Spatial and Spatio-temporal Epidemiology 1 (2010) 169–176

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste

A Bayesian hierarchical model of nontraumatic lower-extremity amputation rates ☆

Xiaoyi Min [a], Dongchu Sun [a,*], Zhuoqiong He [a], Mario Schootman [b]

[a] University of Missouri-Columbia, USA
[b] Washington University – St. Louis, USA

ARTICLE INFO

Keywords:
Linear model
Lower-extremity amputations
Diabetes patient

ABSTRACT

A Bayesian hierarchical generalized linear model is used to estimate the risk of lower-extremity amputations (LEA) among diabetes patients from different counties in the state of Missouri. The model includes fixed age effects, fixed gender effect, random geographic effects, and spatial correlations between neighboring counties. The computation is done by Gibbs sampling using OPENBUGS. DIC (Deviance Information Criterion) is used as a criterion of goodness of fit to examine age effects, gender effect, and spatial correlations among counties in the risks of having LEAs. The Bayesian estimates are also shown to be quite robust in terms of choices of hyper-parameters.

1. Introduction

The nontraumatic lower-extremity amputation (LEA) is a devastating complication of diabetes. According to *a report of the National Diabetes Advisory Board* (1980), as much as 15% of persons with diabetes will have such amputations in their life time. About 50% of all LEAs are performed in persons with diabetes. Also see Most and Sinnock (1983). People with diabetes are 10–20 times more likely to have LEAs than those without diabetes. From Sugerman et al. (1998), people age 65 and older account for about 55% of patients with diabetes who had nontraumatic LEAs. Considering the geographical variation of diabetes-related LEAs, Wrobel et al. (2001) showed that the incidence of LEAs varied ninefold among 306 large hospital regions in the US. However, the variation of LEAs among smaller geographic areas, such as counties, has not been studied much yet, and it is also a major interest of this paper. Specifically, the LEAs data for 100,280 diabetes patients age over 65 from the state of Missouri are analyzed with regard to their gender, age, and the counties of beneficiary.

The largest difficulty that traditional methods have when solving such small-area estimation problems is the relatively small sample sizes for subareas such as counties compared to the large overall sample size. First, some units may have no observations, which makes it impossible to give estimates. For example, in the dataset being studied, there are no male patients aged 90 or more for 3 of 115 counties. Moreover, when subarea sample sizes are small, rates acquired by simply dividing can be non-informative or misleading as the variability will be large, and it is difficult to distinguish the chance variation from the true difference. For example, the average rate of LEA is about 1.5% among all 100,280 diabetes patients during the study period for the entire state while for male patients aged 90 or above from county 9, the rate is 100% since there is only one patient in this group and he had LEA. Meanwhile, pooling of neighboring units often masks important real differences.

Advances in Bayesian hierarchical modeling have made it possible to obtain stable estimates for such small-area problems by using information from all of the areas to obtain estimates for individual areas. Rapidly developing computational tools such as Gibbs sampling procedure by Gelfand and Smith (1990) and other Markov chain Monte

☆ This research was supported in part by NSFGrant SES-0720229 and NIH Grants R21DK067172 and R01-MH071418.
* Corresponding author. Address: Department of Statistics, University of Missouri-Columbia, 134 Middlebush Hall, Columbia, MO, USA.
E-mail address: sund@missouri.edu (D. Sun).

1877-5845/$ - see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.sste.2010.03.008

Carlo (MCMC) methods by Tanner (1993) have greatly promoted the development of this topic. Literatures using this technique have been growing rapidly. Applications to epidemiology and disease mapping can be found in Lawson et al. (1999), Elliott et al. (2000), and Lawson and Williams (2001).

In this paper, we propose several Bayesian hierarchical models for LEA risk. In Section 2, a logistic linear mixed model including fixed age effects, a fixed gender effect, and random spatial effects is proposed as the first-stage prior. Conditional autoregressive prior is used for the random spatial effects. Priors for fixed effects and variance components are also specified.

In Section 3, the computation using Gibbs sampling procedure is implemented. Numerical results for posterior densities are given. The Bayesian estimates are shown to be quite robust in terms of choices of hyper-parameters. The convergence issue is also discussed. In Section 4, maps of estimates of risk for LEA are shown, and the practical meanings of which are discussed. Several alternative models are examined using Deviance Information Criterion (DIC) by Spiegelhalter et al. (2002) to evaluate the significance of parameters in the original model. Some comments are given in Section 5.

2. Bayesian hierarchical model

2.1. Logistic linear mixed model

We first divide the patients into groups according to their age, gender, and county of beneficiary. Let n_{ijk} represent the total number of patients from county i, age group j, and gender k; y_{ijk} be the total number of patients who had LEA and p_{ijk} be the probability of a patient having LEA from such a group, where $i = 1, 2, \ldots, I = 115$ are for the 115 counties in Missouri (including St. Louis City); $j = 1$ when age is between 65 and 70, $j = 2$ when age is between 70 and 80, $j = 3$ when age is between 80 and 90, and $j = 4$ when age is above 90; $k = 1$ stands for male and $k = 2$ stands for female. We first assume that

$$y_{ijk} \sim Binomial(n_{ijk}, p_{ijk}). \tag{1}$$

Alternatively, the model (1) can be replaced with a Poisson distribution with the mean $n_{ijk}p_{ijk}$. It turns out that the estimates for p_{ijk} are quite robust in terms changing the likelihood function. Following Ghosh et al. (1998) and Sun et al. (2000). The first level of hierarchical prior is

$$log\left(\frac{p_{ijk}}{1-p_{ijk}}\right) = \alpha_j + \beta_k + Z_i + \epsilon_{ijk}, \tag{2}$$

where α_j stands for the effect of age group j, β_k is the effect of gender k, Z_i is the spatial effect of county i, and ϵ_{ijk} accounts for extra variation not explained by the additive model with mean 0 and variance δ_e. In particularly, we assume that

$$\epsilon_{ijk} \overset{iid}{\sim} N(0, \delta_e). \tag{3}$$

Notice that β_1 is set to zero so that the parameters are identifiable.

2.2. Distribution of Z_i

The modified conditional autoregressive model is used to capture the spatial correlations among $\boldsymbol{Z} = (Z_1, \ldots, Z_I)'$,

$$\boldsymbol{Z} \sim N\left(\boldsymbol{0}, \delta_z(\boldsymbol{D} - \rho\boldsymbol{C})^{-1}\right), \tag{4}$$

where $\boldsymbol{C}$ is the adjacency matrix defined by: $C_{kl} = 1$ if counties k and l share a common boundary, including $C_{kk} = 0$. Here $\boldsymbol{D} = diag(d_1, d_2, \ldots, d_{115})$, and $d_i = \sum_{k \neq i} C_{ik}$ is the number of neighbors of county i. Note that ρ is a parameter of correlation. If $\rho = 0$, the Z_i's are independent and there is no spatial correlation among county effects. The density of $\boldsymbol{Z}$ exists if ρ is between -1 and 1. The use of this prior can be dated to Besag (1974) and became popular since Besag et al. (1991) and Cressie (1991). It was applied to model the spatial effects for lung cancer mortality in Sun et al. (2000).

2.3. Other priors

To complete the Bayesian analysis, we need the priors for $(\alpha_j, \beta_2, \rho, \delta_e, \delta_z)$. We assume that

$$\alpha_j \sim N(\mu_j, \delta_j), \quad j = 1, \ldots, 4; \tag{5}$$

$$\beta_2 \sim N(\mu_5, \delta_5); \tag{6}$$

$$\rho \sim Uniform(-1, 1). \tag{7}$$

We also assume inverse gamma priors for δ_e and δ_z:

$$\delta_e \sim Inv - Gamma(a_1, b_1), \tag{8}$$

$$\delta_z \sim Inv - Gamma(a_2, b_2). \tag{9}$$

In the computation, α_j and β_2 are given a flat prior, $N(0, 10^6)$ (with mean $\mu_i = 0$ and variance $\delta_i = 10^6$ for $i = 1, \ldots, 5$). For δ_e and δ_z, a set of hyper-parameters is used:

$$(a_1, b_1) = (2.90, 0.39) \quad \text{and} \quad (a_2, b_2) = (2.42, 0.65). \tag{10}$$

Here the mean and standard deviation of δ_e are 0.20 and 0.22, and the mean and standard deviation of δ_z are 0.46 and 0.71. These hyper-parameters were chosen based on the method given in Sun et al. (2000). That is, it was based on some vague priors. The comparison of the prior and posterior densities will be given in Section 3.1. We will see that the prior variance is much bigger then the posterior variance. We will also see the Bayesian estimators are quite robust in terms of changing the hyper-parameters.

3. Computation via MCMC using OPENBUGS

3.1. Numerical results

OPENBUGS is used to run Gibbs sampling to calculate the posterior summaries and densities. Three groups of initial values are used and we burn in the first 20,000 samples so that the parameters become stationary, and we take every 5th sample from the next 50,000 iterations, which gives a total of 30,000 samples for every parameter.

Age effects α_j: The summaries of posterior quantities of α_j are shown in Table 1, and the posterior densities of α_j are compared in Fig. 1(a). From the plot we can see that the age effects α_j are decreasing with j (age) for groups