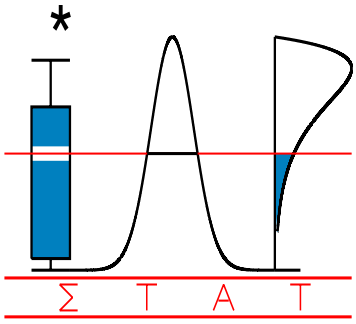
T E C H N I C A L

R E P O R T

0457

A BAYESIAN APPROACH TO JOINTLY ESTIMATE CENTER AND TREATMENT BY CENTER HETEROGENEITY IN A PROPORTIONAL HAZARDS MODEL

Catherine LEGRAND, Vincent DUCROCQ, Paul JANSSEN, Richard SYLVESTER and Luc DUCHATEAU



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

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

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

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

C. Legrand[1], V. Ducrocq[2], P. Janssen[3], R. Sylvester[1] and L. Duchateau[4]

[1] European Organisation for Research and Treatment of Cancer, Brussels, Belgium

[2] Institut National de la Recherche Agronomique,Station de Génétique Quantitative et Applique, Jouy-en-Josas, France

[3] Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, 3590 Diepenbeek, Belgium

[4] Department of Physiology, Biochemistry and Biometrics, Faculty of Veterinary Medicine, Ghent University, Belgium

Keywords: frailty model, treatment outcome, multicenter trial, EM-algorithm, penalized likelihood, REML

Abstract

When multicenter clinical trial data are analyzed, it has become more and more popular to look for possible heterogeneity in outcome between centers. However, beyond the investigation of such heterogeneity, it is also interesting to consider heterogeneity in treatment effect over centers. For time-to-event outcomes, this may be investigated by including a random center effect and a random treatment by center 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 σ_0^2 and σ_1^2 of these random effects, which can be interpreted as a measure of center and treatment effect over centers 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 center and treatment effect over centers heterogeneity in disease free interval was found.

1 Introduction

In previous work [1], we showed that the Cox proportional hazards model including a random center effect (frailty model) can be used to investigate heterogeneity in time-to-event outcome over centers when considering data from large multicenter cancer clinical trials. Assuming a one-parameter gamma distribution (with mean 1) for the random center effect, we fitted this model 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 center interaction. As proposed by Yamaguchi and Ohashi [3] this may be done by adding a random interaction between the treatment and center effect in the frailty model. Assuming a normal distribution of the random center effect and the random treatment by center interaction, they estimated this model using an extension of the McGilchrist approach [4] to accomodate for the two random effects.

We propose in this paper an alternative approach, computationaly less intensive and therefore more convenient, especially for large databases. Assuming that the random center effect and the random interaction are independent and follow a normal distribution with variance, respectively, σ_0^2 and σ_1^2, 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 due to center and treatment by center 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 notation. The estimation technique is summarized in Section 3. The Survival Kit and its extension is discussed in Section 4 while Section 5 presents the results obtained for the bladder data set, with a discussion of the results. Section 6 contains some concluding remarks.

2 Frailty model with a random center effect and a random treatment by center interaction

In the following, we assume that we have data from a total of N patients coming from G different centers, n_i patients coming from center i ($N = \sum_{i=1}^{G} n_i$). For the j^{th} patient in the i^{th} center, we observe $Y_{ij} = \min\left(T_{ij}, C_{ij}\right)$ where T_{ij} is the time-to-event for this patient (possibly right-censored), and C_{ij} is a random censoring time independent of T_{ij}. Additionally, a censoring indicator δ_{ij} is observed, with δ_{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. Without loss of generality, we will assume that $x_{ij} = 0$ 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 center effect and a random treatment by center interaction. With such model, the hazard for the j^{th} patient in the i^{th} center is given by

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

where $\lambda_0(t)$ represents the unspecified baseline hazard at time t, β is the fixed treatment effect coefficient and the random effects b_{0i} and b_{1i} are assumed to follow a particular distribution with mean 0. The variance-covariance matrix of the vector of random effects

$\boldsymbol{b}^T = \left(\boldsymbol{b}_0^T, \boldsymbol{b}_1^T\right) = (b_{01}, \ldots, b_{0G}, b_{11}, \ldots, b_{1G})$ is denoted by

$$\boldsymbol{V}(\boldsymbol{b}) = \boldsymbol{V}\begin{pmatrix} \boldsymbol{b}_0 \\ \boldsymbol{b}_1 \end{pmatrix} = \begin{pmatrix} \sigma_0^2 \boldsymbol{I}_G & \boldsymbol{0}_G \\ \boldsymbol{0}_G & \sigma_1^2 \boldsymbol{I}_G \end{pmatrix}.$$

In this model, b_{0i} can be interpeted as the influence of the i^{th} center on the overall underlying baseline risk, patients treated in a center with a value of b_{0i} above (resp. below) 0 having a higher (resp. lower) risk. Similarily, b_{1i}, the random interaction term, can be interpreted as the influence of the i^{th} center on the overall treatment effect (β). The variance components of the random effects σ_0^2 and σ_1^2 can be interpreted as a measure of center and treatment effect over centers heterogeneity of the hazard.

3 Estimation: a Bayesian approach

For estimation of the model parameters we assume

$$\boldsymbol{b} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{V}(\boldsymbol{b})) \tag{2}$$

and we define $\boldsymbol{\theta} = \left(\sigma_0^2, \sigma_1^2\right)^T$ the vector of the diagonal elements of the variance-covariance matrix.

The Bayesian approach proposed by Ducrocq and Casella [5] to estimate mixed survival models is extended here. The variance components of the random effects are estimated from their marginal posterior distribution after integrating out β and $\boldsymbol{b}^T$. In the case of a normal distribution of the random effects, this integration cannot be performed analytically and we therefore approximate the marginal posterior distribution using the Laplace integration technique [6]. The details are as follows.

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

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

In the Cox model, the likelihood $L(\beta, \boldsymbol{b} \mid \boldsymbol{y})$ is given in terms of the partial likelihood function

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

The second factor is the joint prior distribution of the random effects given by

$$\pi_0\left(\boldsymbol{b} \mid \boldsymbol{\theta}\right) = \prod_{i=1}^{G} \frac{1}{2\pi\sigma_0\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{b_{0i}^2}{\sigma_0^2} + \frac{b_{1i}^2}{\sigma_1^2}\right)\right).$$

The third and fourth factor represents the prior distribution for β and $\boldsymbol{\theta}$ which we assume to be flat

$$\pi_0\left(\boldsymbol{\theta}\right) \propto 1 \quad \text{and} \quad \pi_0\left(\beta\right) \propto 1.$$

The log joint posterior density is then given by

$$\begin{aligned} \ln \pi\left(\beta, \boldsymbol{b}, \boldsymbol{\theta} \mid \boldsymbol{y}\right) &= \sum_{i=1}^{G} \sum_{j=1}^{n_i} \delta_{ij} \left[b_{0i} + \left(\beta + b_{1i}\right) x_{ij} - \ln \sum_{t_{kl} \geq t_{ij}} \exp\left(b_{0k} + \left(\beta + b_{1k}\right) x_{kl}\right)\right] \\ &\quad -G \ln\left(2\pi\sigma_0\sigma_1\right) - \tfrac{1}{2} \sum_{i=1}^{G} \left(\tfrac{b_{0i}^2}{\sigma_0^2} + \tfrac{b_{1i}^2}{\sigma_1^2}\right). \end{aligned}$$

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.

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

$$\pi\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right) = \int\int \pi\left(\beta, \boldsymbol{b}, \boldsymbol{\theta} \mid \boldsymbol{y}\right) d\beta d\boldsymbol{b}.$$

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

$$\int\int \pi\left(\beta, \boldsymbol{b} \mid \boldsymbol{y}, \boldsymbol{\theta}^*\right) d\beta d\boldsymbol{b} = \int\int \exp\left(\ln\left(\pi\left(\beta, \boldsymbol{b} \mid \boldsymbol{y}, \boldsymbol{\theta}^*\right)\right)\right) d\beta d\boldsymbol{b}.$$

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

$$\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^*} = \left(\hat{\beta}_{\boldsymbol{\theta}^*}, \hat{\boldsymbol{b}}_{\boldsymbol{\theta}^*}\right)^T = \mathrm{Arg}_{\boldsymbol{\Psi}} \max \pi\left(\boldsymbol{\Psi} \mid \boldsymbol{y}, \boldsymbol{\theta}^*\right)$$

where $\boldsymbol{\Psi} = \left(\beta, \boldsymbol{b}\right)^T$.

At the mode, the gradient vector equals zero

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

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

$$\hat{\boldsymbol{H}}_{\boldsymbol{\theta}^{*}}=\left(\frac{-\partial^{2} \ln \pi\left(\boldsymbol{\Psi} \mid \boldsymbol{y}, \boldsymbol{\theta}^{*}\right)}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}^{T}}\right)_{\boldsymbol{\Psi}=\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}}}.$$

The integral is then approximately given by

$$\pi\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{y}\right) \approx \int \exp\left(\ln \pi\left(\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}} \mid \boldsymbol{y}, \boldsymbol{\theta}^{*}\right)-\frac{1}{2}\left(\boldsymbol{\Psi}-\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}}\right)^{T} \hat{\boldsymbol{H}}_{\boldsymbol{\theta}^{*}}\left(\boldsymbol{\Psi}-\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}}\right)\right) d \boldsymbol{\Psi}. \tag{3}$$

Furthermore, we can write

$$\int(2 \pi)^{-\frac{1}{2}(2 G+1)}\left|\hat{\boldsymbol{H}}_{\boldsymbol{\theta}^{*}}^{-1}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\boldsymbol{\Psi}-\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}}\right)^{T} \hat{\boldsymbol{H}}_{\boldsymbol{\theta}^{*}}\left(\boldsymbol{\Psi}-\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}}\right)\right) d \boldsymbol{\Psi}=1$$

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

$$\pi\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{y}\right) \approx(2 \pi)^{G+\frac{1}{2}}\left|\hat{\boldsymbol{H}}_{\boldsymbol{\theta}^{*}}^{-1}\right|^{\frac{1}{2}} \pi\left(\hat{\boldsymbol{\Psi}}_{\boldsymbol{\theta}^{*}} \mid \boldsymbol{y}, \boldsymbol{\theta}^{*}\right). \tag{4}$$

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

$$\ln \pi\left(\boldsymbol{\theta}^{*} \mid \boldsymbol{y}\right)=\text{constant}+\ln \pi\left(\hat{\Psi}_{\boldsymbol{\theta}^{*}} \mid \boldsymbol{y}, \boldsymbol{\theta}^{*}\right)-0.5 \ln \mid \hat{\boldsymbol{H}}_{\boldsymbol{\theta}^{*}} \mid .$$

In the two-dimensional space of the two variance components, we use the Simplex algorithm [9] with σ_0^2 and σ_1^2 as parameters and the approximated marginal posterior density (4) as function to identify the values which maximize this approximated marginal posterior distribution. 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 σ_0^2 and σ_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 [10].

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 Surival Kit can fit either semi-parametric (Cox) or parametric (Weibull) models, including continuous and/or discrete covariates (eventually 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 maximisation of the marginal posterior density is a one-dimensional problem solved by the bisection method. The maximisation of the joint posterior density to estimate the fixed and random effects coefficients β and $\boldsymbol{b}$ is implemented through a limited memory quasi-Newton method [11] 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 maximisation of the marginal posterior density now takes place in a two-dimensional space and we therefore implemented the Simplex algorithm [9] 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 % 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 tumors. 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 centers 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)[12, 13, 14, 15, 16, 17]. 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 centers were treated with or without further intravesical treatment after TUR. In total, 1204 patients (45.5%) received no further intravesical treatment while 1445 patients (54.5%) received further intravesical treatment.

Our analysis is based on disease free interval (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 cytoscopy. Considering this endpoint, a total of 1223 (46.2%) 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% CI: 0.76-0.95], p = 0.0053).

However, for our analysis, we have restricted our attention to the centers which accrued more than 20 patients in total. We therefore include in our analysis a total of 2292 patients, 1004 (43.8%) in the no intravesical treatment group and 1288 (56.2%) in the intravesical treatment group from 35 centers in 9 European countries. The number of patients per center varies from 21 to 249 with a median 52 and mean 65. Within this subset of patients, the major baseline characeristics (Table I) were in general well balanced over the two groups, with slightly more patients with multiple tumors and patients with Ta disease in the intravesical treatment group.

A total of 1218 patients (53.1%) were considered as censored for DFI and 1074 patients had a recurrence. The number of events over centers ranged from 7 to 117, with median 21 and mean 31. Using a Cox PH model including only a fixed effect for treatment in this subset of patients, the DFI remained significantly longer in the intravesical treatment

group (HR: 0.83 [95%CI:0.73-0.93], p-value = 0.0020) with a median DFI of 2.2 years (95%CI: 1.9-2.6) in the no intravesical treatment group and of 3.3 years (95%CI: 2.6-4.4) in the intravesical treatment group. Treatment effect estimates when stratifying for centers or when considering centers as fixed effects are presented in Table II. Results of these two models (Models 2 & 3) lead to rather similar results, with in both cases a treatment effect estimate much closer to 0 and a 95% confidence interval for the hazard ratio including 1. However, with such a large number of centers and rather low number of events in most centers these models are probably overparametrized. Furthermore Model 2 does not provide any information about centers and therefore does not help us in investigating a potential center effect.

We finally fit the model discussed in Section 2 with a fixed treatment effect, a random center effect and a random treatment by center interaction. The estimate of the fixed treatment effect is in between the results obtained by ignoring the center effect (Model 1) and by stratifying for center or introducing it as fixed effect (Models 2 & 3) (Table II). In Figure 1, the predicted center baseline risks ($\exp(b_{0i})$) are plotted along the horizontal axis. These values represent the deviation of the i^{th} centers from the overall baseline hazard. The variance of this random center effect, which can be interpreted as a measure of the heterogeneity in DFI over centers induced by the center effect, is estimated to be 0.10854. Similarily, the predicted treatment effect for each particular center, i.e. $\exp(\beta + b_{1i})$ is plotted on the same figure along the vertical axis. The variance of the random treatment by center interaction is estimated to be 0.10860 and can be interpreted as a measure of the heterogeneity due to treatment by center interaction.

These results seem to indicate that there is substantial heterogeneity in DFI, both due to center as to differing treatment effects over centers. However, these results can hardly be interpreted as such as particular values of σ_0^2 and σ_1^2 do not have a straightforward meaning.

To better understand the value we obtained for σ_0^2, one possibility is to look at the impact of such a value on the spread of the median DFI from center to center in the no intravesical treatment group ($x = 0$) by considering the density function of M_c, the median DFI in this group of patients over centers. Assuming a constant baseline hazard λ_0, normally distributed random center effects and $x = 0$, we show in the appendix that

this density function is given by

$$f_{M_c}(m_c) = \frac{1}{m_c\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}\left(\ln\left(\frac{\ln 2}{\lambda_0 m_c}\right)\right)^2\right). \quad (5)$$

By considering this density function and its tails, for example the 5% and the 95% quantiles, we get an immediate interpretation of the impact of particular values of the variance of the random center effect in the no intravesical treatment group. This density function is depicted in Figure 2, assuming a constant yearly baseline hazard of 0.3151 (as observed in our data). With these assumptions, 90% of the centers would have a median DFI in the no intravesical treatment group between 1.3 and 3.8 years.

The same argument as above allows one to consider the impact of a particular value of σ_1^2 on the spread of "treatment effect" or hazard ratio $HR = \exp(\beta + b_{1i})$ over centers. In the case of normally distributed center by treatment interactions b_{1i}, the density function of the hazard ratio HR over centers is given by

$$f_{HR}(h) = \frac{1}{h\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}(\ln h - \beta)^2\right). \quad (6)$$

Considering a β value of -0.1011 as in our database, 90% of the centers would have a HR between 0.53 and 1.56 (Figure 3).

The difficulty is clearly to interpret σ_0^2 and σ_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 centers, 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 center effects, and for these fixed values of b_{0i} consider the median DFI in the no intravesical treatment group and the spread of median DFI over centers induced by the heterogeneity in treatment effects over centers. For example, we can consider the 25^{th}, 50^{th} and 75^{th} quantile of $b_{0i} \sim N(0, \sigma_0^2 = 0.10854)$, i.e. -0.222, 0, and 0.222 as representing respectively "good center outcome", "average center outcome" and "poor center outcome". Assuming, as above, a constant yearly baseline hazard of 0.3151, this would correspond to centers having a median DFI in the no intravesical treatment group ($x = 0$) of respectively 2.7, 2.2 and 1.8 years. With $(\beta + b_{1i}) \sim N(0.3151, 0.10860)$, the density function of M_e,

the median DFI over centers in the intravesical treatment group ($x = 1$) for these three "typical" examples of centers is given by

$$f_{M_e}(m_e) = \frac{1}{m_e\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}\left(\ln\left(\frac{\ln 2}{\lambda_0 m_e}\right) - b_0 - \beta\right)^2\right) \tag{7}$$

and the density functions are plotted in Figure 4.

6 Concluding remarks

Although clinical trial protocols are written with the objective of supressing as much variability as possible, it becomes more and more popular to apply frailty model methodology to consider heterogeneity due to centers within large multicenter clinical trials. However, apart from this source of heterogeneity, one further interesting step is to also consider the potential heterogeneity due to treatment by center interaction. With this work, we demonstrate that data from large cancer multicenter 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 center effect deals with the deviation of each center from the overall baseline hazard while a random treatment by center interaction deals with deviation of each center from the overal treatment effect. This Bayesian approach allows one to estimate the variance of these random effects by maximising the posterior marginal distribution after integrating out the fixed treatment effect and the random effects using the Laplace technique.

Variance components of these random effects can be interpreted as a measure of the heterogeneity respectively in outcome and treatment effect over centers. However we show in this paper one needs to be careful in interpreting a particular value of σ_0^2 and σ_1^2 and we propose several graphical displays to better understand the impact of the variance components. This is illustrated by fitting our model to a database of seven consecutives Ta-T1 bladder cancer phase III randomized trials. Based on our results we can conclude that there exists substantial heterogeneity over centers both in terms of DFI and treatment effect.

Our methodology could be further generalized by adding a correlation term ρ between

the two random effects within a center and then maximise the marginal posterior density in a three-dimensional space $\boldsymbol{\theta} = (\sigma_0^2, \sigma_1^2, \rho)$. 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 performing centers" while lower in "poor performing centers". The plot of predicted values of the $\exp(b_{0i})$ versus $\exp(\beta + b1i)$ (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.

References

[1] Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R, Sylvester R. The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics and Data Analysis* 2002; **40**:603-620.

[2] Therneau TM, Grambsch PM. *Modeling Survival Data. Extending the Cox Model.* New-York: Springer Verlag. 2000.

[3] Yamaguchi T, Ohashi Y. Investigating centre effects in a mutli-centre clinical trial of superficial bladder cancer. *Statistics in Medicine* 1999; **18**: 1961-71.

[4] McGilchrist CA. REML estimation for survival models with frailty. *Biometrics* 1993; **49**: 221-225.

[5] Ducrocq V, Casella G. A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 1996; **28**:509-529.

[6] Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* 1986; **81**:82-86.

[7] Ducrocq V, Sölkner J. "The Survival Kit", a FORTRAN package for the analysis of survival data. In: 5th World Cong. Genet. Appl. Livest. Prod. 1994; **22**: 51-52. Dep. Anim. Poultry. Sci., Univ. of Guelph, Guelph, Ontario, Canada.

[8] Ducrocq V, Sölkner J. "The Survival Kit - V3.0", a package for large analysis of survival data. In: 6th World Cong. Genet. Appl. Livest. Prod. 1998; **27**: 447-8. Anim. Genetics and Breeding Unit, Univ. of New England, Armidale, Australia.

[9] Nelder JA, Mead R. A Simplex method for function minimization. *Computer Journal* 1965; **7**: 308-13.

[10] Gamerman D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall. 1997.

[11] Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 1989; **45**: 503-28.

[12] Newling DW, Robinson MR, Smith PH, Byar D, Lockwood R, Stevens I, De Pauw M, Sylvester R. Tryptophan metabolites, pyridoxine (vitamin B6) and their influence on the recurrence rate of superficial bladder cancer. Results of a prospective randomized phase III study performed by the EORTC GU Group. EORTC Genitourinary Tract CancerCooperative Group. *Eur. Urol.* 1995; **27**: 110.

[13] Bouffioux C, Denis L, Oosterlinck W, Viggiano G, Vergison B, Keuppens F, de Pauw M, Sylvester R, Cheuvart B. Adjuvant chemotherapy of recurrent superficial transitional cell carcinoma: results of a European Organization for Research and Treatment of Cancer randomized trial comparing intravesical instillation of thiotepa, doxorubicin and cisplatin. *J. Urol.* 1992; **148**: 297.

[14] Kurth KH, Schröder FH, Tunn U, Ay R, Pavone-Macaluso, Debruyne F, de Pauw M, Dalesio O, ten Kate F. Adjuvant treatment of superficial transitional cell bladder carcinoma: preliminary results of an European Organization for Research and Treatment of Cancer randomized trial comparing doxorubicin hydrochloride, ethoglucid and transurethral resection alone. *J. Urol.* 1984; **132**: 258.

[15] Bouffioux C, Kurth KH, Bono A, Oosterlinck W, Kruger CB, de Pauw M, Sylvester R. Intravesical adjuvant chemotherapy for superficial transitional cell bladder carcinoma: Results of two European Organization for Research and Treatment of Cancer randomized trials with mitomycin C and doxorubicin comparing early versus delayed

instillations and short-term versus long-term treatment. European Organization for Research and Treatment of Cancer Genitourinary Group. *J. Urol.* 1995; **153**: 934.

[16] Witjes JA, van der Meijden AP, Sylvester RC, Debruyne FM, van Aubel A, Witjes WP. Long term follow up of an EORTC randomized prospecive trial comparing intravesical bacille Calmette-Gurin-RIVM and Mitomycin C in superficial bladder cancer. *Urology* 1998. **52**: 403.

[17] Oosterlinck W, Kurth KH, Schröder FH, Bultinck J, Hammond B, Sylvester R. A prospective European Organization for Research and Treatment of Cancer Genitourinary Group randomized trial comparing transurethral resection followed by a single intravesical instillation of epirubicin or water in single stage TaT1 papillary carcinoma of the bladder. *J. Urol.* 1993; **149**: 749.

Table 1: Baseline characteristics

	No intravesical treatment N = 1004		Intravesical treatment N = 1288	
	N	(%)	N	(%)
Gender				
Male	781	(77.8%)	1047	(81.3%)
Female	219	(81.8%)	224	(17.4%)
Missing	4	(0.4%)	17	(1.3%)
Tumor status				
Primary	545	(54.3%)	689	(53.5%)
Reccurent	455	(45.3%)	582	(45.2%)
Missing	4	(0.4%)	17	(1.3%)
Number of tumors				
Single	615	(61.3%)	653	(50.7%)
Multiple	383	(38.1%)	617	(47.9%)
Missing	6	(0.6%)	18	(1.4%)
T category				
Ta	500	(49.8%)	766	(59.5%)
T1	490	(48.8%)	483	(37.5%)
Missing	14	(1.4%)	39	(3.0%)
Grade				
Grade G1	437	(43.5%)	503	(39.1%)
Grade G2	425	(42.5%)	557	(43.2%)
Grade G3	98	(9.8%)	154	(12.0%)
Missing	44	(4.4%)	74	(5.7%)

Table 2: Treatment effect in various Cox PH models

Model	Estimate	(se)	Hazard ratio	[95%CI]
Model 1	-0.1890	(0.0612)	0.828	[0.734-0.933]
Model 2	-0.0130	(0.0696)	0.987	[0.861-1.131]
Model 3	0.0117	(0.0693)	1.012	[0.883-1.159]
Model 4	-0.1011	(0.0925)	0.904	[0.754-1.084]

Model 1: Fixed treatment effect

Model 2: Fixed treatment effect + stratified by center

Model 3: Fixed treatment effect + fixed center effect

Model 4: Fixed treatment effect + random center and random center*treatment effects

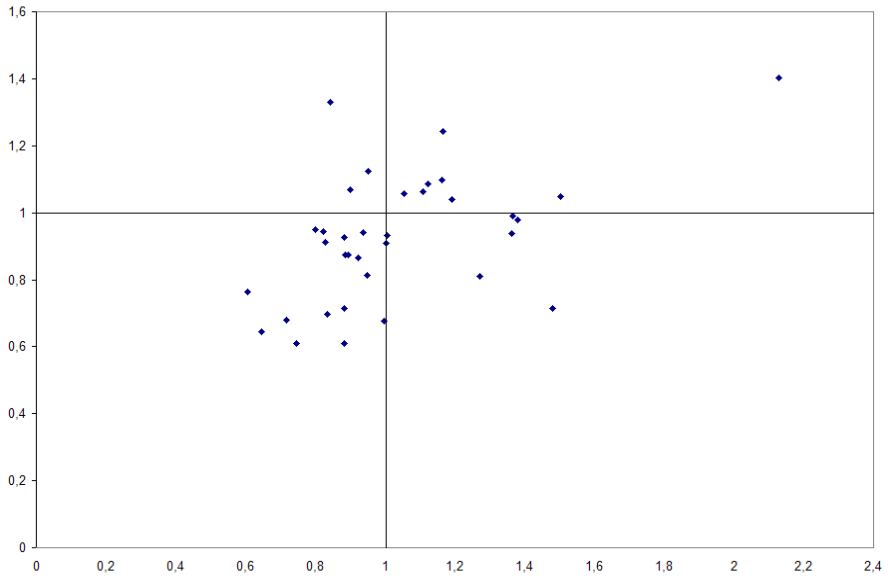


Figure 1. Predicted baseline risk $\exp(b_{0i})$ (horizontal axis) versus predicted treatment effect $\exp(\beta + b_{1i})$ (vertical axis) for each center $(i = 1, ..., 35)$

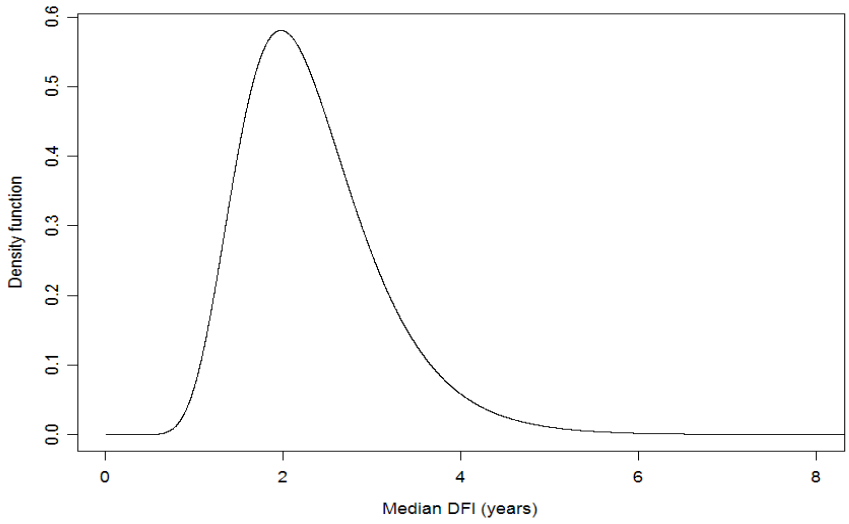


Figure 2. Density function of the median DFI in the no intravesical treatment group over centers.

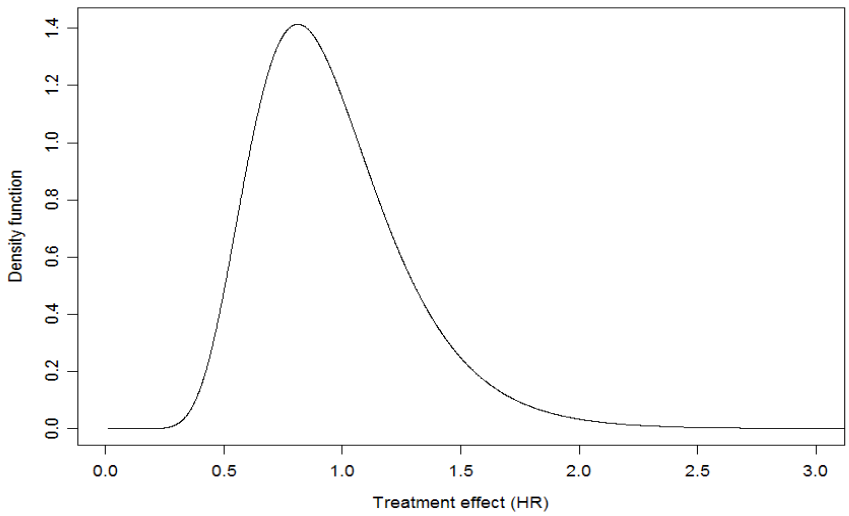


Figure 3. Density function of the Hazard Ratio $HR = \exp(\beta + b_{1i})$ over centers.

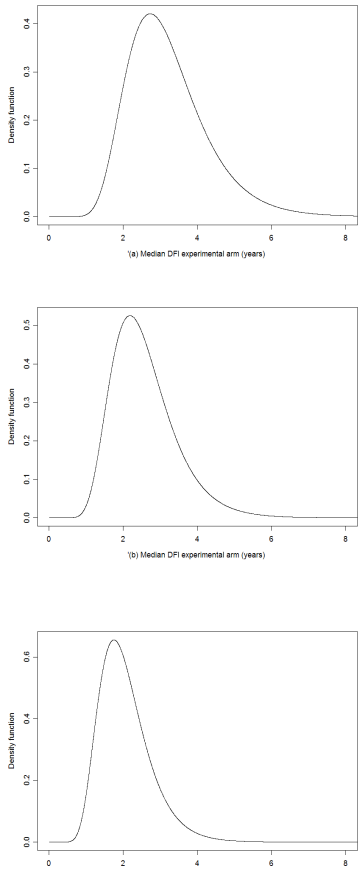


Figure 4. Density function of median DFI in the intravesical treatment group over centers, considering (a) "Good outcome" centers, (b) "Average outcome" centers and (c) "Poor outcome" centers.

Appendix

Considering model (1) with the random effects $\boldsymbol{b}^T = (\boldsymbol{b}_0^T, \boldsymbol{b}_1^T)$ distributed according to (2) and assuming a constant baseline hazard, the density of M_c, the median time to event in the control arm ($x = 0$) over centers, is given by

$$f_{M_c}(m_c) = \frac{1}{m_c\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{1}{2\sigma_0^2}\left(\ln\left(\frac{\ln 2}{\lambda_0 m_c}\right)\right)^2\right). \tag{8}$$

Indeed, for a constant baseline hazard the conditional survival curve is given by

$$S(t \mid b_0, b_1, x) = \exp\left(-\lambda_0 t \exp\left(b_0 + (\beta + b_1)x\right)\right)$$

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)

$$M_c = h(b_0) = \frac{\ln 2}{\lambda_0 \exp b_0}.$$

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

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

with $b_0 \sim \mathrm{N}(0, \sigma_0^2)$ we easily obtain (8).

Similarily, considering $HR = \exp(\beta + b_1)$ 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 HR is given by

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

Under the same assumptions, but fixing b_0 and β to a particular value, we obtain, by noting that $\beta + b_1 \sim \mathrm{N}(\beta, \sigma_1^2)$, that M_e, the median time to event in the experimental arm ($x = 1$) over centers is given by

$$f_{M_e}(m_e) = \frac{1}{m_e\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{1}{2\sigma_1^2}\left(\ln\left(\frac{\ln 2}{\lambda_0 m_e}\right) - b_0 - \beta\right)^2\right).$$