Journal of Data Science **8**(2010), 495-504

The Bayesian Multiple Logistic Random Effects Model for Analysis of Clinical Trial Data

Karan P. Singh[1], Alfred A. Bartolucci[2] and Sejong Bae[1]
[1] *University of North Texas and* [2] *University of Alabama at Birmingham*

Abstract: A prospective, multi-institutional and randomized surgical trial involving 724 early stage melanoma patients was conducted to determine whether excision margins for intermediate-thickness melanomas (1.0 to 4.0 mm) could be safely reduced from the standard 4-cm radius. Patients with 1- to 4-mm-thick melanomas on the trunk or proximal extremities were randomly assigned to receive either a 2- or 4-cm surgical margin with or without immediate node dissection (i.e. immediate vs. later -within 6 months). The median follow-up time was 6 years. Recurrence rates did not correlate with surgical margins, even among stratified thickness groups. The hospital stay was shortened from 7.0 days for patients receiving 4-cm surgical margins to 5.2 days for those receiving 2-cm margins ($p = 0.0001$). This reduction was largely due to reduced need for skin grafting in the 2cm group. The overall conclusion was that the narrower margins significantly reduced the need for skin grafting and shortened the hospital stay. Due to the adequacy of subject follow up, recently a statistical focus was on what prognostics factors usually called covariates actually determined recurrence. As was anticipated, the thickness of the lesion ($p = 0.0091$) and whether or not the lesion was ulcerated ($p = 0.0079$), were determined to be significantly associated with recurrence events using the logistic regression model. This type of fixed effect analysis is rather a routine.

The authors have determined that a Bayesian consideration of the results would afford a more coherent interpretation of the effect of the model assuming a random effect of the covariates of thickness and ulceration. Thus, using a Markov Chain Monte Carlo method of parameter estimation with non informative priors, one is able to obtain the posterior estimates and credible regions of estimates of these effects as well as their interaction on recurrence outcome. Graphical displays of convergence history and posterior densities affirm the stability of the results. We demonstrate how the model performs under relevant clinical conditions. The conditions are all tested using a Bayesian statistical approach allowing for the robust testing of the model parameters under various recursive partitioning conditions of the covariates and hyper parameters which we introduce into the model. The convergence of the parameters to stable values are seen in trace plots which follow the convergence patterns This allows for precise estimation for determining clinical conditions under which the response pattern will change.

We give a numerical example of our results. The major platform for the theoretical development follows the Bayesian methodology and the multiple parameter logistic model with random effects having carefully chosen hyper parameters. We have done the basic infrastructure for the analysis using the commercially available WinBugs software employing the Markov Chain Monte Carlo (MCMC) methodology. The BUGS language allows a concise expression of the parametric model to denote stochastic (probabilistic) relationships and deterministic (logical) relationships.

Key words: Bayesian, covariates, logistic, melanoma, predictive, response.

1. Introduction

High speed computations with user friendly software have facilitated ease in the computations of complex models. This has been the case with Bayesian solutions to problems involving cumbersome analytic calculations in the prior to posterior framework which are often stalled at some point as closed form analytic solutions do not exist. Thus, numeric solutions are called for. This is especially the case in introducing several parameters with random effects having their own hierarchical modeling pattern as is seen in Gelman *et al.* (2004). One makes use of this capability in the present application to a random effects model in a clinical setting. We now describe the clinical setting for our approach. A prospective, multi-institutional, randomized surgical trial involving 724 early stage melanoma patients was conducted to determine whether excision margins for intermediate-thickness melanomas (1.0 to 4.0 mm) could be safely reduced from the standard 4-cm radius. Patients with 1- to 4-mm-thick melanomas on the trunk or proximal extremities were randomly assigned to receive either a 2- or 4-cm surgical margin with or without immediate node dissection (i.e. immediate vs. later-within 6 months). The motivation for this report is that this data is from a large multi-institutional cancer clinical trial in the late 1980's to determine the effectiveness of immediate vs. delayed node dissection in early stage melanoma. There were about 35 participating institutions from the Southeastern Cancer Study Group, a National Cancer Institute funded cohort of Oncologists. The trial did not show any statistical advantage of immediate note dissection in reducing the recurrence rate or maximizing the time to recurrence. Since 1993 patients have been followed for survival with no survival advantage on either therapy. Based on previous studies, a number of the oncology group wanted to confirm the role of ulceration and lesion thickness in influencing the response or appearance of recurrence within ten years time. Balch *et al.* (1981).

The median follow-up time was 6 years. Recurrence rates did not correlate

with surgical margins, even among stratified thickness groups. The hospital stay was shortened from 7.0 days for patients receiving 4-cm surgical margins to 5.2 days for those receiving 2-cm margins ($p = 0.0001$). This reduction was largely due to reduced need for skin grafting in the 2cm group. The overall conclusion was that the narrower margins significantly reduced the need for grafting and shortened the hospital stay. Due to the adequacy of subject follow up, recently a statistical focus was on what factors actually determined recurrence. As was anticipated, the thickness of the lesion ($p = 0.0091$) and whether or not the lesion was ulcerated ($p = 0.0079$), were determined to be significantly associated with recurrence events using the logistic regression model. This type of fixed effect analysis is rather a routine. Following the methodology of Bartolucci *et al.* (2005), the authors have determined that a Bayesian consideration of the results would afford a more coherent interpretation of the effect of the model assuming a random effect of the covariates of thickness and ulceration. Thus, using a Markov Chain Monte Carlo method of parameter estimation with non informative priors, one is able to obtain the posterior estimates and credible regions of estimates of these effects as well as their interaction on recurrence outcome. Graphical displays of convergence history and posterior densities affirm the stability of the results.

The stochastic parameters , however specified, may be given proper but minimally informative prior distributions, while the logical expression for the variance in the model allows the standard deviation (of the random effects distribution) to be estimated. Fixed effect model approaches are also handled rather well with the software. As seen in the WinBugs manual by Spiegelhalter *et al.* (2003) also at www.mrc.bsu.cam.ac.uk/bugs /winbugs/ manual14.pdf, the WinBUGS software uses compound documents, which comprise various different types of information (formatted text, tables, formulae, plots, graphs, etc.) displayed in a single window and stored in a single file for application to the problem at hand. This manual describes the WinBugs software an interactive Windows version of the BUGS program for Bayesian analysis of complex statistical models using MCMC techniques. We have been careful to apply only the models that WinBugs handles thus avoiding possibly spurious results with untested models as one is so warned.

2. The Model

In the past ulceration and thickness were considered as fixed effects in a survival model as seen in Balch *et al.* (1993). Now with an expanded database and the multiple logistic model and ulceration and thickness assumed to be random effects, we will show how these parameters fit into the model. We partition the data into a factorial arrangement of ulceration (yes, no) and thickness (< 2mm, ≥ 2mm). Within this total of four groups we have 19 samples and within each

sample we have ni individuals and r_i recurrences. Let p_i be the probability of failure or recurrence after surgery in the ith group, where $i = 1, \ldots, 19$, where each sample has a minimum of 8 subjects comprising a total sample size of 724. i.e., $\sum_i n_i = 724$ and $\sum r_i = 183$. The models for r_i and p_i can be written as

$$r \sim \text{Binomial}(p_i, n_i),$$

where

$$\begin{aligned} \text{logit}(p_i) &= \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_{12} X_{1i} X_{2i} + b_i \\ B_i &\sim \text{Normal}(0, \sigma^{-2}), i = 1, \ldots, 19. \end{aligned} \tag{2.1}$$

In equation (2.1), omitting the ith subscript for ease of notation we have $X_1 =$ thickness, $X_2 =$ ulceration and $X_1 X_2$ is the interaction term. The parameters, $\alpha_0, \alpha_1, \alpha_2$ and α_{12} (interaction term) are given non informative priors. They are:

$$\begin{aligned} \alpha_k &\sim \text{Normal}(0, 0.001), k = 0, 1, 2 \text{ and } 12, \quad \text{(subscript for the interaction term)} \\ \tau &= \sigma^{-2} \sim \text{Gamma}(0.001, 0.001). \end{aligned} \tag{2.2}$$

Our goal will be to apply this model with parameter estimates to our data and then to model the predicted recurrence based on various robust thickness and ulceration considerations. Sensitivity testing was applied to evaluate the robustness of these priors especially for the shape of the variance parameter of the Gamma from 0.001 to 0.01, and alpha's mean from 0 to 0.01 but is not presented here.

3. The Data

As mentioned above, we have a sample from 724 melanoma patients upon which to build our model. The primary endpoint was recurrence with random effects for thickness and ulceration each subject. The objectives of the study were (1) to examinee the efficacy of the treatment and to build a predictive model based on the covariates of thickness and ulceration. The focus is on the prediction question. Previously untreated patients with diagnoses of early stage melanoma were eligible. The diagnosis was confirmed by pathology. Patients with clinically localized melanoma (i.e. Stages I and II according to the American Joint Committee on Cancer Staging Criteria) and intermediate thickness melanomas located on the trunk or proximal extremity (i.e. proximal to the elbow or knee) were randomly assigned to receive either a 2 or 4cm margin excision. Each subject was also randomized to receive and immediate or delayed (within six months) elective node dissection. The primary endpoints were survival and recurrence. Treatment did not impact either and thus the predictive model came into play

to determine which factors affected survival or recurrence. This work involves recurrence events only.

4. Building the Model

Based on our data set described above, we now detail application of the WinBugs software to obtaining posterior estimates for the parameters of equation (2.1) with standard graphics and summary statistics. Two chains of initial values were incorporated into the data to attempt the conversion to the five estimates in our model. The names in parentheses are the names of the variables in the WinBugs program for ease of interpretation and differentiation from each other when we examine the output and present graphical results. For the first chain we had α_0, α_1 and α_{12} (alpha0, alpha1 and alpha12) all prior means are set to 0 and then are set to 0.01 for the second chain. τ or σ^{-2}, the shape hyper parameter was 0.001 in the first chain and 0.01 for the second chain. The trace plots which map the conversion through the iterations give the pattern for both sets of starting values for each chain. The trace (not shown here) for both chains converged quickly but was carried out to about 15,000 iterations.

It appeared that as this iterating process was carried beyond 15,000 that the estimates proved to be very stable in that they changed at most in the third decimal place. We thus present the results here at 15,000 and note below that at the testing of our model the prediction is very reliable. Below we present the posterior densities of the parameters as well as list the results of the sensitivity testing of the robustness of the procedure. The posterior density of the intercept parameter α_0 is seen here.

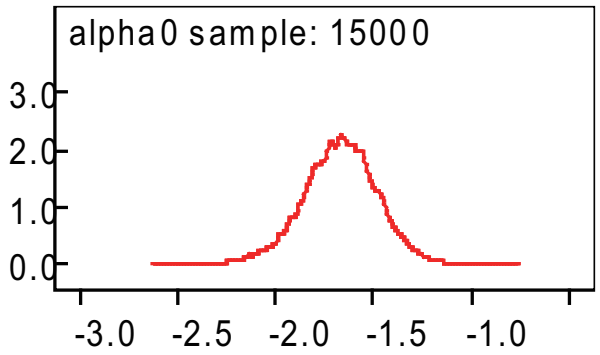

Figure 1: Posterior density plot for α_0.

Note that for the next, parameters, i.e, the main effects α_1 and α_2 for the thickness and ulceration coefficients we also have symmetry in the posterior densities.

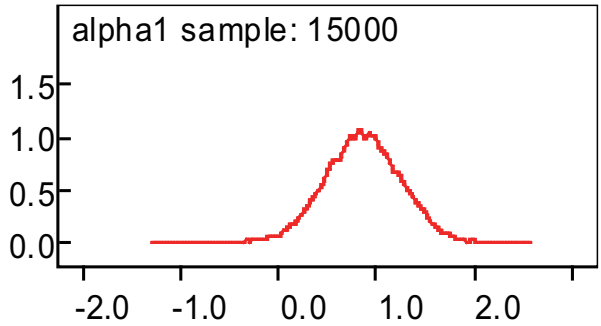

Figure 2: Posterior densities for α_1 and α_2.

Note the positive modes for α_1 and α_2 indicating that as one increases thickness and ulceration then the chance of recurrence will increase. The posterior variance, σ^2 (sigma), remained stable and decreased slightly as the prior shape hyper parameter was increased from 0.001 to 0.01 indicating the stability of the posterior prediction. The posterior gamma density of the variance is seen in the next figure.

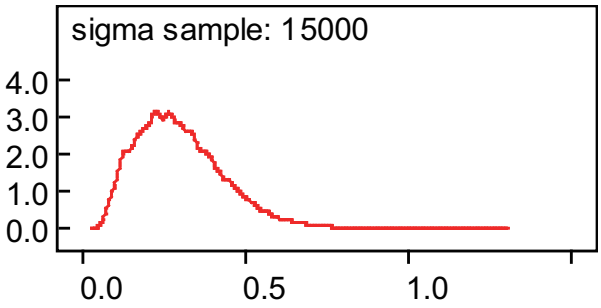

Figure 3: Poster density of the variance parameter.

One can see from Figures 1 to 3 the symmetry and skewness in the posterior densities. The actual posterior mean parameter values with their posterior credible limits (CL) are seen in Table 1.

Table 1: Parameters, posterior means and 95% posterior credible regions

Parameter	Posterior Mean	95% CL
α_0	-1.671	-2.081, -1.299
α_1	0.8613	0.070, 1.636
α_2	0.9112	0.383, 1.449
α_{12}	-0.2927	-1.267, 0.694
σ	0.2483	0.037, 0.588

Note in Table 1 that these values pretty much approximate the modes of the parameter densities in Figures 1 to 3. Also note the width of the confidence limits for all the parameters indicating fairly stable variation in the random components in this investigation. Of note also is the fact that the interaction term has a very plausible posterior value of zero indicating little or no interaction of thickness and ulceration.

5. The Predictive Model

The task now is to take the parameter estimates from Table 1 and insert them into the model of equation (2.1) to determine the prediction at various values of thickness and ulceration. This is seen in the next Tables 2 and 3.

Table 2: Percent of recurrences for thickness category

Recurrence	Thickness < 2mm	Thickness ≥ 2mm
Yes	35.6	64.4
No	62.3	37.7

One thus sees from Table 2 the consistency of the association between the thickness category and the likelihood of recurrence. Those with the thicker lesions (≥ 2mm) were most predictive of recurrence while those with the thinner lesions (< 2mm) were less likely to recur. The next table, Table 3, is the same information for those with and without ulcerated lesions.

Table 3: Percent of recurrences for ulceration category

Ulcerate-recurrence	No	Yes
Yes	35.6	64.4
No	62.3	37.7

One thus sees from Table 3 the consistency of the association between the ulceration category and the likelihood of recurrence. Those with the ulcerated lesions (Yes) were most predictive of recurrence while those without the ulcerated lesions (No) were less likely to recur.

6. Conclusions

We have attempted to show that one can assume an underlying random effects model with a parametric distribution such as the logistic distribution and apply this methodology in a clinical or biological setting. We really set out to do this with the added caveat that assuming vague prior information one can then further

extend the methodologic application to the Bayesian framework. Thus there is a lot to consider when attempting this approach. The data comes from an actual database as described above. The next step was to assume a reasonable model for the data. The logistic model is a natural for response or categorical data. Here we had only two categories.

However, a multinomial response would also apply such a categorical model. The parametric estimation using the random effects was enhanced by the available software, WinBugs, which is specific to Bayesian applications and easy to handle in the random effects environment since random effects automatically assumes some underlying probability distribution. Thus this fits naturally into the Bayesian mindset. Also one is not overly committed to assigning subjective priors which some may consider as unrealistic as we were cautious to place rather largely dispersed vague priors on the parameters of interest. The WinBugs software allows one to break away from the temptation of assigning just normal models to parameters of interest as the use of other distributions such as the Gamma in our case can be easily applied as well. 0ne also has the flexibility of simulating results as one is likely to do in a numerical environment.

Being provided with visuals of the variance pattern and the underlying density structure of the parameters of interests allows one to logically determine if the convergence is following a logical pattern and not deviating wildly as one goes through the iterations for variations of hyper parameter input. Having to conduct this exercise for different initial values or chains in our case is yet a further enhancement of the tools available and another check on the consistency of the functional patterns of the variables over the iterative domain. See Meeker *et al.* (1998). We note our results were consistent with the science in that when accounting for the randomness of the covariates is most explanatory of the prognosis in this group of subjects. After having done all this analysis one wants to be assured that outcomes are consistent with common knowledge of the discipline one is involved in.

A word on the ease of the use of this software may be in order. One has to know the underlying model one would attempt in the analysis, provide the data and then the initial values for the parameters of interest. That is all that is required. However, not to be lulled into a false sense of security, it is wise with larger data sets to take a random sample of the observations as a split test sample and check for the consistency of that result with the remainder of the data set. Jackknifing and bootstrapping samples are also suggested as well. We in fact took a subset of our data here to check for its consistency with not only the previous analysis in 1993 but with our current analysis as well. Also one should do a sensitivity analysis for logical ranges of the hyper parameters in the Bayesian model to check for the robustness of the results. It was done thoroughly

here as we varied the prior hyper parameters on all the normal priors as well as the gamma prior. Space is limited here to go into all the options we attempted. The bottom line is that our results were in fact robust and clinically consistent. We hope that the reader(s) find these last few words of recommendation helpful and we look forward to applying more complex Bayesian models to past and future biological and environmental problems of interest.

References

Balch,C. , Urist, M. M. , Karakousis, C. P. , Smith,T. J. , Temple, W. J. , Drzewiecki, K. , Jewell, W. R. , Bartolucci, A. A. *et al.* (1993). Efficacy of 2-cm surgical margins for intermediate-thickness melanomas (1-4mm): results of a multi-institutional randomized surgical trial. *Annals of Surgery* **218**, 262-269.

Balch, C. M. , Soong, S. J. , Murad, T. M. , Ingalls, A. L. , Maddox, W. A. (1981). A multifactorial analysis of melanoma III: Prognostic factors in melanoma patients with lymph node metastases (Stage II). *Annals of Surgery* **193 3**, 377-388.

Bartolucci, A. A. , Bae, S. J. and Singh, K. P. (2005). Establishing a Predictive Survival ModelIn *Modeling and Simulation Society of Australia and New Zealand* (Edited by A. Zeger and R. M. Argent, 380-385.)

Gelman, A. ,Carlin, J. B. , Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapter 11, 117-156, Chapman and Hall.

Meeker, W. Q. and Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. Wiley.

Spiegelhalter, D. , Thomas, A. , Best,N. and Lunn, D. (2003). *WinBugs User Manual.* Version 1. 4,. Cambridge, UK. http://www.mrc-bsu.cam.ac.uk/bugs.

Received September 3, 2008; accepted March 2, 2009.

Karan P. Singh
Department of Biostatistics
University of North Texas Health Science Center
3400 Camp Bowie Blvd.
Fort Worth, Texas 76107, USA
ksingh@hsc.unt.edu

Alfred A. Bartolucci
Department of Biostatistics
University of Alabama at Birmingham
Birmingham, Alabama 35294 - 0002, USA
Abartolucci@ms.soph.uab.edu

Sejong Bae
Department of Biostatistics
University of North Texas Health Science Center
3400 Camp Bowie Blvd., Fort Worth, Texas 76107, USA
sbae@hsc.unt.edu