mu_{r_{i,l}}, \sigma^2\biggr)\biggr\}, \tag{11}$$

$$p(\boldsymbol{r}\,|\,\boldsymbol{B}, \boldsymbol{\theta}) = \mathrm{P}\bigl(\boldsymbol{r} = (j_{1,1}, \dots, j_{N,n_N})'\,|\,\boldsymbol{\theta}\bigr) = \prod_{i=1}^{N}\prod_{l=1}^{n_i} w_{j_{i,l}}(\boldsymbol{a}), \tag{12}$$

$$p(\boldsymbol{B}\,|\,\boldsymbol{\theta}) = \prod_{i=1}^{N}\varphi_q(\boldsymbol{b}_i\,|\,\boldsymbol{\gamma}, \mathbb{D}), \tag{13}$$

and $p(\boldsymbol{\theta})$ is given by (8). Prior distributions (11)–(13) follow directly from the original expressions for the likelihood, eq. (6) and (7) and their product is in fact equal to the likelihood if the latent data had been observed.

Specifically, the form of $p(\boldsymbol{B}\,|\,\boldsymbol{\theta})$ stems from our assumption of normality of random effects. The form of $p(\boldsymbol{r}\,|\,\boldsymbol{\theta}, \boldsymbol{B})$ follows from the fact that, intrinsicly, we can assume that the (i, l)th residual log-event time belongs to one of the $2K + 1$ normal components, labeled by $r_{i,l}$. Following our model of a penalized Gaussian mixture (7), the probability of belonging to the jth mixture component is equal to $w_j(\boldsymbol{a})$ and hence the form of (12). Finally, the form for $p(\boldsymbol{t}\,|\,\boldsymbol{r}, \boldsymbol{B}, \boldsymbol{\theta})$ follows from the AFT model with the error distribution specified as a PGM. However, the mixture is involved only implicitly by conditioning on the component labels $\boldsymbol{r}$.

Given the full parameter vector $(\boldsymbol{\psi}', \boldsymbol{\theta}')'$, the likelihood has now a trivial form

$$L(\boldsymbol{\psi}, \boldsymbol{\theta}) = p(\mathrm{data}\,|\,\boldsymbol{\psi}, \boldsymbol{\theta}) = p(\mathrm{data}\,|\,\boldsymbol{t}) \propto \prod_{i=1}^{N}\prod_{l=1}^{n_i} I\bigl\{t_{i,l} \in \lfloor t_{i,l}^{L}, t_{i,l}^{U}\rfloor\bigr\}. \tag{14}$$

Marginal posterior characteristics of the original parameter vector $\boldsymbol{\theta}$, used for inference, are indeed the same, irrespectively whether they are obtained from $p(\boldsymbol{\psi}, \boldsymbol{\theta}\,|\,\mathrm{data}) \propto L(\boldsymbol{\psi}, \boldsymbol{\theta}) \times p(\boldsymbol{\psi}, \boldsymbol{\theta})$ or from $p(\boldsymbol{\theta}\,|\,\mathrm{data})$, eq. (9), which is also equal to $\int p(\boldsymbol{\psi}, \boldsymbol{\theta}\,|\,\mathrm{data})d\boldsymbol{\psi}$.

3.4. Markov chain Monte Carlo

To determine the posterior distribution, we used the Gibbs sampling algorithm (Geman and Geman [24]). The majority of the full conditional distributions are identical to those given in Ref. [9] and we refer the reader therein. The remaining full conditional distributions pertain to the random effects $\boldsymbol{b}_i$ $(i = 1, \dots, N)$, the means of random effects $\boldsymbol{\gamma}$ and the covariance matrix $\mathbb{D}$ of the random effects. However, they either have a multivariate normal or an inverse-Wishart distribution. Details are given in the Appendix.

The complexity of the MCMC to be used requires control of the calculation which is difficult to be achieved by a general purpose software (e.g., WinBUGS). For this reason, an R (R Development Core Team [25]) package bayesSurv, freely available from the *Comprehensive R Archive Network* on `http://www.R-project.org`, has been written to sample from the posterior distribution of the model parameters (function bayessurvreg2) and draw the inference (e.g., function predictive2). Detailed description on how to use the package for the analysis shown in Section 4 is included in the documentation to the package.

3.5. Inference on the model parameters

For each component of the parameter vector $\boldsymbol{\theta}$ we derive summary statistics of the posterior distribution $p(\boldsymbol{\theta}\,|\,\mathrm{data})$, obtained from the MCMC sample, $(\boldsymbol{\psi}^{(m)},\,\boldsymbol{\theta}^{(m)})$ $(m = 1, \dots, M)$. For example, the posterior median values are approximated by the MCMC sample medians. Highest posterior density (HPD) intervals are derived to express the uncertainty with which the parameter is estimated.

To draw inference on the transformed parameter (vector) $\zeta(\boldsymbol{\theta})$, we use the posterior distribution $p\bigl\{\zeta(\boldsymbol{\theta})\,\big|\,\mathrm{data}\bigr\}$ and the corresponding MCMC sample $\zeta\bigl(\boldsymbol{\theta}^{(m)}\bigr)$ $(m = 1, \dots, M)$. For example, in the context of the AFT model, rather than reporting the results for the fixed effects $\beta_1, \dots, \beta_s$ or the means $\gamma_1, \dots, \gamma_q$ of the random effects, we prefer reporting of the acceleration factors $e^{\zeta_1}, \dots, e^{\zeta_s}$, or $e^{\gamma_1}, \dots, e^{\gamma_q}$, respectively. Indeed, these quantities express how much a unit change in the covariate accelerates ($e^{\beta} < 1$) or decelerates ($e^{\beta} > 1$) the reference event time.

3.6. Inference on the survival distribution

When interest lies in the survival distribution for a specific combination of covariates $\boldsymbol{x}_{pred}$ and $\boldsymbol{z}_{pred}$, we can compute the predictive survival function $S(t\,|\,\mathrm{data},\,\boldsymbol{x}_{pred},\,\boldsymbol{z}_{pred})$, or the predictive hazard function $\hbar(t\,|\,\mathrm{data},\,\boldsymbol{x}_{pred},\,\boldsymbol{z}_{pred})$ $(t > 0)$ from the MCMC output. The procedure is analogous, with only an obvious change in notation, to that described in Ref. [9] (Section 4.6) and the reader is referred therein for details.

3.7. Inference on random effects

When interest lies in investigating and explaining the heterogeneity, we can use the (marginal) posterior distribution $p(\boldsymbol{B}\,|\,\mathrm{data})$ of the random effects $\boldsymbol{b}_1, \dots, \boldsymbol{b}_N$, which is obtained from the joint posterior distribution $p(\boldsymbol{\psi},\,\boldsymbol{\theta}\,|\,\mathrm{data})$ by integrating out the remaining parameters. When an MCMC sample from the joint posterior distribution is available, integration is achieved by simply ignoring these remaining parameters in the sample.

4. THE ANALYSIS OF THE DFS TIME IN EARLY BREAST CANCER PATIENTS

For the analysis of the DFS time in early breast cancer patients in the EORTC trial 10854, we fitted two random effects AFT models (1). In both models, we included the following covariates: age group (*<40, 40 – 50, >50 years*), type of prior surgery (*mastectomy, breast conserving*), tumor size (*not palpable or <2 cm, ≥2 cm*), axillary nodal status (*negative, positive*), presence of other related disease (*no, yes*). The first AFT model (**Model with region**) contained also dummies for the geographical location, whereas in the second AFT model (**Model without region**), the geographical location was not included in the covariate vector for fixed effects. Since centers are nested within geographical regions it should be possible to reveal, at least partially, the regional structure of the centers from the estimates of the center-specific random effects $b_{1,1}, \ldots, b_{N,1}$ in the model without region.

For inference we sampled a chain of length 125 000 with 1:5 thinning which took about 2.5 hour on a Pentium IV 2 GHz PC with 512 MB RAM. The last 25 000 iterations of the chain were used to derive the summary statistics which are shown in Tables I and II.

<Table I about here.>
<Table II about here.>

4.1. Effect of covariates and the survival distribution

Table I shows the posterior summaries for the acceleration factors revealing the effect of considered covariates in both models. It is seen that the DFS time in the *control* arm is approximately 0.86 times shorter than in the *POP FAC* arm, or reversible, it is approximately $1/0.86 \approx 1.16$ times longer in the *POP FAC* arm compared to the *control* arm. Based on the model with region included, the DFS time for the middle age group (*40 – 50 years*) is by a factor of 1.38 longer than the DFS time of the youngest group (<*40 years*). For the patients from the oldest group (>*50 years*), the DFS time is by a factor of 1.33 longer than the DFS time of the youngest group. In the group where *breast conserving surgery* was used, the DFS time is longer by a factor of 1.26 compared to the *mastectomy* group. Further, in the bigger tumors (≥*2 cm*) group, the DFS time is shorter by a factor of 0.63 compared to the smaller tumors (<*2 cm*) group. In a group with the *positive* pathological nodal status the DFS time is shorter by a factor of 0.55 compared to a group with the *negative* result. In a group where other related disease is *present*, the DFS time is shorter by a factor of 0.72. From the regional effects it is, for example, seen that *South Africa* performs far worse than all remaining regions.

In the model without region, the effect of the included covariates is estimated to be practically the same as in the model with region. This illustrates, among other things, the general property that the AFT model is robust towards omission of important covariates (Hougaard [26]). Although, probably the reason for this phenomenon is that the omitted covariate region is strongly related to the center indicator which is nested within the region. Consequently, region remains partially or completely included in both models. A complete view on the distribution of the DFS time is given in Figure 2, which shows the predictive hazard and survival functions in the POP FAC and control arm when fixing remaining covariates at their reference values.

<Figure 2 about here.>

4.2. Random effects, heterogeneity and the error density

Figure 4 shows the posterior medians and 95% HPD intervals for acceleration factors based on the center-specific random effects in both models considered. For comparison purposes, the plot related to the random intercepts $b_{i,1}$, ($i = 1, \ldots, 14$) takes also into account the fixed effect of a geographical region in the model with region explicitly included. In the left part of Figure 4, *France* serves as a reference region (model with region) whereas an average over all regions serves as a reference in the right part of Figure 4 (model without region). This causes an overall shift when going from left to right in the upper panel of Figure 4. However, besides that shift, the structure of the posterior medians of the random intercepts is quite similar in both models. That is, the random intercepts in the model without region were to a large extent able to capture the effect of the region.

<Figure 4 about here.>

As one could have expected, omission of the covariate region led to the increase of the variability of the random intercept. Namely, its standard deviation, estimated by the posterior median of $\sqrt{d_{1,1}}$, increased from 0.111 to 0.302, the 95% HPD interval for $\sqrt{d_{1,1}}$ changed from (0.015, 0.292) to (0.142, 0.513). The lower panel of Figure 4 shows that the heterogeneity of the treatment effect between centers is of a lower magnitude than the baseline heterogeneity. This is also seen on the posterior medians of the parameter $\sqrt{d_{2,2}}$, standard deviation of $b_{i,2}$ ($i = 1, \ldots, 14$) which equals to 0.057 in the model with region and to a slightly higher value of 0.074 in the model without region, respectively. Most importantly, all increase of the variability caused by the omission of the important covariate (region) was captured by the variance components of the random effects. The residual variability, which has a direct impact on the precision with which the effect of the covariates is evaluated, remains practically the same, see the row labeled as sd(ε) in Table II.

From the negative posterior estimate (median) of the correlation coefficient between the two random effects we might conclude that patients in centers which perform relatively bad, benefit more from the treatment than patients treated in better performing centers. However, the HPD interval for the correlation is very wide, covering practically (−1, 1) and hence not much can be concluded from the analysis including or excluding the region. Finally, Figure 3 shows a pointwise posterior mean of the error density g_ε which was modeled as the penalized Gaussian mixture.

<Figure 3 about here.>

5. DISCUSSION

We have introduced here a possible approach to perform a regression analysis with survival clustered data dealing with a heterogeneity between clusters (centers). Both the baseline heterogeneity, as well as the heterogeneity with respect to the effect of selected covariates has been considered. The heterogeneity has been taken into account by including the random effects in the AFT model. Parametric assumptions concerning the baseline survival distribution have been avoided by using the penalized Gaussian mixture as a model for the error terms in the AFT model. In general, parametric models lead to a higher precision in the estimation of the regression coefficients. Our approach can serve as an exploratory tool to choose an appropriate parametric model. For example, the analysis of Section 4 might continue with an AFT model with a skewed error distribution suggested by Figure 3.

Not surprisingly, we have illustrated also that the random effects are capable to reveal much of the structure in the data arising from an omitted covariate. In fact, the random effects are able to capture practically all the variability caused by omission of important covariates and this leads to an improved precision of the estimated regression coefficients.

Earlier, Legrand et al. [27] analyzed the same clinical outcome of the EORTC trial 10854 using a frailty PH model. By considering a *fixed* treatment effect and a *random* center effect their objective was to quantify heterogeneity in outcome over centers. However, they did not include a treatment by center interaction and therefore did not account for a possible heterogeneity in the treatment effect between centers. With respect to the baseline heterogeneity between centers, we have found, as in Ref. [27], that it is largely explained by the geographical differences. Legrand et al. [27] investigated heterogeneity in DFS between centers and interpreted it in terms of density of predicted 5-year DFS rates over centers. A similar comparison was performed in this paper using the posterior summaries of the acceleration factors based on the center-specific random effects. Similar results were found by both methods for the baseline heterogeneity. Our model can estimate the heterogeneity by inspecting the diagonal elements of the matrix $\mathbb{D}$ (variances of random effects). However, a formal test for heterogeneity is not possible with this approach. In our analysis, Figure 1 reveals that the treatment by center interaction, while not large, cannot be automatically ruled out.

Finally, we have opted for a random effects approach to model the effect of centers. The choice between a fixed effects and a random effects model is still a matter of debate, see, e.g., Anello et al. [3] and both approaches have their merits. However, we do admit that there is no general agreement on this, see also Glidden and Vittinghoff [28]. The choice of a fixed approach, though, allows to test statistically whether the heterogeneity of the baseline and treatment effect by including a center by treatment interaction in the covariate vector $\boldsymbol{x}_{i,l}$ for fixed effects. On the other hand, as mentioned in the CPMP Guideline [29], assessment of interaction terms based on statistical significance tests is of little value since (a) these tests often lack statistical power and the absence of statistical evidence of an interaction is not evidence that there is no clinically relevant interaction, (b) conversely, an interaction cannot be considered as relevant on the sole basis of a significant test for interaction.

ACKNOWLEDGEMENTS

We thank the editor and the anonymous referees for their valuable comments that led to a considerable improvement of the paper.

The research of the first author was performed in the framework of the postdoctoral mandate PDM/06/242 financed by the Research Funds of Katholieke Universiteit Leuven. The authors further acknowledge support from the Interuniversity Attraction Poles Program P6/03 – Belgian State – Federal Office for Scientific, Technical and Cultural Affairs. The authors thank the European Organisation for Research and Treatment of Cancer (EORTC) Breast Cancer Group for permission to use the data from EORTC trial 10854 for this research. This research project was further supported by grant number 5U10 CA11488-35 from the US National Cancer Institute (Bethesda, Maryland, USA). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

REFERENCES

1. P. C. Clahsen, C. J. van de Velde, J. P. Julien, J. L. Floiras, T. Delozier, F. Y. Mignolet, and T. M. Sahmoud. Improved local control and disease-free survival after perioperative chemotherapy for early-stage breast cancer. A European Organization for Research and Treatment of Cancer Breast Cancer Cooperative Group Study. *Journal of Clinical Oncology*, 14:745–753, 1996.
2. J. A. van der Hage, C. J. H. van de Velde, J.-P. Julien, J.-L. Floiras, T. Delozier, C. Vandervelden, and L. Duchateau. Improved survival after one course of perioperative chemotherapy in early breast cancer patients: long-term results from the European Organization for Research and Treatment of Cancer (EORTC) Trial 10854. *European Journal of Cancer*, 37:2184–2193, 2001.
3. C. Anello, R. T. O'Neill, and S. Dubey. Multicenter trials: a US regulatory perspective. *Statistical Methods in Medical Research*, 14:303–318, 2005.
4. D. R. Cox. Regression models and life-tables (with Discussion). *Journal of the Royal Statistical Society, Series B*, 34:187–220, 1972.

5. D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.
6. J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data.* John Wiley & Sons, Chichester, Second edition, 2002.
7. N. Reid. A conversation with Sir David Cox. *Statistical Science*, 9:439–455, 1994.
8. A. Komárek, E. Lesaffre, and J. F. Hilton. Accelerated failure time model for arbitrarily censored data with smoothed error distribution. *Journal of Computational and Graphical Statistics*, 14:726–745, 2005.
9. A. Komárek and E. Lesaffre. Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. *To appear in Journal of the American Statistical Association*, 2007.
10. N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.
11. W. Pan and T. A. Louis. A linear mixed-effects model for multivariate censored data. *Biometrics*, 56:160–166, 2000.
12. W. Pan and J. E. Connett. A multiple imputation approach to linear regression with clustered censored data. *Lifetime Data Analysis*, 7:111–123, 2001.
13. A. Komárek and E. Lesaffre. Bayesian accelerated failure time model for correlated censored data with a normal mixture as an error distribution. *Statistica Sinica*, 17:549–569, 2007.
14. P. Lambert and P. H. C. Eilers. Bayesian proportional hazards model with time-varying regression coefficients: A penalized Poisson regression approach. *Statistics in Medicine*, 24:3977–3989, 2005.
15. T. Kneib. Mixed model based inference in geoadditive hazard regression for interval censored survival times. *Computational Statistics and Data Analysis*, 51:777–792, 2006.
16. T. Kneib and L. Fahrmeir. A mixed model approach for geoadditive hazard regression. *Scandinavian Journal of Statistics*, 34:207–228, 2007.
17. N. Keiding, P. K. Andersen, and J. P. Klein. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16:215–225, 1997.
18. P. Lambert, D. Collett, A. Kimber, and R. Johnson. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine*, 23:3177–3192, 2004.
19. P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties (with Discussion). *Statistical Science*, 11:89–121, 1996.
20. G. Wahba. Bayesian "confidence intervals" for the cross–validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, 45:133–150, 1983.
21. C. P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer-Verlag, New York, Second edition, 2004.
22. H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications.* Chapman & Hall/CRC, Boca Raton, 2005.
23. M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–550, 1987.
24. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayes restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
25. R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0.
26. P. Hougaard. Fundamentals of survival data. *Biometrics*, 55:13–22, 1999.
27. C. Legrand, L. Duchateau, R. Sylvester, P. Janssen, J. A. van der Hage, C. J. H. van de Velde, and P. Therasse. Heterogeneity in disease free survival between centers: lessons learned from an EORTC breast cancer trial. *Clinical Trials*, 3:10–18, 2006.
28. D. V. Glidden and E. Vittinghoff. Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine*, 23:369–388, 2004.
29. Committee for Proprietary Medicinal Products. *Points to Consider on Adjustment for Baseline Covariates.* The European Agency for the Evaluation of Medical Products, Evaluation of Medicines for Human Use, London, May 2003. CPMP/EWP/2863/99.

APPENDIX: MARKOV CHAIN MONTE CARLO

In appendix, we provide the full conditional distributions for the random effects $\boldsymbol{b}_i$ $(i = 1, \ldots, N)$, the means of random effects $\boldsymbol{\gamma}$ and the covariance matrix $\mathbb{D}$ of the random effects. For convenience in the notation, we will assume that $z_{i,l,1} \equiv 1$, in which case γ_1 is constantly equal to zero for identifiability reasons.

Namely,

$$\boldsymbol{b}_i \mid \cdots \sim \mathcal{N}\Bigl(\mathrm{E}(\boldsymbol{b}_i \mid \cdots), \mathrm{var}(\boldsymbol{b}_i \mid \cdots)\Bigr), \qquad i = 1, \dots, N, \tag{15}$$

with

$$\begin{aligned}\mathrm{E}(\boldsymbol{b}_i \mid \cdots) &= \mathrm{var}(\boldsymbol{b}_i \mid \cdots) \times \Bigl[\mathbb{D}^{-1}\boldsymbol{\gamma} + (\sigma\,\tau)^{-2}\sum_{l=1}^{n_i}\boldsymbol{z}_{i,l}\bigl\{\log(t_{i,l}) - \alpha - \boldsymbol{\beta}'\boldsymbol{x}_{i,l} - \tau\,\mu_{r_{i,l}}\bigr\}\Bigr],\\ \mathrm{var}(\boldsymbol{b}_i \mid \cdots) &= \Bigl\{\mathbb{D}^{-1} + (\sigma\,\tau)^{-2}\sum_{l=1}^{n_i}\boldsymbol{z}_{i,l}\,\boldsymbol{z}'_{i,l}\Bigr\}^{-1}.\end{aligned}$$

Further, let $\boldsymbol{\nu}_{(-1)}$ be the vector of prior means of $\boldsymbol{\gamma}_{(-1)} = (\gamma_2, \dots, \gamma_q)'$ and $\mathbb{U}_{(-1)}$ be a diagonal matrix having prior variances of $\boldsymbol{\gamma}_{(-1)}$ on the diagonal. Let $\mathbb{V}_{(-1)}$ and $\mathbb{V}_{(-1,1)}$ be the $(2, \dots, q)$-$(2, \dots, q)$ block and the $(2, \dots, q)$-1 block, respectively, of the matrix $\mathbb{D}^{-1}$. Finally, let $\boldsymbol{b}_{i(-1)} = (b_{i,2}, \dots, b_{i,q})'$ $(i = 1, \dots, N)$. Then

$$\boldsymbol{\gamma}_{(-1)} \mid \cdots \sim \mathcal{N}\Bigl(\mathrm{E}(\boldsymbol{\gamma}_{(-1)} \mid \cdots), \mathrm{var}(\boldsymbol{\gamma}_{(-1)} \mid \cdots)\Bigr), \tag{16}$$

with

$$\begin{aligned}\mathrm{E}(\boldsymbol{\gamma}_{(-1)} \mid \cdots) &= \mathrm{var}(\boldsymbol{\gamma}_{(-1)} \mid \cdots) \times \Bigl(\mathbb{U}_{(-1)}^{-1}\boldsymbol{\nu}_{(-1)} + \mathbb{V}_{(-1)}\sum_{i=1}^{N}\boldsymbol{b}_{i(-1)} + \mathbb{V}_{(1,-1)}\sum_{i=1}^{N}b_{i,1}\Bigr),\\ \mathrm{var}(\boldsymbol{\gamma}_{(-1)} \mid \cdots) &= \Bigl(\mathbb{U}_{(-1)}^{-1} + N\,\mathbb{V}_{(-1)}\Bigr)^{-1}.\end{aligned}$$

Finally,

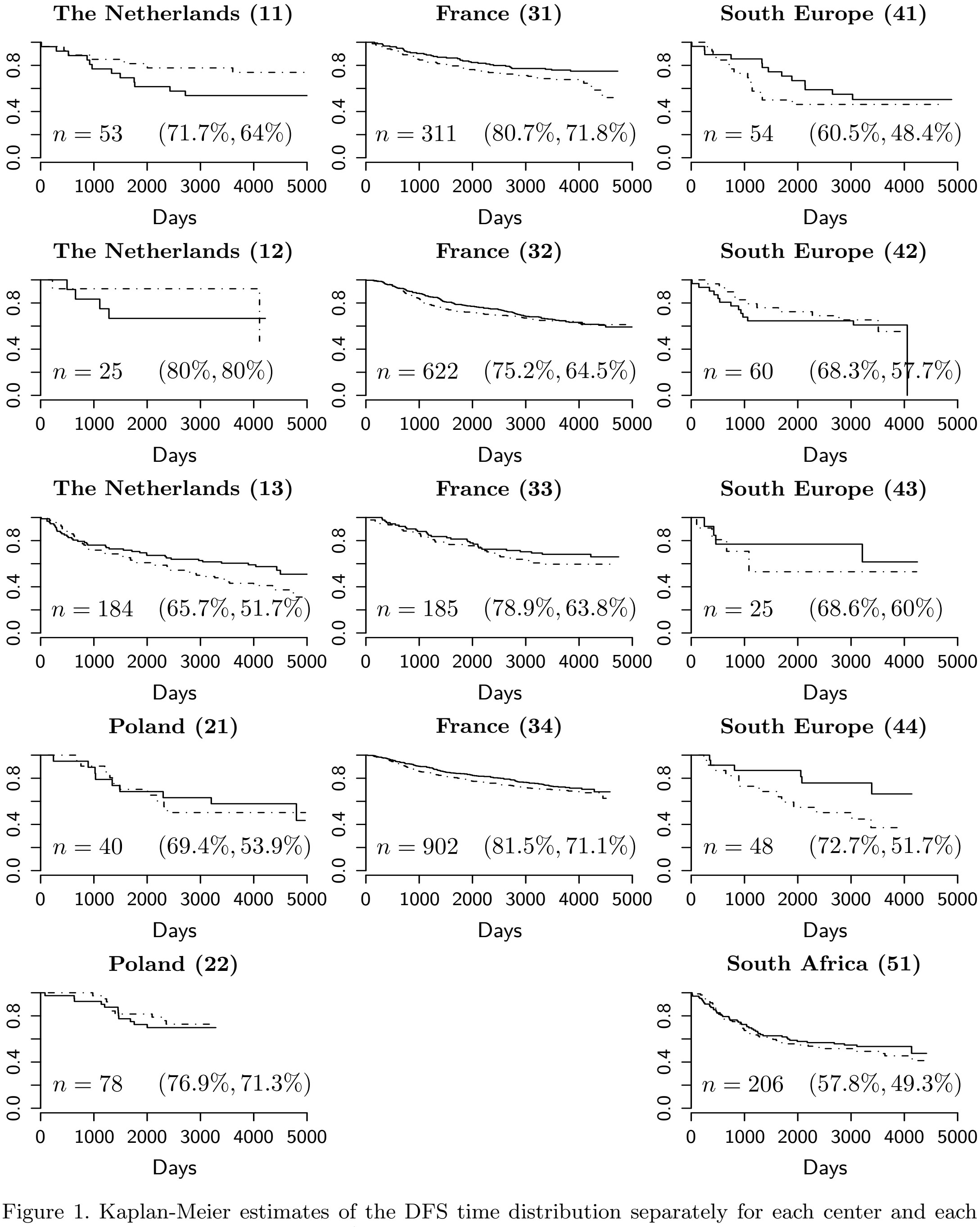
$$\mathbb{D} \mid \cdots \sim \text{inverse-Wishart}\left(df_b + N,\ \mathbb{S}_b + \sum_{i=1}^{N}(\boldsymbol{b}_i - \boldsymbol{\gamma})(\boldsymbol{b}_i - \boldsymbol{\gamma})'\right). \tag{17}$$

Table I. Posterior medians and 95% highest posterior density (HPD) intervals for the acceleration factors ($\exp(\gamma)$ and $\exp(\beta)$ parameters).

	Model with region		Model without region	
Effect	Posterior median	95% HPD interval	Posterior median	95% HPD interval
Treatment group (reference: *POP FAC arm*)				
control arm	0.858	(0.712, 1.010)	0.860	(0.729, 1.009)
Age group (reference: *<40 years*)				
40 – 50 years	1.384	(1.035, 1.762)	1.411	(1.064, 1.819)
>50 years	1.330	(1.019, 1.656)	1.368	(1.061, 1.738)
Type prior surgery (reference: *mastectomy*)				
breast conserving	1.257	(1.041, 1.483)	1.281	(1.070, 1.509)
Tumor size (reference: *<2 cm*)				
≥2 cm	0.630	(0.521, 0.748)	0.625	(0.515, 0.745)
Nodal status (reference: *negative*)				
positive	0.549	(0.461, 0.635)	0.546	(0.459, 0.639)
Other disease (reference: *absent*)				
present	0.724	(0.538, 0.930)	0.716	(0.536, 0.926)
Region (reference: *France*)				
The Netherlands	0.669	(0.457, 0.943)		
Poland	1.417	(0.845, 2.154)		
Southern Europe	0.713	(0.465, 1.007)		
South Africa	0.479	(0.295, 0.700)		

Table II. Posterior medians and 95% highest posterior density intervals for the moments of the error distribution and variance components of random effects.

	Model with region		Model without region	
Effect	Posterior median	95% HPD interval	Posterior median	95% HPD interval
Moments of the error distribution				
$\mathrm{E}(\varepsilon)$	9.155	(8.763, 9.513)	8.967	(8.559, 9.340)
$\mathrm{sd}(\varepsilon)$	1.481	(1.341, 1.640)	1.470	(1.345, 1.628)
Variance components of the random effects				
$\sqrt{d_{1,1}} = \mathrm{sd}(b_1)$	0.111	(0.015, 0.292)	0.302	(0.142, 0.513)
$\sqrt{d_{2,2}} = \mathrm{sd}(b_2)$	0.057	(0.014, 0.180)	0.074	(0.015, 0.212)
$\frac{d_{1,2}}{\sqrt{d_{1,1}\,d_{2,2}}} = \mathrm{corr}(b_1,\,b_2)$	−0.219	(−0.997, 0.939)	−0.675	(−0.998, 0.967)



The Netherlands (11)

France (31)

South Europe (41)

The Netherlands (12)

France (32)

South Europe (42)

The Netherlands (13)

France (33)

South Europe (43)

Poland (21)

France (34)

South Europe (44)

Poland (22)

South Africa (51)

Figure 1. Kaplan-Meier estimates of the DFS time distribution separately for each center and each treatment group. Solid line: POP FAC arm, dotted-dashed line: control arm. Further, we report the sample size and overall DFS proportion after 5 years and 10 years per center.

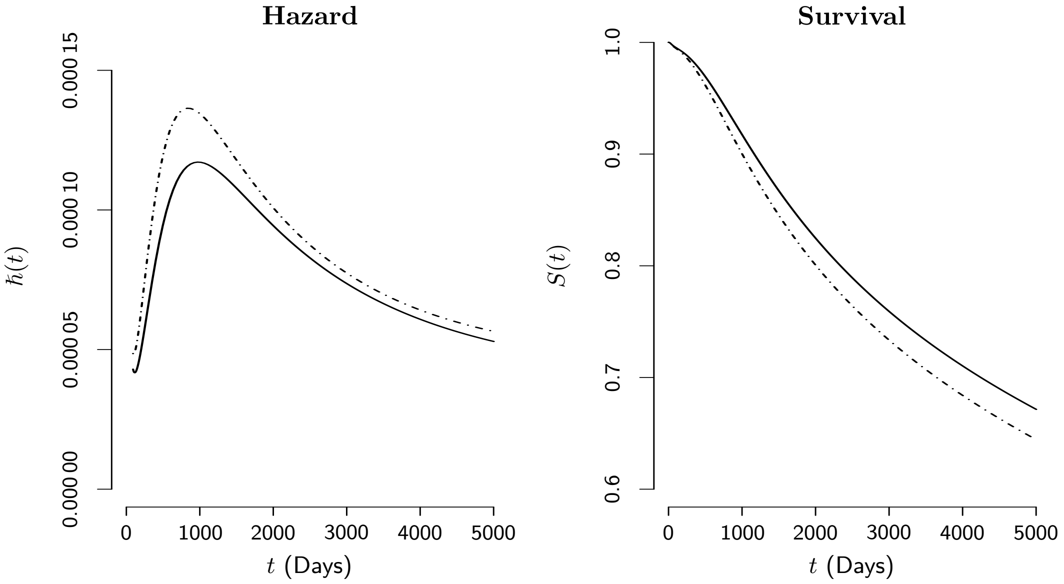


Figure 2. Model with region. Predictive hazard and survival function for the POC FAC arm (solid line) and control arm (dotted-dashed line) and remaining covariates fixed to the reference values.

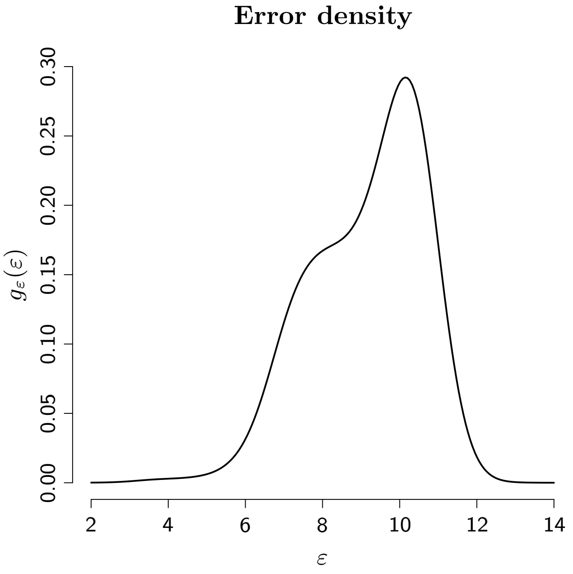


Figure 3. Model with region. Pointwise posterior mean of the error density g_ε. The density g_ε involves a shift by the intercept α and a scaling by the parameter τ and thus it is not standardized.

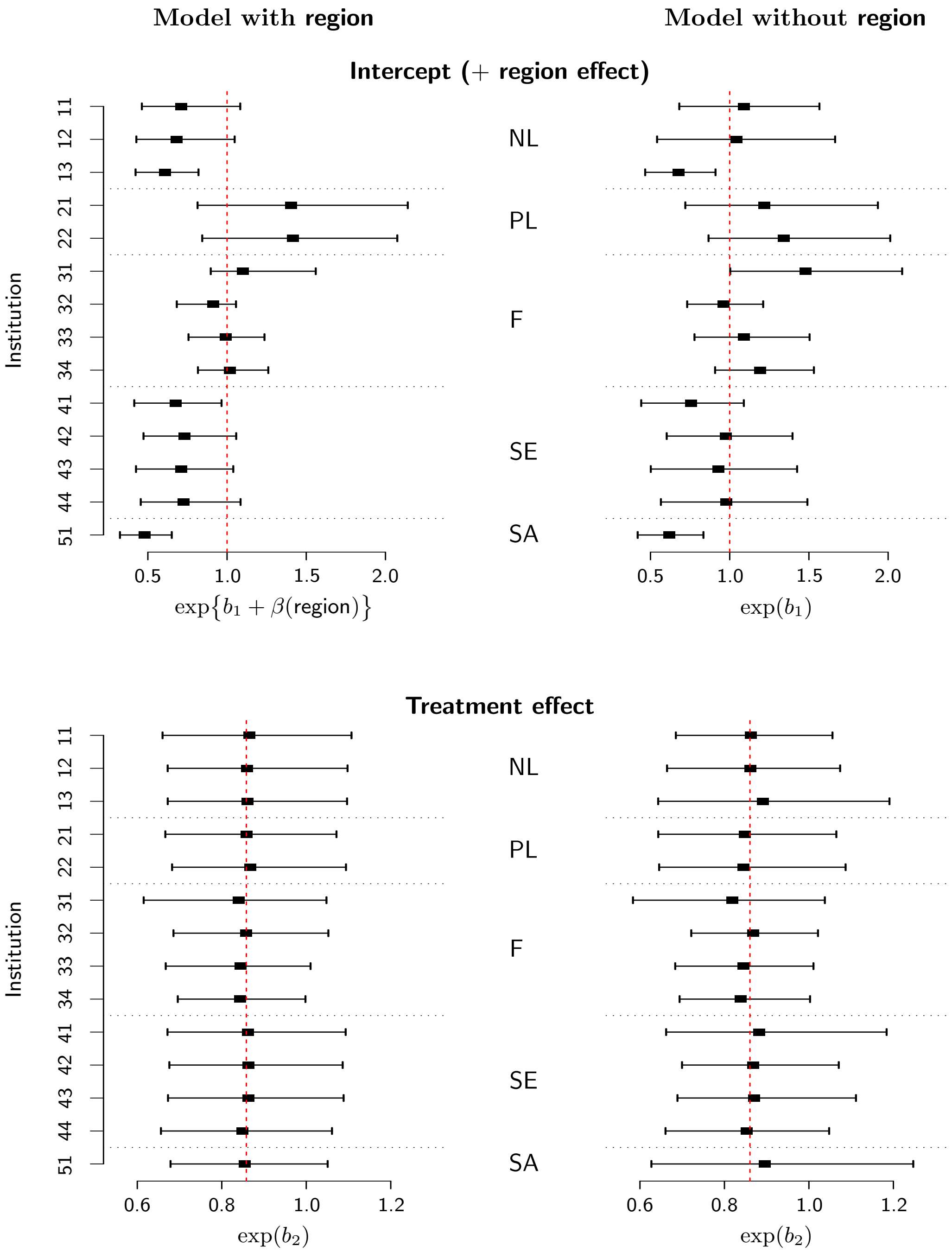


Figure 4. Posterior medians and 95% highest posterior density intervals for center-specific random effects based acceleration factors. Random intercepts in the model with region are further shifted by a corresponding region main effect $\beta(\text{region})$.