Bayesian partitioning for mapping disease risk using a matched case-control approach to confounding

DEBORAH A. COSTAIN

Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YN, UK
d.costain@lancaster.ac.uk

SUMMARY

Disease maps are useful for exploring geographical heterogeneity in health outcomes. Typically interest lies in unearthing atypical regions after adjusting for known confounders. This paper presents a Bayesian partitioning approach for analyses when individual-level matching has been used to control confounding. The model makes few assumptions about the surface form and, in particular, permits discontinuity. The specification is inherently parsimonious and posterior sampling permits direct assessment of surface uncertainty; additional unmatched covariates can also be incorporated. The method is used to investigate spatial variation in perinatal mortality in the North-West Thames region, England.

Keywords: Bayesian partitioning; Disease mapping; Matched case-control data; Reversible jump MCMC.

1. INTRODUCTION

The case-control study is one of the most frequently used study designs in analytical epidemiology, the rationale being to compare a group of diseased individuals to controls with regard to characteristics of interest (Breslow, 1996). In disease mapping interest lies in investigating spatial variation in risk after adjustment for known risk factors so as to yield aetiological clues. Contextually, case-control methods are particularly appealing since they allow for point-georeferencing and yield individual-level covariate information. Variation in disease odds can then be investigated within a continuous spatial framework and, after controlling for known determinants, any residual latent structure revealed.

The most challenging aspect of a case-control study is selecting an appropriate control group (Woodward, 1999). The aim is to yield an unbiased subset of the disease-free otherwise at risk, to maximize efficiency, and to control confounding. Confounding is a bias arising as a result of failure to disassociate the effects of extraneous variables that are independently associated with both the exposure and disease of interest. In disease mapping confounding is manifest as a distortion of a disease-environmental exposure relation due to known spatially varying risk factors. For example, social deprivation, which is spatially structured and independently a disease determinant, may confound environmental exposure. Consequently, maps exhibiting heterogeneous risk may be of little value, simply reflecting the known associations. Note that confounding can attenuate or accentuate a true association.

Confounding can be handled at the analysis stage, for example, by model adjustment, or alternatively by design. Kelsall and Diggle (1998) and Costain (2009), respectively, developed frequentist and Bayesian model-based approaches to environmental confounding based upon the completely random case-control designs.
study. In this design the cases, \( y = 1 \), are all known disease incidences and the controls, \( y = 0 \), are a simple random sample drawn from the disease-free otherwise at risk. Post-selection individuals are classified according to spatial location, \( x \), and relevant confounders, \( z \), and disease mapping is undertaken within a logistic modeling framework, with model-based adjustment for the confounders.

The completely random design can be inefficient since controls contribute little information if they fall within confounder-defined strata containing few cases (Breslow and Day, 1980). An alternative approach known as matching keeps the ratio of cases to controls constant within strata. Frequency-matching involves maintaining constant ratios across broad strata, whereas individual-matching involves selecting a group of matched controls for each case. Matching aims to eliminate confounding and also to increase analytical efficiency (Breslow and Day, 1980). In addition, adjustment is not model reliant and a parametric form is not assumed for the confounder effects.

A consequence of matching is that analyses need to reflect the sampling scheme. Indeed, failure to acknowledge that the matching has forced the controls to be more like the cases, in terms of exposures of interest, than would occur by random selection generally biases toward the null (Breslow and Day, 1980). The extent to which the matching impacts analytically relates to the degree of stratification. When the number of strata is small, compared with the number of cases, a standard Bernoulli likelihood and logistic model formulation augmented to include strata-specific intercepts applies directly. For individually matched designs, however, the number of strata increases with the number of cases leading to inconsistent estimation. In this case, inference typically proceeds based upon the matched conditional likelihood (Breslow and Day, 1980).

Methods for disease mapping using individually matched case-control data have been developed, however, there are potential limitations and scope for methodological development. Based upon the matched conditional likelihood and a logistic model formulation, Jarner and others (2002) use an additive model formulation (Hastie and Tibshirani, 1990) to recover smooth spatial variation in risk: an imposition which may, or may not, be appropriate. For instance, in the presence of spatial discontinuity, a low-risk area surrounded by high-risk areas, the surface may be overly smoothed. Recall that disease mapping seeks to unearth latent risk factors, the spatial structure of which are not known. Consequently, one may seek to increase robustness against mis-specification. In addition, the amount of smoothing is determined by a fixed, user-defined, bandwidth and in practice one might want some degree of spatial adaptiveness over the study region. From an inference perspective, assessment of uncertainty is not direct: evidence of residual spatial variation in risk is assessed using Monte Carlo methods under a composite null. As such, uncertainty in other model parameters may not be fully integrated.

A special case of disease mapping concerns investigating risk about a point or line source. Diggle and others (2000) embed the matched conditional likelihood within a Bayesian framework and model disease risk as a function of distance from the source. Now while a distance–risk relation is conceivable, imposing a strict parametric form may be limiting. Emissions may not be isotropic and, or, centered at the source, for example, due to wind direction or geography, and thus distance may not adequately measure exposure. More generally, point source methods can be problematic since the data are often used to generate and test hypotheses. Blind mapping of the disease risk may thus pose fewer problems; post-analyses surface estimates can be examined about any proposed source(s). For example, Denison and Holmes (2001), based upon regional count data, use Bayesian partitioning to map the risk of leukemia and then examine the posterior surfaces at known waste sites.

More generally, Bayesian partitioning has been used to tackle a variety of related problems including intensity estimation (Heikkinen and Arjas, 1998), cluster detection (Knorr-Held and Rasser, 2000), and disease mapping (Costain, 2009). Major strengths include the potential to detect spatial discontinuity and the inherent parsimony.

This paper describes a new Bayesian partitioning approach to disease mapping using individually matched case-control data. Methodologically, the work stems from Costain (2009) and extends the partitioning
Bayesian partitioning for mapping disease risk

The formulation *a priori* supposes that the region of interest can be decomposed into a number of disjoint subregions in which the residual disease risk, adjusted for known confounders, is a constant. The latent variability in the data dictates the form of the surface, and features due to unobserved covariates (both smooth and continuous) can be flexibly captured.

The matched conditional likelihood has a product multinomial form and the matched groups are the independent entities. This poses particular inferential challenges for the partitioning and also for efficient candidate generation for posterior sampling. The multinomial-Poisson transform is thus used to yield a likelihood that is proportional to the matched conditional likelihood, therewith simplifying the sampling algorithm and reconfiguring the partitioning as a Bayesian generalized linear modeling problem.

Reversible jump Markov chain Monte Carlo (MCMC, Green, 1995) is used to sample from the trans-dimensional posterior distribution; iterations are Gibbs updates or consider candidates drawn from approximations to the full conditional distributions. Posterior realizations can be used variously for Bayesian inference and, in particular, permit both surface estimation and also direct assessment of uncertainty.

The paper is organized as follows. The model is presented in Section 2 and in Section 3 the reversible jump MCMC sampler is described. Section 4 presents details of simulations and a Bayesian partitioning re-analysis of the North-West Thames region, England, perinatal mortality data previously analyzed by Jarner and others (2002). The paper concludes with a brief discussion, Section 5.

2. Model formulation

Suppose that interest lies in investigating spatial variation in disease risk over a particular planar region, $A$. Underpinning the partitioning approach is an assumption that the region can be divided into a number of disjoint subregions, $R_l$, $l = 1, \ldots, k$, in which the responses, $y$, are exchangeable and are derived from the same probability distribution. In the simplest case scenario this corresponds to an assumption that the subregional disease risk, $p_l$, is a constant. In practice, however, known spatially varying risk factors, such as ethnicity and age, impact upon health outcome and are not of intrinsic interest. The risk assumption is thus relaxed and rather the residual disease risk, adjusted for known risk factors, is *a priori* taken to be piece-wise constant over the region. The rationale is to flexibly decompose any residual latent structure via a tessellation on $A$.

Voronoi tessellations are used for the spatial decomposition and are defined by a number of generating points, $c_l$, $l = 1, \ldots, k$, and a distance metric $\|\cdot\|$. The points, $c_l$, define the subregional *centers* and the corresponding polygonal region (namely *tile*) is then given by $R_l = \{x \in A : \|x - c_l\| < \|x - c_i\| \forall i \neq l\}$, which is simply those points $x \in A$ which are *closer* to center $c_l$ than any other center. In high-dimensional problems proximity measures can be problematic due to the “curse of dimensionality”: distance contrasts can tend to zero in many coordinates. Here the case and control locations $x = (x_1, x_2) \in \mathbb{R}^2$ and the Euclidean distance metric is deemed adequate (Holmes and others, 1999). Note that the metric is used to assign observations to centers, $c_l$, and therewith subregions of the partition, as opposed to induce any distance-defined dependency. Indeed, given a tessellation on $A$, subregions are *a priori* independent. The number of centers, their locations, and the subregional risks are unknown and underpin the inference.

2.1 Conditional likelihood for matched cases and controls

Following Breslow and Day (1980), suppose that in an individually matched case-control study we have $i = 1, \ldots, n$ matched groups (*tuples*), each consisting of $j = 1, \ldots, m$ individuals and that group $i$ yields the exposure vectors $E_{i1}, \ldots, E_{im}$, but it is not known which of them relates to the case and which to the controls. By design, group $i$ contains one case and $m - 1$ controls and thus conditioning upon the unordered set of exposures, the probability that the first value, $E_{i1}$, relates to the case, $y_{i1} = 1$, and the
remainder to the controls, $y_{ij} = 0$, $j = 2, \ldots, m$, is given by

$$P(y_{i1} = 1 \mid E_{i1}) \prod_{j=2}^{m} P(y_{ij} = 0 \mid E_{ij}) \over \sum_{j=1}^{m} P(y_{ij} = 1 \mid E_{ij}) \prod_{r \neq j}^{m} P(y_{ir} = 0 \mid E_{ir}).$$

(2.1)

Typically (2.1) is simplified and expressed in the more familiar form

$$\left\{ 1 + \sum_{j=2}^{m} P(y_{ij} = 1 \mid E_{ij}) P(y_{i1} = 0 \mid E_{i1}) \over P(y_{i1} = 1 \mid E_{i1}) P(y_{ij} = 0 \mid E_{ij}) \right\}^{-1}.$$

(2.2)

The partitioning is undertaken within a logistic modeling framework such that the disease risk $p_{ij} = P(Y_{ij} = 1 \mid E_{ij})$ is linked to the exposure variables via the relation

$$\log \left( {p_{ij} \over 1 - p_{ij}} \right) = \log \left( {P(y_{ij} = 1 \mid E_{ij}) \over P(y_{ij} = 0 \mid E_{ij})} \right) = \alpha_i + z_{ij}' \beta + \mu_{l(x_{ij})},$$

(2.3)

where the $\alpha_i$'s represent the tuple-specific intercepts, $\beta$ any non-matched covariate effects, and $\mu_{l(x_{ij})}$ denotes the spatial height associated with region $l = 1, \ldots, k$, to which the spatial location $x_{ij}$ of individual $ij$ belongs. Substituting this logistic model formulation in (2.2) yields tuple-specific probabilities; the conditional probability of the observed data over all $n$ matched sets is then simply the product over the independent entities:

$$\prod_{i=1}^{n} \left\{ 1 + \sum_{j=2}^{m} \exp[(z_{ij} - z_{i1})' \beta + (\mu_{l(x_{ij})} - \mu_{l(x_{i1})})] \right\}^{-1}.$$

(2.4)

The stratum-specific parameters $\alpha_i$ are thus eliminated in the matched conditional likelihood and cannot be estimated. Similarly, covariate terms that are perfectly matched contain no extra information. Although not practicable or of interest here, interactions between matching variables and exposures can, in general, be considered (Breslow and Day, 1980).

### 2.2 Multinomial-Poisson model formulation

The matched groups form the independent entities in (2.4) and its evaluation requires computation of the $m - 1$ exposure differences, $(z_{ij} - z_{i1}, \mu_{l(x_{ij})} - \mu_{l(x_{i1})})$, for each matched group increasing the computational burden when partitioning. In addition, the likelihood is not of standard exponential family form, posing potential challenges for efficient posterior sampling. As a remedy, the multinomial-Poisson transform is used to yield a likelihood that is proportional to the matched conditional likelihood and a priori maintains the conditional independence structure of the partitioning. The rationale is to suppose that the $N = mn$ responses are independent Poisson realizations:

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}), \quad \text{with} \quad \lambda_{ij} = \exp[\alpha_i + z_{ij}' \beta + \mu_{l(x_{ij})}],$$

(2.5)

where, as previously $i = 1, \ldots, n$ denotes the tuple, $j = 1, \ldots, m$, the elements within tuples, the $\alpha_i$'s the tuple-specific intercepts, $\beta$ the covariate effects, and $\mu_{l(x_{ij})}$ the residual spatial height in subregion $R_l$. Adopting the labeling scheme $y_{i1} = 1$ for the case in tuple $i$ and $y_{ij} = 0$, $j = 2, \ldots, m$, for the $m - 1$
controls and defining
\[
\lambda_i = \sum_{j=1}^{m} \lambda_{ij} = e^{\alpha_i} \sum_{j=1}^{m} e^{z_{ij} \beta + \mu_{(x_{ij})}} \quad \text{and} \quad y_i = \sum_{j=1}^{m} y_{ij},
\]
the Poisson formulation in (2.5) yields a likelihood:
\[
\prod_{i=1}^{n} \prod_{j=1}^{m} e^{-\lambda_{ij}} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \prod_{j=1}^{m} \prod_{i=1}^{n} \left( e^{z_{ij} \beta + \mu_{(x_{ij})}} \right)^{y_{ij}}.
\]

Here the data correspond to \(y_{i1} = y_i = 1\) for all \(i\) and the likelihood becomes
\[
\prod_{i=1}^{n} e^{-\lambda_i} \left( 1 + \sum_{j=2}^{m} \exp[(z_{ij} - z_{i1}) \beta + \mu_{(x_{ij})} - \mu_{(x_{i1})}] \right)^{-1},
\]
which corresponds, up to proportionality in \(\lambda_1, \ldots, \lambda_n\), to the matched conditional likelihood in (2.4).

### 2.3 A prior model for spatial variation in disease risk

Exploiting the likelihood formulation in (2.6) requires a prior specification which maintains the product structure between \((\lambda_1, \ldots, \lambda_n)\) and \((\beta, \mu)\). It is thus a priori assumed \(P(\beta, \mu, \lambda_1, \ldots, \lambda_n) = P(\beta, \mu)P(\lambda_1, \ldots, \lambda_n)\) with \(P(\lambda_1, \ldots, \lambda_n)\) arbitrary. In terms of parameterizations, both \((\beta, \mu, \alpha)\) and \((\beta, \mu, \lambda)\) are amenable since
\[
\lambda_i = e^{\alpha_i} \sum_{j=1}^{m} e^{z_{ij} \beta + \mu_{(x_{ij})}} \quad \rightarrow \quad \alpha_i = \log \left( \frac{\lambda_i}{\sum_{j=1}^{m} e^{z_{ij} \beta + \mu_{(x_{ij})}}} \right).
\]

For the MCMC sampler, however, it is easier to work in terms of the original parameterization, namely \((\beta, \mu, \alpha_1, \ldots, \alpha_n)\). Standard rules for transformed variables give
\[
P_{\beta, \mu, \alpha_1, \ldots, \alpha_n}(\beta, \mu, \alpha) = P_{\beta, \mu, \lambda_1, \ldots, \lambda_n}(\beta, \mu, e^{\alpha_i} \sum_{j=1}^{m} e^{z_{ij} \beta + \mu_{(x_{ij})}}, i = 1, \ldots, n) \times |\text{diag}(\lambda_i)|,
\]
and thus assuming a priori \(P(\lambda_i) \propto (\lambda_i)^{-1}, i = 1, \ldots, n\), leads to a joint prior distribution \(P(\beta, \mu, \alpha_1, \ldots, \alpha_n) \propto P(\beta, \mu)\), that is, a flat prior on \((\alpha_1, \ldots, \alpha_n)\), and one can work with the model \(y_{ij} \sim \text{Poisson}(\lambda_{ij} = e^{\alpha_i} e^{z_{ij} \beta + \mu_{(x_{ij})}})\) with \(i = 1, \ldots, n, j = 1, \ldots, m\), independent.

For \(P(\beta, \mu)\) independent priors are specified with \(p(\beta) \sim N(\mu, R)\) and \(\mu_i \sim N(0, \sigma_0^2)\). For \(\sigma_0^2\) a dispersed, but proper, hyperprior is specified such that \(\sigma_0^2 \sim \text{IG}(a, b)\). Given the unknown dimensionality of the partitioning, priors for the number of and the locations of the tile centers need to be specified. A priori the number of tiles, \(k = 1, \ldots, k_{\max}\), assumes a truncated geometric distribution such that \(P(k) \propto (1 - T)^{k} \cdot T^{k_{\max}}\). Conditional upon \(k\), the locations of the tile centers, \(c_l\), are taken to be uniform on \(A\) such that \(P(c_l \mid k) = \frac{1}{A} I_{A}^{-1}\), independent for \(l = 1, \ldots, k\).

### 3. Building an MCMC sampler

A reversible jump MCMC (Green, 1995) scheme was used to sample from the trans-dimensional posterior distribution. Benefits of the multinomial-Poisson transform include the opportunity to exploit standard
implemented move types (outlined below), computational details are provided in Section 3.1. More generally, traversing the joint posterior is based upon the following move types:

- **Height**: changes to the residual spatial heights, \( \mu_l, l = 1, \ldots, k \), are proposed in succession;
- **Birth**: the number of subregions is increased by adding one subregional center;
- **Death**: the number of subregions is decreased by deleting one subregional center;
- **Alpha**: the tuple-specific parameters \( \alpha_i, i = 1, \ldots, n \), are in-turn updated;
- **Beta**: the additional “non-matched” covariate effects \( \beta \) are changed en bloc;
- **Hyper**: the values of the dispersion hyperparameter \( \sigma_0^2 \) are changed.

A random scanner is used for all but the Hyper move, which is updated at each iterative step. More specifically, at each iteration of the algorithm non-Hyper moves are proposed with probability \( \pi_{\text{move}}(k) : \sum \pi_{\text{move}}(k) = 1 \). The summation is over all implemented move types and the dependence upon \( k \) arises since a Birth/Death is not permitted when \( (k = k_{\text{max}})/(k = 1) \), respectively. The proposed state is accepted with probability determined by the Metropolis-Hastings-Green ratio (Green, 1995); parameters, \( \alpha_i, i = 1, \ldots, n \), and \( \sigma_0^2 \) are Gibbs updates. The tessellation is initialized using a single tile, the location of the tile center is drawn from the prior on \( A \). The height of the tile and hyperparameter \( \sigma_0^2 \) are also initialized via draws from their respective prior distributions. Other parameters take the value zero as a starting value. A period of burn-in is adopted and thinning is used to reduce dependency.

### 3.1 Computational details for posterior sampling

#### 3.1.1 Height

Given a current configuration, a **Height** proceeds by successively sampling candidate heights, \( \mu_l \), for each subregion, \( l = 1, \ldots, k \), of the partition. For efficient candidate generation, the conditional independence underpinning the partitioning is exploited and a re-parameterization performed. Specifically, for each subregion the full conditional is

\[
P(\mu_l | \ldots) \propto \prod_{ij : l(x_{ij}) = l} e^{-\lambda_{ij}x_{ij}\gamma_l} \times P(\mu_l) \propto \prod_{ij : l(x_{ij}) = l} e^{-\gamma_l x_{ij} (\beta + \mu)} (e^{\mu\gamma_l})^{y_{ij}} \times P(\mu_l),
\]

where \( ij : l(x_{ij}) = l \) denotes those observations assigned to region \( l \) of the spatial partition. Defining \( \gamma_l = e^{\mu_l} \), the target distribution can be expressed as

\[
P(\gamma_l | \ldots) = e^{-\gamma_l \sum_{ij : l(x_{ij}) = l} e^{\alpha_i + \beta_i}} \gamma_l^{\sum_{ij : l(x_{ij}) = l} y_{ij}} \times P(\gamma_l),
\]

where the first term is a \( \Gamma(1 + \sum_{ij : l(x_{ij}) = l} Y_{ij}, \sum_{ij : l(x_{ij}) = l} e^{\alpha_i + \beta_i}) \) distribution and \( P(\gamma_l) \sim \log-Normal(\mu_0, \sigma_0^2) \).

A benefit of this change of variable is that a log-normal distribution can be readily approximated with a moment-equated \( \Gamma(c, d) \) distribution and candidate heights \( \gamma_{l}^* \) drawn from an **updated** Gamma distribution derived from the Gamma-Poisson product. Specifically, candidates denoted by \( \gamma_{l}^* \) are sampled...
from an approximation, $Q(\cdot \mid \cdot)$, to the full conditional:

$$
\gamma_i^* \sim \text{Gamma}\left( c + \sum_{ij: (x_{ij}) = l} y_{ij}, d + \sum_{ij: (x_{ij}) = l} e^{\alpha_i z_{ij}^*} \right) = \text{Gamma}(e, f),
$$

where $c = (e^{\mu_0 + \frac{1}{2}\sigma_i^2})^{-1}$ and $d = (e^{\mu_0 + \frac{1}{2}\sigma_i^2}(e^{\sigma_i^2}-1))$. The proposed heights are each in turn accepted with probability

$$
\min\left\{ 1, \frac{p(\gamma_i^*)}{p(\gamma_i)} \times \frac{Q(\gamma_i^* \mid c, d)}{Q(\gamma_i^* \mid c, d)} \right\}.
$$

Note this simple form of acceptance probability arises due to the canceling that occurs between the proposal distribution and likelihood when the subregional assignment of observations, $y_{ij}$, is unchanged between states.

3.1.2 Birth and death. Given a current configuration consisting of $k < k_{\text{max}}$ tile centers, $c_l$, and associated heights, $y_l = \mu_l^i$, $l = 1, \ldots, k$, a Birth proceeds by sampling an additional center $c_{k+1}$ uniformly on $A$ and reassigning observations accordingly. A candidate height $y_{k+1}$ is then sampled from a distribution similar to (3.1); the difference is that here the summation now is taken over those observations assigned to center $c_{k+1}$. Conversely, given a current configuration (and providing $k > 1$) a Death proceeds by randomly selecting a tile center $c_l^*$ and corresponding height $y_{l^*}$ to be deleted. Observations are then reassigned according to the reduced partition structure.

A Birth is accepted with probability

$$
\min\left\{ 1, (1 - T) P(\gamma_{k+1}) \times \frac{L_{ij: (x_{ij}) = k+1}(\gamma_{k+1})}{L_{ij: (x_{ij}) = k+1}(\gamma^c)} \times \frac{1}{Q(\gamma_{k+1} \mid e, f)} \right\},
$$

where $L_{ij: (x_{ij}) = k+1}$ denotes the likelihood contribution for the reassigned observations and $\log(\gamma^c) = \mu^c$ represents the current sub-vector of spatial heights recorded at the individual level. In contrast, a Death is accepted with probability

$$
\min\left\{ 1, \frac{1}{(1 - T) P(\gamma_{l^*})} \times \frac{L_{ij: (x_{ij}) = l^*}(\gamma_{l^*})}{L_{ij: (x_{ij}) = l^*}(\gamma^c)} \times Q(\gamma_{l^*} \mid e, f) \right\}.
$$

The Birth/Death moves work diametrically forming pairs of reversible jumps. Acceptance probabilities are adjusted for the boundary cases: $k = 1$ and $k = k_{\text{max}}$.

3.1.3 Alpha. In the Alpha move, the conditional independence of the tuple-specific parameters, $\lambda_i$, is exploited, negating the necessity for high-order matrix inversion. Specifically, the full conditional distribution for $\lambda_1, \ldots, \lambda_n$ has density proportional to

$$
\prod_{i=1}^n \frac{1}{\lambda_i} \text{Poisson}(y_i \mid \lambda_i) \propto \prod_{i=1}^n \frac{1}{\lambda_i} e^{-\lambda_i} \lambda_i^{y_i} = \prod_{i=1}^n e^{-\lambda_i}
$$

corresponding to $n$ independent $\text{Exp}(1)$ distributions. Hence, for each tuple-specific parameter, $\alpha_i$, a draw, $\lambda_i$, is made from an $\text{Exponential}(1)$ distribution and back-transformed:

$$
\alpha_i = \log_e \left( \frac{\lambda_i}{\sum_{j=1}^m e^{\alpha_j + \mu_{i,j}}^*} \right).
$$
3.1.4 Beta move. The Beta move is operational when additional covariate information, $z$, is available. Following Costain (2009), a weighted-least-squares algorithm (Gamerman, 1997) is used to generate a proposal that readily approximates the full conditional posterior distribution. Specifically, a candidate $\beta^*$ is proposed from a $Q_\beta(\cdot | \cdot) \sim N(b^t, P^t)$ density with moments given by $b^t = (R^{-1} + Z'W(\beta)Z)^{-1} \times (R^{-1}u + Z'W(\beta)\tilde{y}(\beta))$ and $P^t = (R^{-1} + Z'W(\beta)Z)^{-1}$. Here $W(\beta)$ is an $N \times N$ diagonal matrix of Poisson weights: $W(\beta) = \text{diag}(\lambda_{ij} = e^{\alpha_i + z_{ij}^\prime \beta + \mu_{ij}})$ and the N-vector $\tilde{y}(\beta)$ consists of values of the so-called “adjusted dependent variate”:

$$\tilde{y}_{ij}(\beta) = z_{ij}^\prime \beta + \frac{y_{ij} - \lambda_{ij}}{\lambda_{ij}},$$

which are both functions of the current value $\beta$. The proposed value of $\beta^*$ is then accepted with probability

$$\min \left\{ 1, \frac{p(\beta^*)L(\beta^*)}{p(\beta)L(\beta)} \times \frac{Q_\beta(\beta | \beta^*)}{Q_\beta(\beta^* | \beta)} \right\}.$$

3.1.5 Hyper move. The Hyper move concerns the scalar hyperparameter $\sigma_0^2$ which is updated at each iterative step according to the full conditional

$$\sigma_0^2 | \ldots \sim \text{IG} \left( a + \frac{k}{2}, b + \sum_{l=1}^k \frac{(\mu_j - \mu_0)^2}{2} \right).$$

4. Applications

4.1 Simulations

To examine the sampler’s capacity for recovering: constant, discontinuous, and smooth spatial structure, four artificial data sets were analyzed. In addition, sensitivity of the results to choices for $p(k)$ and $p(\sigma_0^2)$ was considered. The results are available in Appendix A of supplementary material available at Biostatistics online (www.biostatistics.oxford.journals.org). The posterior surface estimates and marginal posteriors for $k$ showed little sensitivity to $p(\sigma_0^2)$ and $p(k)$. Values of $a = 1$ and $b = 0.01$ are recommended as mixing was a little better. For $k \in (1, 100)$, small but positive values for the tile penalty $T$ are recommended.

4.2 Investigating spatial variation in perinatal mortality in the North-West Thames region, England

4.2.1 Background. Newborns are sensitive to their immediate surroundings, their mother’s pre- and post-natal health, and the quality of health-care provision. Indeed, infants are conceivably the most vulnerable group in terms of the adverse effects of their environment. Infant births and deaths are vital registrations and child and infant mortality rates are important indicators of inequities in a country’s infrastructure, services, and development. Infant deaths are typically subclassified as perinatal (within 7 days), neonatal (within 8–28 days), or post-neonatal (within 29 days to 1 year).

In this subsection, the matched partition model methodology is used to re-investigate spatial variation in perinatal mortality risk in the North-West Thames region of England. The data, previously analyzed by Jarner and others (2002), derive from the period 1980–1991 and consist of the 3425 cases, $y = 1$, of still births or deaths within 7 days of life together with 6850 ($m = 2$) live-birth controls, $y = 0$, who were individually matched to the cases according to gender and date of birth. Georeferencing, $x$, was based upon the mothers’ place of residence at the time of birth. A point-map depicting the case and control locations is provided in Figure 1.
Additional covariate information, $z$, consists of the Carstair's index: a small-area measure of socioeconomic deprivation (Carstairs and Morris, 1991). This index has been shown to be predictive of health outcome and is based upon the percentage of persons (according to the 1991 census) having no car, living in overcrowded housing, and having a head of household in social class IV, V, or unemployed. The scores were calculated at the electoral district level; however, model adjustment is conducted here at the individual level. Lower scores indicate increased affluence, whereas, higher scores indicate increased deprivation. The score is scaled to have zero mean over England; negative and positive scores corresponding to relatively affluent and disadvantaged regions, respectively.

4.2.2 Data analysis. The sampler was run based upon the perinatal mortality data. Since Carstair’s index is typically spatially structured, both unadjusted and adjusted analyses were conducted. If Carstair’s index is spatially neutral, one would expect little change in the adjusted and unadjusted surface estimates. In contrast, if spatially structured, confounding may lead to attenuation or indeed masking of any environmental effects. Recall that a goal in disease mapping is to investigate spatial variation in risk and if it exists, seek to explain it.

In the analysis presented here, the tile penalty was fixed such that $T = 0.02$, the dispersion hyperparameters fixed such that $\sigma_0^2 \sim \text{IG}(1, 0.01)$ and the maximum number of tiles taken to be $k_{\text{max}} = 100$. When adjusting for Carstair’s index, a diffuse but proper prior for $\beta$ was assumed: $p(\beta) \sim N(0, 1000)$. A total of
10 010 000 iterations were performed; the first 10 000 were discarded as burn-in. Thinning was performed by storing every 1000th draw.

Owing to the nature of the sampling, case-control methods yield up to proportionality estimates of disease odds as opposed to risk. Inference thus proceeds based upon the estimation of disease odds ratios taken relative to a baseline. As in Costain (2009), the method of summarizing the stored log-odds surfaces was to map the posterior mean, averaged over the 10 000 piecewise constant surfaces, at pixels $x \in A$ on a

Fig. 2. Posterior mean odds ratio surfaces (relative to the surface mean odds) for perinatal mortality in the North-West Thames region, England for the period 1980–1991. The unadjusted estimate is depicted in (a) and in (b) the surface is adjusted for Carstair’s index; contours in (a) and (b) depict $\rho(x)$. In (c) and (d), the surface estimates are overlaid with the 0.025 and 0.975 of the “probability surface” contours if they exist. All surfaces are on the same scale.
40 × 40 grid, relative to the surface mean. Specifically, the estimated posterior odds ratio surface is given by \( \rho(x) = \exp[\mu(x) - \bar{\mu}] \), where \( \mu(x) \) denotes the posterior mean log-odds at grid location \( x \) and \( \bar{\mu} \) is the overall surface mean log-odds. This choice minimizes squared-error loss and can recover both smooth and continuous functional forms.

To investigate evidence of spatial variation in risk, the posterior stored heights were examined pointwise, again at pixels \( x \in A \) on a 40 × 40 grid (Costain, 2009). This involved running through each of the 10,000 stored surfaces and counting the proportion of values at each of the 1600 grid points that exceed the surface mean. The 0.025 and 0.975 contours, if any, of this resultant proportion surface were then superimposed on the surface estimate, \( \rho(x) \), to identify any regions of atypical risk. Note that any 0.025 contours would indicate regions with lower than average disease odds, whereas any 0.975 contours would indicate regions with higher than average disease odds.

Prior sensitivity with respect to \( p(k) \) and \( p(\sigma_0^2) \) was examined using combinations of tile penalties \( T = 0.0, 0.1 \) and hyperparameters \( (a, b) = (5, 0.125), (0.25, 0.00025) \). The prior sensitivity was minimal. Posterior mean surface estimates and marginal posterior distributions for \( k \) and \( \beta \) were stable. The results are provided in Appendix B of supplementary material available at Biostatistics online (www.biostatistics.oxford.journals.org).

4.2.3 Results. The unadjusted surface estimate (see Figure 2(a)) exhibited some degree of spatial variation in perinatal mortality odds over the North-West Thames region. The odds ratios spanned 1.15 toward the North of the region to 0.85 in the East. On adjusting for Carstair’s index, \( z \), however, the odds ratio surface contours were materially one over the entire study region; see Figure 2(b). In terms of the evidence of spatial variation in perinatal mortality, the resulting 0.975 contours from the unadjusted proportion surface indicated a subregion of increased neonatal mortality in the North of Thames Valley region; see Figure 2(c). On adjusting for Carstair’s index, however, there was no evidence of spatial variation in risk over the study region. The contours of the proportion surface did not precede 0.025 or exceed 0.975 at any grid point.

Histograms depicting the number of tiles, \( k \), in the stored models are given in Figure 3. In the Carstair’s adjusted analyses, Figure 3(b), the number of tiles in the surface realizations was typically lower than in the unadjusted analyses. The median number of tiles was found to be one, the interquartile range (IQR) 1 and the maximum 8, (Table 1). In comparison, the median number of tiles in the unadjusted analysis was 4, the maximum was 22, and the IQR was 4.

![Fig. 3. Posterior distributions for the number of tiles in the stored models for (a) the unadjusted and (b) the adjusted analyses. In (b), adjustment is made for Carstair’s index. The overlaid curve depicts the prior distribution for \( k \).](https://academic.oup.com/biostatistics/article-abstract/14/1/99/250486/14119826068)
Table 1. Marginal posterior summaries for the coefficient of Carstair’s, $\beta$, the hyperparameter, $\sigma_0^2$, and the number of tiles, $k$, for the Carstair’s adjusted analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Median</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carstair’s coefficient ($\beta$)</td>
<td>0.0465</td>
<td>0.0072</td>
<td>0.0464</td>
<td>0.0095</td>
</tr>
<tr>
<td>Hyperparameter ($\sigma_0^2$)</td>
<td>0.0184</td>
<td>—</td>
<td>0.0155</td>
<td>0.0107</td>
</tr>
<tr>
<td>Number of tiles ($k$)</td>
<td>1.2400</td>
<td>—</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Fig. 4. Posterior distribution for the effect, $\beta$, of Carstair’s index, $z$, on the log-odds of perinatal death for the period 1980–1991 in the North-West Thames region, England.

Since Carstair’s exhibits spatial structure and is predictive of perinatal mortality, the finding that there are typically fewer tiles with more consistent log-odds in the adjusted realizations is perhaps not surprising. Bayesian partitioning is data adaptive and the idea is to place more centers in spatially heterogeneous sub-regions (Holmes and others, 1999), the variation arising here due to the confounding effects of deprivation.

The posterior distribution for the effect of Carstair’s index is presented in Figure 4. Numerical summaries are provided in Table 1. An increase in Carstair’s index was found to be associated with an increased perinatal mortality risk, indicating a higher risk in deprived areas. The posterior mean $\bar{\beta}$ was found to be 0.0465, with a posterior standard deviation of 0.0072. Over the study region the recorded Carstair’s scores span $-5.26$ to $15.53$ and thus the point estimate is practically relevant; a 10-unit increase, on average, raising the perinatal mortality odds by 60%.

5. Discussion

A key feature in disease mapping is the ability to accommodate known confounders. Confounding can be controlled analytically or, alternatively, eliminated by design by using confounders as stratifying factors for control to case matching. Matching aims to increase efficiency and is particularly useful when confounder effects are complex. The aim of this paper was to develop a Bayesian partitioning approach for disease mapping using individually matched case-control data.

A consequence of matching is that analyses must reflect the sampling scheme. In the case of individual-matching the dimensionality increases with the number of cases and analysis typically proceeds based upon the matched conditional likelihood, increasing the computational burden for partitioning. The
multinomial-Poisson transform was thus used to yield a likelihood proportional to the matched conditional likelihood, setting the problem within a more amenable Poisson modeling framework.

The methodology was used to examine data previously analyzed by Jarner and others (2002) to investigate spatial variation in perinatal mortality in the North-West Thames region, England. Infants were individually matched based upon their sex and date of birth. A measure of social deprivation, Carstairs’ index, was available.

The results largely agreed with the previous analyses. Social deprivation was found to be predictive of perinatal mortality ($\hat{\beta} = 0.0465, SD(\beta) = 0.0072$), although the coefficient found in the additive modeling was greater in magnitude ($\hat{\beta} = 0.1398, SE(\hat{\beta}) = 0.0194$). The unadjusted surface estimate exhibited some variability and examination of posterior realizations indicated a small region of elevated mortality risk in the North. When adjusted for Carstairs’ index, however, there was no evidence of spatial variation in perinatal risk over the study region. In the previous analyses, the authors present cross-validation results as opposed to disease maps and tolerance contours $per se$. They show that, on adjusting for Carstairs’ index, the cross-validation curve is monotone decreasing (synonymous with a constant residual risk), and say “how including relevant covariate information, in this case social deprivation, can provide an explanation for what might otherwise be treated as unexplained spatial variation in risk”.

The resultant surface estimate appeared smooth and this is a key partitioning feature. Although discontinuity can be detected, posterior averaging allows for the recovery of a smooth functional form and thus is inherently flexible. Moreover, in contrast to existing methods, posterior realizations allow for the direct assessment of uncertainty. Although the method is developed to handle a fixed number of controls per case, the model is readily generalizable to variable numbers of cases and/or controls in each group.

**Supplementary material**


**Acknowledgments**

The author expresses special thanks to Carmen Fernandez and also to Leonhard Held for their guidance and supervision. *Conflict of Interest*: None declared.

**Funding**

This work was funded by the Medical Research Council.

**References**


[Received May 24, 2011; revised June 12, 2012; accepted for publication July 18, 2012]