Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

Bayesian assessment of times to diagnosis in breast cancer screening

Carmen Armero [a], Antonio López-Quílez [a] & Rut López-Sánchez [a]
[a] Departament d'Estadica i Investigació Operativa, Universitat de València, Spain
Published online: 01 Aug 2008.

To cite this article: Carmen Armero, Antonio López-Quílez & Rut López-Sánchez (2008) Bayesian assessment of times to diagnosis in breast cancer screening, Journal of Applied Statistics, 35:9, 997-1009, DOI: 10.1080/02664760802191397

To link to this article: http://dx.doi.org/10.1080/02664760802191397

Journal of Applied Statistics
Vol. 35, No. 9, September 2008, 997–1009

Bayesian assessment of times to diagnosis in breast cancer screening

Carmen Armero*, Antonio López-Quílez and Rut López-Sánchez

Departament d'Estadística i Investigació Operativa, Universitat de València, Spain

Breast cancer is one of the diseases with the most profound impact on health in developed countries and mammography is the most popular method for detecting breast cancer at a very early stage. This paper focuses on the waiting period from a positive mammogram until a confirmatory diagnosis is carried out in hospital. Generalized linear mixed models are used to perform the statistical analysis, always within the Bayesian reasoning. Markov chain Monte Carlo algorithms are applied for estimation by simulating the posterior distribution of the parameters and hyperparameters of the model through the free software WinBUGS.

Keywords: Bayesian statistics; breast cancer screening program; generalized linear mixed models; Markov Chain Monte Carlo

1. Introduction

Breast cancer is one of the diseases with the most profound impact on women's health in developed countries. The success of therapies basically depends on the extent of damage at the time of detection. Consequently, an early diagnosis is the best way of improving the possibilities of a recovery. The most popular method for discovering breast cancer at a very early stage is breast screening through a mammogram, which is an X-ray of each breast. Mammography can detect small alterations in breast tissue which may suggest malignant tumors before they can be identified by clinical or self-breast examination. These radiograms are the main tool of the system designed by health authorities to reduce overall breast cancer mortality rates in the target female population [11,22]

When a woman has a mammogram indicating a well-founded suspicion of malignancy, the quality of pathological services, providing quick and reliable diagnoses and prognostic information, is essential for successful treatment. In addition, a woman in this extreme situation suffers from stress, anxiety, depression, fear, or even more serious psychological effects [12,13]. Consequently, shortening and improving the effectiveness of the different steps involved in the diagnostic process is desirable in order to increase the quality and benefits of the service. For these reasons,

*Corresponding author. Email: Carmen.Armero@uv.es

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760802191397
http://www.informaworld.com

quantitative studies to gain knowledge of the system become necessary to validate theoretical patterns and paths with real data in order to identify delays [5,15,19], assess survival times [10,18], evaluate success in final diagnosis [7,14,17], and so on.

Our reference framework is the *Comunitat Valenciana* Breast Cancer Screening Program (CV-BCSP, from now on). The *Comunitat Valenciana* is an autonomous region in the east of Spain, has a population of about 4.2 million, and has full responsibility for the management of all its health services. The first step for a woman in the CV-BCSP occurs in the breast screening unit and involves a mammogram of each breast, which is independently examined by two radiologists (a double reading). They need to reach a consensus and basically classify the reading into five categories: normal (without pathological signs), benign, findings which are probably benign, those with highly suspicious results of malignancy, and malignant.

In the case of an abnormal mammogram (the last two categories where malignancy is strongly suspected), the woman is transferred to hospital in order to confirm the diagnosis. This is a very complex process with a stringent protocol based on one or several diagnostic tests. In this paper, we concentrate exclusively on women with abnormal mammographies, who have only been exposed to one of these confirmatory procedures: a stereotactic core biopsy (CoreB from now on) and an open biopsy (OpenB from now on). CoreB is performed by inserting a hollow needle through the skin and into the breast in order to remove a tissue sample which is analyzed to confirm malignancy, while in OpenB all or part of the lump is removed. In patients with abnormal mammographies, CoreB was developed to provide a non-surgical means for diagnosing breast cancer, whereas OpenB is more aggressive and expensive, but provides a more definitive diagnosis.

Our main purpose in this paper is to provide and analyze a statistical model for the time period between an abnormal mammography and diagnosis after an OpenB or a CoreB has been conducted. This is done in connection with the qualitative evaluation of the mammographic lesion and by taking into account variability in the different hospitals where the diagnoses are carried out. Generalized linear mixed models (GLMMs) are used to estimate and perform the statistical analysis, always within Bayesian reasoning. Markov Chain Monte Carlo (MCMC) algorithms are applied for estimation by simulating the posterior distribution of parameters and hyperparameters of the model through the free software [21].

The rest of the paper is as follows. Section 2 contains a brief background on the main characteristics of the CV-BCSP, particularly the protocol for mammograms with positive evidence of cancer, introduces the data, and presents a basic description of them. Section 3 contains the main part of the paper. Subsection 3.1 discusses the main features of the modelling process and Subsection 3.2 presents the selected model and reports the principal results. The paper ends with a short section devoted to some concluding remarks.

2. Screening for breast cancer and mammogram data

Although there are drawbacks involved in screening women under the age of 50, public screening programs for the early detection of breast cancer have provided an important and beneficial impact on health [11]. In the *Comunitat Valenciana*, the CV-BCSP [3] was established in 1992 and concentrates on all healthy women aged between 45 and 65. They are routinely invited for breast screening every two years, and since 2001, it has been extended to women up to the age of 69. At present, there are 23 breast screening units across our region, covering the total target population ($>$500,000 women). Participation in this program is over 70% and on average four tumors per 1000 women screened are detected, many of them ($\sim$65%) in the early stages of 0 and I.

Our data bank comes from the CV-BCSP and records all women with an abnormal mammography during 2002 who have had a final diagnosis as a result of a CoreB or an OpenB.

All observations are made anonymously and contain, among other things, information on:

- The time, in days, required to determine the final diagnosis ('time to diagnosis' from now on) and the conclusion (negative or positive for cancer) of the confirmation procedure.
- The age, in years, of the woman at the moment when the positive mammographic reading is discovered.
- The mammographic lesion, which is evaluated for the presence of calcifications, combined and/or multiple damage, nodules and masses, architectural distortion, asymmetric parenchyma and skin changes.
- The hospital (25 throughout the Comunitat Valenciana) where the tests are conducted.

Table 1 presents the total number of women with an abnormal mammogram and the final diagnosis for cancer, positive or negative, after the procedure CoreB or OpenB is carried out. It is worth noting that when a CoreB is conducted, the percentage of women with a final positive diagnosis is 61%, whereas in the case of an open biopsy this quantity is nearly 68%. Note that both values are >50%, which is the objective in Europe [6].

The empirical distribution of the time to diagnosis for both procedures is presented in Figure 1. A heavy right tail can be observed in both distributions as well as a clear concentration of the data on small values. For CoreB times, the median is 23 days with a standard deviation of 21.41 days and 25% (75%) of the women are waiting for >41 (21) days for confirmation of the diagnosis. In the case of the OpenB times, the median is nearly twice as much, 41 days with a standard deviation of 33.22 days, the 25th (75th) percentile being 27.25 (62.75) days.

Table 1. Total number of women with a positive and negative diagnosis of cancer after a CoreB or an OpenB test has been conducted.

Test	♯ Women	Positive diagnosis	Negative diagnosis
CoreB	100	61	39
OpenB	222	150	72
Total	322	211	111

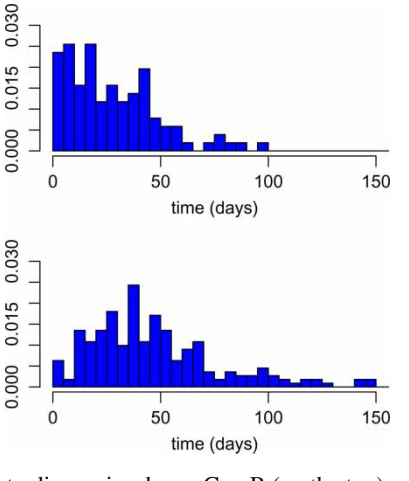

Figure 1. Histogram of the times to diagnosis when a CoreB (on the top) or an OpenB (on the bottom) has been performed.

Table 2. Total frequency and percentage of women according to the mammographic lesion and diagnostic procedure, CoreB or OpenB, conducted.

	Total	CoreB		OpenB	
Mammographic lesion	Frequency	Frequency	Percentage	Frequency	Percentage
1. Calcification	137	46	33.58	91	66.42
2. Combined/multiples	98	24	24.49	74	75.51
3. Nodules and masses	50	20	40.00	30	60.00
4. Architectural distortion	17	3	17.65	14	82.35
5. Asymmetric parenchyma	11	5	45.45	6	54.55
6. Skin changes	1	0	00.00	1	100.00
Missing data	8	2	25.00	6	75.00
Total	322	100	31.06	222	68.94

Table 2 shows the empirical frequency distribution for the mammographical lesion categories for all the women and with regard to the technique, CoreB or OpenB, performed. It can be noted that calcifications, combined/multiples, nodules and masses are the most frequent groups, covering practically 90% of the total. It can also be noted that the choice between CoreB and OpenB procedures varies with the type of lesion. Despite the small quantity of data, we observe that in the case of architectural distortion and combined/multiple lesions, the majority of the women (82.35% and 75.51%, respectively) have been sent for an OpenB procedure.

For CoreB and OpenB times, Figure 2 displays a histogram of the time to diagnosis with respect to the mammographic lesion detected. We have only included the graphs corresponding to the most numerous categories: calcifications, combined/multiples, nodules and masses. All of them exhibit positive skewness and the categories with the longest times always being OpenB. For CoreB times, women in the combined/multiple category have shorter times with a median of 14 days and a standard deviation of 20.47 days. Nodules and masses and calcifications have a median of 20.0 and 29.0 days and a standard deviation of 19.91 and 20.75 days, respectively. When the procedure is OpenB, women with nodules and masses or with combined/multiple lesions exhibit the same median time to diagnosis, 38 days, and similar standard deviations, 40.74 and 35.03 days, respectively. Also, the longer times are associated with calcifications with a median of 46 days and a standard deviation of 30.42 days.

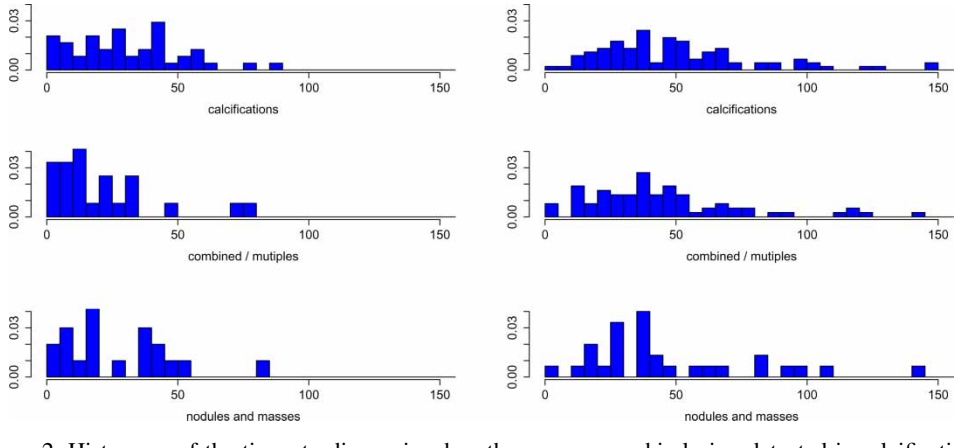

Figure 2. Histogram of the times to diagnosis when the mammographic lesion detected is calcifications, combined/multiples, nodules and masses and the diagnostic procedure is a CoreB (on the left) or an OpenB (on the right).

3. Waiting time for a diagnosis

We are interested in analyzing and comparing CoreB and OpenB procedures with regard to the time to diagnosis where the confirmatory procedures are carried out. The woman's age is also considered in the analysis. It is well known that the incidence of breast cancer increases with age, but also that age should not be linked with the time to diagnosis. Our interest here in this variable is justified by an element of distrust in the management of the program with regard to a prior weak suspicion that older women have longer times to diagnosis than younger. Fortunately, as we will discuss later, the results do not confirm these ideas. We have structured this section into two parts: Subsection 3.1 presents a detailed discussion of the modeling process in all its stages. Subsection 3.2 contains the final fitted model with the main results.

3.1. *Formulating the model*

Times to diagnosis are similar to times to failure in reliability studies and life times in biomedical applications, etc, that is the so-called times to event data [4]. At first, we also examined a logarithmic transformation in order to analyze our times with a normal model, but the results were not satisfactory. We also noticed that there were some zero values in the data. These are quite infrequent and really correspond to non-zero values which are less than or equal to 1 day. They occur in situations where the unit is located inside the hospital and the lesions seem extremely serious. Finally, for practical purposes we have considered them in the data set as near but different from zero. So, with no transformations, we immediately thought about the typical distributions for these data, basically on exponential and gamma distributions. After observing the histogram of the data in Figure 2, the gamma model is assumed to be appropriate for describing the stochastic behavior of the time to diagnosis with each procedure, that is:

$$T_{ijk} \sim \text{Ga}(a_{ijk}, b_{ijk}),$$

where k indexes the woman, j the hospital, i the mammographic lesion and $\text{Ga}(x \mid a, b) = [b^a / \Gamma(a)]x^{a-1}\,\text{e}^{-bx}$, $x > 0$, stands for the density at x of a gamma distribution with expectation a/b and variance a/b^2.

Next, we need to use a statistical model that allows us to connect this distribution with the woman's age, the type of lesion, and the hospital where the diagnostic procedure is performed. Our interest in these variables is not the same for every one: the age of the woman is a covariate, the mammographic lesion is considered as a fixed-effect factor and the hospital a random-effect one. Those in charge of the program do not feel that there are any systematic differences between hospitals and so, we have considered them as a random sample from a common general population.

Generalized linear models (GLMs) are the natural extension of linear models in cases in which the normal distribution is not appropriate for describing the probabilistic behavior of the outcome variable given the explanatory variables. Following this spread process, generalized linear mixed models (GLMMs) are more general models that allow the inclusion of random effects in their structure. This type of model is specially suited to dealing with more complicated error scenarios [2,23].

In our case, for each one of the CoreB and OpenB times, we have first considered the regression equation for the linear predictor as:

$$g(\mu_{ijk}) = \alpha + (l_i - l_1) + h_j + \beta x_{ijk}, \tag{1}$$

where $g(.)$ is a link function, $\mu_{ijk} = a_{ijk}/b_{ijk}$ is the expected mean of T_{ijk} given the predictor, α is the intercept, β the coefficient associated with the covariate x_{ijk} that registers the age of the woman, h_j represents the random effect of the hospital j, $j = 1, \ldots, 25$, and l_i the mammographic lesion effect, where $i = 1$ stands for calcifications that we have chosen as the reference category

because it is the most numerous, $i = 2$ for combined/multiples, $i = 3$ for nodules and masses, $i = 4$ for architectural distortion, $i = 5$ for asymmetric parenchyma and $i = 6$ for skin changes. Note that in Equation (1), the effect of the mammographic lesion is expressed with respect to lesion 1 and consequently, the effect of this category is included in the parameter α.

We have examined various link functions. At first, we chose the canonical link (the reciprocal one in the case of the gamma distribution), but it did not work suitably because it led to non-positive values for some posterior expectations of the mean time to diagnosis. Finally, we decided to use the logarithmic link, $g(\mu_{ijk}) = \log(\mu_{ijk})$, because it operated very well with our data and it is a very usual link for the gamma distribution due to the close relationship with the logarithmic transformation of the response-variable data.

In order to fully manage our statistical modeling we need to elicit a prior distribution for the parameters $\alpha, \beta, l_i - l_1, i = 2, \ldots, 6$ and build a hierarchical model for the random hospital effect h_j, $j = 1, \ldots, 25$. We established a general scenario with non-informative prior knowledge but, in order to avoid any chance of obtaining improper posteriors, we have always worked with proper densities. Particularly, following the idea that a prior distribution can symbolize a population of the possible values of the parameter, we have considered:

$$\begin{aligned} \alpha &\sim N(0, 100) \\ \beta &\sim N(0, 100) \\ l_i - l_1 &\sim N(0, 4), \quad \text{i.i.d.} \quad \text{i} = 2, 3, 4, 5, 6 \end{aligned}$$

where in all cases the parameterization of the normal distribution is in terms of the mean and variance. It is interesting to point out that not only do all these priors have scant information, but they are also especially ample and reasonable. In the case of the parameters α and β, they are a common non-informative choice in these types of models. In fact, the value 100 for the prior variance of α and of β allows the parameter to cover, for at least 95% of the values, the interval $[-19.6, +19.6]$, which is a very wide range of values, especially when we realize that we are working in the logarithmic scale. In the case of the difference for lesion effects $l_i - l_1$, $i = 2, \ldots, 6$, we have also used a wide prior, although with a minor variance. With this prior we can cover, for at least 95% of the values, the interval $[-3.92, +3.92]$ which means that, compared with the effect of the mammographic lesion 1, the effect of the other type of lesions on the expected time to diagnosis can be about 50 times greater and 50 times less, obviously given the rest of predictors being unchanged. This is an extremely generous quantity in the opinion of our medical colleagues in CV-BSCP.

With regard to the random hospital effects, $h_1, \ldots, h_{25}$, we assume that they constitute a random sample from a general population $N(0, \eta^2)$, where η^2 is the hyperparameter describing the variance of this common distribution. The elicitation of hyperprior distributions in this type of models is a very challenging issue in Bayesian statistics that has produced, and is still generating, very fruitful discussions and results [8,16]. In this scenario, Inverse Gamma for variance parameters and Uniform models for standard deviations have been the most popular proposals for dispersion parameters: the first one because of conditional conjugacy and the second one because it seems to play a more neutral role in the analysis. We have been careful with the Inverse Gamma model and we have not worked with extremely small hyperparameter values. Obviously, other alternative choices could also be interesting in a more demanding study. In this paper, we have examined these two possibilities:

$$\begin{aligned} \eta^2 &\sim \text{IGa}(0.01,\ 0.01), \\ \eta &\sim \text{Un}(0.1,\ 3) \end{aligned}$$

where $\mathrm{IGa}(x \mid a,\ b) = [b^a / \Gamma(a)]x^{-(a+1)}\, e^{-b/x}$, $x > 0$ stands for the density at x of an Inverse Gamma distribution with expectation $b/(a-1)$, for $a > 1$ and variance $b^2/[(a-1)^2(a-2)]$, for $a > 2$.

Different assumptions for the variance of the time to diagnosis, Var (T_{ijk}) denoted by σ^2_{ijk}, have also been examined:

PARAMETERIZATION 1 *This is the simplest one. It assumes constant variability of times to diagnosis, independently of the woman, the lesion and the hospital considered, that is* $\sigma^2_{ijk} = \epsilon^2$, *and so*:

$$a_{ijk} = \frac{\mu^2_{ijk}}{\sigma^2_{ijk}} = \frac{\mu^2_{ijk}}{\epsilon^2},$$
$$b_{ijk} = \frac{\mu_{ijk}}{\sigma^2_{ijk}} = \frac{\mu_{ijk}}{\epsilon^2}.$$

PARAMETERIZATION 2 *The coefficient of variation of the times to diagnosis is the same in all hospitals,* $\sigma_{ijk}/\mu_{ijk} = \varphi$. *Consequently, we expect less variability in the hospitals with shorter times*:

$$a_{ijk} = \frac{\mu^2_{ijk}}{\sigma^2_{ijk}} = \frac{1}{\varphi^2},$$
$$b_{ijk} = \frac{\mu_{ijk}}{\sigma^2_{ijk}} = (\varphi^2 \mu_{ijk})^{-1}.$$

PARAMETERIZATION 3 *The variance of the time to diagnosis is proportional to its expected time,* $\sigma^2_{ijk}/\mu_{ijk} = \psi^2$. *This structure also gives more variability to hospitals with longer times, but is weaker than in parameterization 2. In this situation, the relationship between the parameters can be expressed as*:

$$a_{ijk} = \frac{\mu^2_{ijk}}{\sigma^2_{ijk}} = \frac{\mu_{ijk}}{\psi^2}$$
$$b_{ijk} = \frac{\mu_{ijk}}{\sigma^2_{ijk}} = \frac{1}{\psi^2}.$$

We are still working in the same general scenario of vague prior information for eliciting a prior distribution for ϵ, φ, and ψ. Uniform distributions defined in intervals which are wide enough to allow us to capture the more extreme values of the parameters (taking care to avoid values near zero for variance parameters) are selected.

3.2. *The fitted model and the results*

The different model configurations discussed before were estimated by using MCMC and run in WinBUGS. For each model tested, we considered three parallel chains running with different starting values in order to examine convergence with the statistics proposed by Gelman and Rubin [9]. With each run, we have always performed 300,000 iterations with a burn-in of 50,000 iterations. The deviance information criterion (DIC) introduced by Spiegelhalter et al. [20] has been used to compare the different fitted models. This criterion weights up the goodness-of-fit and the complexity of the model and is especially suitable when, as in this paper, comparing models

whose posterior distributions have been approximated by MCMC methods, but there is also some controversy about its performance, especially in mixture and missing value problems [1].

In the final evaluation of CoreB and OpenB times, we hardly discovered any differences in the results when considering Inverse Gamma or Uniform hyperpriors. However, in all the parameterizations, the Uniform hyperpriors provide the smallest DIC values. We also discovered that age does not improve the fitted model, neither in the case of the CoreB nor the OpenB, and consequently we have removed the subindex k from the final model.

Table 3 presents the results of the posterior mean of the deviance (as a measure of model goodness-of-fit), the deviance of the posterior means of the parameters, the so-called effective number of parameters (as a measure of model complexity) and the value of DIC for the different models constructed to study time to diagnosis in relation to the procedure, CoreB or OpenB, and the parameterization examined. Parameterization 2 provides the best results for CoreB times (it has the lowest value for the DIC), whereas parameterization 3 is the more suitable one for OpenB times. However, it is also noteworthy that parameterization 1 provides a very similar DIC to the best one in both CoreB and OpenB models.

So, our fitted model turns out as:

$$T_{ij} \sim \text{Ga}(a_{ij}, b_{ij}), \quad \text{with } \mu_{ij} = \frac{a_{ij}}{b_{ij}}, \sigma_{ij}^2 = \frac{a_{ij}}{b_{ij}^2}$$

$$\log(\mu_{ij}) = \alpha + (l_i - l_1) + h_j, \quad i = 1, 2, \ldots, 6, \quad j = 1, \ldots, 25$$

with the following characteristics depending on the confirmatory procedure, CoreB or OpenB, selected:

1. CoreB:

$$a_{ij} = \frac{1}{\varphi^2}, b_{ij} = (\varphi^2 \mu_{ij})^{-1}$$

$$\alpha \sim N(0,\ 100), l_i - l_1 \sim N(0, 4), \quad i = 2, 3, 4, 5, \quad \varphi \sim \text{Un}(0.1, 2)$$

$$h_j \sim N(0, \eta^2), \quad j = 1, \ldots, 25, \quad \eta \sim \text{Un}(0.1, 3)$$

2. OpenB:

$$a_{ij} = \frac{\mu_{ij}}{\psi^2}, b_{ij} = \frac{1}{\psi^2}$$

$$\alpha \sim N(0,\ 100), \quad l_i - l_1 \sim N(0, 4), \quad i = 2, 3, 4, 5, 6, \quad \psi \sim \text{Un}(0.2, 10)$$

$$h_j \sim N(0, \eta^2), \quad j = 1, \ldots, 25, \quad \eta \sim \text{Un}(0.1, 3)$$

Table 3. Computed posterior expected deviance (Dbar), deviance of posterior means of the parameters (Dhat), effective number of parameters (pD) and deviance information criterion (DIC) for CoreB and OpenB time to diagnosis models with regard to the three different parameterizations examined.

	CoreB				OpenB			
Parameterization	Dbar	Dhat	pD	DIC	Dbar	Dhat	pD	DIC
1	843.341	824.868	18.473	861.814	2038.530	2014.660	23.872	2062.410
2	843.207	824.752	18.455	861.662	2064.160	2042.230	21.920	2086.080
3	852.869	835.575	17.293	870.162	2038.280	2014.320	23.958	2062.230

Table 4 summarizes the main characteristics (mean, standard deviation and 95% credible interval) of both fitted models computed from the sample of the posterior distribution of the parameters and hyperparameters generated by WinBUGS. In the case of the model for CoreB times, this posterior sample is as follows

$$\{(\alpha^{(m)},\ (l_i - l_1)^{(m)},\ i = 2, \ldots, 6,\ h_j^{(m)},\ j = 1, \ldots, 25,\ \varphi^{(m)},\ \eta^{(m)}),\ m = 1, \ldots, 750{,}000\}, \tag{2}$$

and for CoreB procedure, all elements in Table 4 are calculated from this sample. For instance, if we focus on the parameter α, the mean value that appears in Table 4 as 3.498 corresponds to the mean of the sample $\{\alpha^{(m)}, m = 1, \ldots, 750{,}000\}$ of the posterior marginal distribution, $p(\alpha \mid \text{data})$.

In both confirmatory procedures, the effect of the mammographic lesion corresponding to calcifications (the reference one) is important. Note that posterior means and intervals for this effect indicate different values for CoreB and OpenB procedures, generally with fewer values for CoreB (broadly speaking about $e^{3.498} = 33.05$ and $e^{3.888} = 48.81$ days, respectively). Also in both of them, the differential effect of the remaining mammographic categories (except for skin changes) on the expected time to diagnosis with regard to calcifications have negative means. However, also because most of the intervals in the table contain negative and positive values, it seems clear that their corresponding differential effect will be statistically imperceptible. For CoreB procedures, and always comparing with calcifications, we observe that combined/multiple lesions will have a clearly negative effect on the expected times to diagnosis. This is also the situation of nodules and masses, architectural distortion and asymmetric parenchyma in OpenB. What is also remarkable is the striking quantities for OpenB and the difference in the outputs when comparing CoreB and OpenB for asymmetric parenchyma. With respect to skin changes, it is clear that the only observation in OpenB produces a result which is scarcely informative.

All the previous comments about the coefficients of the fitted models are clearer when examining the expected time to diagnosis with regard to mammographic lesions. Irrespective of the hospital where the woman will be sent for confirmation of the diagnosis, Figure 3 displays a 95% credible interval for the expected time to diagnosis according to the lesion registered. The intervals in both pictures are constructed from the corresponding sample of the posterior distribution of parameters and hyperparameters generated by WinBUGS. In particular, for CoreB times, we use the sample in (2) to construct a random sample for the posterior distribution of the expected time to diagnosis for each type of lesion, $p\,(\mu_i, i = 1, \ldots, 6 \mid \text{data})$, as $\{\mu_i^{(m)} = \exp\{\alpha^{(m)} + (l_i - l_1)^{(m)}\},\ i = 1, \ldots, 6,\ m = 1, \ldots, 750{,}000\}$. For each lesion, the interval and median displayed in Figure 3 are the empirical interval and median of this sample.

At first sight, we appreciate quite a regular pattern in the variability of OpenB intervals in all mammographic lesions (except for skin changes). This is not the situation for CoreB intervals: variability corresponding to calcifications, combined/multiples, nodules and mass lesions are similar, whereas the posterior sample for the effects of architectural distortion and asymmetric parenchyma exhibit a positive skewness and a great dispersion. For OpenB, the longest and shortest times to diagnosis correspond to calcifications and asymmetric parenchyma, respectively. For combined/multiple, and nodule and mass categories, the difference between CoreB and OpenB increases and the estimated OpenB mean is about twice the CoreB one (41.49 and 38.07 days versus 20.21 and 20.57 days, respectively). CoreB and OpenB samples for architectural distortion have a similar mean of 36.58 and 33.78 days, respectively, but the distribution of both samples is clearly different. Asymmetric parenchyma also presents a similar feature.

Finally, we deal with the marginal relation between time to diagnosis and medical centers. For CoreB and OpenB procedures, Figure 4 displays a 95% credible interval of the expected time to diagnosis for women with calcifications in the different hospitals, depending on the procedure

Table 4. Mean, standard deviation and 95% interval of the MCMC sample of the posterior distribution of the parameters and hyperparameters of the fitted model for CoreB and for OpenB times. Throughout this table, subindex 1 is for the mammographic lesion corresponding to calcifications, 2 for combined/multiples, 3 for nodules and masses, 4 for architectural distortion, 5 for asymmetric parenchyma and 6 for skin changes.

CoreB				OpenB			
Parameter	Mean	sd	(2.5%, 97.5%)	Parameter	Mean	sd	(2.5%, 97.5%)
α	3.498	0.234	(3.046, 3.971)	α	3.888	0.119	(3.646, 4.113)
$l_2 - l_1$	−0.485	0.219	(−0.913, −0.053)	$l_2 - l_1$	−0.167	0.092	(−0.346, 0.012)
$l_3 - l_1$	−0.368	0.238	(−0.830, 0.108)	$l_3 - l_1$	−0.249	0.126	(−0.501, −0.006)
$l_4 - l_1$	−0.075	0.563	(−1.146, 1.087)	$l_4 - l_1$	−0.368	0.179	(−0.734, −0.033)
$l_5 - l_1$	−0.104	0.418	(−0.887, 0.747)	$l_5 - l_1$	−0.820	0.326	(−1.517, −0.246)
				$l_6 - l_1$	0.321	0.730	(−1.428, 1.454)
Hyperparameter	Mean	sd	(2.5%, 97.5%)	Hyperparameter	Mean	sd	(2.5%, 97.5%)
φ	0.753	0.056	(0.652, 0.872)	ψ	4.020	0.214	(3.630, 4.464)
η	0.732	0.214	(0.370, 1.216)	η	0.431	0.096	(0.272, 0.649)

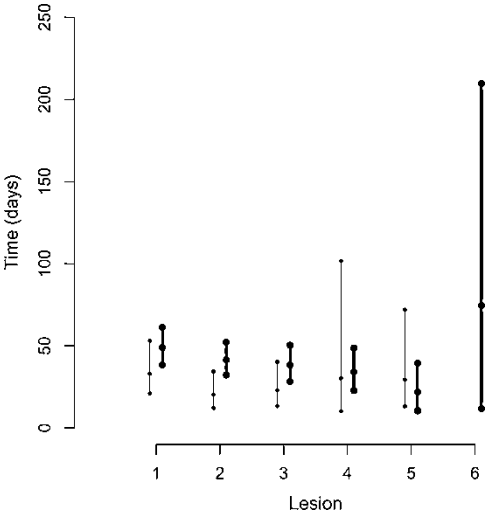

Figure 3. Ninety-five percent credible interval and median for the expected time to diagnosis with respect to the mammographic lesion registered and the diagnostic procedure CoreB (intervals on the left in fine lines) or OpenB (on the right in bold type lines) performed. In the picture, 1 stands for the mammographic lesion corresponding to calcifications, 2 for combined/multiples, 3 for nodules and masses, 4 for architectural distortion, 5 for asymmetric parenchyma and 6 for skin changes.

performed. The intervals in broken lines correspond to hospitals with no data. Also here, all the intervals in both pictures are constructed from the corresponding sample of the posterior distribution of parameters and hyperparameters generated by WinBUGS. In the case of the fitted model to CoreB times, we use the sample in (2) to construct a random sample for the posterior distribution of the expected time to diagnosis in each hospital as $\{\mu_j^{(m)} = \exp\{\alpha^{(m)} + h_j^{(m)}\},\ j = 1, \ldots, 25,\ m = 1, \ldots, 750{,}000\}$. For each hospital, the interval and median displayed in Figure 4 are the empirical interval and median of this sample. It is worth noting that hospitals with no data can only feed the posterior distribution with the prior one and so the corresponding interval (wide) and median is always the same for all of them.

As was to be expected, in both diagnostic procedures, a considerable quantity of the variability detected comes from the centers: the sample mean of the standard deviation η of the effect of the hospitals is 0.732 in the case of CoreB times and 0.431 for OpenB times (remember that we are in the logarithmic scale). If we only focus our attention on the CoreB procedure, we observe

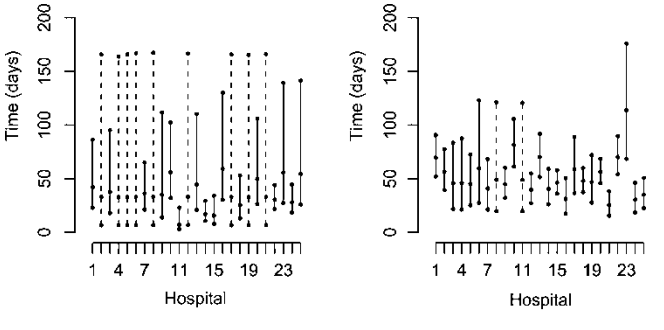

Figure 4. Ninety-five percent credible interval and median for the expected time to diagnosis for the 25 hospitals participating in the CV-BCSP when the diagnostic procedure is a CoreB (on the left) and an OpenB (on the right).

remarkable differences among hospitals: there are centers with expected times of under 20 days and hospitals with mean times between 50 and 75 days. A quick look at OpenB intervals gives the same general impression, although with smaller variability.

4. Conclusion

Bayesian GLMMs are used to analyze the time to diagnosis for CoreB and OpenB procedures with regard to the mammographic lesion and the hospital where the confirmatory intervention is carried out. Different models are examined and compared using the DIC statistics. The fitted model for each procedure shows different behavior for the time to diagnosis, particularly in relation to the differential effect of the different lesions in relation to the calcification lesion, that is chosen as the reference category. All the results indicate that CoreB presents the lower waiting times than OpenB, with the type of mammographic lesion being a relevant factor and with substantially added variability due to the medical institution where the confirmatory tests are carried out.

Acknowledgements

This research is partially supported by the Ministerio de Educación y Ciencia, under grant MTM2007-61554 and the Conselleria de Sanitat de la Generalitat Valenciana. The authors would like to thank the working group of the Programa de Prevención del Cáncer de Mama de la Comunidad Valenciana for their collaboration. We also acknowledge the constructive comments and suggestions on the original version of this paper by two anonymous referees.

References

[1] G. Celeux, F. Forbes, C.P. Roberts and D.M. Titterington, *Deviance information criteria for missing data models*, Bayesian Anal. 4 (2006), pp. 651–674.
[2] D.G. Clayton, *Generalized linear mixed models,* in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson and D.J. Spiegelhalter eds., Chapman and Hall, London, 1996, 275–301.
[3] Conselleria de Sanitat de la Generalitat Valenciana, *Programa de prevención de cáncer de mama en la Comunidad Valenciana*. Monografies Sanitaries, serie E, núm. 25. Conselleria de Sanitat, Valencia, 1998.
[4] M. Crane, M.C Newman, P.F. Chapman, and J. Fenlon, *Risk Assessment with Time to Event Models*, Lewis Publishers, Boca Raton, 2002.
[5] L. Duijm, J.H. Groenewoud, F.H. Jansen, J. Fracheboud, M. van Beek, and H.J. de Koning, *Mammography screening in the Netherlands: delay in the diagnosis of breast cancer after breast cancer screening*, Br. J. Cancer 91 (2004), pp. 1795–1799.
[6] European Comission *European Guidelines for Quality Assurance in Mammography Screening*, 3rd ed. and 2001, ISBN 92-894-1145-7.
[7] A.E. Gelfand and F. Wang, *Modelling the cumulative risk for a false-positive under repeated screening events*, Stat. Med. 19 (2000), pp. 1865–1879.
[8] A. Gelman, *Prior distributions for variance parametrs in hierarchical models*, Bayesian Anal. 3 (2006), pp. 515–534.
[9] A. Gelman and D.B. Rubin, *Inference from iterative simulation using multiple sequences*, Stat. Sci. 7 (1992), pp. 457–472.
[10] K. Hemming and J.E.H. Shaw, *A parametric dynamic survival model applied to breast cancer survival times*, J. Roy. Stat. Soc. C Appl. Stat. 51(4) (2002), pp. 421–435.
[11] S. Lee, H. Huang, and M. Zelen, *Early detection of disease and scheduling of screening examinations*, Stat. Methods Med. Res. 13 (2004), pp. 443–456.
[12] I.M. Lipkus, S. Halabi, T.S. Strigo, and B.K. Rimer, *The impact of abnormal mammograms on psychosocial outcomes and subsequent screening*, Psycho-Oncol. 9(5) (2000), 402–410.
[13] J.B. Lowe, K.P. Balanda, C. Del Mar, and E. Hawes, *Psychologic distress in women with abnormal findings in mass mammography screening*, Cancer 85(5) (1999), pp. 1114–1118.
[14] J. McCann, D. Stockton, and S. Godward, *Impact of false-positive mammography on subsequent screening attendance and risk of cancer*, Breast Cancer Res. 4(5) (2002), p. R11.
[15] M. Montella, A. Crispo, M. D'Aiuto, G. de Bellis, G. Fabbrocini, M. Pizzorusso, M. Tamburini, and P. Sivestra, *Determinant factors for diagnostic delay in operable breast cancer patients* Eur. J. Cancer Prev. 10 (2001), pp. 53–59.

[16] R. Natarajan and R.E. Kass, *Reference Bayesian methods for generalized linear mixed models*, J. Am. Stat. Assoc. 95 (2000), pp. 227–237.

[17] P. Paliwal, A.E. Gelfand, L. Abraham, W. Barlow, and J.G. Elmore, *Examining accuracy of screening mammography using an event order model*, Stat. Med. 25 (2006), pp. 267–283.

[18] R. Pérez-Ocón, J.E. Ruiz-Castro, and M.L. Gámiz-Pérez, *Non-homogeneous Markov models in the analysis of survival after breast cancer*, J. Roy. Stat. Soc. C (Appl. Stat.) 50(1) (2001), pp. 111–124.

[19] K. Robin Yabroff, N. Breen, S.W. Vernon, H.I. Meissner, A.N. Freedman, and R. Ballard-Barbash, *What factors are associated with diagnostic follow-up after abnormal mammograms? Findings from a US National Survey*, Cancer Epidemiol. Biomarkers Prev. 13(5) (2004), pp. 723–731.

[20] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. van der Linde, *Bayesian measures of model complexity and fit (with discussion)*, J. Roy. Stat. Soc. B (Stat. Methodol.) 64(4) (2002), pp. 583–639.

[21] D.J. Spiegelhalter, A. Thomas, N. Best and D. Lunn, WinBUGS Version 1.4 User Manual, 2003. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf

[22] J. Warwick and S.W. Duffy, *A review of cancer screening evaluation technique with some particular examples in breast cancer screening*, J. Roy. Stat. Soc. A 168 (2005), pp. 657–677.

[23] Y. Zhao, J. Staudenmayer, B.A. Coull, and M.P. Wand, *General design Bayesian generalized linear mixed fodels*, Stat. Sci. 21 (2006), pp. 35–51.