Research Article

Statistics in Medicine

Received 14 May 2009, Accepted 12 July 2010 Published online 22 September 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4046

Cure fraction model with random effects for regional variation in cancer survival

Karri Seppä,[a,b]*† Timo Hakulinen,[a] Hyon-Jung Kim[b] and Esa Läärä[b]

Assessing regional differences in the survival of cancer patients is important but difficult when separate regions are small or sparsely populated. In this paper, we apply a mixture cure fraction model with random effects to cause-specific survival data of female breast cancer patients collected by the population-based Finnish Cancer Registry. Two sets of random effects were used to capture the regional variation in the cure fraction and in the survival of the non-cured patients, respectively. This hierarchical model was implemented in a Bayesian framework using a Metropolis-within-Gibbs algorithm. To avoid poor mixing of the Markov chain, when the variance of either set of random effects was close to zero, posterior simulations were based on a parameter-expanded model with tailor-made proposal distributions in Metropolis steps. The random effects allowed the fitting of the cure fraction model to the sparse regional data and the estimation of the regional variation in 10-year cause-specific breast cancer survival with a parsimonious number of parameters. Before 1986, the capital of Finland clearly stood out from the rest, but since then all the 21 hospital districts have achieved approximately the same level of survival. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: breast cancer survival; regional variation; cure fraction; mixture model; random effect; Bayesian hierarchical model

1. Introduction

Equity in cancer care should be a desirable aim within a national health service. Regional variation in the survival of cancer patients may reflect differences in the effectiveness of cancer care. However, inequalities in cancer care may not be the only reason for apparent differences in the overall survival. Variations may also be due to confounding variables or chance [1]. Dickman *et al.* [2] estimated that 2.7 per cent of excess deaths in breast cancer patients could be prevented in Finland by eliminating the systematic regional variation in cancer survival.

Mortality from other causes of death than the cancer of interest is a source of confounding that has to be taken into account. Relative survival or cause-specific survival are suitable measures to describe the mortality due to the cancer itself [3]. Traditional parametric survival models that assume that all patients are susceptible to eventually die from the disease itself are often inadequate in describing the survival experience of cancer patients. It is conceivable that many patients are in fact cured in the sense that their lifetime is not shortened by the cancer but their mortality rates remain the same as if they had avoided the cancer. At an individual level it is practically impossible to determine for sure whether a patient is cured or not. However, for many cancers it appears to be possible in principle to identify the cure fraction [4], i.e. the proportion of patients, whose mortality will not be elevated as compared with the mortality rates in a similar cancer-free population. The cure fraction is operationalized as a value between 0 and 100 per cent at which level the relative or cause-specific survival function of the whole patient population stabilizes in a finite time. This kind of survival curve can be represented as a mixture of two components: a constant line at level 1 for the cured subset and a curve representing the relative or cause-specific survival in the non-cured subset of patients, the cure fraction determining the mixing proportions.

Cure fraction models have become increasingly popular in population-based studies on cancer survival performed for individual countries, but also in international comparisons, and their usefulness is motivated, e.g. by the EUROCARE Working Group [5]. The proportion cured and the mean survival time for the non-cured patients can be useful summary

[a] *Finnish Cancer Registry, Pieni Roobertinkatu 9, 00130 Helsinki, Finland*
[b] *Department of Mathematical Sciences, University of Oulu, Finland*
* *Correspondence to: Karri Seppä, Finnish Cancer Registry, Pieni Roobertinkatu 9, 00130 Helsinki, Finland.*
† *E-mail: karri.seppa@cancer.fi*

parameters to assess the differences in survival in more detail. Cooner *et al.* [6] proposed a unifying class of cure fraction models for the estimation of the cause-specific survival. Cure fraction models can also be extended for analyses of relative survival [7, 8].

When separate regions are small or sparsely populated, crude region-specific estimates become less precise. Different approaches have been proposed to shrink the variation in these estimates [2, 9]. Many unmeasured covariates are likely to be contributing to variability between regions, and random effects may be used to capture such a variation. The Bayesian modelling framework allows us to incorporate cure fraction models with random effects. Cooner *et al.* [10] considered cure fraction models with random effects and illustrated them with breast cancer data.

In this paper we apply a parametric mixture model on the cause-specific survival of female breast cancer patients in Finland diagnosed since 1953. Our focus is on assessing regional variation and its changes over five decades in the survival across the 21 central hospital districts in the country. The model incorporates both a cure fraction and district-specific random effects. A similar model but without regional random effects was previously used for the cause-specific survival of Finnish breast cancer patients in 1987–2001 [11]. Regional variation and temporal changes in it is assessed by model-based estimates of district-specific cure fractions, mean survival times of the non-cured patients, as well as 10-year cause-specific survival proportions, the latter being used to summarize the overall prognosis of the whole patient population in each district.

2. Cure fraction model with regional random effects

2.1. Finnish breast cancer data

The study population consists of women diagnosed with a breast cancer at the ages of 40–69 years in Finland during 1953–2000 and reported to the Finnish Cancer Registry. The patients were followed up for breast cancer death up to the end of 2007. Cases based only on death certificate or autopsy report ($n=128$) were excluded from the data.

In Finland, specialized hospital care is currently provided by 20 central hospital districts. The number of districts was 21 until year 2000, when the, thus far separate, hospital districts of Helsinki (the capital area) and Uusimaa (the surrounding province) were merged into one big hospital district. Our analysis is based on the old division into 21 districts (Figure 1). Each central hospital district is subsumed under one of five larger cancer control regions, each being lead by a university hospital. A university hospital serves also the central hospital of its own hospital district. In the national health-care system, inhabitants of a given municipality needing hospital treatment are primarily treated in their 'own' central hospital and further referred to the pertinent university hospital, if more advanced hospital care is needed. The survival of the patients living in the catchment area of a given hospital district may be more similar than the survival of the patients who are from different districts. Random effects were used to make allowance for possible systematic variation between the central hospital districts.

The year of diagnosis was categorized into three periods: 1953–1969, 1970–1985 and 1986–2000 this division being motivated as follows: The central hospitals did not yet exist in 1953 but were built-up mainly during the first period. In the 1970s, the regional health-care system was developed to what it currently is (see above) [12, 13]. The last period covers the era of mammography screening. A nationwide screening programme for breast cancer was initiated in 1987 [11]. From the early 1990s, mammographic screening has been free of charge to all 50–59 years old women. More recently, many municipalities have offered screening to women aged 60–69 years, too.

Statistical models were fitted separately for the three periods. Age at diagnosis was used as a covariate and it was categorized into three classes: 40–49, 50–59, 60–69.

2.2. Mixture cure fraction model

Suppose that the relevant patient population can be latently divided into the subsets of cured and non-cured patients, respectively [8]. The proportion π of cured patients would not experience excess mortality due to the cancer of interest, i.e. the cause-specific survival function of the cured patients equals 1. Let $S_D(t)$ be the cause-specific survival function for the non-cured patients. Now the cause-specific survival function for the entire patient population can be written as a mixture of the two survival functions

$$S(t)=\pi+(1-\pi)S_D(t).$$

The generalized gamma distribution introduced by Stacy [14] was assumed to model the survival function for the non-cured patients. We used a reparametrized and extended form of the generalized gamma distribution with location (μ),

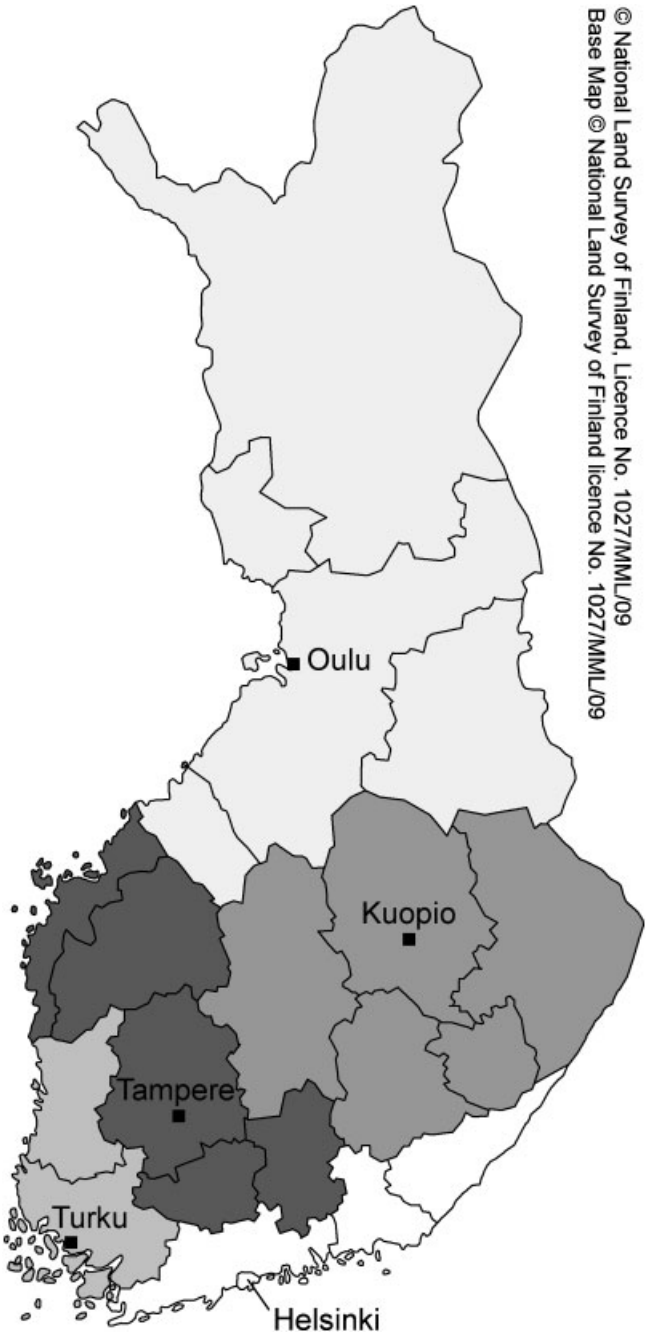

Figure 1. Capital district, Helsinki, and 20 other central hospital districts in Finland. Their division into the five cancer control regions is shown by varying grey tones. The university hospitals are located in Helsinki, Kuopio, Oulu, Tampere and Turku.

scale (σ) and shape (λ) parameters [15, 16]:

$$S_D(t)=\begin{cases}1-G[\lambda^{-2}(te^{-\mu})^{\lambda/\sigma};\lambda^{-2}] & \text{if } \lambda>0\\ G[\lambda^{-2}(te^{-\mu})^{\lambda/\sigma};\lambda^{-2}] & \text{if } \lambda<0\\ 1-\Phi[(\log(t)-\mu)/\sigma] & \text{if } \lambda=0\end{cases}$$

where $G(t;a)=\int_0^t x^{a-1}e^{-x}\,dx/\Gamma(a)$ is the cumulative distribution function for the particular case of the gamma distribution with mean and variance equal to a, and Φ is the cumulative distribution function of the standard normal distribution. The generalized gamma distribution covers, as special cases, the lognormal ($\lambda=0$), Weibull ($\lambda=1$) and gamma ($\lambda=\sigma$)

distributions. The mean survival time for non-cured patients is given by:

$$E(T_D)=\int_0^\infty t f_D(t)\,\mathrm{d}t=\begin{cases}\dfrac{\Gamma(\lambda^{-2}+\sigma/\lambda)}{\Gamma(\lambda^{-2})}\mathrm{e}^{\mu}(\lambda^2)^{\sigma/\lambda} & \text{if } \lambda\neq 0\\ \mathrm{e}^{\mu+\sigma^2/2} & \text{if } \lambda=0\end{cases}.$$

Based on the mixture model, the 10-year cause-specific survival proportion $\theta=S(10\text{ years})$ is used as a single measure to summarize the survival of the whole patient group. Note that $S(t)$ and $S_D(t)$ are estimating the so-called net survival probabilities [3] at time t for the entire patient group and for the non-cured patients, respectively, because only deaths certified as due to the cancer of interest are considered as events. Therefore, θ and $E(T_D)$ are interpreted as hypothetical quantities that would be observed, if the cancer of interest were the only possible cause of death and no competing risks existed.

2.3. *Entering fixed and random effects into the model*

The proportion cured π and the parameters μ, σ and λ in the generalized gamma distribution may vary by age category. Upon that, random effects $\omega_{\pi,r}$ are used to model the possible systematic variation of the cure fraction between the hospital districts ($r=1,\ldots,R$). In our application $R=21$. The proportion cured in district r is modelled using the logistic model, i.e.:

$$\pi_r(\boldsymbol{x}_i)=\frac{\exp(\boldsymbol{\alpha}'\boldsymbol{x}_i+\omega_{\pi,r})}{1+\exp(\boldsymbol{\alpha}'\boldsymbol{x}_i+\omega_{\pi,r})}$$

where $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3)'$ is a vector of fixed effect parameters and $\boldsymbol{x}_i$ is a vector of indicator variables of age for a patient i:

$$\boldsymbol{\alpha}'\boldsymbol{x}_i=\alpha_1 I^{\text{age}_i}_{[40,50)}+\alpha_2 I^{\text{age}_i}_{[50,60)}+\alpha_3 I^{\text{age}_i}_{[60,70)}.$$

The systematic regional variation in the survival of the non-cured patients is modelled by including a district-specific random effect $\omega_{\mu,r}$ in the location parameter μ of the generalized gamma distribution:

$$\mu_r(\boldsymbol{x}_i)=\boldsymbol{\beta}'\boldsymbol{x}_i+\omega_{\mu,r}.$$

The scale (σ) and shape (λ) parameters are assumed to be the same in every district, but are allowed to vary by age category:

$$\sigma(\boldsymbol{x}_i)=\exp(\boldsymbol{\gamma}'\boldsymbol{x}_i),$$
$$\lambda(\boldsymbol{x}_i)=\boldsymbol{\delta}'\boldsymbol{x}_i.$$

The linear predictors $\boldsymbol{\beta}'\boldsymbol{x}_i$, $\boldsymbol{\gamma}'\boldsymbol{x}_i$ and $\boldsymbol{\delta}'\boldsymbol{x}_i$ of the fixed effects are defined similarly as the predictors $\boldsymbol{\alpha}'\boldsymbol{x}_i$.

The random effects $\omega_{\pi,r}$ and $\omega_{\mu,r}$ are assumed to be drawn independently from the normal distributions:

$$\omega_{\pi,r}\sim \mathrm{N}(0,\sigma^2_\pi)\quad\text{and}$$
$$\omega_{\mu,r}\sim \mathrm{N}(0,\sigma^2_\mu).$$

We also considered an alternative specification, in which the random effect pairs $(\omega_{\pi,r},\omega_{\mu,r})'$ were assumed to be drawn independently from a bivariate normal distribution with a correlation coefficient ρ_ω.

2.4. *The magnitude of regional variation in survival*

The variation in the random effects of the two survival components is measured by the standard deviations σ_π and σ_μ of the distributions of the random effects $\omega_{\pi,r}$ and $\omega_{\mu,r}$, respectively. The magnitude of the regional variation in the survival of the whole patient group is measured by the standard deviation of the distribution of the district-specific 10-year survival proportions and the variance of that was approximated using the delta method:

$$\sigma^2_\theta=\mathrm{Var}\left(\theta_r(\omega_{\pi,r},\omega_{\mu,r})\right)\approx\left(\frac{\partial\theta_r}{\partial\omega_{\pi,r}}(0,0)\right)^2\sigma^2_\pi+\left(\frac{\partial\theta_r}{\partial\omega_{\mu,r}}(0,0)\right)^2\sigma^2_\mu+2\frac{\partial\theta_r}{\partial\omega_{\pi,r}}(0,0)\frac{\partial\theta_r}{\partial\omega_{\mu,r}}(0,0)\rho_\omega\sigma_\pi\sigma_\mu$$

where θ_r is the 10-year cause-specific survival proportion in district r written as a function of $\omega_{\pi,r}$ and $\omega_{\mu,r}$, and the last term is omitted, if the two sets of random effects are assumed to be independent.

As the absolute value of the standard deviation of any proportion type of measure strongly depends on the level of the proportion itself, being highest at the level of 50 per cent and tending to zero when the proportion approaches either 0 or 100 per cent, we decided to standardize the standard deviation σ_θ of the estimated 10-year cause-specific survival proportions θ_r for each of the three periods by dividing them by $\sqrt{\bar{\theta}(1-\bar{\theta})}$, this quantity indicating the relative level of a standard deviation of a proportion, when the mean proportion is $\bar{\theta}$. Hence, we used this standardized standard deviation $\sigma_\theta(\bar{\theta})=\sigma_\theta/\sqrt{\bar{\theta}(1-\bar{\theta})}$ in comparisons of the regional variability of the 10-year proportions across the three calendar periods.

2.5. *Posterior simulation*

The hierarchical random effects model specified above can be implemented in a Bayesian framework using Markov chain Monte Carlo methods [17]. Based on the model where the district-specific effects are modelled by the two independent normal distributions, the joint posterior distribution of the unknown parameters is given by

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\omega}_\pi, \boldsymbol{\omega}_\mu, \sigma_\pi, \sigma_\mu) \propto L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\omega}_\pi, \boldsymbol{\omega}_\mu) p(\boldsymbol{\omega}_\pi \,|\, \sigma_\pi) p(\boldsymbol{\omega}_\mu \,|\, \sigma_\mu) p(\boldsymbol{\alpha}) p(\boldsymbol{\beta}) p(\boldsymbol{\gamma}) p(\boldsymbol{\delta}) p(\sigma_\pi) p(\sigma_\mu), \tag{1}$$

where $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\omega}_\pi, \boldsymbol{\omega}_\mu)$ is the likelihood function of the cure fraction model, $p(\boldsymbol{\omega}_\pi \,|\, \sigma_\pi)$ and $p(\boldsymbol{\omega}_\mu \,|\, \sigma_\mu)$ are the joint distribution functions of the random effects $\omega_{\pi,r}$ and $\omega_{\mu,r}$, and the remaining terms are prior distributions for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, σ_π and σ_μ.

Based on the individual data records, the likelihood function on all patients $i=1,\ldots,n$ can be written as

$$L(\pi, \mu, \sigma, \lambda) = \prod_{i=1}^{n} \{[\pi + (1-\pi) S_D(t_i \,|\, \mu, \sigma, \lambda)]^{1-d_i} [(1-\pi) f_D(t_i \,|\, \mu, \sigma, \lambda)]^{d_i}\},$$

where the quantities π, μ, σ and λ are the functions of the fixed effects $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ plus the random effects $\omega_{\pi,r}$ and $\omega_{\mu,r}$ as specified in Section 2.3, t_i is the observed follow-up time, $d_i=1$ for a patient who was observed to die from breast cancer and $d_i=0$ for a censored patient (due to death from other cause, or end of follow-up). The function $f_D(t)$ is the probability density function of the generalized gamma distribution (see [16]).

The Gibbs sampler [17] cannot be used directly to draw values from the joint posterior distribution, because direct sampling from the full conditional distributions is not possible. The posterior distribution can be simulated using the Metropolis algorithm [17] within the Gibbs sampler. However, if either of the hierarchical standard deviations σ_π or σ_μ is small, the Markov chain for the standard deviation can easily get stuck close to zero because of the dependence in the posterior distribution between the random effects and their variance parameters. To avoid the slow mixing, parameter expansion [18–20] was used. In a parameter-expanded model, each set of random effects is multiplied by an additional parameter. The random effects $\omega_{\pi,r}$ and $\omega_{\mu,r}$ in (1) correspond to $\xi_\pi \eta_{\pi,r}$ and $\xi_\mu \eta_{\mu,r}$ in the parameter-expanded model. In addition, the hierarchical standard deviations σ_π and σ_μ correspond to $|\xi_\pi|\tau_\pi$ and $|\xi_\mu|\tau_\mu$ in the parameter-expanded model. The joint posterior distribution based on the parameter-expanded model is given by

$$\begin{aligned} p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\eta}_\pi, \boldsymbol{\eta}_\mu, \tau_\pi, \tau_\mu, \xi_\pi, \xi_\mu) \propto\ & L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\eta}_\pi, \boldsymbol{\eta}_\mu, \xi_\pi, \xi_\mu) p(\boldsymbol{\eta}_\pi \,|\, \tau_\pi) p(\boldsymbol{\eta}_\mu \,|\, \tau_\mu) \\ & \times p(\boldsymbol{\alpha}) p(\boldsymbol{\beta}) p(\boldsymbol{\gamma}) p(\boldsymbol{\delta}) p(\tau_\pi) p(\tau_\mu) p(\xi_\pi) p(\xi_\mu), \end{aligned}$$

where the likelihood function $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\eta}_\pi, \boldsymbol{\eta}_\mu, \xi_\pi, \xi_\mu)$ is otherwise similar as in (1), but $\omega_{\pi,r}$ and $\omega_{\mu,r}$ terms in $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\omega}_\pi, \boldsymbol{\omega}_\mu)$ are replaced with $\xi_\pi \eta_{\pi,r}$ and $\xi_\mu \eta_{\mu,r}$ terms.

To complete the Bayesian model, prior distributions for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$, τ_π, τ_μ, ξ_π and ξ_μ have to be specified. Vague $\mathrm{N}(0, 100^2)$ priors were assumed for the fixed effects $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$. For the hierarchical standard deviations τ_π and τ_μ we assumed $\mathrm{Unif}(0, 100)$ priors. The auxiliary parameters ξ_π and ξ_μ were assigned to have improper uniform prior distributions on $(-\infty, \infty)$.

The random walk Metropolis algorithm was used within the Gibbs sampler to sample from the full conditional distributions. Normal distributions $\mathrm{N}(\alpha_k^{\mathrm{cur}}, c_{\alpha_k})$, $\mathrm{N}(\beta_k^{\mathrm{cur}}, c_{\beta_k})$, $\mathrm{N}(\gamma_k^{\mathrm{cur}}, c_{\gamma_k})$ and $\mathrm{N}(\delta_k^{\mathrm{cur}}, c_{\delta_k})$ centred on the current points α_k^{cur}, β_k^{cur}, γ_k^{cur} and δ_k^{cur} were used as the proposal distributions in the Metropolis algorithm to sample candidate values for the fixed effects α_k, β_k, γ_k and δ_k. Candidate values for $\eta_{\pi,r}$ $(r=1,\ldots,R)$, τ_π and ξ_π were drawn from $\mathrm{N}(\eta_{\pi,r}^{\mathrm{cur}}, c_{\eta_{\pi,r}}/(\xi_\pi^{\mathrm{cur}})^2)$, $\mathrm{N}(\tau_\pi^{\mathrm{cur}}, c_{\tau_\pi}/(\xi_\pi^{\mathrm{cur}})^2)$ and $\mathrm{N}(\xi_\pi^{\mathrm{cur}}, c_{\xi_\pi}/((1/R)\sum_{r=1}^{R} |\eta_{\pi,r}^{\mathrm{cur}}|)^2)$ distributions, respectively (similarly for $\eta_{\mu,r}$, τ_μ and ξ_μ). Here, we let the variances of the proposals to depend on the current state of the chain, because the parameters $\eta_{\pi,r}$, τ_π and ξ_π (as well as the parameters $\eta_{\mu,r}$, τ_μ and ξ_μ) heavily depend on each other in the posterior distribution. For each proposal, the constant c was tuned during the burn-in period so that the acceptance rate of the random walk Metropolis algorithm was close to 44 per cent. After the burn-in period, c was fixed and the inference was based on a sample that excludes the adaptive phase.

We also fitted the model where a bivariate normal distribution was assumed for the district-specific random effects. The parameter-expanded model was defined similarly and the correlation coefficient ρ_ω corresponds to $\text{sign}(\xi_\pi \xi_\mu)\rho_\eta$ in the parameter-expanded model. For ρ_η we assumed a Unif$(-1, 1)$ prior. For the other unknown parameters we assumed the same vague priors as in the earlier model.

The C++ language was used to program the simulation algorithm. Further processing of the simulated samples was performed using R [21]. The posterior simulation of the parameter-expanded model was not applicable in the current version of WinBUGS [22], because in WinBUGS, the variance of the proposal distribution of the Metropolis algorithm is the same constant in every iteration (after an adaptive phase) and it cannot depend on the current state of the chain.

3. Results

By the end of 2007, there were 5736, 7238 and 6675 deaths from breast cancer observed among 9671, 15 665 and 28 028 women diagnosed in 1953–1969, 1970–1985 and 1986–2000, respectively. Figure 2 shows the estimated cause-specific cumulative survival proportions of breast cancer patients by age class and period of diagnosis. The posterior mean estimates of the parameter λ in the generalized gamma distribution were ranged from 0.24 to 0.97. The estimates were closest to the value 1 (the Weibull model) in the older age classes.

The Bayesian hierarchical model was fitted separately to the three calendar periods of diagnosis. For each analysis, two series of 110 000 iterations were run. Running the two chains simultaneously for the three periods of diagnosis took 50 h on a standard desktop computer with a dual-core processor (Intel Core 2 Duo E8400, 3 GHz). The first 10 000 iterations were discarded as burn-in in each run. Every 10th of the remaining 100 000 iterations was stored and the posterior inferences were based on the pooled sample of 20 000 values. Figure 3 shows the trace plots of 10 000 samples (after the burn-in and thinning) for π, $E(T_D)$ and θ in district 1 in the last period, when the chain mixed slowest due to high negative correlation between the two survival components. To assess the convergence of the Markov chain, we calculated the Gelman–Rubin statistic, as modified by Brooks and Gelman [23]. Based on the two independently simulated chains with overdispersed initial values, the ratio between the pooled 80 per cent credible interval and the average 80 per cent credible interval tended to 1 (Figure 4). Even though the regional variation in the survival of the non-cured patients was close to zero in the last period, the hierarchical standard deviation σ_μ mixed nicely, when the parameter-expanded model was used (Figure 5).

We fitted two models in each diagnosis period. District-specific effects were modelled either by two independent normal distributions or by a bivariate normal distribution. The correlation structure did not turn out to be necessary in any period of diagnosis. The estimated variance of the correlation coefficient ρ_ω was very large, and the estimates of the target parameters were almost identical whether correlation was allowed for or not. Hence, only the results based on the model with two independent normals are shown.

Figure 6 shows the posterior mean estimates and the 95 per cent credible intervals of the two survival parameters for the 21 central hospital districts. The districts are arranged according to the cure fraction estimates for the last period. The

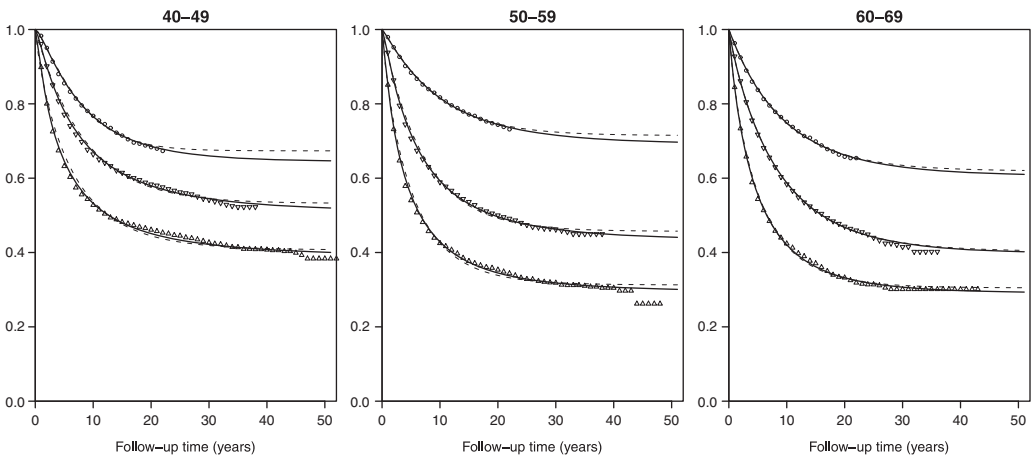

Figure 2. Cumulative cause-specific survival proportions by age class for the breast cancer patients diagnosed in 1953–1969 (triangle up), 1970–1985 (triangle down) and 1986–2000 (circle). Posterior mean estimates of the cumulative survival proportions based on the mixture cure fraction model with the generalized gamma distribution (solid line) and the Weibull distribution (dashed line).

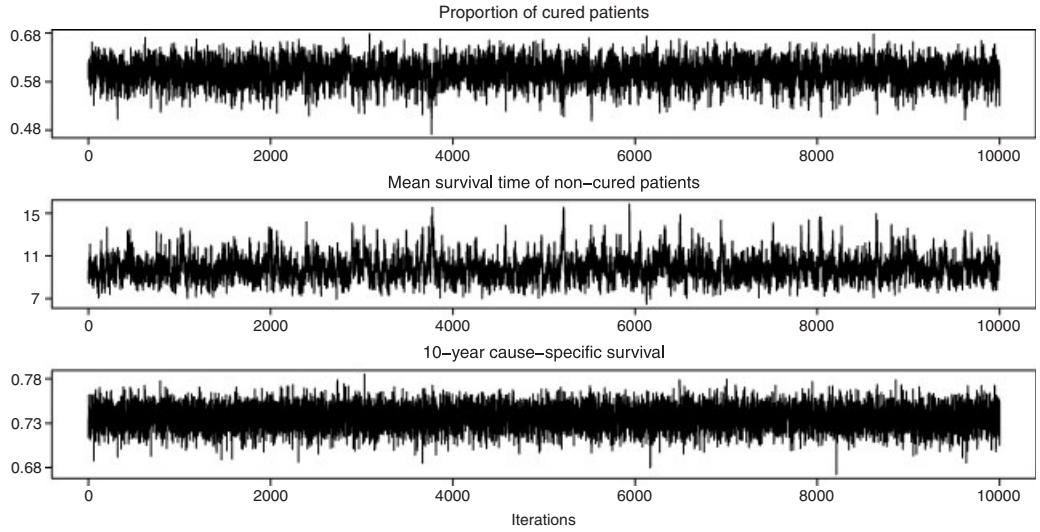

Figure 3. Trace plots after the burn-in period and thinning for the proportion of cured patients, the mean survival time of non-cured patients and the 10-year cause-specific survival proportion in district 1 in the last period of diagnosis (1986–2000).

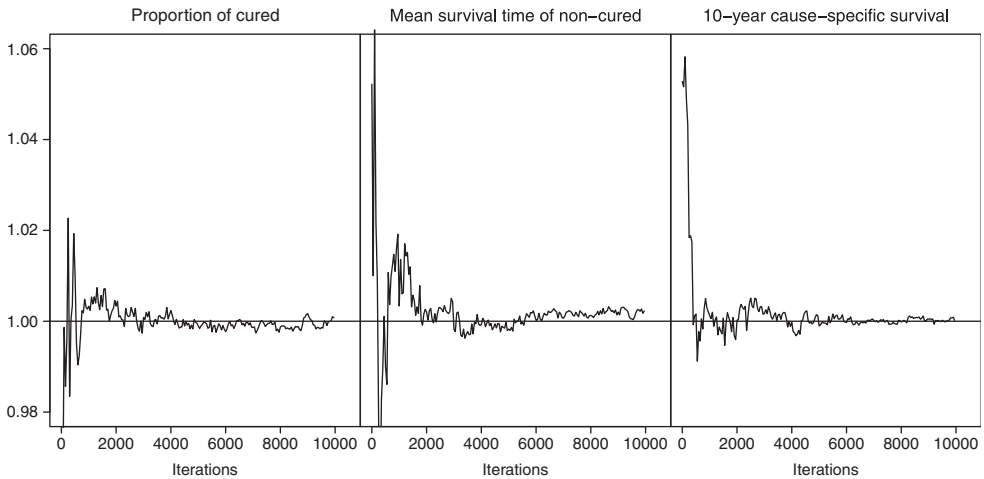

Figure 4. The ratio of pooled to average 80 per cent credible interval widths calculated over iterations for the proportion of cured patients, the mean survival time of non-cured patients and the 10-year cause-specific survival proportion in district 1 in the last period of diagnosis.

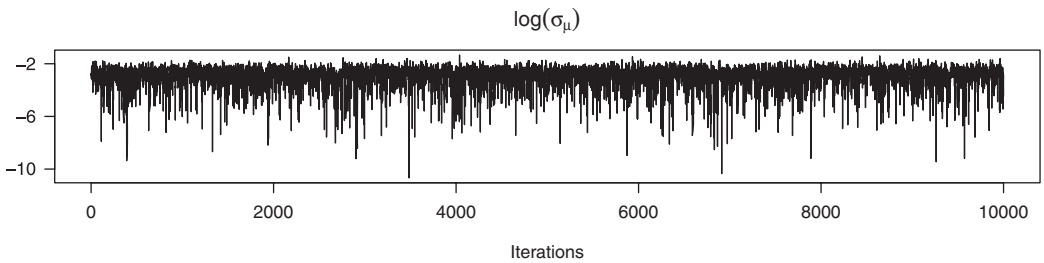

Figure 5. Trace plot after the burn-in period and thinning for the logarithm of the standard deviation σ_μ of the normal distribution from which the random effects $\boldsymbol{\omega_\mu}$ were drawn.

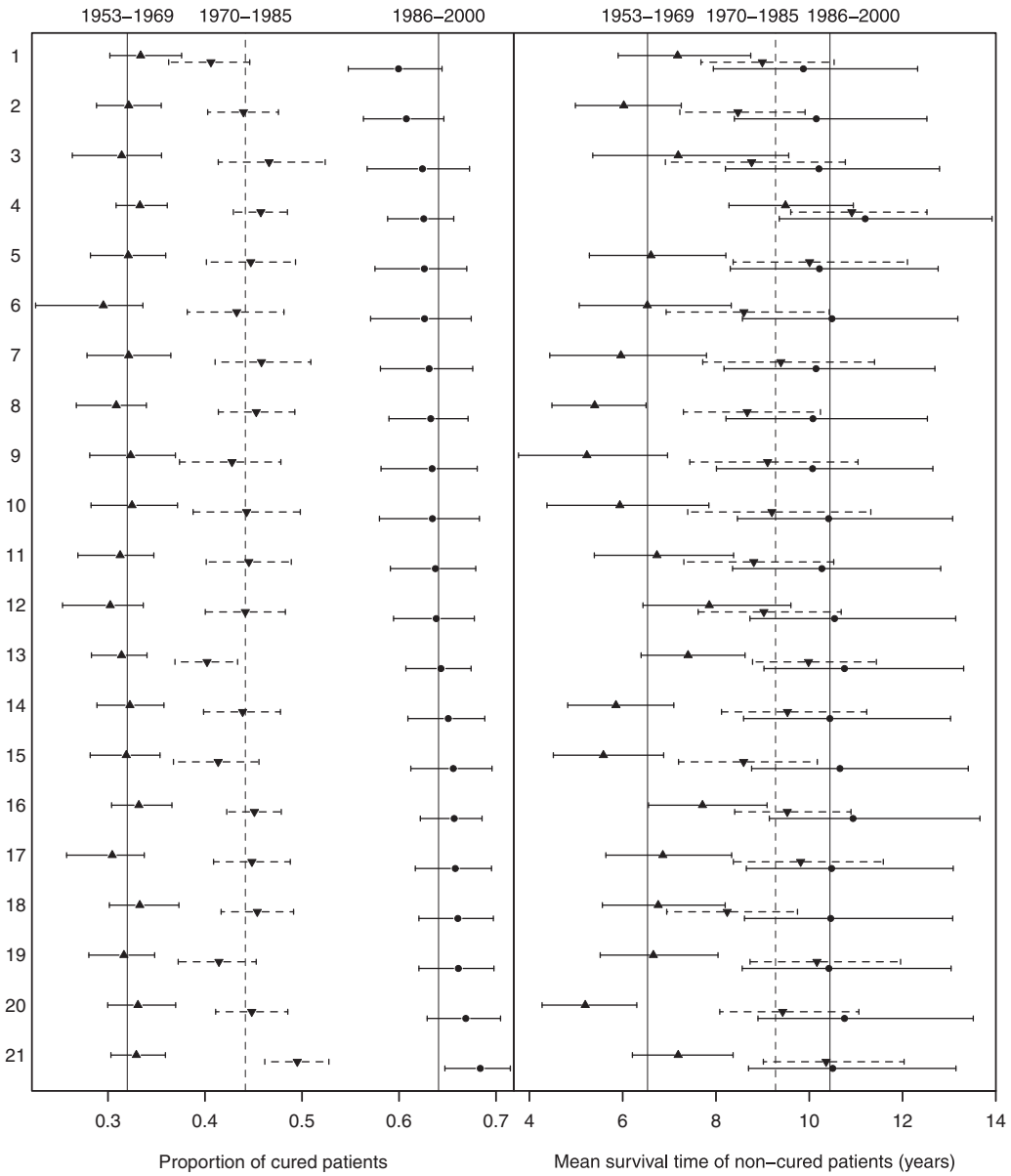

Figure 6. Posterior mean estimates and 95 per cent credible intervals of the two survival components in 21 central hospital districts (4 = Helsinki) by period of diagnosis (triangle up, 1953–1969; triangle down, 1970–1985; circle, 1986–2000). The vertical lines show the posterior mean estimates of the overall cure fractions and mean survival times for the three periods.

posterior mean estimate of the average cure fraction increased from 32 per cent (for patients diagnosed in 1953–1969) to 44 per cent (1970–1985) and further to 64 per cent (1986–2000). None of the districts differed essentially from the average cure fraction in the first period. In addition, the estimated standard deviation σ_π of the random effects $\omega_{\pi,r}$ was smallest in the first period (Table I). The estimate for district 21 was larger than the average all the time. In the last period, the lowest cure fraction was estimated to be in district 1, being 8.4 percentage points smaller than that in district 21.

The estimated mean survival time in the non-cured patients overall increased clearly from the period 1953–1969 to the period 1970–1985 (Figure 6). The average change was smaller between the two last periods than from the 1st to the 2nd period. The standard deviation σ_μ of the random effects $\omega_{\mu,r}$ was largest in the first period (Table I), and the largest difference between the district-specific estimates of the mean survival was 4.3 years. In the last period, there were no big differences in these estimates any more. The largest mean lifetime was estimated for the capital district, Helsinki, throughout. The estimate of district 21 was larger than the global mean estimate all the time, whereas the two latest estimates of district 1 were smaller than the overall mean estimate.

Figure 7 shows the district- and period-specific estimates of the two survival components on a plane. In addition, 95 per cent highest posterior density regions are displayed for hospital districts 1, 4 and 21. The cure fraction and the mean

Table I. Posterior median estimates and 95 per cent credible intervals of the standard deviations σ_π and σ_μ of the random effects $\omega_{\pi,r}$ and $\omega_{\mu,r}$, the standard deviations σ_θ and the standardized standard deviation $\sigma_\theta(\bar{\theta})$ of the 10-year cause-specific survival proportions by the period of diagnosis.

	σ_π		σ_μ		σ_θ		$\sigma_\theta(\bar{\theta})$	
Diagnosis period	Posterior median	95 per cent credible interval	Posterior median	95 per cent credible interval	Posterior median	95 per cent credible interval	Posterior median	95 per cent credible interval
1953–1969	0.09	(0.01, 0.21)	0.19	(0.12, 0.29)	0.033	(0.022, 0.051)	0.067	(0.044, 0.103)
1970–1985	0.12	(0.07, 0.21)	0.11	(0.06, 0.18)	0.029	(0.019, 0.042)	0.059	(0.040, 0.087)
1986–2000	0.12	(0.07, 0.20)	0.06	(0.00, 0.13)	0.020	(0.013, 0.031)	0.048	(0.030, 0.074)

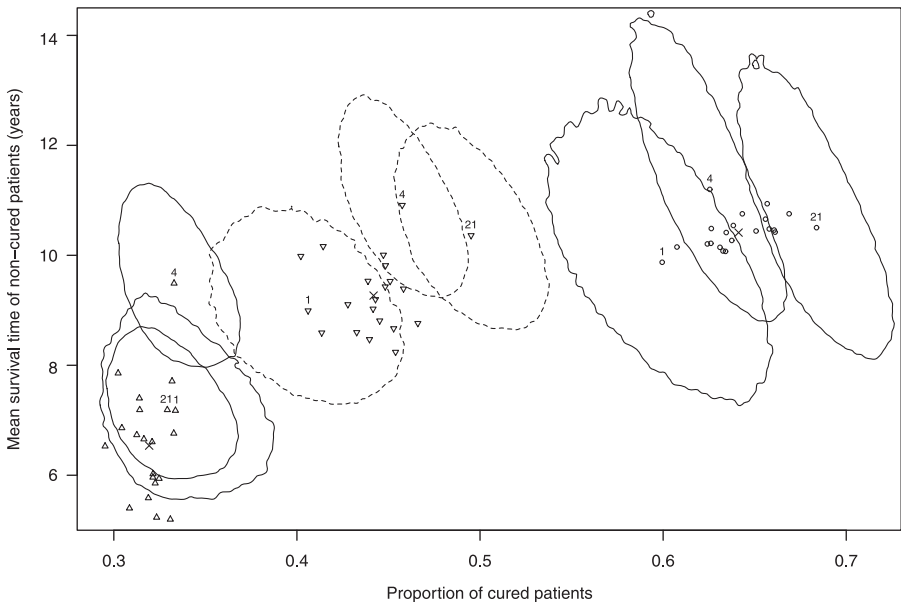

Figure 7. Posterior mean estimates of the two survival components in 21 central hospital districts by period of diagnosis (triangle up, 1953–1969; triangle down, 1970–1985; circle, 1986–2000). For hospital districts 1, 4 (Helsinki) and 21, 95 per cent highest posterior density regions are shown (in the first period, the larger one of the overlapping regions belongs to region 1). Crosses mark the overall mean values for the whole country.

survival time of non-cured patients were negatively correlated in the posterior distribution. The negative correlation between the survival components increased when the maximum follow-up time decreased. For instance, in district 21 the absolute value of the estimated correlation coefficient increased from 0.34 in the first period, to 0.51, and further to 0.74 in the last period. When the correlation increases, it becomes more difficult to identify the two components. Here, however, these correlations were still quite benign.

Ten-year cause-specific survival proportions were estimated for each district, combining the cure fraction with the survival of the non-cured patients (Figure 8). The posterior mean estimate of the national average increased from 44 per cent (1953–1969) to 61 per cent (1970–1985) and further to 77 per cent (1986–2000). The estimate for Helsinki was clearly larger than the overall mean in the whole country in the first two periods: by 7.4 percentage points (95 per cent credible interval (CI) 5.2 to 9.8) in the first period, and by 4.1 percentage points (95 per cent CI 2.1 to 6.0) in the second period.

The regional variation between the 10-year cause-specific survival proportions decreased during the study period in terms of the standard deviation estimates: the posterior median estimate of the standard deviation of the average district-specific 10-year proportions was 3.3 percentage points in 1953–1969, 2.9 percentage points in 1970–1985 and 2.0 percentage points in 1986–2000 (Table I). The standard deviation σ_θ decreased by 40 per cent (95 per cent CI −11 to 68 per cent) from the first to the last period. The standardized standard deviation $\sigma_\theta(\bar{\theta})$ decreased by 29 per cent (95 per cent CI −31 to 62 per cent).

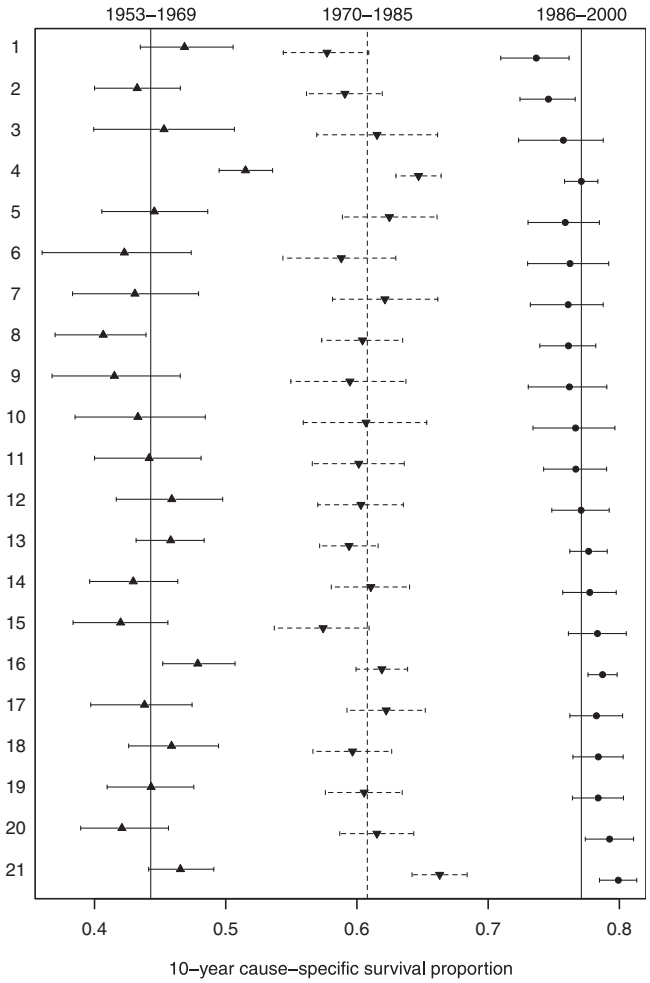

Figure 8. Posterior mean estimates and 95 per cent credible intervals of the 10-year cause-specific survival proportions in 21 central hospital districts (4 = Helsinki) by the period of diagnosis (triangle up, 1953–1969; triangle down, 1970–1985; circle, 1986–2000). The vertical lines show the posterior mean estimates of the overall 10-year cause-specific survival proportions for the three periods.

4. Discussion

This paper demonstrates the use of a mixture cure fraction model with random effects in monitoring the regional variation in cause-specific breast cancer survival across various regions in a country.

In breast cancer, latent division into cured and non-cured patients is particularly problematic, because breast cancer deaths are still observed for a very long time since diagnosis and the point of statistical cure may not be well achieved during the lifespan of the patients [24, 25]. Rather than the actual proportion of cured patients, the cure fraction should be interpreted as a measure of long-term survival: a level nearby which the long-term cause-specific or relative cumulative survival curve stabilizes at a constant level. In practice, the estimated proportion cured may be strongly affected by the parametric form of the survival function $S_D(t)$ chosen for the non-cured patients. Because the cure fraction is negatively correlated with the survival of the non-cured patients, the estimates of the parameters can sometimes be very unstable and the model may be non-identifiable even within the same family of survival functions [26]. To improve the accuracy of cure fraction estimates, either the number of patients should be increased or the follow-up time should be extended. To moderate the identifiability problems, we decided to collect patients from a long period covering several decades and not to include patients whose potential follow-up times were less than 7 years.

Even if the fraction of cured patients among breast cancer patients were difficult to establish, our cure fraction model fitted quite well and allowed the estimation of the 10-year cause-specific survival proportions for all districts with a parsimonious number of parameters, when the districts were assumed to have separate effects on the short- and long-term survival, respectively. This single summary estimate circumvents problems caused by the negative correlation between the separate summary quantities for the cured and the non-cured patients. We must remind, though, that this quantity

cannot be interpreted as a qenuine survival proportion, because deaths from competing causes are treated as censoring events. Keeping this in mind, the 10-year proportion θ is a concise description of the fatality of the disease in question in the whole patient group, in terms of which one may informatively describe regional variation and also to compare the variability across the calendar periods of diagnosis.

The estimates from the cure fraction models would become even more unstable in regional analysis, if a separate fixed effect parameter was included in the model for each of the districts. Random effects can be used instead of the fixed ones to capture the systematic variation caused by several unknown or unmeasured factors that differ between the regions. Comparing with the estimates of the fixed effects, the estimates of the random effects tend to be more precise as the estimates are shrinked towards the overall mean value.

We observed severe convergence problems in the simulation of the posterior distribution, when either of the hierarchical standard deviations (σ_π or σ_μ) was small. The Markov chain for the standard deviation got easily stuck close to zero and the chain mixed slowly. The mixing was particularly poor in the first and in the last period, when either the systematic variation of the cure fraction π or the parameter μ was small. The convergence problems were solved by parameter expansion [18–20] that greatly improved the mixing of all the random effects and the hierarchical standard deviations.

In the mixture cure fraction model, regressing the cure fraction with flat improper priors on regression coefficients produces improper posteriors [6]. However, because we were using vague but still proper priors, proper posteriors should arise for the regression coefficients. This may sometimes lead to poor convergence, but we had no problems in convergence, when the parameter-expanded model was used.

When the random effects are drawn from the same distribution, the districts are assumed to be exchangeable. Suitable adjustment has to be done, if the assumption is not valid. First of all, either the cause-specific or the relative survival should be used for measuring survival in the patient population, because we know that the total mortality varies by region in Finland. Besides, we know that the age at diagnosis affects cancer survival, and that the age distributions differ across districts. Hence, age at diagnosis must at least be taken into account. On the other hand, stage of cancer should not be considered as a confounding factor but a link in the causal chain that depends, for instance, on access to health care and on diagnostic delay.

Cause-specific survival analysis requires reliable death certification, whereas for relative survival it would be important to take the regional differences in general population mortality into account [27]. At the Finnish Cancer Registry the cause of death has been evaluated separately for each patient. This evaluation is based on all available information, and results a specific code on the relationship between the cancer in question and the official cause of death [28]. Because there cannot be a fully satisfactory method to classify whether a given death is caused by the cancer of interest or by the competing causes, the cause of death remains to an extent unreliable after all.

If cancer survival were dependent on geographical proximity, too, CAR models might be more appropriate [10]. However, the possibility of such a dependence between the 21 hospital districts was considered irrelevant. Inhabitants of a given municipality are treated in a designated central hospital and further referred to the pertinent university hospital, if more advanced care is needed. Thus, patients do not cross-over to other districts for cancer treatment, and we do not expect much spatial correlation between the central hospital districts. Some kind of spatial structure would be expected within each district, though, because both more central and more isolated areas exist within each district. However, in this study, we are not addressing variations within districts, nor considering survival at the level of municipalities.

On the other hand, the survival of the patients might systematically vary across the five broad cancer control regions or within them, e.g. such that an essential contrast would be between the university central hospital districts and the more remote districts. Hence, we also fitted a model, in which each of the five cancer control regions were given a fixed intercept parameter in π and μ, and also a model where each university central hospital district was given such a fixed parameter in π and μ. Own fixed intercept parameters were also tried for the capital district, Helsinki. Nevertheless, the estimates of the parameters of interest remained very similar across the different models. We decided to use the simplest model, where the random effects of the 21 districts were drawn from the same distribution with a common mean and variance. The penalized expected deviance [29] or the posterior predictive L-measure [6] could be used for more formal comparison of the cure fraction models. However, the computation of these measures is time consuming, if there are tens of thousands of patients in the data set, because replicate data sets need to be simulated.

We were mainly interested in the broad features in the regional variation and its development since 1950s until 2000s. Therefore, instead of modelling the effect of calendar time of diagnosis by some smooth continuous function we simplified the analysis by categorizing the time span into three relatively long subperiods, this division being motivated by certain major steps in the development of the regional health-care system and introduction to the national screening programme. Splines have been used to model the trends in the two survival components in a cure fraction model [30]. Another possibility would be to split the diagnosis periods into narrower subperiods and assume autoregressive priors for the period effects to obtain smoother estimates [31]. Those approaches can be used to model the overall trends. In addition, the district-specific effects can be splitted into narrower periods by the time of diagnosis. Then time-independent random effects might not be appropriate and temporal correlation between the district-specific effects

should be modelled. As another simplification we assumed the effect of district to be the same in every age group, which may not be realistic. For example, if there were differences in screening practices between the hospital districts, the differences would have affected the survival of 50–69 years old patients but that would presumably not have affected the survival of the younger patients.

Explicit ranking of the districts according to the point estimates of the target parameters should be considered with due caution. The ranks tend to have very wide interval estimates and a league table would vainly judge the districts into good and bad performers without telling anything about the absolute differences between the districts [32]. Thus, we decided to show the point estimates of the cure fraction and the mean survival time together with their interval estimates for the three diagnosis periods in the same figure, since it is easier to assess the amount of regional differences with respect to the variation between the diagnosis periods. By comparing the point and the interval estimates with the global mean of the period, it is possible to detect extreme cases that happen to lie in the tails of the random effect distribution. However, when multiple comparisons are made, the probability that at least one of the district-specific random effects does not belong to the estimated interval increases.

The present analysis suggested that both the proportion of cured patients and the mean survival time of non-cured patients increased over time. The posterior mean estimates of the national averages of the two summary parameters increased from 32 per cent and 6.5 years (in 1953–1969) to 64 per cent and 10.4 years (in 1986–2000). Along with the advances in oncological therapies, also the nationwide screening programme that was gradually initiated in 1987 has apparently affected the survival times observed in patients aged 50–69 years at diagnosis in the last period [33]. Active screening can be seen as a part of effective cancer control. At best, screening detects slow-growing tumours while still localized; hence, these patients can be cured. Screening will not benefit so much persons with more aggressive breast cancers that are more likely to be diagnosed between screens. Hence, the mean survival time of the non-cured patients might even decrease as a consequence of screening, because a larger proportion of patients will be detected earlier and will be cured so that relatively more aggressive cancers will be concentrated in the group of the non-cured patients. Besides true benefits, it is reasonable to believe that screening also artificially affected the survival components in the last period [34]. Overdiagnosis has contributed to elevating the cure fraction, and the survival times of some non-cured patients have been prolonged by the lead time effect. In the last period, the possible differences between the municipalities in the screening practices and in the attendance to screening may have affected the regional differences in survival, also artificially. However, this should not be a major problem, because all the municipalities have been involved in the screening programme and the attendance rate has been quite high in the 1990s [35].

The regional variation in breast cancer survival decreased in Finland during the three diagnosis periods. The capital of Finland, Helsinki, clearly stood out from the other districts in 1953–1969, but in 1986–2000 the patients all over the country achieved on average the same level of 10-year survival as the patients in the capital. The current figures support a good equity in the survival of breast cancer patients in Finland.

Acknowledgements

Karri Seppä was supported by a grant from the Cancer Society of Finland and the grant No. 122150 from the Academy of Finland (to T. Hakulinen). Esa Läärä was supported by the Research Grant for Senior Scientists No. 120146 from the Academy of Finland.

References

1. Karjalainen S. Geographical variation in cancer patient survival in Finland: chance, confounding, or effect of treatment? *Journal of Epidemiology and Community Health* 1990; **44**:210–214. DOI: 10.1136/jech.44.3.210.
2. Dickman PW, Gibberd RW, Hakulinen T. Estimating potential savings in cancer deaths by eliminating regional and social class variation in cancer survival in the Nordic countries. *Journal of Epidemiology and Community Health* 1997; **51**:289–298. DOI: 10.1136/jech.51.3.289.
3. Estève J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* 1990; **9**:529–538. DOI: 10.1002/sim.4780090506.
4. Verdecchia A, De Angelis R, Capocaccia R, Sant M, Micheli A, Gatta G, Berrino F. The cure for colon cancer: results from the EUROCARE study. *International Journal of Cancer* 1998; **7**:322–329.
5. Francisci S, Capocaccia R, Grande E, Santaquilani M, Simonetti A, Allemani C, Gatta G, Sant M, Zigon G, Bray F, Janssen-Heijnen M, the EUROCARE Working Group. The cure of cancer: a European perspective. *European Journal of Cancer* 2009; **45**:1067–1079. DOI: 10.1016/j.ejca.2008.11.034.
6. Cooner F, Banerjee S, Carlin BP, Sinha D. Flexible cure rate modeling under latent activation schemes. *Journal of the American Statistical Association* 2007; **102**:560–572. DOI: 10.1198/016214507000000112.
7. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine* 1999; **18**:441–454.
8. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 2007; **8**:576–594. DOI: 10.1093/biostatistics/kxl030.

9. Yu XQ, O'Connell DL, Gibberd RW, Smith DP, Dickman PW, Armstrong BK. Estimating regional variation in cancer survival: a tool for improving cancer care. *Cancer Causes Control* 2004; **15**:611–618. DOI: 10.1023/B:CACO.0000036165.13089.e8.
10. Cooner F, Banerjee S. Modelling geographically referenced survival data with a cure fraction. *Statistical Methods in Medical Research* 2006; **15**:307–324. DOI: 10.1191/0962280206sm453oa.
11. Seppänen J, Heinävaara S, Hakulinen T. Predicting impacts of mass-screening policy changes on breast cancer mortality. *Statistics in Medicine* 2008; **27**:5235–5251. DOI: 10.1002/sim.3345.
12. Teperi J, Vuorenkoski L. Health and health care in Finland since the Second World War. In *Health in Finland*, Koskinen S, Aromaa A, Huttunen J, Teperi J (eds). National Public Health Institute: Helsinki, 2006.
13. Tahvanainen A. Health Politics in Helsinki and Uusimaa. *Ph.D. Thesis* (*in Finnish*), University of Tampere, Tampere, 2004.
14. Stacy EW. A generalization of the gamma distribution. *Annals of Mathematical Statistics* 1962; **33**:1187–1192. DOI: 10.1214/aoms/1177704481.
15. Prentice LR. A log gamma model and its maximum likelihood estimation. *Biometrika* 1974; **61**:539–544. DOI: 10.1093/biomet/61.3.539.
16. Cox C, Chu H, Schneider MF, Muñoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine* 2007; **26**:4352–4374. DOI: 10.1002/sim.2836.
17. Robert CP, Casella G. *Monte Carlo Statistical Methods*. Springer: New York, 1999.
18. Liu JS, Wu YN. Parameter expansion for data augmentation. *Journal of the American Statistical Association* 1999; **94**:1264–1274. DOI: 10.2307/2669940.
19. Gelman A. Parameterization and Bayesian modeling. *Journal of the American Statistical Association* 2004; **99**:537–545. DOI: 10.1198/016214504000000458.
20. Gelman A, van Dyk DA, Huang Z, Boscardin JW. Using redundant parameterizations to fit hierarchical models. *Journal of Computational and Graphical Statistics* 2008; **17**:95–122. DOI: 10.1198/106186008X287337.
21. R Development Core Team. R: A Language and Environment for Statistical Computing, version 2.9.0. R Foundation for Statistical Computing, Vienna, Austria, 2009. Available from: http://www.R-project.org, accessed 29 April, 2009.
22. Spiegelhalter DJ, Thomas A, Best NG, Lunn D. *WinBUGS*, *Version 1.4*, *User Manual*. MRC Biostatistics Unit: Cambridge, 2003.
23. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**:434–455. DOI: 10.2307/1390675.
24. Brenner H, Hakulinen T. Are patients diagnosed with breast cancer before age 50 years ever cured? *Journal of Clinical Oncology* 2004; **22**:432–438. DOI: 10.1200/JCO.2004.04.067.
25. Woods LM, Rachet B, Lambert PC, Coleman MP. 'Cure' from breast cancer among two populations of women followed for 23 years after diagnosis. *Annals of Oncology* 2009; **20**:1331–1336. DOI: 10.1093/annonc/mdn791.
26. Yu B, Tiwari RC, Cronin KA, Feuer EJ. Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine* 2004; **23**:1733–1747. DOI: 10.1002/sim.1774.
27. Dickman PW, Hakulinen T. Adjusting for region of residence in relative survival analysis. *Journal of Epidemiology and Biostatistics* 1996; **1**:213–218.
28. Heinävaara S, Teppo L, Hakulinen T. Cancer-specific survival of patients with multiple cancers: an application to patients with multiple breast cancers. *Statistics in Medicine* 2002; **21**:3183–3195. DOI: 10.1002/sim.1247.
29. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics* 2009; **9**:523–539. DOI: 10.1093/biostatistics/kxm049.
30. Lambert PC, Dickman PW, Österlund P, Andersson T, Sankila R, Glimelius B. Temporal trends in the proportion cured for cancer of the colon and rectum: a population-based study using data from the Finnish Cancer Registry. *International Journal of Cancer* 2007; **121**:2052–2059. DOI: 10.1002/ijc.22948.
31. Berzuini C, Clayton D. Bayesian analysis of survival on multiple time scales. *Statistics in Medicine* 1994; **13**:823–838. DOI: 10.1002/sim.4780130804.
32. Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A* 1996; **159**:385–443. DOI: 10.1111/1467-985X.00087.
33. Sarkeala T, Heinävaara S, Anttila A. Organised mammography screening reduces breast cancer mortality: a cohort study from Finland. *International Journal of Cancer* 2008; **122**:614–619. DOI: 10.1002/ijc.23070.
34. Saxén E, Hakama M. Cancer illness in Finland. With a note on the effects of age adjustment and early diagnosis. *Annals Medicinae Experimentalis et Biologiae Fenniae* 1964; **42**(Suppl.2):1–28.
35. Sarkeala T, Anttila A, Forsman H, Luostarinen T, Saarenmaa I, Hakama M. Process indicators from ten centres in the Finnish breast cancer screening programme from 1991 to 2000. *European Journal of Cancer* 2004; **40**:2116–2125. DOI: 10.1016/j.ejca.2004.06.017.