Comparing the performance of two indices for spatial model selection: application to two mortality data

Chuhsing Kate Hsiao[1,*,†], Jung-Ying Tzeng[2] and Chih-Hao Wang[3]

[1] *Divison of Biostatistics, Institute of Epidemiology, National Taiwan University, No. 1, Jen-Ai Rd, Sec. 1, Rm 1542, Taipei 100, Taiwan, R.O.C.*
[2] *Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, U.S.A.*
[3] *Taipei Municipal Chung-Hsiao Hospital, No. 87, Tong-Teh Rd., Taipei 115, Taiwan, R.O.C.*

SUMMARY

The statistical analysis of spatially correlated data has become an important scientific research topic lately. The analysis of the mortality or morbidity rates observed at different areas may help to decide if people living in certain locations are considered at higher risk than others. Once the statistical model for the data of interest has been chosen, further effort can be devoted to identifying the areas under higher risks. Many scientists, including statisticians, have tried the conditional autoregressive (CAR) model to describe the spatial autocorrelation among the observed data. This model has greater smoothing effect than the exchangeable models, such as the Poisson gamma model for spatial data. This paper focuses on comparing the two types of models using the index LG, the ratio of local to global variability. Two applications, Taiwan asthma mortality and Scotland lip cancer, are considered and the use of LG is illustrated. The estimated values for both data sets are small, implying a Poisson gamma model may be favoured over the CAR model. We discuss the implications for the two applications respectively. To evaluate the performance of the index LG, we also compute the Bayes factor, a Bayesian model selection criterion, to see which model is preferred for the two applications and simulation data. To derive the value of LG, we estimate its posterior mode based on samples derived from the BUGS program, while for Bayes factor we use the double Laplace–Metropolis method, Schwarz criterion, and a modified harmonic mean for approximations. The results of LG and Bayes factor are consistent. We conclude that LG is fairly accurate as an index for selection between Poisson gamma and CAR model. When easy and fast computation is of concern, we recommend using LG as the first and less costly index. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION AND MOTIVATION

Modelling spatially collected, possibly correlated, data has become an important scientific research topic for scientists such as environmental engineers, hydrogeologists and epidemiologists. Whether the spatial correlation exists may lead to different decision making for environmental remediation,

---

public policy, or inference on disease aetiology. For instance, analysis of the mortality rates observed at different areas may help to decide if certain locations should be considered high-risk areas. One interesting question therefore is the selection of a proper model to represent the underlying distribution of the observed spatial data. The classical Poisson regression model for counts data may not be applicable for the observed number of deaths since the true but unknown rates may not be identical across areas. In addition, the disease numbers based on the observed rates at areas on the same map may not be comparable since the variance may be larger than expected under the Poisson model. The observed rates may range greatly if the population sizes or other demographic characteristics are not of similar magnitudes. Two alternative models counting for the extra-Poisson variability are the Poisson gamma [1] and conditional autoregressive models [2].

The Poisson gamma (PG) model assumes the rates at various areas are independent and identically distributed from a gamma distribution, namely the rates are exchangeable. Consequently, the estimates resulting from the Poisson gamma model will shrink toward the overall mean. Under the conditional autoregressive (CAR) model, however, the rates are assumed to be correlated with each other. In other words, the variability among the observed rates can be smoothed via the correlation due to the spatial structure. Therefore, the CAR model has larger smoothing effect than the Poisson gamma model. The CAR model has, in fact, received great attention lately, especially in applied research areas. Most research so far has concentrated mainly on deriving estimates [3–5]. It has not yet been discussed in the literature how to choose among the Poisson gamma and CAR models.

This paper was motivated by a study [6] on mortality rates of asthma in Taiwan. Although it was suspected that asthma mortality rates from neighbouring areas were similar, the researchers asked for evidence in choosing between the Poisson gamma and CAR models. Inference based on different models may lead to different public health policy or various aetiological studies. For instance, because areas identified as high-risk areas may require special attention, such as a well designed study for levels of pollution, it is important to select an appropriate statistical model for the observed rates and for pointing the priority areas. The resulting areas may be different when fitting the PG and CAR. When these two models are both of interest and a better one needs to be selected for making statistical inference later, a quantity for the purpose of selection should be employed. In this paper we propose to use the ratio of local to global (LG) variabilities as an indicator of which model fits the data better. With help from the Markov chain Monte Carlo method [7], this quantity is easy to estimate. We will demonstrate its usage in Taiwan asthma mortality and Scotland lip cancer data [6, 8]. The first mortality data are simpler in the sense that no explanatory variable is included, while the lip cancer data contain a covariate. In order to examine the accuracy of LG in model selection, we use Bayes factor, a measurement for Bayesian model selection and hypotheses testing [9, 10], as a confirmation.

Specifically, our purpose is to test the Poisson gamma versus CAR model. In other words, to test whether the exchangeable model or the correlated model describes better the observed spatial variability. The main index we investigate is LG, the ratio of local to global variabilities under the CAR model. This quantity has been mentioned briefly in other places [4, 5] to be used as an indication of the relative magnitudes of spatial heterogeneity. When LG is large, it indicates that the local spatial variation is larger than the global variation. An estimate of LG can be easily obtained from their posterior samples when fitting the CAR model using the BUGS [11] program. Once its value under CAR is obtained first, it can indicate whether CAR, the model with spatial correlation, is a good model for the data. If CAR is not proper for the observations, then fitting

an exchangeable model such as Poisson gamma may be necessary. The other index, Bayes factor, for checking the accuracy of LG, is a measure of evidence supported by the data [9]. It takes account of not only any prior knowledge regarding the parameters but also the uncertainty in each competing model [10, 12, 13]. Because the exact evaluation of the Bayes factor in this setting is not analytically feasible due to the complexity in computations, we use the posterior samples and the double Laplace–Metropolis method for approximation. It is essentially a variant of the Laplace–Metropolis estimate [14] but is more suitable in testing PG versus CAR model for the difficulty in computation. We also discuss the use of Schwarz criterion and a harmonic mean estimate for approximations [12]. Some modifications will be necessary when applied to the CAR model.

Following this section, we discuss the backgrounds of two mortality data. In Section 3, we give notations for the two models and specify quantities of interest. We will define LG and discuss how to obtain the approximations to Bayes factor. In Section 4, we investigate the LG when applied to two examples. In Section 5 we examine the operating characteristics and performance of LG and compare it with Bayes factor using data simulated from CAR model. All computations were based on the BUGS output for computational convenience. We summarize and discuss our results in Section 6.

2. TWO CASE STUDIES: TAIWAN ASTHMA MORTALITY AND SCOTLAND LIP CANCER

Asthma mortality and morbidity have been on the rise in Taiwan and world-wide [15, 16]. Failure to prevent episodes of asthma attacks and to decrease the mortality may elevate the medical costs. Some studies point out that this recent increase may relate to environmental and genetic factors [16–18]. Some studies search for the implications for public health preventive strategies [19, 20] or focus on the evaluation of the effectiveness of existing intervention programs [21–23]. Before the formation and advocation of public health policy to residents in certain regions, it must be identified that certain areas are at higher risks than others and thus deserve more attention. The first example we discuss is the asthma mortality rates for the 11 administration areas in Taipei city in 1991 (Figure 1(a)). The goal is to investigate whether there exists any area with higher mortality rates. Once the statistical evidence is confirmed, the public health organization may need to evaluate the availability of timely medical attention for that priority area or consider an intervention programme for the whole study region if no difference is found among the 11 areas. In statistical terminology, we need to construct models to describe the spatial pattern of the mortality data and to identify the special areas.

The two spatial models, Poisson gamma (PG) and conditional autoregressive (CAR), will be used to analyse these asthma mortality rates. Under the PG model, the mortality in area i will not affect the magnitude of the mortality in area j due to independence. This model is simple and easy to understand conceptually but the independence assumption rules out the possibility that nearby areas may share similar medical resources and thus result in mortality of similar magnitudes. The alternative CAR model considers the correlation among neighbouring areas. In other words, the mortality rates of adjacent areas i and j are allowed to be similar and correlated to a certain degree. The two models may result in different estimates and thus provide different areas of higher risks.
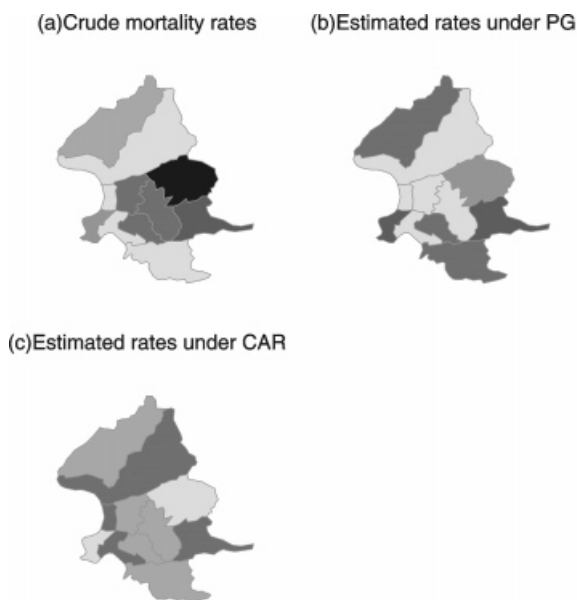
Figure 1. Maps of the (*a*) crude mortality rates, (*b*) estimated rates under Poisson gamma model, and (*c*) estimated rates under CAR, for Taipei city. Darker colours indicate higher rates.

To simplify the computation and illustrate the methods discussed later, we chose deliberately the raw counts of deaths for males aged from 25 to 59 for the areas under investigation. The observed numbers were $2, 1, 1, 1, 4, 1, 2, 0, 1, 3$ and 0, respectively. In this example, we consider no other variables and thus the computation would be less complicated. It should be borne in mind, however, that the methodology discussed in this paper should not be limited to simple cases only. In the following example, we discuss another data set which contains an explanatory variable in the model. The more variables that are included, the more intensive the computation becomes.

The second example considers the lip cancer rates for $N = 56$ areas in Scotland [8, 24]. A description and map for the data can be found in other references [8, 24]. The lip is the most common site for oral cancer in western countries. It is known to occur mostly in ageing White males and is extremely rare among Blacks and Orientals [25, 26]. Two major risk factors for lip cancer are rural residence and outdoor occupations related to sun exposure [26, 27]. In this data, the observed number of cases and a covariate measuring environmental risk for each area are available. This covariate measures the percentage of labourers employed in outdoor occupations such as agriculture, fishing and forestry for each area. It was noted that the covariate seems to aggregate spatially and the pattern is similar to that of the lip cancer rates. We are interested in knowing whether the cancer rates correlate spatially and whether this correlation can be explained by the covariate. Again, we use Poisson gamma to represent the model without spatial correlation. We use CAR to represent the model with spatial correlation among neighbouring areas. The covariate relevant to sun exposure is included in both models. If CAR is preferred over the PG model, it may imply that we need to identify factors for the type of correlation or ex-

amine similar characteristics shared by those areas. If the PG model is favoured, it implies that the percentage of the workforce engaged in outdoor occupations can explain the observed local correlation.

3. NOTATIONS AND METHODS

For the Poisson gamma model, let y_i denote the observed number of diseased, m_i be the population size, p_i be the disease rate (morbidity), $\tau_i \, (= m_i p_i)$ be the Poisson parameter for area i where $i = 1, \ldots, N$, and N be the total number of areas under investigation. The model can be written as $(y_i | \tau_i) \sim \text{Poisson}(\tau_i)$ with $(\tau_i | \alpha, \beta) \sim \text{gamma}(\alpha, \beta)$. When covariates x_i are of interest, we decompose $\log p_i$ into $\alpha_1 x_i / 10$ and b_i where b_i is a random error and thus assumed to follow a normal distribution. These two models (Poisson-gamma and Poisson-normal) are commonly used to describe extra-Poisson variability [1, 23]. Because these two are both uncorrelated random effects models, we refer them to the PG model for simplicity. In these models, the expected number τ_i of diseased for area i is assumed independently and identically distributed from a distribution. In other words, the τ_i's in different areas are exchangeable and the gamma distribution on τ_i is the source of the extra-Poisson variability. The prior density functions on the hyperparameters α and β can be assumed gamma (a, b) and gamma (s, t) distributions, respectively. By changing the values of a, b, s and t, the magnitudes of the spatial variability among y_i's may be different.

3.1. CAR Model and Index LG

The Poisson gamma model, although it smoothes the spatial variability through the gamma distribution, may not be ideal for spatially correlated observed data. An alternative is the conditional autoregressive (CAR) model. The observation y_i follows a Poisson distribution with parameter $(m_i p_i)$ where $\log(p_i) = \mu + \theta_i$. That is, the logarithm of p_i is decomposed to two parts. The components $\theta_{j, j \neq i}$ are rates of the first-order neighbours of area i, and follow an intrinsic normal prior [8, 28] with the constraint $\sum_i \theta_i = 0$ for identifiability. The conditional mean of θ_i given its neighbours and λ is $\bar{\theta}_i$, and the variance is $1/(n_i \lambda)$. The other component μ stands for the global spatial heterogeneity. It is assumed that μ is from a normal distribution with mean c and precision σ. This model is sometimes referred to as an intrinsic Gaussian conditional autoregressive (ICAR) [28, 29]; here we will call it simply the CAR model. The spatial correlation in CAR is modelled through the intrinsic normal prior [8, 28] on θ_i and hence the expected numbers of diseased are correlated with each other. Because the mean of θ_i depends only on rates from adjacent neighbours rather than all others, it is sometimes referred to as the Gaussian intrinsic autoregression. Writing θ as a vector $(\theta_1, \theta_2, \ldots, \theta_N)$ and using multivariate normal distribution for notation, its precision matrix Q, $Q = D \times (I - C)$, would be positive semi-definite [30]. In other words, it is not invertible due to the local correlation. The I above is the identity matrix, D is a diagonal matrix with $d_{ii} = n_i \lambda$, and the (i, j)th entry in C is $c_{ij} = w_{ij}/n_i$ where $w_{ij} = 1$ if areas i and j are neighbours, $w_{ii} = 0$, and 0 otherwise.

The two hyperparameters, σ and λ, represent, respectively, the precision of the global spatial effect (μ) and that of the local spatial effect (θ_i). The σ and λ are further assumed from two gamma distributions, respectively, where each distribution may represent vague prior information for the parameter. The LG is defined as σ/λ, namely the ratio of local variability (λ^{-1}) to global variability

(σ^{-1}). Although the local spatial variation for each area is $1/(n_i\lambda)$, here we have dropped n_i and used simply the inverse of λ for local variability for convenience. It was argued [4] that when LG is close to the average value of numbers of neighbours ($\bar{n} = \sum n_i/N$), both local and global spatial variability are equally important. When LG is larger than $\bar{n}$, the locally correlated spatial variation is more important. In other words, larger LG indicates greater local variation and thus the CAR model may be favoured over unstructured spatial models such as Poisson gamma. When LG is small, it may indicate that it is inappropriate to fit the locally correlated spatial variation. When LG is close to $\bar{n}$, the simpler PG model may be preferred because of no strong support for either model. Based on the Gibbs output from the BUGS program, obtaining the estimates of LG is fairly straightforward; we will show later that one can simply look at its value and decide whether CAR is a proper model and if fitting an exchangeable model such as Poisson gamma is necessary.

3.2. Three Approximations to Bayes Factor

When there are two competing models for a data set, the Bayesian approach is to use the Bayes factor as a model selection criterion. The Bayes factor, defined as the posterior odds divided by the prior odds in favour of H_0, originally was used as a measure of strength of evidence in testing the null hypothesis H_0 versus alternative H_1 [9, 31]. It can be written as the ratio of two marginal probabilities of data, $\Pr(y|H_0)/\Pr(y|H_1)$, where the representation of each probability can be found in Appendix A. The Bayes factor can also serve as a measure for model selection when each competing model is *a priori* equally likely [10]. In our settings, the two competing models are Poisson gamma and CAR. By convention, we put the simpler PG model as the null H_0 and CAR as the alternative. We now need to evaluate the value of Bayes factor to decide which model fits the data better. Because the analytical forms of these two marginal probabilities are not tractable, we will use approximations. In the following, we describe the double Laplace–Metropolis method, Schwarz criterion, and the Markov chain Monte Carlo method for approximations. The three approximations to the marginal probability are all based on the output obtained via the Markov chain Monte Carlo (MCMC) method [7]. The MCMC method uses the Markov chains to perform Monte Carlo integration. One special case, the Gibbs sampler technique [32, 33], is often used to obtain posterior samples of quantities of interest from the set of the fully conditional distributions. We have used the BUGS [11] program to obtain the posterior samples of α and β under the Poisson gamma model as well as the samples of μ, θ, c, σ and λ under the CAR model.

The first approximation uses the concept of Laplace's method [34–36] which is accurate with order $O(N^{-1})$ where N is the sample size. Owing to the difficulty in evaluating the MLE under CAR, we will use the double Laplace–Metropolis approximation. The first step is

$$\Pr(y|\text{CAR}) = \int f(y|\phi)\pi_1(\phi)\,\mathrm{d}\phi \approx L_1(\hat{\phi})\pi_1(\hat{\phi})(2\pi)^{3/2}|\Sigma_1|^{1/2} \tag{1}$$

where $\phi = (c, \sigma, \lambda)$ and Σ_1 is the observed Fisher information matrix of c, σ and λ. Lewis and Raftery [14] have suggested using estimates from the posterior samples to approximate the marginal probability under the random effect model and called it the compound Laplace–Metropolis method.

However, the L_1 in our case is more complicated and its analytical form is not available. Because

$$L_1 = f(y|c, \sigma, \lambda) = f(y|\phi) = \int f(y|\mu, \theta)\pi(\mu, \theta|\phi)\, d(\mu, \theta)$$

the integration over μ and θ is not feasible, and we will approximate L_1 based on posterior samples. The details of its derivation is given in Appendix A.

The Schwarz criterion S, also called Bayes information criterion (BIC) [37], is an approximation to the logarithm of the Bayes factor [38,39] with constant order. It can be written as

$$S = \max \log f(y|H_0, \eta_0) - \max \log f(y|H_1, \eta_1) - \frac{d_0 - d_1}{2} \log N \qquad (2)$$

where $f(y|H_i, \eta_i)$ is the likelihood of η_i, d_i is the dimension of η under model H_i, and N is the sample size. In theory, if the form of likelihood is available, its maximum can be obtained either analytically or numerically by choosing the largest value among all the ones evaluated at various posterior samples. For Poisson gamma model, the likelihood is $f(y|\alpha, \beta)$ and is analytically tractable due to the conjugacy, while for CAR, the maximum likelihood can be estimated by $L_1(\hat{c}, \hat{\sigma}, \hat{\lambda})$. Substituting this quantity for $\max \log f(y|H_1, \eta_1)$ in equation (2) we obtain the Schwarz criterion, an approximation to Bayes factor.

The third approximation is based on the importance sampling method. To estimate the marginal distribution of the data, $\Pr(y|H_i)$, using the K posterior samples from the BUGS output, we take the posterior distribution as the importance sampling function when estimating $\Pr(y|H_i)$ [40]. In other words

$$\Pr(y|H_i) \approx \left[\frac{1}{K} \sum_{k=1}^{K} L_i(\eta^{(k)})^{-1} \right]^{-1}$$

where L_i is the likelihood of η, η is the vector of parameters under H_i, and $\eta^{(k)}$ are posterior samples. This is in fact the harmonic mean of the likelihoods evaluated at various parameter values.

The likelihood L_1 under the CAR model evaluated at K various points $\phi_i (= (c_i, \sigma_i, \lambda_i))$ requires taking integrations over μ and θ for K times, as indicated in equation (A2) in Appendix A. If Laplace's method is applied for the K fixed ϕ_i points to derive $L_1(\phi_i)$, the error may contribute to the unstability of the harmonic mean estimator due to the occasional small likelihoods [13]. Instead we interchange the integration signs under regularity conditions to obtain

$$\Pr(y|H_1) = \int f(y|\mu, \theta)\pi^*(\mu, \theta)\, d(\mu, \theta)$$

where π^* is the 'marginal prior' after integrating out ϕ (see Appendix B). In particular, π^* can be derived analytically when $(\mu|c, \sigma)$ and $(\theta|\lambda)$ are from normal, c from uniform, and σ and λ from gamma distribution, as assumed in the following sections. Therefore, the harmonic mean of $f(y|\mu_i, \theta_i)$ with (μ_i, θ_i) being the posterior samples is an approximation to $\Pr(y|CAR)$ and mathematically equivalent to the harmonic mean of $L_1(\phi_i)$. This approximation is easier in

computation especially when a general form of $L_1(\phi)$ does not exist. In the next section, we apply the methodology to the two applications introduced earlier.

4. EXAMPLES REVISITED

For the asthma data, the PG model implies that the difference in the observed mortality rates occurs by chance alone. No area has unusual mortality rates and neighbouring areas do not have similar rates. On the other hand, the CAR model implies clustering of mortality rates. Specifically, this model indicates that neighbuoring areas would have mortality rates of similar magnitude. When fitting the PG model to the asthma data, we chose the exponential (0.5) distribution and gamma (0.5,0.5) for the hyperparameters α and β, respectively. When fitting the CAR model, we used the uniform $(-3,1)$ for c, gamma (4,1) for σ, and gamma (4,1) for λ as prior distributions. These distributions indicated fairly vague prior information about parameters. Moreover, slightly stronger local $(1/\lambda)$ than global $(1/\sigma)$ variability was assumed for the priors of σ and λ under CAR. The mean of c under the uniform was -1 corresponding to the average value $\exp(\mu) = 0.37$ of deaths among the population of size 100 000. The range of c was between $\exp(-3) = 0.05$ and $\exp(1) = 3$. These numbers were considered reasonable for a not-too-rare disease and were selected based on the asthma mortality rates of Taiwan in 1991 [41]. The population sizes m_i for the 11 areas are 123 529, 91 710, 39 717, 62 526, 57 614, 32 006, 79 667, 63 000, 46 252, 60 859 and 59 353 [41]. Dispersed starting values were used in the Gibbs sampler and posterior samples were collected from multiple runs. Figure 1 displays the estimated rates under each model. Darker colour indicates higher rates.

The estimate mode of LG based on the BUGS output was only 2.7, far smaller than the average number of neighbours, 3.64, implying that the local spatial variability was not significant as compared with the global spatial variation. In other words, the CAR model may be too complicated for the asthma mortality data. To approximate the logarithm of Bayes factor, the Laplace–Metropolis estimate gave $-13 - (-49) = 36$, the Schwarz criterion was $-15 - (-48) = 33$, and the modified harmonic mean estimate was $-34 - (-57) = 24$. All three approximations indicate very strong evidence that PG model fits better than the CAR model. The conclusions based on LG and Bayes factor are consistent and the same as in the previous literature [6]. Although the area with the highest estimated rate under the PG model remains the same as under the CAR, other areas with higher estimated rates under PG differ from those areas identified under the CAR model. For instance, area 6 ranks the second highest estimated rate under PG but it ranks only the eighth under CAR. Therefore, if the intervention programme can only be applied in a few areas, the choice will depend heavily on the correct model used. Once the areas at higher risks are identified, further studies on its medical resources, availability of timely emergency care, and design of intervention programmes can be carried out [21–23].

For the Scotland lip cancer data, the PG model implies that the rates are exchangeable after adjusting for the covariate x_i of percentage of outdoor occupations. That is, when taking x_i into consideration, all areas are under the same risk and the difference in the observations is due to chance only. On the other hand, the CAR model implies that even when x_i is considered, there may exist correlation among neighbuoring areas. In this data, a term of $\alpha_1 x_i/10$ was added in the decomposition of log p_i for each area to stand for the covariate under both PG and CAR models. The prior distribution for c, σ, λ and α_1 were uniform $(-1,1)$, gamma (2,1), gamma (2,1), and normal $(0, \text{var} = 1 \times 10^5)$, respectively. While under the PG model, the priors for α_1 and

b_i were normal $(0, 1 \times 10^5)$ and gamma $(1 \times 10^{-3}, 1 \times 10^{-3})$. Again, these prior distributions all have large variances indicating no preferable values in the range based on previous available knowledge.

The estimated LG based on the BUGS output was 0.007, indicating very weak local variability as compared to the global one (the average number of neighbours was 4.71). We also computed the approximation to logarithm of the Bayes factor based on the modified harmonic mean estimate. The value was 11 representing strong support of the PG over CAR model. In other words, with the presence of the covariate of percentage of outdoor occupations, there does not seem to be spatially local correlation for lip cancer mortality rates in Scotland. As a consequence of model selection, the Poisson gamma would be a better choice than the CAR model. It also indicates that much of the spatial similarity was explained once the covariable was included in the model. A similar conclusion pointed out [8,24] that there did not seem to be spatial correlation for these 56 areas once the covariate was contained in the statistical model. In other words, this explanatory variable of workforce engaged in outdoor occupation removes most of the spatial correlation. Note that the Laplace and Schwarz criterion approximations are not available for this data set because the calculation involves computing the covariance matrix of the rates whose dimensionality is 56×56. In the following we will consider fewer numbers of areas and carry out all approximations.

5. SIMULATION

In this section we describe the results based on simulations and compute the LG and the above three approximations to Bayes factor for each replication. Because the previous two data sets have demonstrated the use of LG when indeed the Poisson gamma is a better model than the CAR, here in this section we focus on data generated from CAR only. We would like to first examine the behaviour of LG when the spatial local correlation among data ranges from strong to weak, and next to confirm its indication of the better model using the approximations to Bayes factor as the standard.

There were three sets of data generated from CAR model, denoted as CAR1, CAR2 and CAR3. When generating the CAR data, because the intrinsic autoregressive prior on $\theta = (\theta_1, \theta_2, \ldots, \theta_N)$ was a multivariate normal distribution with a zero mean vector and a non-invertible precision matrix Q, the θ cannot be generated directly from a non-singular multivariate normal distribution as commonly done in most computing packages. Based on the result in Besag and Kooperberg [30], we performed some transformations along with data generation. More details were listed in Appendix C. For the three CAR data sets, the hyperparameters σ, λ and c were fixed at $(1, 0.1, -1)$, $(0.5, 0.1, -1)$ and $(1, 1, -1)$, respectively. These values were specified to control the level of autocorrelation. For each area, the population size m_i was chosen deliberately to be 100 000 for the sake of comparison. In other words, the numbers of diseased now stood for the observations from 100 000 people. For each simulation, there were 1000 replications. For each replication, an estimate of LG was calculated first to see if spatial local correlation is significant as compared with global variability. All replications were fitted with both PG and CAR models and the three approximations to Bayes factor were computed to see which model fits better and whether the conclusion is consistent with that from LG. Under the PG model, we have assigned gamma $(4,1)$ distribution to α and gamma $(4,4)$ to β. Under the CAR model, the priors on c, σ and λ were assumed uniform $(-3, 1)$, gamma $(4,1)$, and gamma $(4,4)$, respectively. With reference

Table I. Summary statistics (the 5th and 95th percentiles, quartiles, median and mean) of LG when data were generated from CAR model.

Data	5th percentile	25th percentile	median	mean	75th percentile	95th percentile	Number $\geqslant 3.64$
CAR1	1.66	6.09	11.69	13.99	18.77	35.99	836
CAR2	1.28	4.59	9.74	12.38	17.11	32.84	799
CAR3	0.66	1.16	1.39	1.53	1.74	2.99	17

to the application, we have fixed the numbers of neighbours to be 5, 4, 4, 6, 5, 6, 10, 5, 7, 5 and 9 for each area, respectively, and the total number of areas was $N = 11$. The data were all generated from the built-in subroutines in FORTRAN and then analysed with BUGS to produce posterior samples.

5.1. Behaviour of LG

The ratio of local to global variability LG was suggested [4, 5] to be compared with the average number of neighbours $\bar{n}$ to see whether the local or global spatial variability is stronger. Table I lists the summary statistics of LG. The estimated LG for each simulation was obtained using the posterior samples of σ and λ when data were fitted with the CAR model. In CAR1 and CAR2 data sets, almost 80 per cent of the LGs were larger than 3.64, indicating greater local spatial variability than the global variance. It implies that CAR should be a fairly good model for the data. On the other hand, the LG of CAR3 was relatively small and the mean was around 1.5. These values indicate that CAR may not be a proper model for the data. In fact, when checking the values of hyperparameters specified for each data set, the ratio of local to global variance (σ/λ) were 10, 5 and 1, respectively. That is, in CAR3 the local variance was much smaller than the global variance. In this case, the choice of a simpler model such as Poisson gamma may be necessary. In summary, the estimated LGs showed positive evidence for the existence of local correlation for data from CAR1 and CAR2 but not for the data from CAR3 which is closer to the Poisson gamma model. We observed that LG does correctly imply the better model. Next, we used the approximations to Bayes factor for further confirmation.

5.2. Comparison of LG and Bayes factor

For each replication we used the modified harmonic mean of likelihoods, Schwarz criterion, and double Laplace–Metropolis method described in Section 3 for approximations. When comparing the behaviour of LG with various approximations to Bayes factor (BF), we found that LG does agree with Bayes factor and favours the same model. The columns in Table II were the percentages of data whose LG and BF satisfied LG $\geqslant 3.64$ and $\log(\mathrm{BF}) \leqslant -3$. The number 3.64 is simply the average number of neighbours and the cut-off number 3 for log BF is chosen because it has been suggested as a value of having positive evidence against H_0 [9, 31]. These numbers indicated that both LG and BF favoured CAR over the PG model. The numbers were around 80 per cent for CAR1 and CAR2, showing strong agreement. For CAR3, more than 95 per cent of LG were less than 3.64 and more than 80 per cent of log BF were greater than -5; both provided support for the PG model. The consistency between LG and BF provides strong support for using either one or both for model selection. Another way to demonstrate that LG agrees with Bayes factor is to look at the scatter plot. Figure 2 was the plot of LG versus approximated Bayes factor. The patterns

Table II. The agreement of LG and three approximations to Bayes factor for simulation data. L–M is the Laplace–Metropolis estimate, SC is the Schwarz criterion, and harm is the harmonic mean estimate. Numbers are shown in percentages.

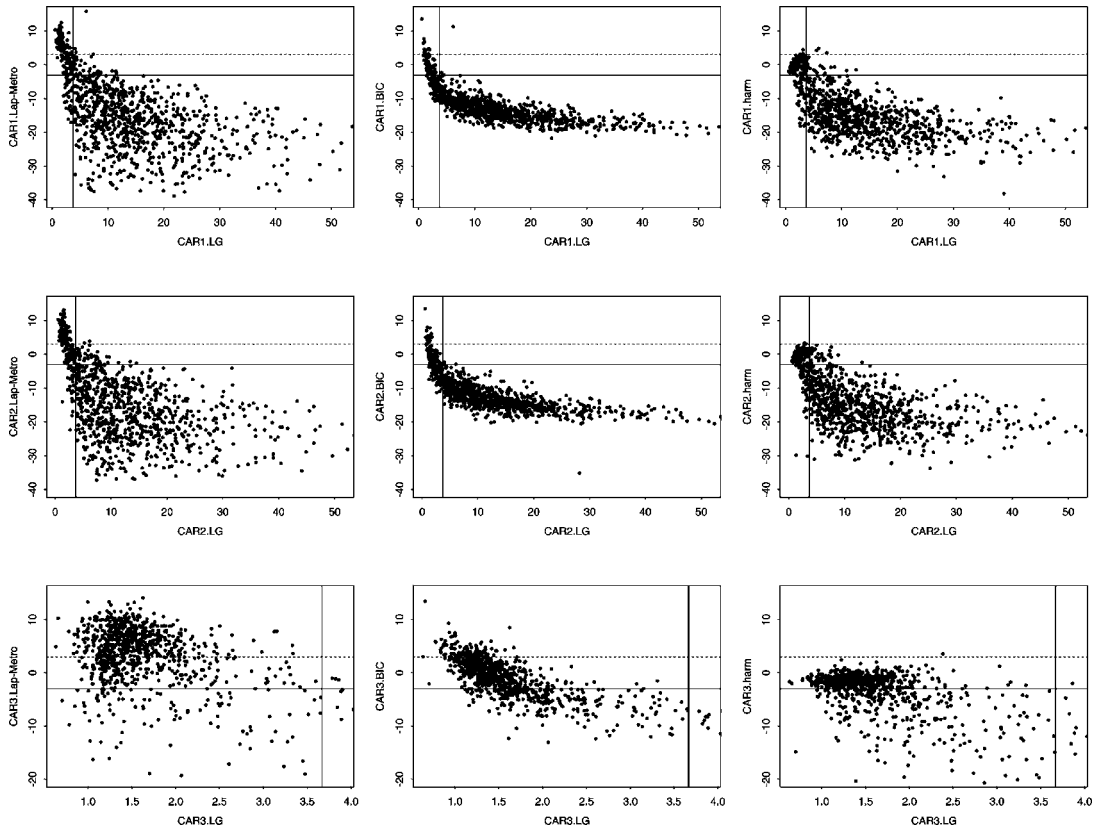|	LG $\geqslant 3.64$ $\log(\text{L–M}) \leqslant -3$	LG $\geqslant 3.64$ $\log(\text{SC}) \leqslant -3$	LG $\geqslant 3.64$ $\log(\text{harm}) \leqslant -3$
CAR1	80.49	84.23	82.00
CAR2	76.18	79.80	77.49
CAR3	1.10	1.70	1.60

Figure 2. Plots of three approximations to Bayes factor versus LG. The horizontal solid line is $\log \text{BF} = -3$ the horizontal dotted line is $\log \text{BF} = 3$ and the vertical line is LG = 3.64.

of points (LG,BF) were all similar. The horizontal lines in each graph were $\log(\text{BF}) = -3$ and $\log(\text{BF}) = 3$, and the vertical line in each graph was LG = 3.64. Most of the points (LG, $\log(\text{BF})$) lay in the two opposite-diagonal dimensions (the second and forth). Again, the two indices agreed with high probability and showed strong linear correlation.

6. DISCUSSION

In this paper we proposed to use the local to global variation (LG) as an indicator for spatial model selection, particularly when choosing between Poisson gamma and CAR models. We have illustrated its usage and applied it to Taiwan asthma mortality and Scotland lip cancer data. Both results imply that the Poisson gamma would fit the data better than CAR. We have also computed Bayes factor, a measure commonly used in Bayesian model selection and hypotheses testing, for the two data sets, respectively, to see if the conclusions confirm the finding based on LG. Both indicators agreed in favouring the Poisson gamma model. In addition, to illustrate the behaviour when data indeed contain spatial local correlation, we generated data from the CAR model and examined the choice for a better model using LG and Bayes factor. As the results show, the LG is indeed a good quantity for selection between the two competing models. Although Bayes factor may be commonly used in the setting of model selection, its computation can be complex, depending on the structure of model. On the other hand, an estimate of LG based on the posterior samples can be derived easily using the BUGS program. We recommend using LG because it is intuitive and also for its easy computation.

For the Taiwan asthma mortality, the Poisson gamma model fits better and thus it implies that there exists no significant local correlation among rates from neighbouring areas. Our results raise several points and questions for further investigation. The asthma mortality rates of the whole region are similar and thus the next step of research on identifying risk factors for individuals may focus on all the areas instead of some particular ones. The asthma mortality may contain deaths of patients having early-onset asthma or other factors including hypoxic seizures and a history of respiratory failure. It is possible that these two types of mortality may have different models. In that case, the autopsy or medical charts may help to distinguish these two types. Furthermore, although the current statistical model supports the independence among areas, it is not known whether the independence would disappear when taking into consideration the asthma prevalence and morbidity. In addition, to prevent deaths from asthma, an effective medical intervention should be available for all the areas under study. The officers from the health department may need to examine whether the medical resources are available and effective equally for the whole region.

For the Scotland lip cancer rates, we conclude that the Poisson gamma fits the data better than the CAR model once the explanatory variable related to exposure to sunlight is included in the model. That is, the observed spatial pattern of rates among areas is explained by the covariate. It is possible, however, that the sun exposure may not be the only explanatory variable [26,27]. Collecting other information for each area may be useful in identifying unusual areas. After that, further investigation of risk factors for people living there would be necessary. For instance, the individual's occupation and amount or time of exposure to sunlight may be possible risk factors. However, the assess of effect and interactions of factors remains important because various factors may act interactively or independently when developing the carcinogenesis [27]. Another issue regards the values of m_i in this study. When Breslow and Clayton [8] analysed this data, they decomposed $\log(m_i p_i)$ to $\log E_i$ and θ_i where E_i is the expected number of deaths for the ith area. The $\log E_i$ is similar to the sum of μ and $\log m_i$ in our model but not exactly the same because μ is random and E_i is constant. Since the information about m_i is not available, we derive m_i based on the values of E_i in their paper with m_1 set to 100 000.

There are some points worth mentioning in the process of model fitting for the two data sets. First, there are other hierarchical models for these spatial data. For instance, instead of assuming p_i follows a gamma distribution when no covariate is included, its logarithm, $\log p_i$, can follow

a normal distribution as an uncorrelated random effects model. An advantage of this model would be the easy and direct comparison with the CAR assuming correlated random effects model. Other forms of distribution can also serve as alternatives. However, it should be kept in mind that these models should interpret the data properly. Second, the prior distributions for hyperparameters are either non-informative or based on national estimates of mortality or morbidity rates. When one has other belief in these parameters, different priors can be used and the computation should proceed in the same manner. For instance, the Taiwan asthma data have been examined previously but different priors were used [6]. In fact, our conclusion is consistent with theirs. This implies that the results are robust to these chosen prior distributions. Third, all the computations covered in this paper were based on the posterior samples generated from the BUGS program. BUGS is a friendly software and can cover a variety of models beyond the two spatial ones considered here. However, the convergence of the posterior outputs has to be monitored by the user and a companion software CODA of BUGS can serve the purpose [42]. Fourth, it is a custom in BUGS that the prior distribution is assumed for the inverse of variance rather than the variance. However, when one has an informative prior for the variance, it should not matter which kind is used as long as the transformation works through. In both the asthma and lip cancer data, the gamma prior for the inverse of variance can be transformed to inverse-gamma distribution for variance. It can be seen after transformation that these inverse-gamma are proper and have large means. This again supports our prepossession of a large variance. Fifth, the modified harmonic mean estimate discussed at the end of Section 2 uses $f(y|\mu, \theta)$ as the likelihood function and π^* as the modified prior information. It is due to the difficulty in integration over μ and θ to get $L_1(c, \sigma, \lambda)$. We have called it a mathematically equivalent estimate. For our computations, several plots of functions $f(y|\mu, \theta)$ and $\pi^*(\mu, \sigma)$ (figures not shown here) appeared to have flatter π^* than $f(y|\mu, \theta)$, which implied dominant information contained in $f(y|\mu, \theta)$. Consequently, using mainly $f(y|\mu, \theta)$ in the harmonic mean estimate should be a reliable and acceptable choice. Finally, as an alternative to examine whether the spatial correlation is important, one may prefer to include both correlated and uncorrelated terms in one model and compare their contribution. For instance, to add a random term in $\log p_i$ for each area under CAR model may be considered. In that case, the number of parameters will increase, and the computation and convergence may not be achieved easily. A better approach would be to replace the correlated random effects with a fixed structure for correlation under the type of Poisson-gamma model. The model construction and fitting would be an interesting topic and more research is expected.

APPENDIX A

Letting $\phi = (c, \sigma, \lambda)$, the marginal probabilities $\Pr(y|H_0)$ and $\Pr(y|H_1)$ mentioned in Section 3.2 are formulated as

$$\Pr(y|H_0) = \int \prod_{i=1}^{N} \left[\int f(y_i|\tau_i) f(\tau_i|\alpha, \beta) \, d\tau_i \right] \pi_0(\alpha, \beta) \, d(\alpha, \beta)$$

$$\Pr(y|H_1) = \int \left(\int f(y|\mu, \theta) \pi(\mu, \theta|\phi) \, d(\mu, \theta) \right) \pi_1(\phi) \, d\phi \qquad (A1)$$

where $y = (y_1, \ldots, y_N)$, $\theta = (\theta_1, \ldots, \theta_N)$, and π_0, π_1 are priors on hyperparameters. We now estimate the likelihood value L_1 using the posterior samples of (c, σ, λ) and Laplace's method. Define

$L(\phi) = L_1(c, \sigma, \lambda)$, $f_1 = \max f(y|\mu, \theta)$, and $\pi_1(\phi) = \max \pi(\mu, \theta|\phi)$ for notational convenience, then

$$\max L(\phi) \cong \max_{\phi_i} l(\phi_i) = \max_{\phi_i} \int f(y|\mu, \theta) f(\mu, \theta|\phi_i) \mathrm{d}(\mu, \theta) \tag{A2}$$

$$\approx \max_i (2\pi)^{N/2} |\Sigma_1|^{1/2} f_1 \pi_1(\phi_i) \tag{A3}$$

with $\phi_i = (c_i, \sigma_i, \lambda_i)$ being the posterior samples, $(\hat{\mu}, \hat{\theta})$ the MLE of (μ, θ), and Σ_1 the observed Fisher information matrix of (μ, θ). The first maximum over ϕ was obtained by evaluating L at various posterior samples of ϕ. One reminder for practical users is that we have used in equation (A3) the posterior mode of (μ, θ) instead of the mode conditioning on (y, c, σ, λ) and assumed the two are close enough. The estimate of $\Pr(y|\mathrm{CAR})$ is now complete after substituting the values of equation (A3) into equation (1). For the parameters at two different stages, the Metropolis estimator of the mode was used in the Laplace approximation. This estimate is a variant of the compound Laplace–Metropolis method [14] where it is possible to evaluate analytically the mode in one of the two stages but not the other, while the analytical modes are not available for both stages under CAR model.

APPENDIX B

To evaluate the modified harmonic mean estimate, we write out the integration of $\Pr(y|H_1)$, letting $\phi = (c, \sigma, \lambda)$, and interchange the integration signs under regularity conditions to obtain

$$\Pr(y|H_1) = \int \left(\int f(y|\mu, \theta) \pi(\mu, \theta|\phi) \, \mathrm{d}(\mu, \theta) \right) \pi_1(\phi) \, \mathrm{d}\phi$$

$$= \int f(y|\mu, \theta) \left(\int \pi(\mu, \theta|\phi) \pi_1(\phi) \, \mathrm{d}\phi \right) \mathrm{d}(\mu, \theta)$$

$$= \int f(y|\mu, \theta) \pi^*(\mu, \theta) \, \mathrm{d}(\mu, \theta) \tag{B1}$$

where π^* is the 'marginal prior' after integrating out $\phi = (c, \sigma, \lambda)$. In particular, π^* can be derived analytically when $(\mu|c, \sigma)$ and $(\theta|\lambda)$ are from normal, c from uniform, and σ and λ from gamma distributions, as assumed in the applications.

APPENDIX C

Based on the result in Besag and Kooperberg [30], we generated $x_1, \ldots, x_{N-1}$ from a multivariate normal with zero mean and a variance matrix equal to the inverse of Q^* where Q^* was the upper left $(N - 1)$ by $(N - 1)$ matrix of Q. Next, defining

$$\theta_N = \sum_{i=1}^{N-1} x_i/N \quad \text{and} \quad \theta_i = x_i - \theta_N \quad \text{for } 1 \leqslant i \leqslant N - 1$$

these θ_i then were used together with μ to generate y_i, the number of diseased. The matrix $Q = D \times (I - C)$ contained the information about neighbours where the matrices D, I and C were

defined in Section 2. The matrix $I - C$ is

$$
\begin{bmatrix}
1 & -1/5 & 0 & -1/5 & -1/5 & -1/5 & 0 & 0 & 0 & 0 & -1/5 \\
-1/4 & 1 & 0 & -1/4 & 0 & 0 & 0 & 0 & -1/4 & 0 & -1/4 \\
0 & 0 & 1 & -1/3 & 0 & 0 & -1/3 & 0 & -1/3 & 0 & 0 \\
-1/6 & -1/6 & -1/6 & 1 & -1/6 & 0 & -1/6 & 0 & -1/6 & 0 & 0 \\
-1/4 & 0 & 0 & -1/4 & 1 & -1/4 & -1/4 & 0 & 0 & 0 & 0 \\
-1/3 & 0 & 0 & 0 & -1/3 & 1 & 0 & 0 & 0 & 0 & -1/3 \\
0 & 0 & -1/4 & -1/4 & -1/4 & 0 & 1 & -1/4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\
0 & -1/5 & -1/5 & -1/5 & 0 & 0 & 0 & 0 & 1 & -1/5 & -1/5 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
-1/4 & -1/4 & 0 & 0 & 0 & -1/4 & 0 & 0 & -1/4 & 0 & 1
\end{bmatrix}
$$

where the number in the denominator is the number of neighbours.

REFERENCES

1. Tsutakawa RK. Mixed models for analyzing geographic variability in mortality rates. *Journal of the American Statistical Association* 1988; **83**:37–42.
2. Besag J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* 1974; **36**:192–236.
3. Bernardinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* 1992; **11**:983–1007.
4. Mollié A. Bayesian mapping of disease. In *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, 1996.
5. Richardson S, Monfort C, Green M, Draper G, Muirhead C. Spatial variation of natural radiation and childhood leukaemia incidence in Great Britain. *Statistics in Medicine* 1995; **14**:2487–2501.
6. Tzeng JY, Hsiao CK, Chen CJ. Spatial model selection using Bayes factor and ratio of variabilities for asthma mortality data. *Chinese Journal of Public Health* (*Taipei*) 1998; **17**:158–169.
7. Gilks WR, Richardson R, Spiegelhalter DJ (eds). *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, 1996.
8. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**:9–25.
9. Kass RE, Raftery AD. Bayes factors. *Journal of the American Statistical Association* 1995; **90**: 773–795.
10. Carlin BP, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 1995; **57**:473–484.
11. Spiegelhalter DJ, Thomas A, Best NG, Gilks, WR. BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50. MRC Biostatistics Unit: Cambridge, 1995.
12. Raftery AE. Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice*. Chapman and Hall: London, 1996.
13. Raftery AE. Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika* 1996; **83**:251–266.
14. Lewis SM, Raftery AE. Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association* 1997; **92**:648–655.
15. Weiss K, Wagener D. Changing patterns of asthma mortality. *The Journal of the American Medical Association* 1990; **264**:1683–1687.
16. Yang CY, Lin MC, Hwang KC. Childhood asthma and the indoor environment in a subtropical area. *Chest* 1998; **114**:393–397.

17. Packe G, Archer P, Ayres J. Asthma and the weather. *Lancet* 1983; **2**:281–283.
18. Barbee RA, Murphy S. The natural history of asthma. *Journal of Allergy and Clinical Immunology* 1998; **102**:S65–72.
19. 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**:179–188.
20. Weitzman JB, Kanarek NF, Smialek JE. Medical examiner asthma death autopsies: a distinct subgroup of asthma deaths with implications for public health preventive strategies. *Archives of Pathology and Laboratory Medicine* 1998; **122**:691–699.
21. Wobig EK, Rosen P. Death from asthma: rare but real. *Journal of Emergency Medicine* 1996; **14**:233–240.
22. Sherman JM, Capen CL. The Red Alert Program for life-threatening asthma. *Pediatrics* 1997; **100**:187–191.
23. Cote J, Cartier A, Robichaud P, Boutin H, Malo JL, Rouleau M, Fillion A, Lavallee M, Krusky M, Boulet LP. Influence on asthma morbidity of asthma education programs based on self-management plans following treatment optimization. *American Journal of Respiratory and Critical Care Medicine* 1997; **155**:1509–1514.
24. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987; **43**:671–678.
25. Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, Powell J. *Cancer Incidence in Five Continents, Volume VI*. IARC Scientific Publication No. 120. International Agency for Research on Cancer: Lyon, 1992.
26. Douglass CW, Gammon MD. Reassessing the epidemiology of lip cancer. *Oral Surgery* 1984; **57**:631–642.
27. de Visscher JG, van der Waal I. Etiology of cancer of the lip. A review. *International Journal of Oral and Maxillofacial Surgery* 1998; **27**:199–203.
28. Besag J, York J, Mollié A. Bayesian image restoration, with applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 1991; **43**:1–59.
29. Richardson S. Statistical methods for geographical correlation studies. In *Geographical and Environmental Epidemiology*, Elliott P, Cuzick J, English D, Stern R (eds). Oxford University Press: London, 1992; 182–204.
30. Besag J, Kooperberg C. On conditional and intrinsic autoregressions. *Biometrika* 1995; **82**:733–746.
31. Jeffrey H. *Theory of Probability*, 3rd edn. Oxford University Press: Oxford, U.K., 1961.
32. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; **85**:398–409.
33. Geman S, Geman D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; **6**:721–741.
34. Kass RE, Tierney L, Kadane JB. The validity of posterior asymptotic expansions based on Laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics*, Geisser S, Hodges JS, Press SJ, Zellner A (eds). North-Holland: New York, 1990.
35. Hsiao CK. Approximate Bayes factors when a mode occurs on the boundary. *Journal of the American Statistical Association* 1997; **92**:656–663.
36. Tierney L, Kass RE, Kadane JB. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association* 1989; **84**:710–716.
37. Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978; **6**:461–464.
38. Kass RE, Vaidyanathan SK. Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society, Series B* 1992; **54**:129–144.
39. Raftery AE. A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society, Series B* 1986; **48**:249–250.
40. Newton MA, Raftery AE. Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B* 1994; **56**:3–48.
41. Department of Health, the Executive Yuan. *Health and Vital Statistics 1991, 2. Vital Statistics, Republic of China*. Department of Health, the Executive Yuan: Taipei, 1992.
42. Best NG, Cowles MK, Vines SK. *CODA Manual version 0.30*. MRC Biostatistics Unit: Cambridge, U.K., 1995.