

## SPATIAL MODEL SELECTION USING BAYES FACTOR AND RATIO OF VARIABILITIES FOR ASTHMA MORTALITY DATA

JUNG-YING TZENG<sup>1</sup>, CHUHSING KATE HSIAO<sup>2</sup>,  
CHIEN-JEN CHEN<sup>2</sup>

*Hierarchical models are commonly used in analyzing geographical data. They take account of the random variation in addition to the systematic variability among observations. Through specifying a distribution for rates at different areas, various kinds of random mechanism for variability can be considered. The exchangeable (EX) priors and conditional autoregressive (CAR) priors are the two most common approaches. However, it is unclear about how to choose between these two mechanisms. In this study, motivated by looking for the true pattern of the asthma mortality data for Taipei City, we adopt the two competing EX and CAR models to investigate the spatial pattern. With the two hypotheses (the EX or CAR model), we not only need to obtain estimates of quantities of interest but also need to choose an appropriate model since the final decision may result in different etiologic studies. In this paper, we use the fully Bayesian approach with the Monte Carlo Markov Chain to obtain estimates. Then, we focus on two model selection indices—the Bayes factor and the ratio of the variances (the local effect to the global effect) for the asthma study. Based on the study results, we conclude: (1) Both the Bayes factor and the ratio of the local variance to the global variance should be used together for choosing an appropriate model. The Bayes factor offers a direct answer for which model is favored by the data, while the ratio of variances reflects the characteristic of the data and provides a way to evaluate whether it is necessary to consider the area-specific effect. (2) According to the two indices, the EX model is considered more appropriate for the asthma mortality data, and the rates at Neihu and Nankang are higher than other areas. The remaining variation among areas for the EX model may be caused by some spatial-independent variables rather than spatial-correlated variables. (Chin J Public Health. (Taipei): 1998; 17(2): 158-169)*

**Key words:** Bayes factor, CAR model, EX model, spatial modelselection.

<sup>1</sup> Institute of Statistical Science, Academia Sinica.

<sup>2</sup> Graduate Institute of Epidemiology, College of Public Health, National Taiwan University.

Corresponding author: Chuhsing Kate Hsiao.

Division of Biostatistics, Graduate Institute of Epidemiology, College of Public Health, National Taiwan University.

Address: 1 Jen-Ai Road, Section 1, Room 1542, Taipei, Taiwan 10018, R.O.C.

Phone: (02)397-0800#8372

Fax: (02)351-1955

E-Mail: ckhsiao@ha.mc.ntu.edu.tw.

### INTRODUCTION

Spatial data is frequently encountered in many scientific fields, such as agriculture, mining engineering, meteorology, epidemiology and environmental health. According to Cressie's classification [1], the scientific interests of modeling are often focused on three categories: (1) for geostatistical data, (2) for lattice data,

and (3) for spatial pattern and spatial processes. In this article we focus on the mortality rates, which are one type of the lattice data. These data, also known as regional data or geographical data, are often encountered in the environmental health and epidemiological studies.

Through analyzing the geographical data, it helps the scientists such as epidemiologists to investigate the spatial pattern of diseases, look for risk factors, or explore the associations between diseases and potential risk factors. To reach the goal, researchers may like to construct maps showing the magnitudes of rates at each area under study. However, the crude rates may not be compatible between various areas if the variability among areas vary greatly. The variabilities could come from two sources, the systematic and the stochastic [2]. The systematic variability is contributed from the effect of covariates, such as the exposure status, age, and sex. The stochastic variability is the random variation of the data. For the geographical data, the counts of events are often assumed distributed as the Poisson distribution. In other words, the probability of having a disease for each individual within each area is assumed the same from a Poisson distribution. However, both the variability between these individuals within each area and the variability among various areas are not considered. In that case, the observed variation may be greater than the expected under the Poisson assumption, which is called extra-Poisson variation.

Standardized estimation methods (such as the standardized mortality rates, SMR) are commonly used for analyzing geographical data. However, these methods can only deal with the systematic variation due to covariates. For

instance, the standardized approach, though adjusting demographic variables measured in individuals (such as age, sex, and socioeconomic status), does not deal with the random variation related to population nor neighboring effect, and thus the adjusted map of rates may remain unstable. The Poisson regression, although considers the Poisson random variation for each area and controls the group covariates (such as pollution, urbanization and climate), may not explain enough the phenomenon of extra-Poisson variation. Moreover, there may exist spatial correlation and thus the general regression methods are not appropriate.

Alternative modeling techniques have been proposed to solve these problems. One of them was the hierarchical spatial model[3,4]. We will consider this model in the paper as well. Particularly, we consider the conditional autoregressive (CAR) model and the Poisson-Gamma model, called the exchangeable (EX) model hereafter. The former considers the spatial correlation among areas, while the latter assumes rates at all areas are independent. These models contain: (1) the observed data which can be used to model the systematic heterogeneity and stochastic variability, and (2) a prior distribution which takes account of the researchers' prior knowledge for the events.

In practice, it may be difficult to decide which model is better in explaining the data. Researchers generally rely on their prior belief to choose an appropriate model. Nevertheless, in most cases we may know very little about the underlying mechanism for the study events and anticipate to gain more knowledge, through modeling, for the etiology of the disease. Therefore, criteria for selecting models may be necessary in spatial data analysis. Classical model selection index which

may be useful for nested models is not so helpful in this situation. Here, we will use the Bayes factor to evaluate the two non-nested competing models. In addition, by comparing the magnitudes of the local and global variations, we can learn more about how to choose a better model.

## STATISTICAL METHODS

### Modeling Spatial Data

Two types of hierarchical models are applied here. The first model, used by Besag and Mollie [5], considers the spatial dependence among areas. The second model, used by Clayton and Kaldor [4], assumes areas are independently distributed and therefore the random variation among areas are exchangeable.

#### 1. Conditional Autoregressive (CAR) Model (1)Model Specification

Let  $y_i, i=1,2,\dots,N$  denote the number of deaths due to asthma in the district  $i$ ,  $m_i$  denote the corresponding population size, and  $p_i$  be the true unknown mortality rate needed to be estimated. Conditioning on the rate  $p_i$  and population size  $m_i$ ,  $y_i$  is assumed to follow a Poisson distribution with mean  $m_i p_i$ :

$$y_i | m_i, p_i \sim \text{Poisson}(m_i p_i).$$

The rate  $p_i$  is next decomposed into the global effect  $\mu$  and an area-specific value (local effect)  $\theta_i$ :

$$\log(10^4 p_i) = \mu + \theta_i, \dots \quad (1)$$

The common value  $\mu$  for each area follows a Normal ( $c, 1/\sigma$ ) distribution, with  $\sigma$  representing the precision of  $\mu$ . For the area-specific effect  $\theta_i$ , it follows a conditional autoregressive (CAR) distribution. That is, the spatial autocorrelation is assumed to exist

among areas.

The equation (1) is an "intermediate" stage between the independent structure (the prior of the global effect  $\mu$ ) and the dependent structure (the prior of the local effect  $\theta_i$ ). It accounts for the heterogeneity beyond the global effect through the first-order neighboring spatial structure. Defining  $w_{ij}$  as an index for neighboring relationship of area  $i$  and area  $j$ :

$$\begin{aligned} w_{ii} &= 0, \\ w_{ij} &= 1, \text{ if area } i \text{ and area } j \text{ are contiguous,} \\ w_{ij} &= 0, \text{ otherwise,} \end{aligned}$$

and introducing a hyperparameter  $\lambda$  as a precision parameter for  $\theta_i$ , the conditional autoregressive prior is:

$$\theta_i | \theta_{j \neq i}, \lambda \sim \text{Normal}(\bar{\theta}_i, \frac{1}{\lambda_{n_i}})$$

where  $n_i = \sum_{j=1}^N w_{ij}$  represents the number of neighbors for the area  $i$  and  $\bar{\theta}_i = \sum_{j=1}^N w_{ij} \theta_j / n_i$  is the mean of the area-specific effect given neighbors of area  $i$ . In other words, the mean of  $\theta_i$  at area  $i$  is affected by all its neighbors and the variance is affected by the number of its neighbors. More the neighbors it has, more information we have about  $\theta_i$  and less variation for  $\theta_i$ . The neighbor information is included in the CAR model through  $\bar{\theta}_i$ . The constraint  $\sum_{i=1}^N \theta_i = 0$  is used further to ensure identifiability.

To express the joint distribution in the multivariate form, the CAR prior can also be written as [6,7,8]:

$$\begin{aligned} \theta | \lambda &\propto \left( \prod_{i=1}^N \sqrt{n_i} \right) \lambda^{\frac{N}{2}} \exp \left\{ -\frac{\lambda}{2} \sum_{i=1}^N n_i (\theta_i - \bar{\theta}_i)^2 \right\} \\ &\propto \left( \prod_{i=1}^N \sqrt{n_i} \right) \lambda^{\frac{N}{2}} \exp \left\{ -\frac{\lambda}{2} \theta' \Sigma \theta \right\} \end{aligned}$$

where  $\Sigma = \text{diag}(n_1 \lambda, n_2 \lambda, \dots, n_N \lambda) (I - A)$  and  $A =$

( $a_{ij}$ ) with  $a_{ij}=w_{ij}/n_i$ .

Other prior distributions needed to be specified for the mean  $c$  of the global effect, the precision  $\sigma$  of the global effect, and the precision  $\lambda$  of the local effect are:

$$\begin{aligned}c &\sim \text{Uniform}(u,v), \\ \sigma &\sim \text{Gamma}(a,b), \\ \text{and } \lambda &\sim \text{Gamma}(r,d)\end{aligned}$$

with  $a$ ,  $b$ ,  $r$ , and  $d$  fixed.

## (2) Estimation and Computation

We use the fully Bayesian method to estimate the parameters of interest  $\mu$ ,  $\theta$ ,  $c$ ,  $\sigma$  and  $\lambda$ . Inferences for parameters are based on their marginal posterior distributions:  $f(\mu | y)$ ,  $f(\theta | y)$ ,  $f(c | y)$ ,  $f(\sigma | y)$ , and  $f(\lambda | y)$ . However, obtaining these posteriors requires multiple integration which are not analytically feasible. We will use the Gibbs sampling technique to obtain samples from the marginal posteriors. Then, to derive the estimations of  $\mu$ ,  $\theta$ ,  $c$ ,  $\sigma$  and  $\lambda$  by calculating the sample mean, median and mode of these posterior samples.

In the case of CAR model, we use the BUGS program [9,10,11] to perform the Gibbs sampling. By generating from the fully conditional distributions,  $f(\mu | y, \theta, c, \sigma, \lambda)$ ,  $f(\theta_i | y, \mu, \theta_{jj \neq i}, c, \sigma, \lambda)$ ,  $i=1,2,\dots,N$ ,  $f(c | y, \mu, \theta, \sigma, \lambda)$ ,  $f(\sigma | y, \mu, \theta, c, \lambda)$ , and  $f(\lambda | y, \mu, \theta, c, \sigma)$ , the Gibbs sampler produces posterior samples from  $f(\mu | y)$ ,  $f(\theta_i | y)$ ,  $i=1,2,\dots,N$ ,  $f(c | y)$ ,  $f(\sigma | y)$ , and  $f(\lambda | y)$ . All the fully conditional distributions can be derived from the joint distribution of  $(y, \mu, \theta, \sigma, \lambda)$  [12].

In addition to specifying these full conditional distributions, we also need to provide the initial values ( $\mu_0$ ,  $\theta_0$ ,  $\sigma_0$ ,  $\lambda_0$ ) for all parameters to start the Gibbs sampling. In principle, the starting values should have little in-

fluence on the results since the first several iterations will be dropped as the "burn in" process and the remaining Gibbs sequence would be irrelevant to the starting values [13]. When using Gibbs sampling, the Gibbs output must be monitored to decide the length of the burn-in and the total run length.

## 2. Exchangeable (EX) Model – Poisson-Gamma Model

### (1) Model Specification

Notations for  $y_i$ ,  $m_i$ , and  $p_i$  have the same meaning as in the CAR model. Given  $m_i$  and  $p_i$ , the number of deaths  $y_i$  is assumed to follow a Poisson distribution with mean  $\tau_i = m_i p_i$ :

$$y_i | m_i, p_i \sim \text{Poisson}(m_i p_i).$$

Alternatively, it can be written as:

$$y_i | \tau_i \sim \text{Poisson}(\tau_i).$$

To account for the variability of  $\tau_i$ , a Gamma distribution is specified for the prior distribution of  $\tau_i$ . It is assumed that  $\tau_i$ 's are identically and independently distributed conditioning on  $\alpha$  and  $\beta$ :

$$\tau_i | \alpha, \beta \sim \text{Gamma}(\alpha, \beta),$$

where  $\alpha$  and  $\beta$  are assumed the exponential distribution and Gamma distribution respectively,

$$\begin{aligned}\alpha &\sim \text{Exp}(a) \\ \beta &\sim \text{Gamma}(s, t)\end{aligned}$$

where  $a$ ,  $s$ , and  $t$  are given constants. In this Poisson-Gamma model, the variance beyond the Poisson variability is explained by the Gamma prior. This Gamma distribution assumes independent and exchangeable random variation among areas and thus considers no particular geographical characteristics. Therefore, such a type of model is called "ex-

changeable" (EX) model.

## (2) Estimation and Computation

To estimate through the fully Bayesian approach, we need to construct the posterior distributions  $f(\alpha | y)$ ,  $f(\beta | y)$ , and  $f(\tau | y)$ , where  $\tau = (\tau_1, \tau_2, \dots, \tau_N)$ . The Gibbs sampler technique is used again to get these posteriors. Through the BUGS software, we obtain the marginal posterior samples of  $f(\tau_i | y)$ ,  $i=1, 2, \dots, N$ ,  $f(\alpha | y)$ , and  $f(\beta | y)$  by generating from the fully conditional distributions:  $f(\tau_i | y, \tau_{j \neq i}, \alpha, \beta)$ ,  $i=1, 2, \dots, N$ ,  $f(\alpha | y, \tau, \beta)$ , and  $f(\beta | y, \tau, \alpha)$ . Again, these fully conditional distributions can be derived directly from the joint distribution of  $(y, \tau, \alpha, \beta)$ .

As mentioned above for the CAR model, here we also have to specify the starting values  $\tau_0$ ,  $\alpha_0$ , and  $\beta_0$  to start the Gibbs sampler. A "burn-in" process and long iterations may be necessary to obtain a set of stable samples from the posterior distributions.

## Model Selection

Choosing either the EX or CAR models may result in fairly different follow-up etiological study of the target diseases. If we can decide which model is the proper one in explaining the random mechanism, it helps to understand better the sources of heterogeneity. We therefore use two indices to evaluate the performance of the two competing spatial models. The two indices are Bayes factor (BF) and ratio of two area variations for compar-

ing the CAR and EX models.

### 1. Bayes Factor

The Bayes factor, proposed by Jeffereys [14], summarizes the evidence provided by the data for one model against the other, and offers a way to compare competing models (nested or not). Suppose there are two candidate models  $M_1$  and  $M_2$ , and they have parameters  $\eta_1$  and  $\eta_2$ , respectively. The Bayes factor, denoted by BF, is defined [14] as the ratio of the posterior odds to prior odds:

$$\begin{aligned} BF &= \frac{P(M_1|y) / P(M_2|y)}{P(M_1) / P(M_2)} \\ &= \frac{\left[ \frac{P(y|M_1)P(M_1)}{P(y)} \right] / \left[ \frac{P(y|M_2)P(M_2)}{P(y)} \right]}{P(M_1) / P(M_2)} \\ &= \frac{P(y|M_1)}{P(y|M_2)} \dots \dots \dots (2) \end{aligned}$$

In the above equation (2), the Bayes factor can be written as the ratio as two marginal probabilities of  $y$  under model  $M_1$  and  $M_2$ , respectively. The marginal probability of  $y$  represents the probability of the observed data under the model  $M_i$ . The Bayes factor can be evaluated on the natural log scale and different values indicate different strengths of evidence, as listed in Table 1.

To obtain the Bayes factor, one needs to calculate the marginal probabilities (here the index is dropped for simplicity):

**Table 1. Interpretation of the Bayes factor  $BF_{12}$  on log<sub>e</sub> scale**

| $\log_e(BF_{12})$ | $BF_{12}$ | Evidence for $M_1$         |
|-------------------|-----------|----------------------------|
| < 0               | < 1       | negative (supports $M_2$ ) |
| 0 to 1            | 1 to 3    | barely worth mentioning    |
| 1 to 3            | 3 to 20   | positive                   |
| 3 to 5            | 20 to 150 | strong                     |
| > 5               | > 150     | very strong                |

$$P(\mathbf{y} | M) = \int P(\mathbf{y} | \eta, M) P(\eta | M) d\eta.$$

In most situations, it is not easy to compute the integral analytically. The simple Monte Carlo method can approximate  $P(\mathbf{y} | M)$  when  $P(\eta | M)$  is a density:

$$\begin{aligned} P(\mathbf{y} | M) &= \int P(\mathbf{y} | \eta, M) P(\eta | M) d\eta \\ &= E\eta[P(\mathbf{y} | \eta, M)] \\ &\approx \frac{1}{K} \sum_{k=1}^K P(\mathbf{y} | \eta^{(k)}, M) \end{aligned}$$

where  $\eta^{(1)}, \eta^{(2)}, \dots, \eta^{(K)}$  are samples from  $P(\eta | M)$ . That is, the marginal probability of the data  $\mathbf{y}$  is approximated by the average of the likelihoods evaluated at the samples generated from the prior  $P(\eta | M)$ . However, since our Gibbs outputs contain the samples from the posterior rather than prior distributions, the importance sampling method was used. In particular, the posterior distribution is treated as the importance-sampling function, and the harmonic mean of the corresponding likelihood values is computed to estimate the marginal probability under each model,  $P(\mathbf{y} | M)$  [15,16],

$$\hat{P}(\mathbf{y} | M) = \frac{1}{\frac{1}{K} \sum_{k=1}^K \frac{1}{P(\mathbf{y} | \eta^{(k)}, M)}}$$

## 2. Ratio of the Local to Global Variances

In the CAR model, the rate  $p_i$  is decomposed into two independent parts: the global effect ( $\mu$ ) and the local neighboring effect ( $\theta_i$ ). For each area, the variance of  $p_i$  can be written as

$$\begin{aligned} \text{Var}[\log(10^4 p_i)] &= \text{Var}(\mu) + \text{Var}(\theta_i) \\ &= \frac{1}{\sigma^2} + \frac{1}{n_i \lambda^2}, \end{aligned}$$

in which one part is the global variance and

the other is the local spatial structured variation. Thus, an intuitive approach for checking the relative magnitude of the area-specific geographic effect is to evaluate the ratio  $\text{Var}(\theta_i) / \text{Var}(\mu)$  [17,18]. This index gives information about whether it is necessary to include the regional neighboring structure for the heterogeneity among the regions.

In the CAR model, the ratio of variances is

$$\begin{aligned} \text{Var}(\theta_i) / \text{Var}(\mu) &= \frac{1}{n_i \lambda^2} / \frac{1}{\sigma^2} \\ &= \frac{\sigma^2}{n_i \lambda^2} \end{aligned}$$

When the ratio is close to 1, both sources of variation are equally important. If the ratio is smaller than 1, the global effect dominates; otherwise, the local effect dominates. For convenience, we can also use  $\sigma/\lambda$ , and compare it with  $\bar{n}_i$ , the average of numbers of neighbors.

## A CASE STUDY: THE ASTHMA MORTALITY DATA

### Background

This data set is about the asthma mortality rates for 11 administrative units (Sungshan (including the Sungshan and Hsinyi), Taan, Chungshan, Chungcheng, Tatung, Wanhua, Wenshan, Nankang, Neihu, Shilin, and Peitou) in Taipei metropolitan area in 1991 [19]. Figure 1(a) gives the positions of the areas. The raw counts for the 11 areas were 5, 4, 3, 6, 4, 3, 4, 1, 5, 1, and 4, respectively. Figure 1(b) gave the corresponding rates. The map of crude rates shows great variability among areas. The incidence of asthmatic attack is associated with the environmental condition and therefore the geographically close areas may tend to have similar probability of asthma episode and pos-

sibly the subsequent death. For instance, epidemiologists consider the time needed for getting proper medical care when asthma episode occurs is influential on fatality [20,21]. The availability of the emergency system and the traffic situation of local areas are important in the determination of the asthma fatality. Some of these factors at neighboring areas correlate with each other and some are independent. Therefore, we consider the two models, with and without spatial autocorrelation for the asthma mortality data.

The asthma mortality rates and certain demographic information such as sex and age are available. In this paper, we consider only the males who are 25-59 years old, excluding variations due to demographic variables for simplicity. Here we use the CAR and the EX models to investigate the spatial patterns.

### Fitting the Models

In the CAR model, we assume Normal  $(-3,1)$  for the prior of  $\sigma$  and Gamma(2,1) for the prior of  $\lambda$ . The BUGS program was used to run the Gibbs sampler and chose  $\mu_0=0$ ,  $\theta_0=(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$ ,  $c_0=-1$ ,  $\sigma_0=1$ , and  $\lambda_0=1$  for the starting values. For the EX model, the hyperpriors of  $\alpha$  and  $\beta$  were specified respectively as Exp(0.5) and Gamma(0.5, 0.5), and the starting values chosen were  $\tau_0=(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$ ,  $\alpha_0=0.5$ , and  $\beta_0=0.5$ . For both the CAR and EX models, the first 500 posterior samples were dropped to assure convergence, and the following 5000 iterations generated from the full conditional distributions were collected. In addition, the traces of each parameter's marginal posterior samples were graphically summarized to check convergence. There were at least five different initial values used for both

models to make sure that the summary statistics (mode, median and mean) for each parameter are stable enough (numbers not shown here).

Because the posterior samples are highly skewed, we choose the posterior modes as our estimators for parameters of interest. The estimates and the corresponding crude rates under the CAR and the EX models respectively are shown in Table 2. Figure 1(c) and (d) show the map of the fitted rates under each model. Figure 2 displays the difference in crude and two fitted rates.

The CAR model fitted rates are shrunk to lie in a smaller range (1.16~2.49 per  $10^5$ ). The fitted rates under EX model range from 1.02 to 3.71 per  $10^5$ , which is smaller than that of the crude rate (0 to 7.6 per  $10^5$ ), but wider than that of the CAR model. All the maps of rates in Figure 1 show a consistent result that the area 5 has the highest rate. However, the sharp contrast between black and white shown in Figure 1(b) for crude rates disappears. This is because most of the stochastic variability is removed after modeling. The CAR model includes two kinds of smoothing effect: the global level and the local level. The global smoothing effect also appears in the EX model but the local smoothing effect was constructed only through the neighboring relationship. In other words, the CAR model has greater smoothing effect than EX model does. Therefore, the colors become almost uniform under CAR model, while in EX model, the contrast of colors remains though is not strong.

Figure 2 shows the order of the fitted rates under the EX model and the order is more similar to that of the crude rates than that of the CAR results. Two areas have larger changes in CAR model—area 9 (rising up) and area 6 (falling down). However, it should be noticed

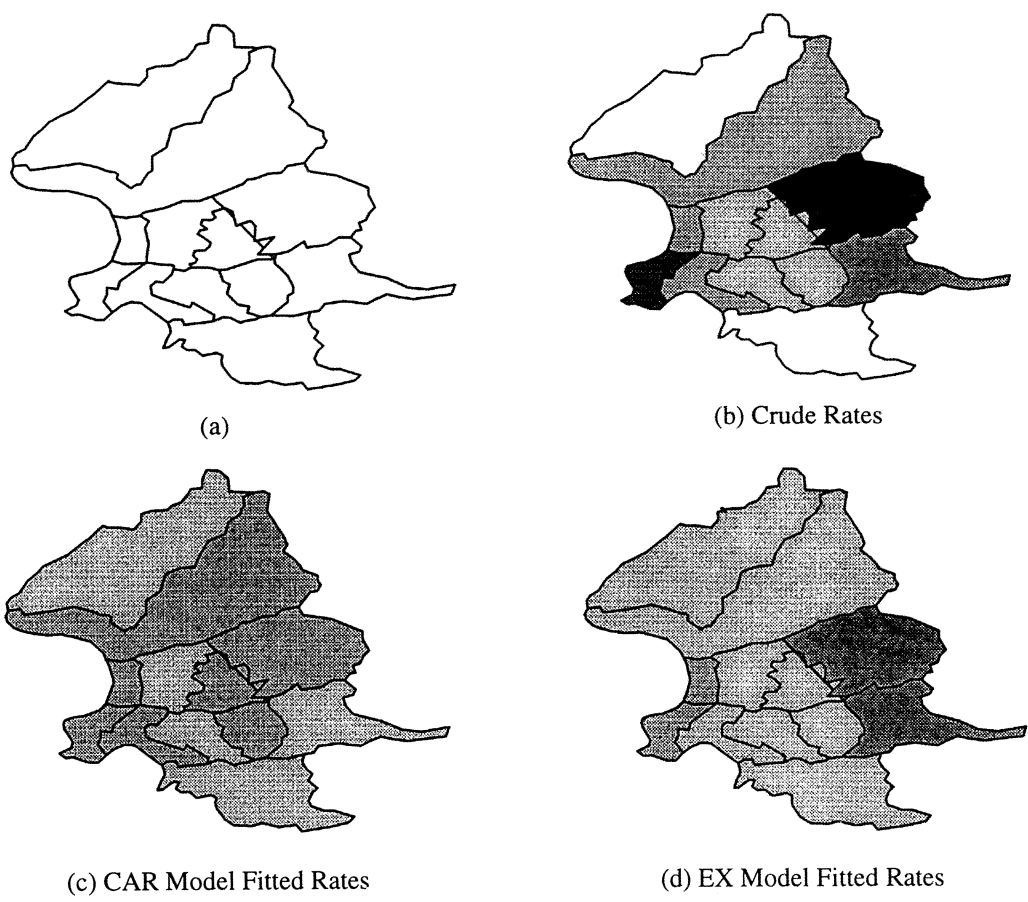
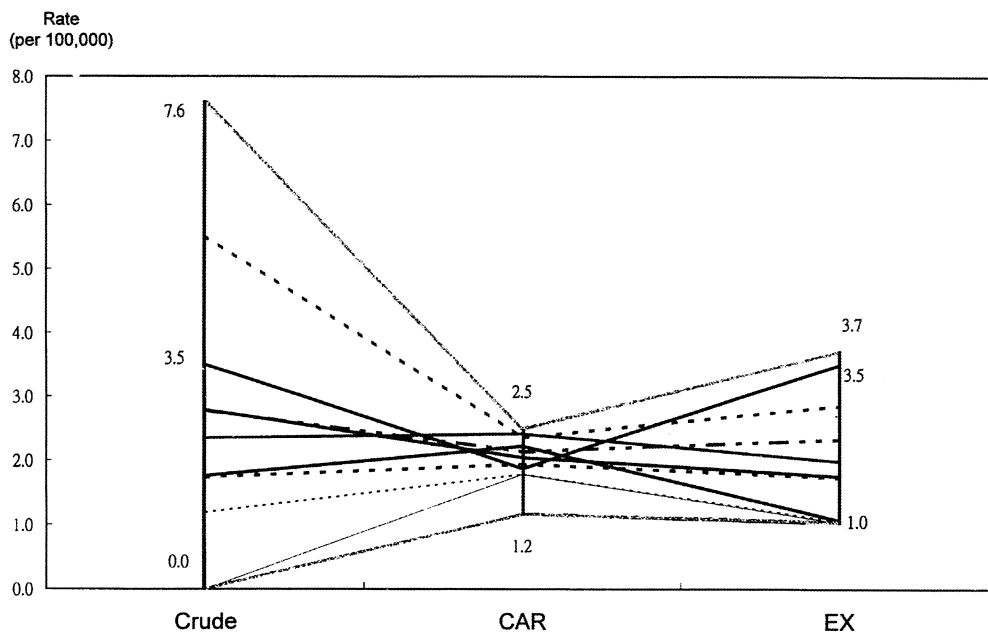


Figure 1. The Crude, CAR and EX Model Fitted Rates

Table 2. Estimation of Mortality Rates by CAR Model and EX Model Respectively

|      |            | Crude Rate   |      |            | CAR          |       |           | EX           |      |
|------|------------|--------------|------|------------|--------------|-------|-----------|--------------|------|
| Area |            | Rate/ $10^5$ | Rank | $\theta_i$ | Rate/ $10^5$ | Rank  | $\tau_i$  | Rate/ $10^5$ | Rank |
| 1    | Sungshan   | 1.75         | 7    | -0.018     | 2.22         | 4     | 1.12      | 1.07         | 8    |
| 2    | Taan       | 1.18         | 9    | -0.050     | 1.78         | 9     | 0.89      | 1.05         | 9    |
| 3    | Chungshan  | 2.77         | 5    | -0.022     | 2.13         | 5     | 0.84      | 2.32         | 4    |
| 4    | Chungcheng | 1.73         | 8    | 0.021      | 1.94         | 7     | 1.00      | 1.73         | 7    |
| 5    | Tatung     | 7.59         | 1    | 0.188      | 2.49         | 1     | 1.95      | 3.71         | 1    |
| 6    | Wanhua     | 3.50         | 3    | 0.058      | 1.87         | 8     | 1.00      | 3.50         | 2    |
| 7    | Wenshan    | 2.78         | 4    | 0.028      | 2.04         | 6     | 1.26      | 1.75         | 6    |
| 8    | Nankang    | 0.00         | 10   | -0.230     | 1.16         | 11    | 0.58      | 1.02         | 11   |
| 9    | Neihu      | 2.34         | 6    | 0.027      | 2.41         | 2     | 0.85      | 1.98         | 5    |
| 10   | Shilin     | 5.50         | 2    | 0.309      | 2.35         | 3     | 1.56      | 2.85         | 3    |
| 11   | Peitou     | 0.00         | 10   | -0.124     | 1.78         | 10    | 0.55      | 1.03         | 10   |
|      |            |              |      |            | $\mu=$       | -1.51 |           |              |      |
|      |            |              |      |            | $c=$         | -1.50 | $\sigma=$ | 1.59         |      |
|      |            |              |      |            | $\sigma=$    | 0.34  | $\beta=$  | 1.23         |      |
|      |            |              |      |            | $\lambda=$   | 2.37  |           |              |      |



**Figure 2. Rate Change among the Curude, CAR and EX Model Fitted Results**

that the rates are all very close in the CAR model.

Both models seem to explain most part of the variation. Since the CAR model has smoother results, one may think that the spatial autocorrelation is a better mechanism to explain the variation beyond the Poisson distribution than the model with exchangeable Gamma effect. However, we should be cautious because the CAR may also represent an over-smoothing effect. It is necessary to investigate if there indeed exists spatial autocorrelation or the data are simply being overly smoothed. Once we are able to choose a "suitable" model, we may be in a better position to understand the underlying etiological factors of asthma death.

### Model Selection

The Gibbs output for the CAR model and the EX model are used to calculate the mar-

ginal probability for the asthma data and we obtain

$$\begin{aligned}\log(\text{BF}) &= \log\{f(y|M_{\text{CAR}})\} - \log\{f(y|M_{\text{EX}})\} \\ &\approx -32.48 - (-18.56) = -13.87\end{aligned}$$

Comparing with the values in Table 1 we have a "very strong" evidence in favor of the EX over the CAR model.

An alternative selection criterion is  $\sigma/\lambda$ . Based on Gibbs samples of the posterior distribution of  $\sigma/\lambda$ , we took the mode 0.157 to be the estimator. This is far from the  $\bar{\pi}_i = 3.64$ , implying a minor effect of the spatial neighboring variation. In other words, the local area-specific effect is only  $0.157/3.64 \approx 1/23$  of the global effect. This number also indicates that the EX model which does not include the spatial correlation structure may be more suitable for the asthma data.

The estimator of  $\sigma/\lambda$  implies that the lo-

cal area effect is fairly weak so that it can be ignored. The Bayes factor also favors the exchangeable Poisson-Gamma model which does not consider the area-specific effect. This example illustrates that jumping to the conditional autoregressive model without carefully examining the magnitude of the local area-specific effect and the contrast of evidence of CAR and EX models may be risky. It is really important to develop an informative method to judge if the area-specific effect is worth considering in real applications.

## DISCUSSION

In the asthma study, we deliberately chose subjects with the same age and sex, and hence the crude rate can be treated as the age-sex adjusted rate. When comparing the fitted results, the maps of rates under either hierarchical model have less variabilities and more stabilities than observed rates. This is because part of the random variation is removed from the total variation. However, these two models assumed different kinds of random mechanism. For the EX model, the rates are considered to be independent and identical following the same probability distribution. For the CAR model, rates are assumed to be correlated and information from neighbors is included to help estimate the rate of the target area. Therefore, the results from the CAR model have greater smoothing effect than that from the EX model. This is reflected not only on the zero-rate-areas (areas 8 and 11), but also on the ranking pattern. Also because of the neighboring effect, the CAR fitted rates have much smaller range than the EX. There can be several choices in defining the "neighboring areas" besides the adjacent neighbors we used here. For instance, one can use a

fixed distance to decide which areas have contribution effect to the target area, or one can simply include all other areas as neighbors. These approaches can be done by modifying the formulations of mean and variance of  $\theta_i$  when defining its prior distribution.

For the asthma study, the  $\log_e$  BF is -13.87 and the  $\sigma/\lambda$  is 0.157. The small value of  $\sigma/\lambda$  implies a weak local effect over the global effect. Therefore, the local area specific effect may be ignored for the model simplicity and one can focus on the results under the EX model. According to the meanings of the EX model, the heterogeneity among areas may come from the spatially independent variables rather than spatially correlated ones. One can also consult with the expert opinions to see whether the EX model is supported by epidemiological information. For the asthma data introduced in this paper, the results from both models indicate that the mortality rates in area 5 (Neihu) and area 6 (Nankang) are higher than other areas. These areas deserve more attention for possible etiological cause.

For spatial model selection, the Bayes factor and the ratio of the local to global variance ( $\sigma/\lambda$ ) provide similar information but their meanings are different. The Bayes factor assesses two competing models by comparing their marginal probabilities of the data. Through evaluating the strength of evidence supported by the observed data, it offers a direct answer for choosing the better model. On the other hand, the ratio of the local and global variances provides a way to evaluate the necessity to consider the area-specific effect when modeling.

There still exist some issues which need further study. First, in order to investigate the performance of the two model selection

indices, simulation studies that have samples from the CAR and the EX models are needed. Second, the prior distribution may be influential to our results. In our study, although the recommended Gamma distribution is used for hyperparameters, other priors have not been used. With simulation, the effect of the prior in the variance of the estimates can be examined. Finally, there are other ways to approximate the Bayes factor. We have used in this paper an importance-sampling approximation to the Bayes factor based on the Gibbs output. Other sampling functions than the posterior distributions can be alternatives. An evaluation of efficiency and accuracy of these alternatives may be necessary and useful for further research.

## REFERENCES

1. Cressie NAC. *Statistics for Spatial Data*. 2nd. NY: John Wiley & Sons inc, 1991; 10-3.
2. Lindsey JK. *Models for Repeated Measurements*. Oxford: Clarendon Press, 1993; 22.
3. Tsutakawa RK. Estimation of cancer mortality rates: a Bayesian analysis of small frequencies. *Biometrics* 1985; **41**:69-79.
4. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; **43**:671-81.
5. Besag J, Mollié A. Bayesian mapping of mortality rates. *Bull Int Statist Inst* 1989; **53**:127-8.
6. Richardson S. Statistical methods for geographical correlation studies. In: Elliott P, Cuzick J, English D, Stern R eds. *Geographical and Environmental Epidemiology*. London: Oxford University Press. 1992; 181-204.
7. Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika* 1995; **82**:733-46.
8. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**:9-25.
9. Gilks WR, Thomas A, Spiegelhalter DJ. A language and program for complex Bayesian modeling. *The Statistician* 1994; **43**:169-78.
10. Spiegelhalter DJ, Best NG, Gilks WR, Inskeep H. Hepatitis B: a case study in MCMC methods. In: Gilks WR, Richardson S, Spiegelhalter DJ eds. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall, 1996; 21-43.
11. Best NG, Spiegelhalter DJ, Thomas A, Brayne, CEG. Bayesian analysis of realistically complex models. *J Roy Statist Soc A* 1996; **159**:323-42.
12. Gilks WR. Full conditional distributions. In: Gilks WR, Richardson S, Spiegelhalter DJ eds. *Markov Chain Monte Carlo in Practice* 1st. London: Chapman & Hall. 1995; 75-88.
13. Gilks WR, Richardson S, Spiegelhalter DJ. In: Gilks WR, Richardson S, Spiegelhalter DJ eds. *Markov Chain Monte Carlo in Practice* 1st. London: Chapman & Hall. 1995; 1-19.
14. Jeffreys H. *Theory of Probability*. 3rd. London: Oxford University Press, 1961; 432.
15. Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. 1st. London: Chapman & Hall, 1996; 222.
16. Raftery AE. Hypothesis testing and model selection. In: Gilks WR, Richardson S, Spiegelhalter DJ eds. *Markov Chain Monte Carlo in Practice* 1st. London: Chapman & Hall. 1995; 169-76.
17. Mollié A. Bayesian mapping of disease. In: Gilks WR, Richardson S, Spiegelhalter DJ eds. *Markov Chain Monte Carlo in Practice* 1st. London: Chapman & Hall. 1995; 419-40.
18. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* 1995; **21/22**:2433-45.
19. Department of Health, the Executive Yuan. *Health and vital statistics 1991*, Republic of China. Department of Health, the Executive Yuan. 1991.
20. Wobig EK and Rosen P. Death from asthma: rare but real. *Journal of Emergency Medicine* 1996; **14**: 233-40.
21. Tough Sc, Green FH, Paul JE, Wigle DT, Butt JC. Sudden death from asthma in 108 children and young adults. *Journal of Asthma* 1996; **33**:179-88.

# 使用貝氏因子與變異項比值進行氣喘死亡率 地理空間資料模型之選擇

曾仲瑩<sup>1</sup> 蕭朱杏<sup>2</sup> 陳建仁<sup>2</sup>

以階層式模型分析空間地理資料，除了可考慮有規則性的變異量之外，亦可針對隨機變異量加以處理。在認為各地死亡率係受另一隨機分佈(即事前分佈)支配的前提下，階層式模型透過所假設的隨機分配以解釋隨機變異，其中最常用的兩種模式即為conditional autoregressive (CAR) 模式與 exchangeable (EX) 模式。然而，欲在這兩種代表不同意義的模式間進行選擇，卻往往欠缺判斷的準則。本研究以台北地區氣喘死亡率資料為例，在面對兩種不同死亡原因的假說的情況下，我們兼採CAR模式與EX模式來探討氣喘死亡率的地理趨勢，並利用Fully Bayesian方法與Monte Carlo Markov Chain原理(如Gibbs sampling)來估計參數。除了須選出一適當的模式來描

述其地理分佈狀況外，更希望透過所選用模式的意義來了解氣喘死亡的可能成因。我們選採兩個指標—貝氏因子與局部作用之變異量與整體作用之變異量比值—作為模式選擇的依據。研究結果發現：(1)由貝氏因子可直接看出資料較支持何種模式；而變異量比值則顯示資料本身的性質與是否應將局部作用放入模式中加以考慮，在模式的選擇上，二者應相互配合使用。(2)由指標來判斷，對氣喘死亡率資料而言，以EX模式較為適恰。由此亦推論內湖區、南港區有較高死亡率；而仍存在於地區間的變異則較可能由地區間彼此獨立的因素所引起。(中華衛誌1998；17(2)：158-169)

關鍵字：貝氏因子，條件式自我相容模式，可交換模式，地理空間資料模型選擇。

<sup>1</sup> 中央研究院統計科學研究所

<sup>2</sup> 國立台灣大學流行病學研究所

聯絡人：蕭朱杏

聯絡地址：臺北市仁愛路一段一號15樓1542室

電話：(02)2397-0800轉8372

傳真：(02)2351-1955