Giudici, Knorr-Held, Rasser:

Modelling categorical covariates in Bayesian disease mapping by partition structures

Sonderforschungsbereich 386, Paper 152 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

Modelling categorical covariates in Bayesian disease mapping by partition structures

Paolo Giudici

Dipartimento di Economia Politica e Metodi Quantitativi

University of Pavia

Via San Felice 5, I-27100 Pavia

Italy

Leonhard Knorr–Held and Günter Raßer

Institute of Statistics

University of Munich

Ludwigstr. 33, D-80539 München

Germany

April 23, 1999

Abstract

We consider the problem of mapping the risk from a disease using a series of regional counts of observed and expected cases, and information on potential risk factors. To analyse this problem from a Bayesian viewpoint we propose a methodology, which extends a spatial partition model by including categorical covariate information. Such an extension allows to detect clusters in the residual variation, reflecting further, possibly unobserved, covariates. The methodology is implemented by means of reversible jump Markov chain Monte Carlo sampling. An application is presented, in order to illustrate and compare our proposed extensions with a purely spatial partition model. Here we analyse a well-known dataset on lip cancer incidence in Scotland.

1 Introduction

Statistical methods for analysing the geographical variation of disease rates have received increasing interest in the recent literature. Bayesian approaches to this problem typically introduce parameters, which may or may not have a spatial structure a priori, and can be seen as surrogates for unobserved or latent covariates. The inclusion of *observed* covariates in such analyses has been recognized as very crucial, see e.g. Clayton and Bernardinelli (1992). Indeed, the ultimate goal of statistical disease mapping would be to include all relevant causal factors and, thus, remove the need for covariate surrogates (Knorr-Held and Besag, 1998). However, in practice, this is not possible. Furthermore, if available, often such information is either incomplete or very imprecise so it becomes very important to insert such information properly into the statistical model.

The aim of the present paper is to present flexible methodologies suited to incorporate information on categorical covariates into Bayesian analysis of disease maps. We propose to combine the data not only into spatial clusters, but also into clusters according to the

observed levels of each potential explanatory factor. Three different models are introduced, taking into account the order of the covariate categories in more or less restrictive ways.

Our specification is a natural generalization of an approach proposed in Knorr-Held and Rasser (1999a) to model spatial variation in disease maps. Basically their model assumes that the area considered can be partitioned into several clusters, that is, sets of contiguous regions, where each cluster has constant relative risk. Our aim is to adjust the spatial risk surface for categorical covariates. We compare the risk estimates with and without inclusion of such covariates.

The plan of the paper is as follows. Section 2 introduces the Scotland lip cancer dataset with emphasis on the relevant details for our application. Section 3 specifies the proposed model. Section 4 gives some details on the implementation. Section 5 describes results from several analyses of the Scotland dataset with and without a relevant covariate. We close with some comments and possible extensions in Section 6.

2 The dataset

The Scotland lip cancer example is well-known in the literature and has been analysed several times, since Clayton and Kaldor (1987) introduced it. However, a short summary of the data will be given in this section, as the implementation of our models will be described with reference to this dataset.

In Scotland there are $n = 56$ regions. For each region the number of observed and expected cases is given. These numbers are not presented here, but can be found, for instance, in Clayton and Bernardinelli (1993). Figure 1 shows a map of the standard mortality ratios (SMR), calculated for each region by simply dividing the number of observed cases by the number of expected cases. Note that the SMR's vary between 0 and 6.52.

Besides this basic information, for each region the percentage of population employed in agriculture, fishing and forestry is given. This covariate (*agr* for brevity) has $m = 6$ categories with levels L_l, $l = 1, \ldots, m$. In Table 1 for each category such percentage (L_l) is shown, along with the SMR's calculated by summing up the observed (Y_l) and expected (E_l) cases of all regions within each category l. The number of regions in one category, referred to as the size of the category, is also presented.

Category	L_l	Y_l	E_l	SMR	Size
1	0	48	122.28	0.39	5
2	1	85	157.55	0.54	11
3	7	126	91.11	1.38	14
4	10	120	76.81	1.56	12
5	16	129	77.14	1.67	10
6	24	28	11.12	2.52	4

Table 1: Categories for the Scotland lip cancer data.

From Table 1, note that the categories are not homogeneous with respect to the expected cases E_l nor to the size and, furthermore, the regions within the categories are not geographically contiguous, as can be seen from Figure 1, which also displays the geographical variation of the SMR's.

The covariate *agr* is suspected to be related to sunlight exposure, a known risk factor for lip cancer, so with higher values of L_l, that is higher category l in our notation, the relative risk of lip cancer is also supposed to be higher. The crude SMR's within the six categories in Table 1 do in fact support this. Incidentally, the values L_l seem to be not the exact values of *agr* for each district, but are chosen as approximate centers of underlying intervals < 0.5, $0.5 - 1.5$, $1.5 - 8.5$, $8.5 - 12.5$, $12.5 - 20$, > 20, see Figure 5a) in Clayton, Bernardinelli and

Montomoli (1993).

In former analyses of this dataset (see e.g. Breslow and Clayton, 1993, Clayton, Bernardinelli and Montomoli, 1993, Breslow, Leroux and Platt, 1998), the influence of L_l is supposed to be linear on the log relative risk. This so-called exponential relative risk model is standard in epidemiology and has certain advantages in interpretation. However, it might also be interesting to allow for deviations from linearity. In our models we therefore refrain from this assumption and propose several less restrictive formulations.

3 The proposed model

Let y_i and e_i represent, respectively, the observed and the expected number of events (e.g. deaths or illnesses) in a certain region i, for $i = 1, \ldots, n$. The target of our investigation is the unknown relative risk λ_i in region i. Observations y_i are assumed to be realizations from a Poisson distribution with mean $e_i\lambda_i$. Conditional independence of responses $y = (y_1, \ldots, y_n)$, given $\lambda = (\lambda_1, \ldots, \lambda_n)$, leads to the likelihood function

$$L(y|\lambda) = \prod_{i=1}^{n} \frac{(e_i\lambda_i)^{y_i}}{y_i!} \exp(-e_i\lambda_i).$$

To include covariate information in the model, we propose a multiplicative decomposition of the relative risk λ_i as follows:

$$\lambda_i = \theta(i) \times \nu(i), \qquad i = 1, \ldots, n.$$

The above equation factorizes the relative risk into two components, which will be assumed to be independent; a *spatial* component $\theta(i)$, and a *covariate* component $\nu(i)$.

Following the Bayesian paradigm, we shall now describe the prior distribution for both components. Regarding the spatial component we apply the method of Knorr-Held and Rasser (1999a), which will be recalled briefly. The main assumption is that $\theta(i)$ is *constant*

over a set of one or more contiguous regions. This defines clusters C_j, $j = 1, \ldots, k$, a partition of regions with constant relative risk θ_j. More formally,

$$\theta(i) = \theta_j, \text{ for } i \in C_j.$$

The number of clusters k is treated as unknown with $k \in \{1, \ldots, n\}$. The clusters are defined by marking k randomly drawn regions as so–called cluster centers $z_1, \ldots, z_k$. Assigning each of the remaining regions to the "nearest" cluster center leads to a unique *cluster configuration*, given $Z_k = (z_1, \ldots, z_k)$. Here, the distance between two regions is defined as the minimal number of boundaries that have to be crossed to move from one to the other.

We now specify the prior distributions for the unknown parameters. For the number of clusters k we apply a uniform distribution on $\{1, \ldots, n\}$ in this paper. Other choices may also be appropriate, for example a geometric distribution, as thoroughly discussed in Knorr-Held and Rasser (1999b). For a given k, we assume that each vector of cluster centers Z_k has equal probability $Pr(Z_k|k) = \frac{(n-k)!}{n!}$. Components of $\Theta_k = (\theta_1, \ldots, \theta_k)$ are treated independently a priori, with $\log(\theta_j)$ normally distributed with mean μ and variance σ^2, that is:

$$\theta_j \sim \text{LN}(\mu, \sigma^2).$$

Both hyperparameters are also random. For μ, we take an uniform prior on the whole real line and for σ^2 a highly dispersed, but proper, inverse gamma distribution, that is, $\sigma^2 \sim \text{IG}(a, b)$ with suitable values for the hyperparameters a and b.

Without including any covariate this will be our model 1, giving in some sense a reference for comparing the estimates of the relative risks λ.

We now introduce our proposed prior model for the covariate component. We assume to consider the effects of one categorical covariate, which is observed in m distinct levels. This

assumption is only convenient, and does not imply any loss of generality, as, if more than one covariate is available, we can simply consider the observed joint levels.

Let $L_1, \ldots, L_m$ be the observed levels in m categories of the covariate. Each level corresponds to a certain covariate effect. In a first step, we estimate the influences $\nu_1, \ldots, \nu_m$ of all categories separately. In this model, which will be referred to as model 2, the dimension of $N_m = (\nu_1, \ldots, \nu_m)$ is fixed and no further restrictions on N_m are imposed. Therefore this model allows the estimated influences of two or more categories to be nearly equal. If so, it might be interesting to test whether the effects of all m categories are really different. Generalizing this model, we allow to combine two or more "similar" categories into one group, in which the contained categories have the same effect. This will be done in two different ways, which will now be described in more detail.

First, we consider a collection of partial exchangeability structures on N_m, each corresponding to a partition of the regions, as induced by the observed levels of the covariate. Obviously, the partition with most groups will be that with exactly m groups, and the simplest will be that with only one group, corresponding to complete exchangeability. In the latter case, the covariate will clearly have no effect. It may be the case that only one of such partitions is to be considered. This occurs when strong prior opinion on how the covariate may affect the disease pattern is known. In this case inference can proceed conditionally on such a partition, as in model 2.

More generally, when such a prior information is not available or weak, consider the index set $\{1, \ldots, m\}$ and assume it can be partitioned in G different ways. Let g be one of such partitions, including t_g subsets $S_1(g), \ldots, S_{t_g}(g)$. Thus, a partition g, for $g = 1, \ldots, G$, reflects a clustering pattern of the sites which depends on a common level grouping of a covariate. To ease the notation, let $t = t(g)$ be the number of groups of one partition. Let $L_1, \ldots, L_t$ be the corresponding groupings of the observed levels of a covariate, say X. We

assume that,

$$\nu(i) = \nu_h \text{ for } x_i = L_h,\ i = 1, \ldots, n.$$

This model, which will be referred to as model 3, is the most general of those presented in this paper. Note that any order of the categories of the covariate, as in the Scotland example, is totally ignored. It might be more appropriate to choose partitions respecting this order. One possibility would be to allow only *ordered* partitions consisting of subsets $S_h,\ h = 1, \ldots, t$, containing subsequent indices out of $1, \ldots, m$. This will be our proposed model 4, which makes just slightly stronger assumptions than model 3.

Since g is treated as unknown, we need to assign to it a prior probability, according to a probability mass function $p(g)$. In this paper we let $p(g)$ be a discrete uniform distribution on $\{1, \ldots, G\}$.

The total number G of possible partitions will be much smaller when only ordered partitions are considered. In Table 2 those numbers are compared for the situation of the Scotland data. Here the number of groups t_g ranges between 1 and 6.

t_g	#g (unordered)	#g (ordered)
1	1	1
2	31	5
3	90	10
4	65	10
5	15	5
6	1	1
G	203	32

Table 2: Number of possible partitions in model 3 and 4.

Note, that the prior distribution for t_g, which is implicitly modelled with the prior for g, is

not uniform. With this choice of the prior distribution for g, the number of partitions, where two arbitrary categories are within the same group is 52 in the unordered case. Therefore, this prior probability is approximately 0.26. Whereas, in the ordered case, the prior probability that two adjacent categories are within the same group is exactly 0.5.

We now need to specify a prior model for the covariate effects $N_t = (\nu_1, \ldots, \nu_t)$. It is possible to proceed similarly as for the spatial components, but for the covariate effects to be identifiable, the influences have to be centered. We therefore take the $\nu_h,\ h = 1, \ldots, t$ independent a priori, with $\log(\nu_h)$ normally distributed with mean 0 and variance τ^2, that is:

$$\nu_h \sim \mathrm{LN}(0, \tau^2), \quad h = 1, \ldots, t.$$

Again, we assume $\tau^2 \sim \mathrm{IG}(c, d)$ with positive hyperparameters c and d.

In both models 3 and 4, the dimension t of N_t is variable and therefore a reversible jump MCMC algorithm (Green, 1995) is required, allowing to sample from a varying dimension space depending on the considered partition of the covariate. This will smooth the effects of the covariate, as the outcome of model averaging over the class of all possible partitions.

We finally remark that the idea of a Bayesian partition model for the analysis of spatial covariates has been introduced in Borroni and Giudici (1996), yet in the context of extending the approach of Besag, York and Mollie (1991). For a more general discussion on Bayesian partition models see also Consonni and Veronese (1995) and Green (1995).

4 The proposed reversible jump MCMC methodology

We first give a brief description of our reversible jump MCMC (see Green, 1995) strategy for sampling from the posterior distribution. In each iteration of the algorithm, one of the following nine moves is proposed. The first five concern the spatial component of the model.

The next three involve the covariate component. Finally, the last one is related to updating all the hyperparameters of the model.

Birth: The number of spatial clusters is increased by introducing an additional cluster center.

Death: The number of spatial clusters is decreased by deleting one of the cluster centers.

Shift: One of the spatial cluster centers is moved.

Switch: The positions of two cluster centers in Z_k are switched.

Spatial Height: The parameters $\theta_j, j = 1, \ldots, k$, are changed.

Merge: The number of groups of the partition induced by the covariate is decreased, by merging two groups which were previously disjoint.

Split: The number of groups of the partition induced by the covariate is increased, by splitting up one previous group into two new groups.

Covariate Height: The parameters $\nu_h, h = 1, \ldots, t$, are changed.

Hyper: The values of the hyperparameters μ, σ^2 and τ^2 are changed.

In our model 1, only the first five moves and the last one (without τ^2) are used. For a fixed partition, as in model 2, the split and merge moves are omitted. Only in our models 3 and 4 all nine moves are included.

For details on the implementation of the first five moves see Knorr-Held and Rasser (1999a). The same paper contains a description of the strategy for the hyperparameters update for μ and σ^2. We have adopted the same treatment for τ^2. Concerning the additional

covariate height step, the proposal distribution is chosen in analogy to the spatial height move.

The merge and split moves, which only concern the covariate part of the algorithm, and are the main implementational novelty of our approach, will now be described in detail. For reasons of easier explanation and illustration, the moves will be introduced in terms of the Scotland data example as described in Section 2. We remark that it is straightforward to apply the algorithm to any other data.

We propose to combine the categories in groups of variable sizes. The number of groups is also variable. Therefore the minimum number of groups is 1 and the maximum is 6. First we will focus on the general case of model 3, ignoring the order of the categories. The necessary changes leading to model 4 will be introduced later.

The current grouping is represented in a partition g. For example, let there be a partition with $t = 3$ groups $g = (S_1, S_2, S_3) = (\{1, 4, 5\}, \{2\}, \{3, 6\})$. Each group is supposed to have a certain effect, given in the corresponding vector $N_t = (\nu_1, \nu_2, \nu_3)$.

Consider now a merge step attempting to combine two existing groups. Here two groups are chosen randomly, so we have $\binom{t}{2}$ possibilities for this choice, each with equal probability $\frac{2}{t(t-1)}$. If we have chosen to merge the groups S_2 and S_3, this will lead to a new partition $\tilde{g} = (\{1, 4, 5\}, \{2, 3, 6\})$. Simultaneously, the number of groups is decreased from $t = 3$ to $\tilde{t} = 2$. Now all categories in the new group $\tilde{S}_2 = \{2, 3, 6\}$ are assumed to have the same effect. Therefore, the old values ν_2 and ν_3 are replaced with a new value $\tilde{\nu}_2$, drawn randomly from a gamma distribution

$$\tilde{\nu}_2 \sim G\left(y + \frac{\eta^2}{\kappa^2}, e + \frac{\eta}{\kappa^2}\right),$$

where

$$\eta = \exp(0.5\tau^2) \quad \text{and} \quad \kappa^2 = \exp(\tau^2)(\exp(\tau^2) - 1)$$

denote the mean and the variance of the lognormal prior distribution for the covariate-specific risks and

$$y = \sum_{i:\tilde{\nu}(i)=\tilde{\nu}_2} y_i = \sum_{j\in\{2,3,6\}} Y_j \quad \text{and} \quad e = \sum_{i:\tilde{\nu}(i)=\tilde{\nu}_2} e_i \cdot \theta(i).$$

The number of observed cases of all categories belonging to the new subset can simply be taken from Table 1, while for calculating e the expected cases for each region belonging to the new group $\tilde{S}_2$ are needed.

This proposal is built in analogy to those used in Knorr-Held and Rasser (1999a) for the birth and death moves. It approximates the fixed dimension full conditional of $\tilde{\nu}_2$ so the corresponding acceptance rates become rather high and any tuning of the algorithm becomes unnecessary. Note that $\theta(i)$ is the current height of the spatial cluster region i is assigned to and, therefore, our covariate proposal depends on the current values of the spatial component. Correspondingly, the values $\nu(i)$ enter in the proposal of the birth, death and spatial height move.

The reverse move would be to split up one group into two of variable size. To continue with the example, performing the exact reverse transition to the merge move just described, let $\tilde{g} = (\tilde{S}_1, \tilde{S}_2) = (\{1,4,5\},\{2,3,6\})$ and $\tilde{t} = 2$. First of all only those groups with a size larger than 1 are considered to avoid empty groups. Depending on the current partition $\tilde{g}$, there are $k(\tilde{g})$ groups with more than one category in it, here $k(\tilde{g}) = 2$. A random variable uniformly distributed on $\{1,\ldots,k(\tilde{g})\}$ determines the group which will be split up, say $\tilde{S}_2 = \{2,3,6\}$ of size $s = 3$. Depending on the size s there is a total number r of possible ways to split up this group as given in Table 3. One of them is drawn randomly with probability $\frac{1}{r}$. The sizes of the two new groups are just a side effect in this implementation. One part of the categories is kept in the new group S_2, say $S_2 = \{2\}$, and the other categories are placed in a new group $S_{\tilde{t}+1} = S_3 = \{3,6\}$. Now, the new partition $g = (\{1,4,5\},\{2\},\{3,6\})$ has $t = 3$ groups. For both new groups S_2 and S_3 new values ν_2 and ν_3 are proposed, again

s	r (unordered)	r (ordered)
2	1	1
3	3	2
4	7	3
5	15	4
6	31	5

Table 3: Number of possible ways r to split up a group of size s.

from a gamma distribution as in the merge move with y and e changed correspondingly.

To complete details of our implementation, we need to calculate the acceptance probabilities of the two changing-dimension moves concerning the covariate effects. Let $\mathcal{L}$ indicate the likelihood-ratio, $\mathcal{P}$ the proposal-ratio and $\mathcal{A}$ the prior-ratio. The acceptance probabilities for both moves are

$$\alpha = \min\{1,\ \mathcal{L} \cdot \mathcal{A} \cdot \mathcal{P}\}.$$

Let p denote the density of the lognormal prior distribution for the effects and q the density of the gamma proposal distribution as introduced above.

According to the model described in Section 3, the prior ratio for the split move, the transition from $\tilde{t} = t - 1$ to t turns out to be

$$\mathcal{A} = \frac{p(\nu_2)p(\nu_3)}{p(\tilde{\nu}_2)}$$

and the proposal ratio is

$$\mathcal{P} = \frac{q(\tilde{\nu}_2)}{q(\nu_2)q(\nu_3)} \cdot \frac{2\ k(\tilde{g})\ r}{t\ (t-1)}.$$

For a merge move, decreasing the number of groups form t to $\tilde{t} = t - 1$, all terms have to be inverted.

The concept of both moves can be easily transmitted to our model 4. To take into account the order of the categories of the covariate, just two small changes are necessary. First, in a merge move, two adjacent groups have to be selected. Therefore, we now have $t-1$ possibilities for this choice. In the reverse split move the number r of possible splits changes according to Table 3. Also, in a split move, it is no longer possible to add one new group $S_{\tilde{t}+1}$, but both new groups have to be neighbors, so that, possibly, some indices have to be changed. In our example, this is not necessary, as the groups S_2 and $S_3 = S_{\tilde{t}+1}$ are indeed next to each other. The acceptance probabilities just slightly change with the same prior ratio as above and the proposal ratio

$$\mathcal{P} = \frac{q(\tilde{\nu}_2)}{q(\nu_2)q(\nu_3)} \cdot \frac{k(\tilde{g})\ r}{t-1}.$$

for the split move and $\mathcal{P}^{-1}$ for the reverse merge move. These changes insure that the order of the categories is maintained, as long as the initial state of the Markov chain is chosen correctly. We use an initial partition $g^{(0)} = (\{1\},\{2\},\{3\},\{4\},\{5\},\{6\})$, so that $t^{(0)} = 6$, which is suitable both for model 3 and 4.

Finally, we remark, that for the extreme case $t = 1$ a merge move is not possible, as well as a split move for $t = m$. Therefore, a further factor, omitted here for easier presentation, may appear in the proposal ratios, depending on the implementation of those moves.

5 Application

In this section we re-analyse the Scotland lip cancer dataset, using the different model formulations described in Section 3. All results are based on $20,000,000$ iterations, following a burn-in period of $1,000,000$ iterations. In order to calculate posterior quantities of interest, we have stored every 2000th iteration which gives a total of $10,000$ samples. The prior hyperparameters for the spatial component are held fixed at $a = 1$ and $b = 0.01$. The

corresponding values for the covariate component are chosen slightly more informative with values $c = 5$ and $d = 0.5$.

We shall produce two main inferences: the first one is a structural one, and concerns the detection of the number of clusters in the data, both at the spatial and at the covariate level. The second inference is more specific and concerns the Bayesian estimate of the spatial variation of the disease risk, possibly adjusted for the effect of the covariate. As a Bayesian estimator of disease incidence we shall compare the posterior median of the relative risks λ in the different model formulations.

We first present, as a diagnostic on the performance of the MCMC method, the cumulative occupancy fractions for both the number of spatial clusters and the number of covariate groups from model 4. From Figure 2, note the good stability of the output, especially for the covariate part, which is indeed simpler for the application considered. The corresponding plots for the other models look very similar.

Consider now structural learning. Figure 3 gives the posterior probability of the parameter k for the four models. In model 1, there is a mode around $k = 15$ cluster centers. The inclusion of the covariate seems to generally increase the uncertainty about k. The posterior distribution is shifted towards smaller values in model 2 and 4, with a mode around 10.

Figure 4 gives the histogram describing the prior and posterior probability of the number of groups induced by the covariate. It shows support for a number of groups between 2 and 4 both in the ordered and unordered case. Note that for model 4 the posterior variance is clearly smaller than the prior variance, while in model 3 the uncertainty about the number of groups has slightly increased.

Figure 5a) gives the posterior median estimates pattern for the purely spatial model 1. A look at the map identifies a large cluster of eleven districts in the north with estimated relative risk in the range of 3.0 to 3.5. As a side comment we remark here that the estimates for these

eleven districts, reported in Breslow and Clayton (1993) with a conditional autoregressive prior for the log relative risks, are in the range of 2.6 to 4.5, hence show considerably more variation. For a more thorough comparison of our piecewise constant model formulation with the more traditional Markov random field autoregressive prior see the application in Knorr-Held and Rasser (1999a).

Seven districts have estimates below 0.4. Six of them form a cluster around the Glasgow district, one is the Dundee district. There is strong evidence that the latter forms a cluster by itself, as the corresponding posterior probability is 0.96, compared to a prior probability of 0.35. The prior probability has been calculated by simulations as in Knorr-Held and Rasser (1999b). All other regions have much lower posterior probabilities of being alone in a cluster with a median value of 0.06.

We turn now to the extended analyses, in which the estimated relative risks are calibrated for the effect of the covariate *agr*. Some care must be taken in general in interpreting the results, as the effect of the covariate might be estimated overly conservative. The reason is that location might act as a confounder in our model formulation, because we assume that the residual variation has spatial structure. For a thorough discussion of this issue see Clayton, Bernardinelli and Montomoli (1993).

Figure 5b), c) and d) display the estimated spatial residual component, adjusted for the effect of the covariate from model 2, 3 and 4, respectively, while Figure 6 compares the corresponding estimates of the covariate components $\nu_h,\ h = 1, \ldots, 6$. Consider first model 2, the one with a separate covariate effect for each category. The estimates of the spatial component are now in the range of 0.5 to 3.0 and have less variation than in model 1. This is to be expected, as, ideally, the residual variation should vanish, if all relevant factors enter properly in the model. In practice, however, this is not to be expected, but the maps of spatial residual variation in Figure 5b) may provide clues pointing to further risk factors. In

particular, in the south-west of Scotland, the inclusion of the covariate seems to explain a lot of the spatial variation. Also, the Dundee district is now less distinct with an adjusted relative risk of 0.89 and a posterior probability of 0.23 for forming a cluster on its own. However, in the north and east of Scotland, the inclusion of the covariate does only partially explain the spatial variation from Figure 5a). The estimated effects of the covariate levels in Figure 6 show the expected shrinkage effect towards one, in comparison with the SMR's from Table 1.

We now turn to model 3, in which we allow the effects of different categories to be identical, ignoring the order in the category levels. Consider first Table 4, which gives the estimated posterior probabilities that two categories are within the same group, regardless of the the underlying partition. Recall from Section 3 that the corresponding *prior* probability is $52/203 \approx 0.26$. All posterior probabilities are larger with values between 0.34 and 0.79. Correspondingly, the estimated covariate effects in Figure 5 are close to one, and the estimated spatial pattern (Figure 5c)) has much similarity with the purely spatial analysis from model 1. Note however, that Table 4 indicates some structure with lower probabilities between any combination of the categories 1 or 2, with one of the remaining categories 3,4,5 or 6.

In model 4, we now try to fix the obvious deficiencies of model 3 by allowing only for partitions which maintain the order of the category levels. In Figure 5d), the estimated spatial pattern from model 4 can be seen to be rather similar to the one obtained from model 2. The covariate model identifies a change point between categories 2 and 3 with posterior probability of 0.92. In contrast, for all other adjacent categories, the corresponding probabilities are below 0.3. Among the different partitions, partition $\{1,2\},\{3,4,5,6\}$ has the highest posterior probability and, interestingly, all 10 subsequent partitions, ranked in the order of posterior probability, do also separate the category levels 2 and 3, see Table 5.

Category level	1	2	3	4	5	6
1	1.0	0.75	0.50	0.51	0.49	0.46
2		1.0	0.41	0.39	0.37	0.34
3			1.0	0.78	0.78	0.74
4				1.0	0.79	0.76
5					1.0	0.75
6						1.0

Table 4: The posterior probability for two covariate categories to be within the same group, model 3.

Correspondingly, the estimated effects are smooth versions of those obtained from model 2, retaining the change point between categories 2 and 3, see Figure 6. This suggests that the usual assumption of a log linear model for the effect of the covariate is not fully supported by the data.

Finally, Figure 7 compares the median estimates of the fitted log relative risks $\log(\lambda)$. The fit is generally rather similar, and this can also be observed from a look at the mean posterior deviance (Spiegelhalter *et al.*, 1998), which has only for model 1 a slightly lower value. Among the covariate models, some differences can be seen in model 2, where districts belonging to the highest category levels have larger fitted values.

6 Discussion

We have proposed a novel methodology to include information on categorical covariates into the analysis of spatially correlated disease data. Such methodology is based on the notion of partial exchangeable patterns of covariate effects, depending on the observed levels of such

Partition	Posterior Probability
$\{1,2\},\{3,4,5,6\}$	0.28
$\{1,2\},\{3,4,5\},\{6\}$	0.15
$\{1,2\},\{3\},\{4,5,6\}$	0.10
$\{1\},\{2\},\{3,4,5,6\}$	0.10
$\{1,2\},\{3,4\},\{5,6\}$	0.08
$\{1,2\},\{3\},\{4,5\},\{6\}$	0.05
$\{1,2\},\{3,4\},\{5\},\{6\}$	0.05
$\{1,2\},\{3\},\{4,5,6\}$	0.04
$\{1\},\{2\},\{3,4\},\{5,6\}$	0.04
$\{1\},\{2\},\{3,4,5\},\{6\}$	0.02
$\{1\},\{2\},\{3\},\{4,5,6\}$	0.02
$\{1,2,3\},\{4,5,6\}$	0.02

Table 5: The twelve most likely partitions in model 4.

covariate. Our methodology is Bayesian, and has been implemented by means of a reversible jump MCMC algorithm, which has been described in detail.

Our model can be extended in different ways. However, given the complexity of the computations, a great deal of care is needed in checking the correctness of the code, and the convergence of the algorithms. For example, one could consider continuous covariates. This would imply resorting to a linear model formulation for the covariate effect. For instance, if the covariate effect can be assumed to be linear on the log relative risk, an alternative model for ν may be:

$$\nu(i) = \exp(\alpha + x_i\beta)$$

with α and β distributed as a flat prior. However, this would introduce some idiosyncrasy

between the spatial and the covariate effect; one may need to adopt, instead of the spatial clustering approach adopted here, a linear smoothing formulation, as in Besag, York and Mollié (1991).

Several assumptions underlying our application are critical from an epidemiological point of view. For example, the fact that *agr* is only a error-prone proxy for sunlight exposure somewhat conflicts with the assumption that the covariate *agr* is assumed to be measured without any measurement error. Measurement error models for continuous covariates in disease mapping have been proposed in Bernardinelli *et al.* (1997) and might be useful here as well. Due to the categorical scale of the covariate, a misclassification model might also be considered.

Acknowledgments

This research was supported by the German Science Foundation (DFG), SFB 386. The first author acknowledges additional support from the EU TMR network ERB-FMRX-CT96-0096 on: Computational and statistical methods for the analysis of spatial data. The authors express thanks to Michel Smans for access to the coordinates of the Scotland districts.

References

Besag, J., York, J., and Mollié, A. (1991), Bayesian image restoration, with two applications in spatial statistics, *Ann. Inst. Stat. Math.* **43**, 1-59.

Bernardinelli, L., Pascutto, C., Best, N. G. and Gilks, W. R. (1997), Disease mapping with errors in covariates, *Statistics in Medicine* **16**, 741-752.

Borroni, C. and Giudici, G. (1996), Analysis of spatial data by Hierarchical Partition Models (In Italian). Proceedings of the XXXVIII conference of the Italian Statistical Society

vol. 2, pp. 161-167.

Breslow, N.E. and Clayton, D.G. (1993), Approximate inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association* **88**, 9-25.

Breslow, N., Leroux, B. and Platt, R. (1998), Approximate hierarchical modelling of discrete data in epidemiology, *Statistical Methods in Medical Research* **7**, 49-62.

Clayton, D. and Bernardinelli, L. (1993), Bayesian methods for mapping disease risks, in *Small Area Studies in Geographical and Environmental Epidemiology*, edited by J. Cuzick and P. Elliot, Oxford University Press, 205–220

Clayton, D., Bernardinelli, L. and Montomoli, C. (1993), Spatial correlation in ecological analysis, *International Journal of Epidemiology* **22**, 1193-1202.

Consonni, G. and Veronese, P. (1995), A Bayesian method for combining results from several binomial experiments, *Journal of the American Statistical Association* **90**, 935-

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**, 711-732.

Knorr-Held, L. and Besag, J. (1998). Modeling Risk from a Disease in Time and Space. *Statistics in Medicine* **17**, 2045-2060.

Knorr-Held, L. and Rasser, G. (1999a), Bayesian detection of clusters and discontinuities in disease maps, *Biometrics*, revised.

Knorr-Held, L. and Rasser, G. (1999b), Bayesian detection of clusters and discontinuities in disease maps: Simulations, Discussion paper No. 142, Sonderforschungsbereich 386 der Ludwig-Maximilians-University Munich.

Spiegelhalter, D. J., Best, N. G. and Carlin, B. P. (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. Technical Report, MRC Biostatistics Unit, Cambridge, UK.

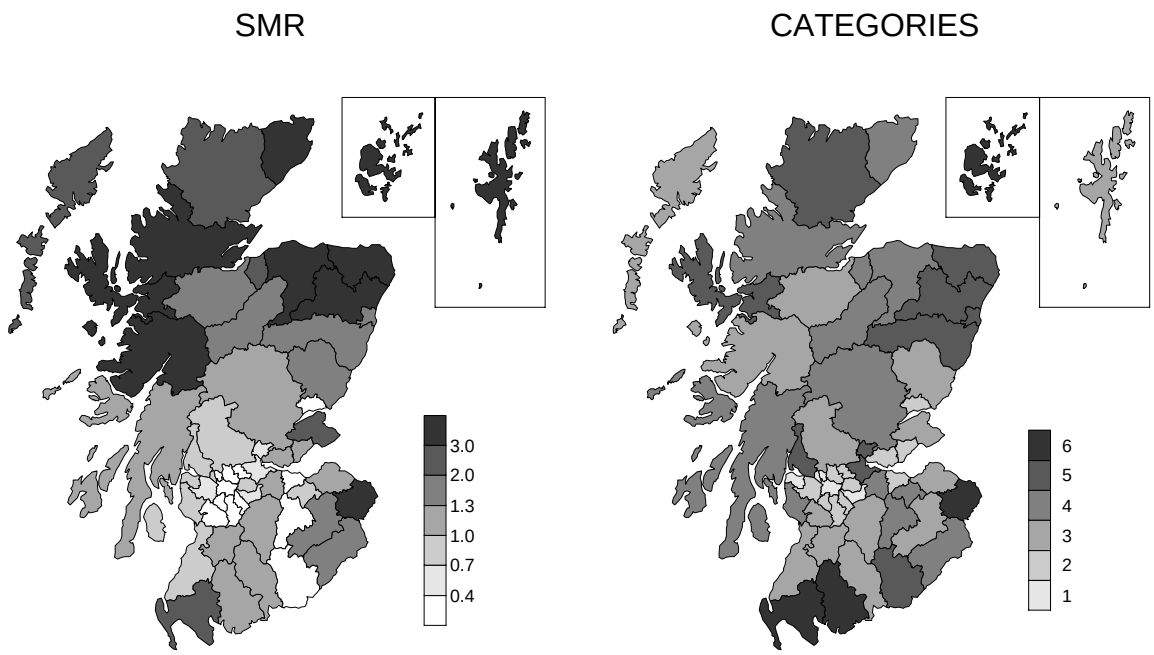

Figure 1: A map of the SMR's and a map showing the assignment of the regions to the categories of the covariate.

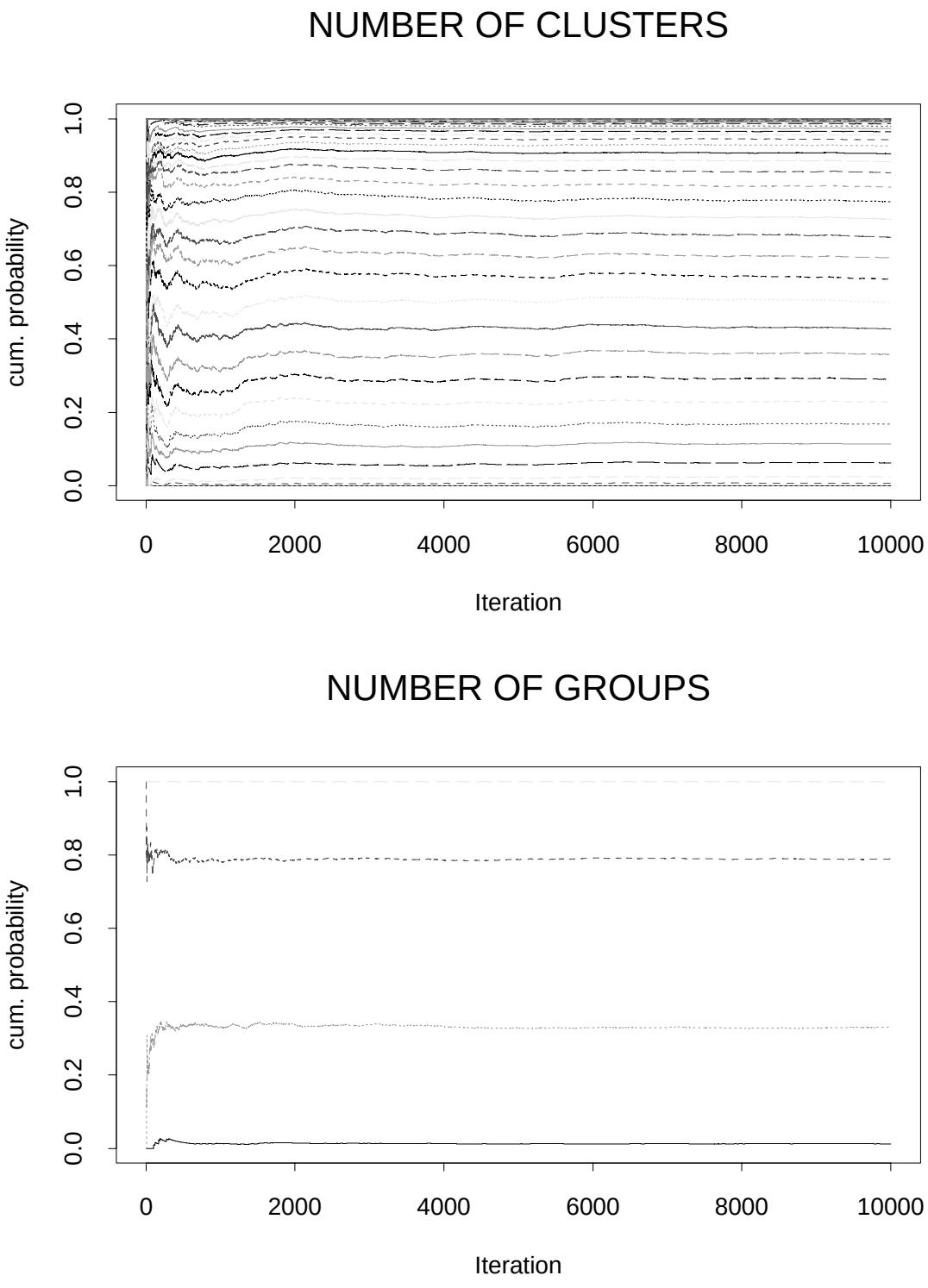

Figure 2: Cumulative occupancy fractions of number of clusters and number of groups.

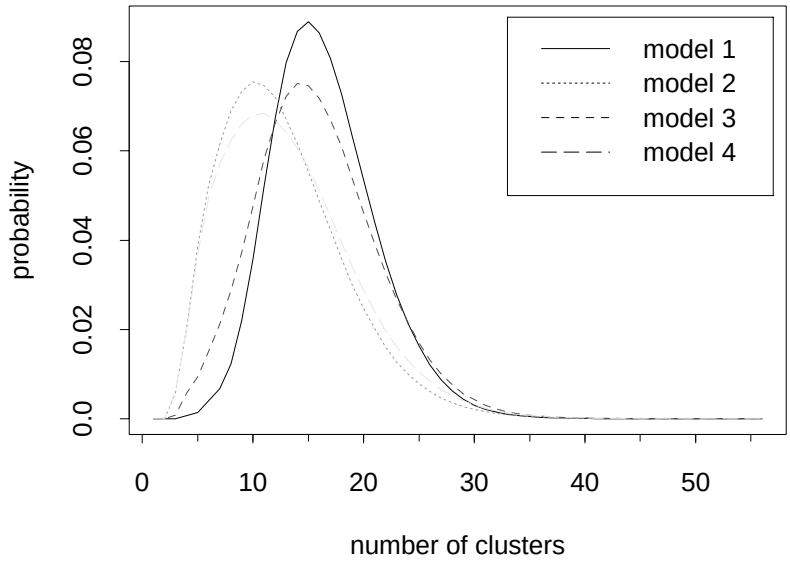

Figure 3: Posterior distributions of the number of clusters for the spatial component.

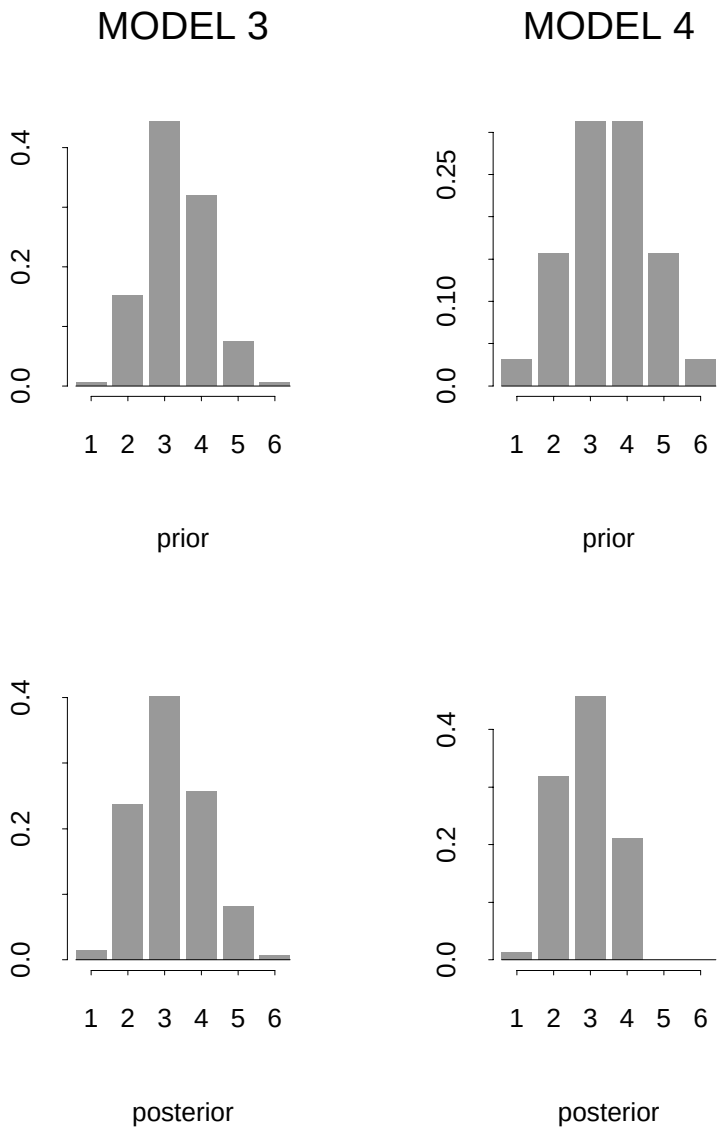

Figure 4: Prior and posterior distributions for the number of groups for the covariate component.

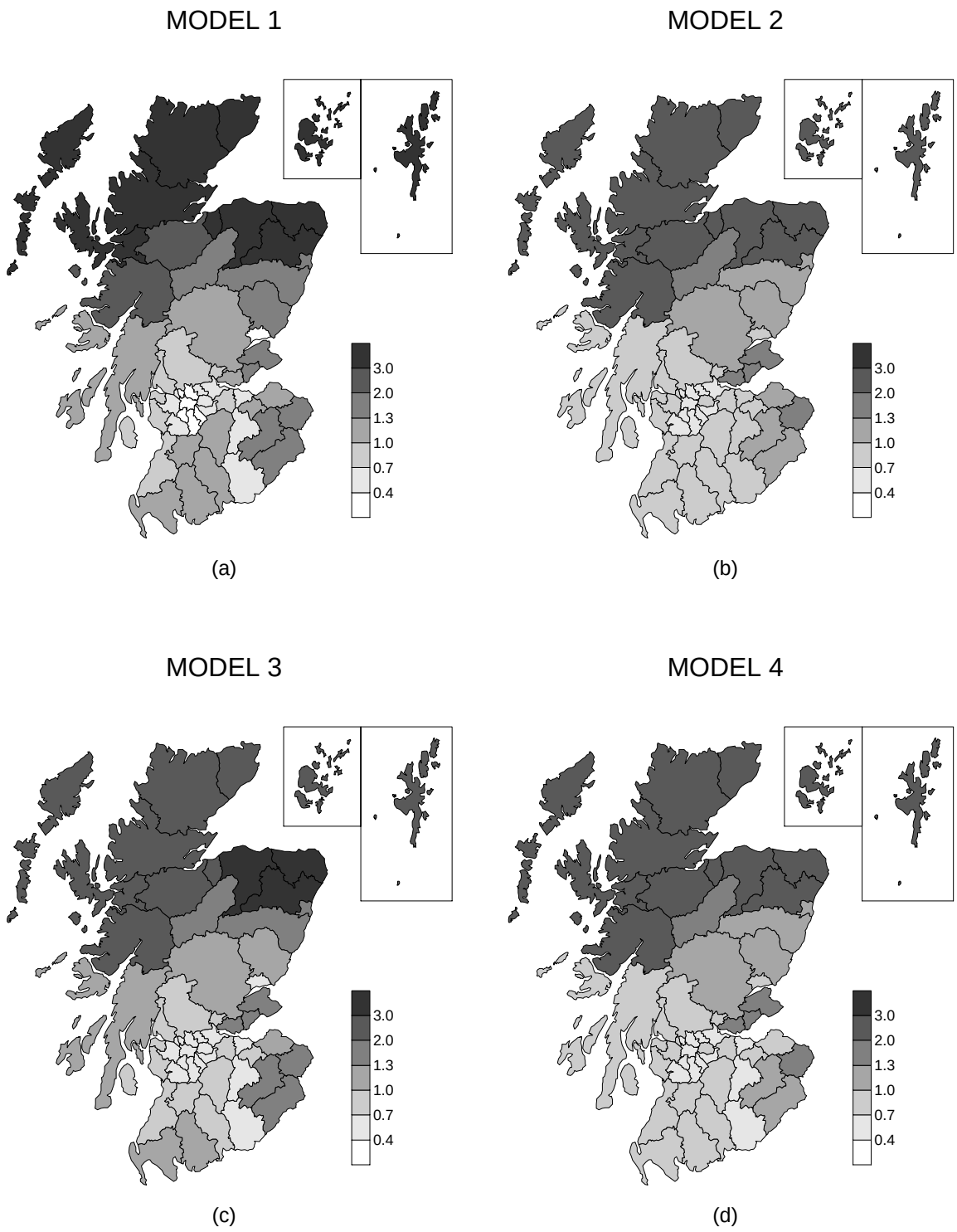

Figure 5: Posterior median estimates of the spatial component.

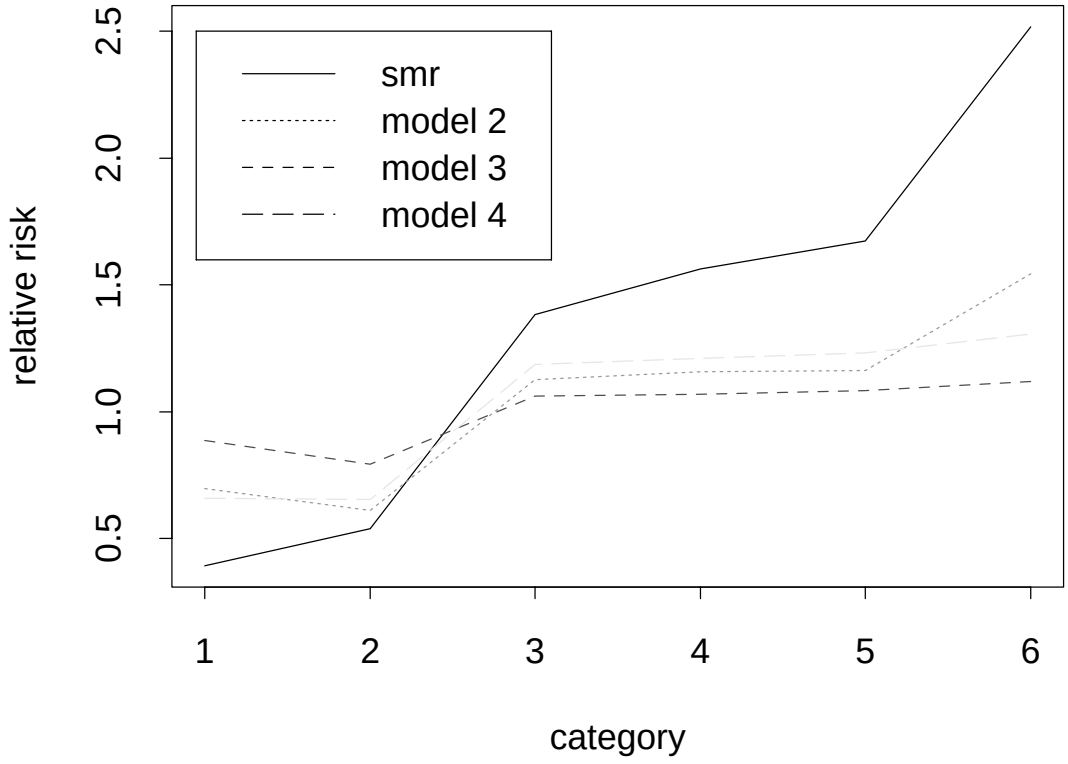

Figure 6: Comparison of the estimated effects of the covariate and the SMR's within the categories.

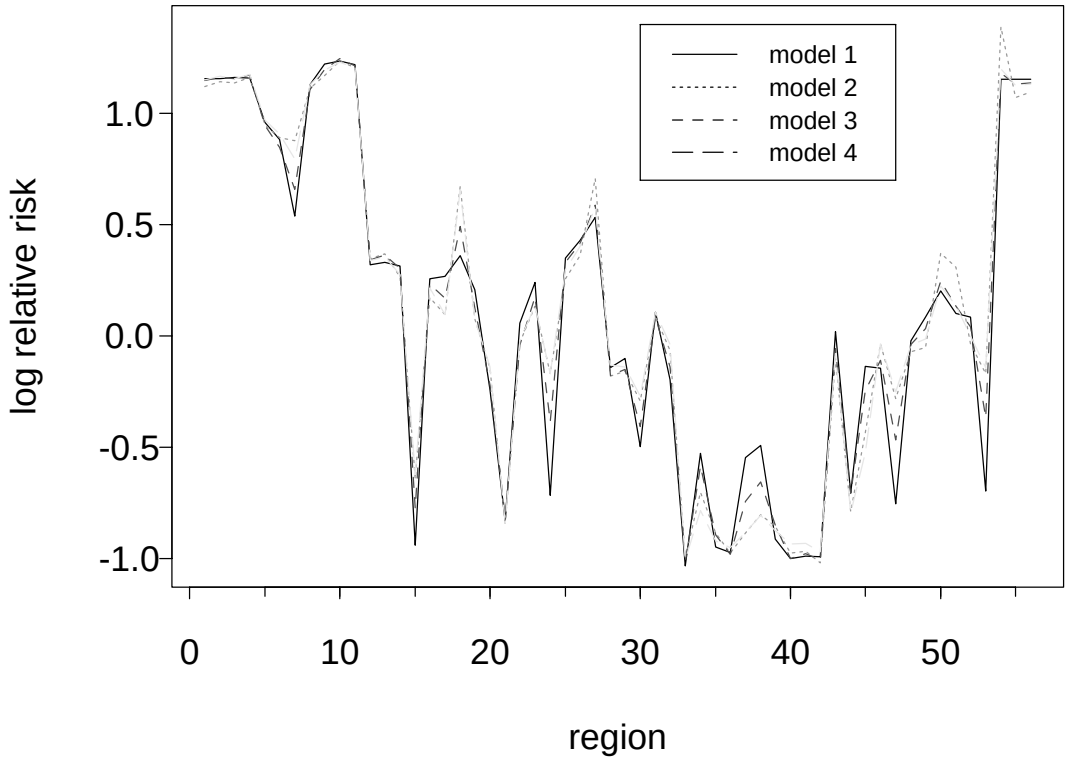

Figure 7: Comparison of the posterior median estimates of the log relative risks.