A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model

Catherine Legrand¹,*,†, Vincent Ducrocq², Paul Janssen³, Richard Sylvester¹ and Luc Duchateau⁴

¹European Organisation for Research and Treatment of Cancer, Brussels, Belgium
²Institut National de la Recherche Agronomique, Station de Génétique Quantitative et Appliquée, Jouy-en-Josas, France
³Center for Statistics, Hasselt University, Agoralaan, 3590 Diepenbeek, Belgium
⁴Department of Physiology, Biochemistry and Biometrics, Faculty of Veterinary Medicine, Ghent University, Belgium

SUMMARY

When multicentre clinical trial data are analysed, it has become more and more popular to look for possible heterogeneity in outcome between centres. However, beyond the investigation of such heterogeneity, it is also interesting to consider heterogeneity in treatment effect over centres. For time-to-event outcomes, this may be investigated by including a random centre effect and a random treatment by centre interaction in a Cox proportional hazards model.

Assuming independence between the random effects, we propose a Bayesian approach to fit our proposed model. The parameters of interest are the variance components $\sigma_0^2$ and $\sigma_1^2$ of these random effects, which can be interpreted as a measure of centre and treatment effect over centres heterogeneity of the hazard. These variance components are estimated from their marginal posterior density after integrating out the fixed treatment effect and the random effects. As this integration cannot be performed analytically, the marginal posterior density is approximated using the Laplace integration technique. Statistical inference is then based on the characteristics of the posterior marginal density, such as the mode and the standard deviation. We demonstrate the proposed technique using data from a pooled database of seven EORTC bladder cancer clinical trials. Substantial centre and treatment effect over centres heterogeneity in disease-free interval was found. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: frailty model; proportional hazards; random treatment by centre interaction; Bayesian inference; Laplace integration; bladder cancer

1. INTRODUCTION

The Cox proportional hazards model with a random centre effect (frailty model) can be used to investigate heterogeneity in time-to-event outcome over centres when considering...
data from large multicentre cancer clinical trials [1]. Assuming a one-parameter gamma distribution (with mean 1) for the random centre effect, such model can be fitted using the expectation–maximization (EM) algorithm, or equivalently, using the penalized partial likelihood approach [2]. It is also of interest to look at the heterogeneity in hazard due to treatment by centre interaction. As proposed by Yamaguchi and Ohashi [3] this may be done by adding a random interaction between the treatment and centre effect in the frailty model. Assuming a normal distribution of the random centre effect and the random treatment by centre interaction, they estimated this model using an extension of the McGilchrist approach [4] to accommodate for the two random effects.

We propose in this paper an alternative approach, computationally less intensive, and therefore more convenient, especially for large databases. Assuming that the random centre effect and the random interaction are independent and follow a normal distribution with variance, respectively, $\sigma^2_c$ and $\sigma^2_{ij}$, we extend the Bayesian approach originally proposed by Ducrocq and Casella [5] to allow the joint estimation of the variance components of the two random effects. Following the Bayesian paradigm, the marginal posterior distribution, obtained after integrating out the fixed and random effects from the joint posterior density, is considered to contain all the information on the parameters of interest. As this integration cannot be performed analytically, we use the Laplace integration technique [6] to approximate the marginal posterior density. The estimates of the variance components of the random effects are then provided by the mode of this approximate marginal posterior density. If needed, further information, such as the standard deviation or the skewness of this marginal posterior density can then be obtained from this approximate marginal posterior density.

This Bayesian estimation approach has been implemented by extending The Survival Kit [7, 8], a package of Fortran programs developed in the field of animal genetics, to jointly estimate the variance components of two normally distributed random effects and the first three moments of the approximate marginal posterior density. Using this software we investigate heterogeneity in disease-free interval (DFI) due to centre and treatment by centre interaction in a large bladder cancer database including data from seven randomized clinical trials conducted by the Genito-Urinary Group of the European Organisation for Research and Treatment of Cancer (EORTC).

Section 2 describes the model and introduces some notations. The estimation technique is summarized in Section 3. The Survival Kit and its extension are discussed in Section 4 while Section 5 presents the results obtained for the bladder data set, with a discussion of the results. In Section 6, we present results of simulations run in a setting similar to the bladder data set. Section 7 contains some concluding remarks.

2. FRAILTY MODEL WITH A RANDOM CENTRE EFFECT AND A RANDOM TREATMENT BY CENTRE INTERACTION

In the following, we assume that we have data from a total of $N$ patients coming from $G$ different centres, $n_i$ patients coming from centre $i (N = \sum_{i=1}^{G} n_i)$. For the $j$th patient in the $i$th centre, we observe $Y_{ij} = \min(T_{ij}, C_{ij})$ where $T_{ij}$ is the time-to-event for this patient (possibly right censored), and $C_{ij}$ is the censoring time independent of $T_{ij}$. Additionally, a censoring indicator $\delta_{ij}$ is observed, with $\delta_{ij}$ equal to 1 if $Y_{ij} = T_{ij}$, otherwise 0.
For each patient, we also observe the binary variable $x_{ij}$ representing the treatment arm to which the patient has been randomized with $x_{ij} = 0$ if the patient is in the standard arm and $x_{ij} = 1$ if the patient is in the experimental arm.

We consider a Cox proportional hazards model including a fixed treatment effect, a random centre effect and a random treatment by centre interaction. With such model, the hazard for the $j$th patient in the $i$th centre is given by

$$\lambda_{ij}(t) = \lambda_0(t) \exp(b_0i + (\beta + b_{1j})x_{ij})$$  \hspace{1cm} (1)

where $\lambda_0(t)$ represents the unspecified baseline hazard at time $t$, $\beta$ is the fixed treatment effect coefficient and the random effects $b_{0i}$ and $b_{1j}$ are assumed to follow a particular distribution with mean 0. The variance–covariance matrix of the vector of random effects $\mathbf{b}^T = (b_{01}^T, b_{11}^T) = (b_{01}, b_{0G}, b_{11}, \ldots, b_{1G})$ is denoted by $\mathbf{V}(\mathbf{b})$.

In this model, $b_{0i}$ can be interpreted as the influence of the $i$th centre on the overall underlying baseline risk, patients treated in a centre with a value of $b_{0i}$ above (resp. below) 0 having a higher (resp. lower) risk. Similarly, $b_{1j}$, the random interaction term, can be interpreted as the influence of the $i$th centre on the overall treatment effect ($\beta$). The variance components of the random effects $\sigma_0^2$ and $\sigma_1^2$ can be interpreted as a measure of centre and treatment effect over centres heterogeneity of the hazard.

3. ESTIMATION: A BAYESIAN APPROACH

The Bayesian approach proposed by Ducrocq and Casella [5] to estimate mixed survival models is extended here. The variance components $\mathbf{\theta}^T = (\sigma_0^2, \sigma_1^2)$ of the random effects are estimated from their marginal posterior distribution after integrating out $\beta$ and $\mathbf{b}$. This integration cannot be performed analytically and we therefore approximate the marginal posterior distribution using the Laplace integration technique. Tierney and Kadane [6] provide a good insight in the accuracy of the Laplace approximation method; see also Monahan [9, Section 12.6].

The details are as follows.

Applying the Bayes theorem, the joint posterior density for model (1) is proportional to

$$\pi(\beta, \mathbf{b}, \mathbf{\theta} \mid \mathbf{y}) \propto L(\beta, \mathbf{b} \mid \mathbf{y}) \times \pi_0(\mathbf{b} \mid \mathbf{\theta}) \times \pi_0(\beta) \times \pi_0(\mathbf{\theta})$$

Considering a Cox model, the likelihood $L(\beta, \mathbf{b} \mid \mathbf{y})$ is given in terms of the partial likelihood function [10, 11]

$$L(\beta, \mathbf{b} \mid \mathbf{y}) = \prod_{i=1}^G \prod_{j=1}^{n_i} \left[ \frac{\exp(b_{0i} + (\beta + b_{1j})x_{ij})}{\sum_{k \geq t_{ij}} \exp(b_{0k} + (\beta + b_{1k})x_{kl})} \right]^{\delta_{ij}}$$

The second factor is the joint prior distribution of the random effects. This joint prior distribution of the random effects is assumed to be multivariate normal with independence between all random effects and is thus given by

$$\pi_0(\mathbf{b} \mid \mathbf{\theta}) = \prod_{i=1}^G \frac{1}{2\pi\sigma_0\sigma_1} \exp \left( -\frac{1}{2} \left( \frac{b_{0i}^2}{\sigma_0^2} + \frac{b_{1j}^2}{\sigma_1^2} \right) \right)$$  \hspace{1cm} (2)
The third and fourth factors represent the prior distribution for $\beta$ and $\theta$ which we assume to be flat

$$\pi_0(\theta) \propto 1 \quad \text{and} \quad \pi_0(\beta) \propto 1$$

The log-joint posterior density is then given by

$$\ln \pi(\beta, \theta | y) \propto \sum_{i=1}^G \sum_{j=1}^{m_i} \delta_{ij} \left[ b_{0i} + (\beta + b_{1i})x_{ij} - \ln \sum_{k \neq j} \exp(b_{0k} + (\beta + b_{1k})x_{k_j}) \right]$$

$$-G \ln(2\pi \sigma_0 \sigma_1) - \frac{1}{2} \sum_{i=1}^G \left( \frac{b_{0i}^2}{\sigma_0^2} + \frac{b_{1i}^2}{\sigma_1^2} \right)$$

Remark at this point that the term posterior ‘density’ is in fact used here for convenience, acknowledging that it is obtained using the partial likelihood and not the full likelihood. While inspired by Bayesian reasoning, our approach however leaves the baseline hazard totally unspecified, and based on the justification provided by Ibrahim et al. [10] and Sinha et al. [11], we use a partial-likelihood-based expression for the likelihood of the observations.

According to the Bayesian principle, statistical inference on $\theta$ should be based on its marginal posterior density obtained by integrating out the nuisance parameters $\beta$ and $b$ from the joint posterior density

$$\pi(\theta | y) = \int \int \pi(\beta, \theta, b | y) \, d\beta \, db$$

This integral cannot be solved analytically. Ducrocq and Casella [5] proposed to approximate the integral for a particular value $\theta^*$ of $\theta$ by Laplacian integration. Fixing the value of $\theta^*$, we thus denote $\pi(\beta, b, \theta^* | y) = \pi(\beta, b | y, \theta^*)$ and we can write

$$\int \int \pi(\beta, b | y, \theta^*) \, d\beta \, db = \int \int \exp(\ln(\pi(\beta, b | y, \theta^*))) \, d\beta \, db$$

In short, the Laplacian integration consists of replacing $\ln(\pi(\beta, b | y, \theta^*))$ by the first terms of its Taylor series expansion around the mode of the joint posterior density function $\pi(\beta, b | y, \theta^*)$ given by

$$\hat{\Psi}_{\theta^*} = (\hat{\beta}_{\theta^*}, \hat{b}_{\theta^*}^T) = \text{Arg}_{\Psi} \max \pi(\Psi | y, \theta^*)$$

where $\Psi = (\beta, b^T)^T$.

At the mode, the gradient vector equals zero

$$\left( \frac{\partial \ln \pi(\Psi | y, \theta^*)}{\partial \Psi} \right)_{\Psi=\hat{\Psi}_{\theta^*}} = 0$$

and the second term in the Taylor series expansion therefore cancels. For the third term of the expansion we need the negative Hessian at $\hat{\Psi}_{\theta^*}$

$$\hat{\mathbf{H}}_{\theta^*} = \left( \frac{-\partial^2 \ln \pi(\Psi | y, \theta^*)}{\partial \Psi \partial \Psi^T} \right)_{\Psi=\hat{\Psi}_{\theta^*}}$$
The integral is then approximately given by

\[
\pi(\theta^* | y) \approx \int \exp \left( \ln \pi(\Psi_0^* | y, \theta^*) - \frac{1}{2}(\Psi - \hat{\Psi}_0^*)^T \hat{H}_0^* (\Psi - \hat{\Psi}_0^*) \right) \, d\Psi
\]

(4)

Furthermore, we can write

\[
\int (2\pi)^{-(1/2)(G+1)}|\hat{H}_0^{-1}|^{-1/2} \exp \left( -\frac{1}{2}(\Psi - \hat{\Psi}_0^*)^T \hat{H}_0^* (\Psi - \hat{\Psi}_0^*) \right) \, d\Psi = 1
\]

as it corresponds to the kernel of a multivariate normal density function with mean \(\hat{\Psi}_0^*\) and variance \(\hat{H}_0^{-1}\), and thus the approximation of the integral can be rewritten as

\[
\pi(\theta^* | y) \approx (2\pi)^{G+1/2}|\hat{H}_0^{-1}|^{1/2} \pi(\Psi_0^* | y, \theta^*)
\]

(5)

Taking the logarithm on both sides, we obtain the following approximation of the log-marginal posterior distribution

\[
\ln \pi(\theta^* | y) = \text{constant} + \ln \pi(\Psi_0^* | y, \theta^*) - 0.5 \ln |\hat{H}_0^*|
\]

In the two-dimensional space of the two variance components, we use the Simplex algorithm [12] with \(\hat{\sigma}_0^2\) and \(\hat{\sigma}_1^2\) as parameters and the approximated marginal posterior density (5) as function to identify the values which maximize this approximated marginal posterior density. Once these values are found, they are used as estimates of the variance components \(\hat{\sigma}_0^2\) and \(\hat{\sigma}_1^2\) of the two random effects.

Apart from the mode, which will provide us with estimates of \(\hat{\sigma}_0^2\) and \(\hat{\sigma}_1^2\), other characteristics based on the marginal posterior density, such as the skewness or credible sets, might be of interest. As mentioned above, the Simplex algorithm is used to obtain the mode of the marginal posterior density. Other characteristics such as the first three moments can be obtained by numerical integration based on the Gauss–Hermite quadrature [13].

4. THE SURVIVAL KIT

The Survival Kit V3.12 [7, 8] is a package of Fortran programs developed by Ducrocq and Sölkner in the field of animal genetics to analyse survival models with random effect(s) on large databases. Freely available from internet (http://www.boku.ac.at/nuwi/software/sofskit.htm), The Survival Kit can fit either semi-parametric (Cox) or parametric (Weibull) models, including continuous and/or discrete covariates (possibly time dependent) as well as random effects (normal, multivariate normal, or log-gamma distributed) and can be used on extremely large databases.

However, the present version of The Survival Kit is not capable of estimating two variance components simultaneously. Due to this restriction the maximization of the marginal posterior density is a one-dimensional problem solved by the bisection method. The maximization of the joint posterior density to estimate the fixed and random effects coefficients \(\beta\) and \(\mathbf{b}\) is implemented through a limited memory quasi-Newton method [14] which only requires the computation of the vector of first derivatives. The approximate Cholesky factor of the Hessian
of the function to maximize, used to determine the next quasi-Newton step, is stored in a very sparse form.

To fit the extended model discussed in the previous section, we needed to extend the existing software so that joint estimation of the two variance components is possible. The maximization of the marginal posterior density now takes place in a two-dimensional space and we therefore implemented the Simplex algorithm [12] to seek the mode of this function. We also extended The Survival Kit so that information on the moments of the approximated posterior marginal density is provided.

5. BLADDER CANCER DATABASE

Bladder cancer is a common urological malignancy and about 70–80 per cent of all bladder cancers are superficial (stage Ta-T1). Standard treatment typically consists of transurethral resection (TUR) conducted with the aim of removing all the tumours. However, a high proportion of patients will experience recurrences or progression to muscle invasive disease, even after complete resection. Therefore, randomized controlled phase III trials have been conducted over the last decades to investigate the use of prophylactic treatment following TUR. The objective of such treatment is both to remove residual, unresectable lesions and to prevent recurrence after complete resection.

In this case study, we consider the individual patient data from 2649 eligible bladder cancer patients randomized by 63 European centres in 7 consecutive phase III randomized clinical trials conducted by the Genito-Urinary Group of the European Organization for Research and Treatment of Cancer (EORTC 30781, 30782, 30791, 30831, 30832, 30845 and 30863) [15–20]. All these patients had Ta-T1 bladder cancer, approximately half with primary bladder cancer and half with recurrent disease. Within the context of these trials, patients in each of these participating centres were treated with or without further intravesical treatment after TUR. In total, 1204 patients (45.5 per cent) received no further intravesical treatment while 1445 patients (54.5 per cent) received further intravesical treatment. Our analysis is based on DFI defined as time from randomization to the date of the first bladder recurrence, censoring the patients without recurrence at the date of last available follow-up cytscopy. Considering this endpoint, a total of 1223 (46.2 per cent) events were observed, with an overall median DFI of about 2.8 years. The DFI was significantly longer in the intravesical-treatment group (HR: 0.85 [95 per cent CI: 0.76–0.95], \( p = 0.0053 \)).

To investigate heterogeneity in outcome and in treatment effect, we have restricted the analysis to the centres which accrued more than 20 patients (although The Survival Kit experiences no difficulty in fitting the model using all centres), because small centres do not contribute much information. We therefore include in our analysis a total of 2292 patients, 1004 (43.8 per cent) in the no-intravesical-treatment group and 1288 (56.2 per cent) in the intravesical-treatment group from 35 centres in nine European countries. The number of patients per centre varies from 21 to 249 with a median 52 and mean 65. Within this subset of patients, the major baseline characteristics were in general well balanced over the two groups, with slightly more patients with multiple tumours and patients with Ta disease in the intravesical-treatment group.

A total of 1218 patients (53.1 per cent) were considered as censored for DFI and 1074 patients had a recurrence. The number of events over centres ranged from 7 to 117, with
Table I. Treatment effect in various Cox PH models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate (SE)</th>
<th>Hazard ratio [95%CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>−0.1890 (0.0612)</td>
<td>0.828 [0.734–0.933]</td>
</tr>
<tr>
<td>Model 2</td>
<td>−0.0130 (0.0696)</td>
<td>0.987 [0.861–1.131]</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.0117 (0.0693)</td>
<td>1.012 [0.883–1.159]</td>
</tr>
<tr>
<td>Model 4</td>
<td>−0.1011 (0.0925)</td>
<td>0.904 [0.754–1.084]</td>
</tr>
</tbody>
</table>

Model 1: Fixed treatment effect.
Model 2: Fixed treatment effect + stratified by centre.
Model 3: Fixed treatment effect + fixed centre effect.
Model 4: Fixed treatment effect + random centre and random centre*treatment effects.

median 21 and mean 31. Using a Cox PH model including only a fixed effect for treatment (Model 1) in this subset of patients, the DFI remained significantly longer in the intravesical-treatment group (HR: 0.83 [95 per cent CI:0.73–0.93], \textit{p}-value = 0.0020) with a median DFI of 2.2 years (95 per cent CI: 1.9–2.6) in the no-intravesical-treatment group and of 3.3 years (95 per cent CI: 2.6–4.4) in the intravesical-treatment group. Treatment effect estimates when stratifying for centres (Model 2) or when considering centres as fixed effects (Model 3) are presented in Table I. These two models lead to rather similar results, with in both cases a treatment effect estimate much closer to 0 and a 95 per cent confidence interval for the hazard ratio including 1. However, with such a large number of centres and rather low number of events in most centres these models are probably overparametrized. Furthermore, Model 2 does not provide any information about centres, and therefore does not help us in investigating a potential centre effect.

We finally fit the model discussed in Section 2 including a fixed treatment effect, a random centre effect and a random treatment by centre interaction (Model 4, corresponding to equation (1)). The estimate of the fixed treatment effect is in between the results obtained by ignoring the centre effect (Model 1) and by stratifying for centre or introducing it as fixed effect (Models 2 and 3) (Table I). Model 4 makes the best use of the available information while Model 1 does not consider the centre effect and Models 2 and 3 are only based on intra-centre comparisons. In Figure 1, the predicted centre baseline risks (exp(\(b_{0i}\))) are plotted along the horizontal axis. Note that the predicted values in Figure 1 are obtained from (3). These values represent the deviation of the \(i\)th centre from the overall baseline hazard. The variance of this random centre effect, which can be interpreted as a measure of the heterogeneity in DFI over centres induced by the centre effect, is estimated to be 0.10854. Similarly, the predicted treatment effect for each particular centre, i.e. exp(\(\beta + b_{1i}\)) is plotted on the same figure along the vertical axis. The variance of the random treatment by centre interaction is estimated to be 0.10860 and can be interpreted as a measure of the heterogeneity due to treatment by centre interaction.

These results seem to indicate that there is substantial heterogeneity in DFI, both due to centre as to differing treatment effects over centres. However, these variance components are difficult to interpret as they refer to variability of the log-hazard rate [21].

To better understand the value we obtained for \(\sigma_{\sigma}^2\), one possibility is to look at the impact of such a value on the spread of the median DFI from centre to centre in the no-intravesical-treatment group (\(x = 0\)) by considering the density function of \(M_c\), the median DFI in this
Figure 1. Predicted baseline risk $\exp(b_{0i})$ (horizontal axis) versus predicted treatment effect $\exp(\beta + b_{1i})$ (vertical axis) for each centre ($i = 1, \ldots, 35$).

Figure 2. Density function of the median DFI in the no-intravesical-treatment group over centres, assuming either a constant baseline hazard (straight line) or a Weibull baseline hazard (dotted line).

As can be expected, the expression of this density function depends on the form of the baseline hazard. Assuming a constant baseline hazard $\lambda_0(t) = \lambda$, normally distributed random centre effects and $\alpha = 0$, we show in the appendix that this density

The assumption of constant baseline hazard is however often not realistic for cancer clinical trial data and can be relaxed by considering, for example, a Weibull baseline hazard \( \lambda_0(t) = \alpha \lambda t^{\gamma - 1} \). In such case, the density function of \( M_c \) is given by

\[
 f_{M_c}(m_c) = \frac{1}{m_c \sqrt{2\pi \sigma_0^2}} \exp \left( -\frac{1}{2 \sigma_0^2} \left( \ln \left( \frac{2}{\lambda m_c} \right) \right)^2 \right)
\]  

(6)

This formula provides more insight in the effect of a particular value of \( \sigma_0^2 \). By considering this density function and its tails, for example, the 5 and 95 per cent quantiles, we get an immediate interpretation of the impact of particular values of the variance of the random centre effect in the no-intravesical-treatment group. This density function is depicted in Figure 2, assuming either a constant yearly baseline hazard of 0.3151 (as estimated from the bladder cancer data) or a Weibull baseline hazard with parameter \( \alpha = 0.711 \) and a yearly \( \lambda \) equal to 0.3383 (as estimated from the bladder cancer data). When allowing the baseline hazard to vary over time, we see that 90 per cent of the centres would have a median DFI in the no-intravesical-treatment group between 1.3 and 5.9 years.

The same argument as above allows one to consider the impact of a particular value of \( \sigma_1^2 \) on the spread of ‘treatment effect’ or hazard ratio \( \exp(\beta + b_{1i}) \) over centres. In the case of normally distributed centre by treatment interactions \( b_{1i} \), the density function of the hazard
Figure 4. Density function of median DFI in the intravesical-treatment group over centres, considering: (a) ‘good outcome centres’; (b) ‘average outcome centres’; and (c) ‘poor outcome centres’.
ratio HR over centres is given by

\[
    f_{HR}(h) = \frac{1}{h \sqrt{2\pi \sigma_1^2}} \exp \left( -\frac{1}{2\sigma_1^2} (\ln h - \beta)^2 \right)
\]

(8)

Considering a \( \beta \) value of \(-0.1011\) as in our database, 90 per cent of the centres would have a HR between 0.53 and 1.56 (Figure 3).

The difficulty is clearly to interpret \( \sigma_0^2 \) and \( \sigma_1^2 \) simultaneously. One possibility is to ‘sum up’ the impact of these two normally distributed random effects and plot the density of the median DFI over centres, considering then heterogeneity from these two sources of variability together. However, this would not be very informative for physicians who are clearly interested in distinguishing variability at the level of the baseline risk and of the treatment effect. We could rather choose ‘typical’ values for the random centre effects, and for these fixed values of \( b_0 \) consider the median DFI in the no-intravesical-treatment group and the spread of median DFI over centres induced by the heterogeneity in treatment effects over centres. For example, we can consider the 25th, 50th and 75th quantile of \( b_0 \sim N(0, \sigma_0^2) \), i.e. \(-0.222, 0, \) and \(0.222)\) as representing, respectively, ‘good outcome centre’, ‘average outcome centre’ and ‘poor outcome centre’. Assuming, as above, a yearly Weibull baseline hazard with parameters \( \lambda = 0.711 \) and \( \lambda = 0.3383 \), this would correspond to centres having a median DFI in the no-intravesical-treatment group (\( x=0 \)) of, respectively, 3.7, 2.7 and 2.0 years. With \((\beta + b_1) \sim N(-0.1011, 0.10860)\), the density function of \( M_e \), the median DFI over centres in the intravesical-treatment group (\( x=1 \)) for these three ‘typical’ examples of centres is given by

\[
    f_{Me}(m_e) = \frac{m_e}{m_e \sqrt{2\pi \sigma_1^2}} \exp \left( -\frac{1}{2\sigma_1^2} \left( \ln \left( \frac{2}{\lambda m_e} \right) - b_0 - \beta \right)^2 \right)
\]

and the density functions are plotted in Figure 4.

6. EVALUATION OF THE PERFORMANCE OF THE PROPOSED METHOD

We performed simulations to evaluate the performance of our proposed method in a setting similar to the one encountered in the previous section, namely large multicentre bladder cancer clinical trials. To simulate data, values of the baseline event rate \( \lambda_0(t) \) (assumed to be constant over time \( \lambda_0(t) = \lambda \), of the treatment effect, \( \beta \), and the accrual and follow-up times were chosen to resemble the bladder cancer data set. We also considered the same number of centres, \( G \), and patients per centres, \( n_i = 1, \ldots, G \), as in our database and assume an equal balance of treatment groups over centres. Different values of the heterogeneity parameters were considered in these simulations, including the case where no heterogeneity is present for one or both random effects.

Considering disease-free interval, we used a yearly constant baseline event rate of \( \lambda = 0.315 \) and a treatment effect of \( \beta = -0.1890 \). We considered an accrual period of 1065 days (±35 months) and a follow-up time of 3440 days (±113 months). Time at risk, \( rt_{ij} \), for a particular patient consists of the time at risk before the end of the accrual period (assuming a constant entry rate over the accrual period) plus the follow-up time. This results in approximately 90 per cent of the patients having an event at the time of the analysis. We considered 35 centres of
Table II. Results of the simulations: bias, median estimated values over the 250 fits, empirical standard error over the 250 fits and median model-based standard errors over the 250 fits.

<table>
<thead>
<tr>
<th>Real value</th>
<th>Bias</th>
<th>Median</th>
<th>Empirical std</th>
<th>Model-based std</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = -0.189$</td>
<td>0.0013</td>
<td>$-0.1858$</td>
<td>0.0466</td>
<td>0.0443</td>
</tr>
<tr>
<td>$\sigma_0^2 = 0$</td>
<td>0.0012</td>
<td>0.0000</td>
<td>0.0024</td>
<td>0.0056</td>
</tr>
<tr>
<td>$\sigma_1^2 = 0$</td>
<td>0.0026</td>
<td>0.0000</td>
<td>0.0050</td>
<td>0.0115</td>
</tr>
<tr>
<td>$\beta = -0.189$</td>
<td>$-0.0093$</td>
<td>$-0.1742$</td>
<td>0.0704</td>
<td>0.0635</td>
</tr>
<tr>
<td>$\sigma_0^2 = 0$</td>
<td>0.0021</td>
<td>0.0000</td>
<td>0.0033</td>
<td>0.0083</td>
</tr>
<tr>
<td>$\sigma_1^2 = 0.08$</td>
<td>$-0.0128$</td>
<td>0.0625</td>
<td>0.0258</td>
<td>0.0383</td>
</tr>
<tr>
<td>$\beta = -0.189$</td>
<td>0.0098</td>
<td>$-0.1793$</td>
<td>0.0677</td>
<td>0.0683</td>
</tr>
<tr>
<td>$\sigma_0^2 = 0.04$</td>
<td>$-0.0016$</td>
<td>0.0362</td>
<td>0.0180</td>
<td>0.0206</td>
</tr>
<tr>
<td>$\sigma_1^2 = 0.08$</td>
<td>0.0016</td>
<td>0.0792</td>
<td>0.0360</td>
<td>0.0447</td>
</tr>
<tr>
<td>$\beta = -0.189$</td>
<td>0.0032</td>
<td>$-0.1975$</td>
<td>0.0466</td>
<td>0.0454</td>
</tr>
<tr>
<td>$\sigma_0^2 = 0.08$</td>
<td>$-0.0013$</td>
<td>0.0757</td>
<td>0.0228</td>
<td>0.0323</td>
</tr>
<tr>
<td>$\sigma_1^2 = 0$</td>
<td>0.0054</td>
<td>0.0000</td>
<td>0.0089</td>
<td>0.0200</td>
</tr>
<tr>
<td>$\beta = -0.189$</td>
<td>$-0.0029$</td>
<td>$-0.1890$</td>
<td>0.0622</td>
<td>0.0582</td>
</tr>
<tr>
<td>$\sigma_0^2 = 0.08$</td>
<td>0.0024</td>
<td>0.0785</td>
<td>0.0286</td>
<td>0.0334</td>
</tr>
<tr>
<td>$\sigma_1^2 = 0.04$</td>
<td>0.0007</td>
<td>0.0375</td>
<td>0.0255</td>
<td>0.0330</td>
</tr>
<tr>
<td>$\beta = -0.189$</td>
<td>0.0104</td>
<td>$-0.1850$</td>
<td>0.0728</td>
<td>0.0567</td>
</tr>
<tr>
<td>$\sigma_0^2 = 0.08$</td>
<td>0.0002</td>
<td>0.0787</td>
<td>0.0283</td>
<td>0.0342</td>
</tr>
<tr>
<td>$\sigma_1^2 = 0.08$</td>
<td>0.0012</td>
<td>0.0743</td>
<td>0.0372</td>
<td>0.0455</td>
</tr>
</tbody>
</table>

respective size 21, 23, 24, 26, 30, 30, 34, 35, 35, 35, 39, 42, 42, 43, 44, 44, 42, 52, 52, 55, 56, 61, 62, 63, 66, 73, 85, 86, 92, 104, 117, 120, 155, 183, 249 to reflect the patient distribution in the actual bladder cancer data. In each centre, half of the patients were randomly assigned to be in the no-treatment group.

For each parameter, 250 data sets were generated in S-plus-2000 from model (1), assuming a constant baseline hazard. Given a particular setting $(G; n; \lambda, \sigma_0^2, \sigma_1^2, \beta)$, the observations of a data set were generated in the following way. First, $G$ random centre effects, $b_{01}, \ldots, b_{0G}$ and $G$ interaction random effects $b_{11}, \ldots, b_{1G}$ were independently generated from a normal distribution (2). The time-to-event outcome for each patient, $t_{ij}$, was randomly generated from an exponential family distribution with parameter $\lambda_{ij}$ given by (1). A patient for which the time to event was longer than the time at risk was censored with time to censoring equal to time at risk so that $t_{ij} = \min (rt_{ij}, et_{ij})$ and $\delta_{ij} = I (rt_{ij} > et_{ij})$ is the censoring indicator.

For each parameter setting $(G; n; \lambda, \sigma_0^2, \sigma_1^2, \beta)$, our model was fitted using the modified version of The Survival Kit described in Section 4, allowing for joint estimation of the two variance components. In Table II, we report for each setting, the bias, the median, the empirical and the model-based standard deviation computed over the 250 fits for the fixed treatment effect and variance of each random effect.

These simulations show a good performance of our model, with only a small bias for all estimates whatever the setting. Model-based standard errors, obtained as the square root of the second moments of the joint posterior density, should be interpreted, according to the Bayesian paradigm, as one of the parameters characterizing this joint posterior density.
this distribution appears to be substantially skewed, the model-based standard error (in fact the standard error of the marginal posterior density of \( \theta \)) overestimates the standard error of the heterogeneity parameter as can be seen from the empirical standard error. Therefore, we advice to use the whole marginal posterior density rather than the model-based standard deviation for the construction of credible sets.

These results seems to indicate that our estimation approach, based on the Laplace approximation, is sufficiently accurate when applied to setting similar to the one we considered in our real data application.

7. CONCLUDING REMARKS

Although clinical trial protocols are written with the objective of suppressing as much variability as possible, it becomes more and more popular to apply frailty model methodology to consider heterogeneity due to centres within large multicentre clinical trials. However, apart from this source of heterogeneity, one further interesting step is to also consider the potential heterogeneity due to treatment by centre interaction. With this work, we demonstrate that data from large cancer multicentre clinical trials can be used to investigate heterogeneity due to these two sources.

We describe a Bayesian approach implemented in The Survival Kit for the estimation of a frailty model with two random effects. One random centre effect deals with the deviation of each centre from the overall baseline hazard while a random treatment by centre interaction deals with deviation of each centre from the overall treatment effect. This Bayesian approach allows one to estimate the variance of these random effects by maximizing the posterior marginal distribution after integrating out the fixed treatment effect and the random effects using the Laplace technique.

Our approach is quite simple and uses Laplace approximation instead of numerical integration techniques, which are usually time consuming. Results of the simulations show a good performance of the methods while only limited computation time is required. In their application of such a model in the meta-analysis setting, Smith et al. [22], who used an improved version of the SAS macro made available by Prof. Yamaguchi [3], mentions a computation time of 16 or 96 h (depending on the method used to handle ties) while considering a simpler data set than the one we consider (1225 patients and 5 clusters while we have 2292 patients and 35 clusters). Our approach, as implemented in our modified version of The Survival Kit, typically takes only a few minutes to fit such model on large clinical trial databases.

Another advantage of this methodology is that it can be easily adapted to the assumption of the log-gamma distribution for the random effects. In that case, the log-gamma distribution needs to be plugged into the joint posterior density and exactly the same approach can be used for estimation.

Considering a database of seven consecutive Ta-T1 bladder cancer phase III randomized trials, our analysis shows that there exists substantial heterogeneity over centres both in terms of DFI and treatment effect. We present the results obtained when assuming a normal distribution of the centre and interaction random effects. However, assuming a log-gamma distribution leads to very similar results.

Our methodology could be further generalized by adding a correlation term \( \rho \) between the two random effects within a centre and then maximize the marginal posterior density in a
three-dimensional space \( \Theta = (\sigma_0^2, \sigma_1^2, \rho) \). Such a model might better reflect clinical reality as we might expect that the beneficial effect of a new treatment will be larger in ‘good outcome centres’ while lower in ‘poor outcome centres’. The plot of predicted values of \( \exp(b_0) \) versus \( \exp(\beta + b_1i) \) (Figure 1), as well as the correlation between these values equal to 0.51, give some hints that such a correlation between the random effects might indeed exist in our data. We are currently working on this extension of our methodology and on implementing this into The Survival Kit.

**APPENDIX A**

Considering Model (1) with the random effects \( b^i = (b_0^i, b_1^i) \) distributed according to (2), the density of \( M_c \), the median time to event in the control arm \((x = 0)\) over centres, is given by

\[
f_{M_c}(m_c) = \frac{1}{m_c \sqrt{2\pi \sigma_0^2}} \exp \left( -\frac{1}{2\sigma_0^2} \left( \ln \left( \frac{2}{\lambda m_c} \right) \right)^2 \right) \tag{A1}\]

when assuming a constant baseline hazard \((\lambda_0(t) = \lambda)\) and by

\[
f_{M_c}(m_c) = \frac{\lambda}{m_c \sqrt{2\pi \sigma_0^2}} \exp \left( -\frac{1}{2\sigma_0^2} \left( \ln \left( \frac{2}{\lambda m_c} \right) \right)^2 \right) \tag{A2}\]

when assuming a Weibull baseline hazard \((\lambda_0(t) = \lambda t^{\alpha-1})\).

Indeed, the conditional survival curve is given by

\[
S(t \mid b_0, b_1, x) = \exp(-\lambda t \exp(b_0 + (\beta + b_1)x)) \quad \text{when assuming a constant baseline hazard}
\]

\[
S(t \mid b_0, b_1, x) = \exp(-\lambda t^\alpha \exp(b_0 + (\beta + b_1)x)) \quad \text{when assuming a Weibull baseline hazard}
\]

and therefore the median time to event in the control arm satisfies \( S(M_c \mid b_0, b_1, x) = 0.5 \), or (with \( x = 0 \) in the control group)

\[
M_c = h(b_0) = \frac{\ln 2}{\lambda \exp(b_0)} \quad \text{when assuming a constant baseline hazard}
\]

\[
M_c = h(b_0) = \left( \frac{\ln 2}{\lambda \exp(b_0)} \right)^{1/\alpha} \quad \text{when assuming a Weibull baseline hazard}
\]

Since \( M_c \) is a monotone transformation of \( b_0 \), we have for \( m_c \geq 0 \)

\[
f_{M_c}(m_c) = f_{b_0}(h^{-1}(m_c)) \left| \frac{d}{dm_c} h^{-1}(m_c) \right|
\]

with \( b_0 \sim N(0, \sigma_0^2) \) we easily obtain (A1) and (A2).

Similarly, considering \( \text{HR} = \exp(\beta + b_1) \) as a monotone transformation of the random variable \( b_1 \), and having in mind that \( b_1 \sim N(0, \sigma_1^2) \), the density function of \( \text{HR} \) is given by

\[
f_{\text{HR}}(h) = \frac{1}{h \sqrt{2\pi \sigma_1^2}} \exp \left( -\frac{1}{2\sigma_1^2} \left( \ln h - \beta \right)^2 \right)
\]
Under the same assumptions (and considering a Weibull baseline hazard), but fixing \( b_0 \) and \( \beta \) to a particular value, we obtain, by noting that \( \beta + b_1 \sim N(\beta, \sigma^2_1) \), that \( M_e \), the median time to event in the experimental arm \((x = 1)\) over centres is given by

\[
f_{M_e}(m_e) = \frac{\alpha}{m_e \sqrt{2\pi \sigma^2_1}} \exp \left( -\frac{1}{2\sigma^2_1} \left( \ln \left( \frac{2}{m_e} \right) - b_0 - \beta \right)^2 \right)
\]

ACKNOWLEDGEMENTS

The authors thank the European Organization for Research and Treatment of Cancer (EORTC) Genito-Urinary Group for permission to use the data from EORTC trials 30781, 30782, 30791, 30831, 30832, 30845, 30863 for this research.

This publication was supported by grant number 5U10CA11488-34 from the National Cancer Institute (Bethesda, MD, U.S.A.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.


The authors also thank the referees for helpful comments on an earlier version of this paper.

REFERENCES


