STATISTICS IN MEDICINE, VOL. 16, 741–752 (1997)

DISEASE MAPPING WITH ERRORS IN COVARIATES

L. BERNARDINELLI, C. PASCUTTO

Dipartimento di Scienze Sanitarie Applicate, University of Pavia, Italy

AND

N. G. BEST* AND W. R. GILKS

Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, U.K.

SUMMARY

We describe Bayesian hierarchical-spatial models for disease mapping with imprecisely observed ecological covariates. We posit smoothing priors for both the disease submodel and the covariate submodel. We apply the models to an analysis of insulin Dependent Diabetes Mellitus incidence in Sardinia, with malaria prevalence as a covariate. © 1997 by John Wiley & Sons, Ltd.

1. INTRODUCTION

Maps that show relative risks of a disease in small geographical areas are important for generating aetiological hypotheses, and for identifying areas that deserve closer scrutiny.[1–5] Maps may show relative risks of death from the disease, or relative risks of disease incidence or prevalence.

When the disease is rare or when geographical areas are small, and when the disease is non-contagious (that is, cases may be considered to occur independently), we may assume a Poisson model for death or disease incidence within each area i

$$d_i \sim \text{Poisson}(\rho_i E_i) \tag{1}$$

where d_i is the number of events, ' $\sim$ ' means 'is distributed as', ρ_i denotes the underlying true area-specific relative risk, and E_i is the 'expected' number of events with control (usually) for age and sex. Thus

$$E_i = \sum_j r_j n_{ij}$$

where j indexes age-sex subgroups, r_j denotes a known reference rate for subgroup j, and n_{ij} denotes the size of subgroup j in area i. If the $\{\rho_i\}$ are not all equal, then the data $\{d_i\}$ display *extra-Poisson variation.*

* Now at Department of Epidemiology and Public Health, Imperial College School of Medicine at St. Mary's, London, U.K.

CCC 0277–6715/97/070741–12$17.50

Received September 1995
Revised May 1996

To investigate whether extra-Poisson variation is geographically related, we would ideally like to map the true relative risks $\{\rho_i\}$. Since these are unobserved, the most obvious strategy is to estimate ρ_i by the empirical relative risk:

$$\hat{\rho}_i = \frac{d_i}{E_i} \tag{2}$$

which is the maximum likelihood estimate of ρ_i. When events are deaths, $\hat{\rho}_i$ is the *standardized mortality ratio* (SMR). Mapping the $\{\hat{\rho}_i\}$, however, can be misleading because sampling variability in the $\{\hat{\rho}_i\}$ can dominate the map and obscure genuine trends. In particular, areas that have exceptionally high or low $\hat{\rho}_i$ tend to be those that have smaller E_i, where sampling variability is most pronounced ($\text{var}\,\hat{\rho}_i = \rho_i/E_i$).

There have been several strategies proposed for dealing with sampling variability in maps. The current state-of-the-art is to adopt a fully-Bayesian hierarchical-spatial model.[6–9] An important feature of this model is that the prior distribution for the $\{\rho_i\}$ incorporates spatial correlation, allowing the estimate of ρ_i to 'borrow strength' formally from neighbouring areas. In this way one smooths the empirical map, and geographical trends and inferences become more reliable.

Mapping Bayesian estimates of relative risk may reveal geographical trends across the map, or may suggest links with area-specific covariates x_i. To incorporate these covariates into the model, a natural assumption, in conjunction with the Poisson assumption (1), is

$$\log \rho_i = \alpha_i + \beta x_i \tag{3}$$

where α_i represents the covariate-adjusted area-specific log relative risk. Such a model is an *ecological regression model*. One can then effect spatial smoothing of the $\{\rho_i\}$ via a smoothing prior on the $\{\alpha_i\}$.

In practice, one rarely observes ecological covariates x_i directly. Available data z_i may be either imperfect measurements of, or proxies for, x_i. The simplest approach to this problem is to estimate x_i from z_i for each area independently, using this estimate $\hat{x}_i$ in place of x_i in the ecological regression (3). When z_i is an accurate measure of x_i, this approach is reasonable. When, however, the correspondence between x_i and z_i is not so close, this approach has several disadvantages. First, the estimate of the regression coefficient β is probably underestimated (see for example reference 10). Second, the precision in parameter estimates or in projections is overestimated, through failure to take account of uncertainty in the $\{\hat{x}_i\}$. Third, when it is reasonable *a priori* to expect spatial correlation in the x_i, one obtains improved estimates of the $\{x_i\}$ and other unknowns through a Bayesian procedure that incorporates a spatial smoothing prior on the $\{x_i\}$.

In this paper we develop Bayesian models with spatial smoothing priors for both relative risks and for imprecisely observed covariates, with estimation using Gibbs sampling.[11,12] We illustrate the models with an analysis of Insulin Dependent Diabetes Mellitus (IDDM) incidence in Sardinia. In this analysis the covariate of interest is malaria prevalence, which historically has varied widely across the island, and is known to have caused genetic selection in the inhabitants. Thus, we investigate the hypothesis that genetic selection has affected susceptibility to IDDM, and is responsible for geographical variation in IDDM incidence. We estimate the models using the Gibbs sampling software BUGS.[13]

These models have been applied in previous papers[1,6] and extended to allow for space–time interaction.[14] Bernardinelli *et al.*[15] discuss issues about the choice of the prior distribution for the dispersion parameter of the log relative risk. The original contribution of the present paper is the introduction of covariate measurement error.

2. A MARKOV RANDOM FIELD PRIOR

In this section we describe a *Markov random field* prior distribution for the $\{\alpha_i\}$ parameters in (3). The development follows that used by Bernardinelli et al.[15]. This prior tends to produce similar estimates for α_i and α_j if areas i and j are geographically close.

The Gaussian Markov random field prior that we employ assumes, for each i, that α_i is normally distributed with a mean and variance that depend on its neighbours. We consider two areas as 'neighbours' if they share a portion of boundary. The conditional prior distribution of α_i given values for $\{\alpha_j, j \neq i\}$ is

$$\alpha_i \sim \mathrm{N}(\mu_{\alpha_i}, \sigma^2_{\alpha_i}) \tag{4}$$

where

$$\mu_{\alpha_i} = \frac{\sum_{j \neq i} w_{ij}\alpha_j}{\sum_{j \neq i} w_{ij}} \tag{5}$$

$$\sigma^2_{\alpha_i} = \frac{1}{\gamma_\alpha \sum_{j \neq i} w_{ij}} \tag{6}$$

where the *adjacency weights* $\{w_{ij}\}$ are fixed constants. In the present example, we set $w_{ij} = 1$ if areas i and j are neighbours; otherwise $w_{ij} = 0$. Other values for w_{ij} are possible (for example, $w_{ij} =$ distance between the 'centres' of regions i and j) but we do not consider such alternatives here.

To ensure that the Gaussian Markov random field model (4), (5) is internally consistent, $\sigma^2_{\alpha_i}$ must depend upon the number of adjacent areas and their adjacency weights. Jointly, the $\{\alpha_i\}$ have an intrinsic multivariate normal prior distribution with *inverse* variance-covariance matrix Λ given by

$$\Lambda_{ij} = \begin{cases} -\gamma_\alpha w_{ij} & j \neq i \\ \gamma_\alpha \sum_j w_{ij} & j = i \end{cases}.$$

Matrix Λ is not full rank. Thus the prior on the $\{\alpha_i\}$ is imporper; adding an arbitrary constant to each α_i does not change the probability (4). This need not concern us since the data d_i contain information on the location of the $\{\alpha_i\}$.

The amount of smoothing in the random effects $\{\alpha_i\}$ is controlled by parameter γ_α in (6). A small value of γ_α induces little smoothing, whilst an infinite value forces all the $\{\alpha_i\}$ to be equal. Since we do not wish to impose any fixed amount of smoothing on these parameters, but rather we wish to let the data themselves determine how much smoothing to induce, we treat γ_α as a model parameter. It is computationally convenient to choose a gamma (a, b) prior distribution for γ_α, for fixed constants a and b. For the present application, we assume a just-proper prior gamma (0·001, 0·001). We also specify a vague normal prior for β.

With these hyperpriors, equations (1) and (3)–(6) specify a full probability model for the data $\{d_i, E_i, x_i\}$, assuming the $\{x_i\}$ are accurately observed. We next consider the problem of imperfectly observed covariates.

3. ERRORS IN COVARIATES

Frequently, ecological covariates x_i are not accurately observed. Sometimes one may use epidemiological data concerning another disease as a proxy variable, z_i. For example, if d_i records deaths from heart disease in area i, an important covariate is the proportion of area residents who smoke. Such data on smoking are generally unavailable, so the incidence of lung cancer as recorded by the cancer registry for the area might be a useful proxy. In Section 5, x_i is related to underlying malaria prevalence, and z_i is observed malaria prevalence.

When the proxy variable z_i relates to another disease, it may be natural to use a spatial smoothing prior for the unobserved covariate x_i, such as a Gaussian Markov random field prior of the form described in Section 2. In anticipation of the analysis of IDDM and malaria in Section 5, assume that

$$z_i \sim \text{binomial}(n_i, \theta_i) \tag{7}$$

where n_i is the population size of area i, and

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = x_i. \tag{8}$$

Thus we take the covariate x_i in (3) as the logistic-transformed expectation of z_i. We assume for x_i the same form of spatial smoothing prior used in Section 2:

$$x_i \sim \mathrm{N}(\mu_{x_i}, \sigma^2_{x_i}) \tag{9}$$

where

$$\mu_{x_i} = \frac{\sum_{j \neq i} w_{ij} x_j}{\sum_{j \neq i} w_{ij}}$$

$$\sigma^2_{x_i} = \frac{1}{\gamma_x \sum_{j \neq i} w_{ij}}. \tag{10}$$

We specify a gamma(0·001, 0·001) prior for γ_x.

Note that the amount of smoothing in the two Markov random field priors (4)–(6) and (9)–(10) may differ, since the smoothing is controlled by different parameter γ_α and γ_x.

4. ESTIMATION

We have specified two arms to the model: the disease model (1), (3)–(6) and the covariate model (7)–(10). One approach to estimation is first to estimate the covariate model, and then to substitute resulting estimates of the $\{x_i\}$ into the disease model. This two-stage approach, however, suffers from some of the deficiencies noted in Section 1. In particular, inferences and predictions from the disease model tend to be over-confident, through failure to acknowledge uncertainty in the $\{x_i\}$. A better approach is to estimate all parameters (including β) simultaneously, treating equations (1), (3)–(10) as a single large model. This is quite feasible using Markov chain Monte Carlo methods such as Gibbs sampling. Indeed, it is probably simpler to do this than to attempt the two-stage procedure indicated above. This can be done using the BUGS software.[13]

Figure 1 shows structural relationships between the model quantities in the form of a conditional independence graph.

5. IDDM AND MALARIA IN SARDINIA

5.1. Background

There is scientific interest in studying the association between IDDM and malaria, since both are associated with the HLA system. Recent evidence suggests that resistance to Plasmodium

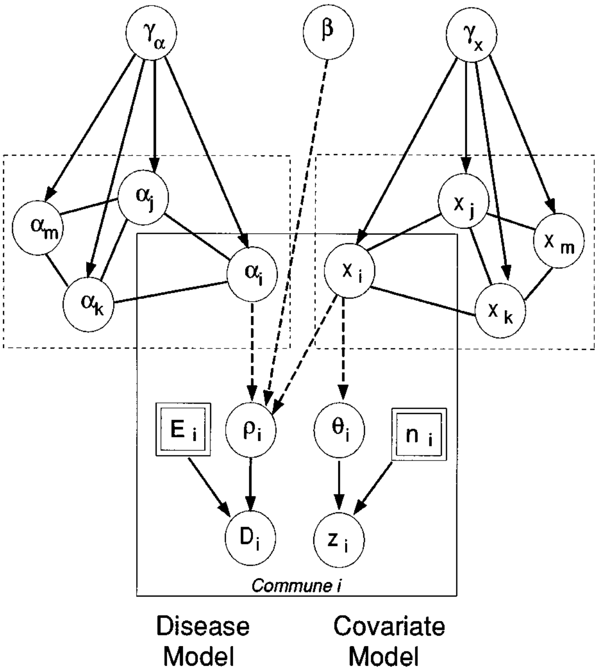

Figure 1. A conditional independence graph corresponding to equations (1), (3)–(10). Square nodes denote constants; round nodes denote model parameters and data; solid arrows denote stochastic dependence as specified in the model equations; dashed arrows denote deterministic relationship; undirected links denote bidirectional stochastic dependence between parameters of neighbouring communes

Falciparum malaria in West Africa is associated with a human class I allele, HLA-B53, and an unusual class II haplotype.[16] Family studies have found genetic linkage between HLA and IDDM, suggesting that genes within or near the HLA region are involved in susceptiblity to IDDM.[17]

Sardinia is a particularly suitable place to investigate the association between IDDM and malaria. Malaria spread gradually all over Sardinia after the Carthaginian conquest, became established after Roman occupation and remained stable until the end of the 19th century. It was completely eradicated in 1950. Malaria has for centuries been a major cause of death in the island. Population genetic studies carried out by Piazza[18] suggest that, in the plains of Sardinia where malaria has been endemic, some genetic traits have adapted to provide greater resistance to the haemolysing action of Plasmodium. In the hilly and mountainous areas, where malaria has been absent, such adaptation has not occurred.

IDDM incidence in Sardinia is the second highest in Europe (35 per 100,000 person years) after Finland (40 per 100,000). Sardinia is a striking exception to the north-south downward trend of IDDM in Europe, being quite atypical of other Mediterranean countries.[19] From studies carried out on the cumulative prevalence of IDDM in military conscripts, the risk for IDDM in Sardinia began to increase with the male birth cohort of 1950.[20]

5.2. The data

We calculated IDDM incidence from a case registry that has operated in Sardinia since 1989. The incidence data refer to the period 1989–1992 and cover the population aged 0–29 years. We let

d_i denote the number of IDDM cases in commune i ($i = 1, 2, \dots, N = 366$), and E_i denote the expected number of IDDM cases based on Sardinian national rates.

Fermi[21,22] has recorded the number of individuals affected by malaria, z_i, for each commune during the period 1938–1940. We obtained the population n_i for each commune from the 1936 census.

5.3. Results

We estimated the models using BUGS. In each case, we ran the Gibbs sampler for 7500 iterations. We checked convergence of the simultions with a variety of diagnostics implemented in the CODA[23] software, and discarded the first 2500 iterations of each run as 'burn-in' or 'precovergence' samples. Posterior inference was thus based on empirical summaries of the final 5000 samples in each run. Computation times ranged from 40–110 minutes per run on a SPARCcentre 2000.

The maps in Figures 2(*a*) and (*b*) show ρ_i, the estimated relative risk of IDDM, obtained by maximum likelihood (2) and by the Bayesian approach with model (1), (3)–(6). In Figure 2(*b*), relative risks are less variable than in Figure 2(*a*), and show considerable smoothing.

The map in Figure 3(*a*) shows the observed prevalence of malaria (z_i/n_i) in each commune during 1938–1940. This map has been smoothed in Figure 3(*b*) by plotting θ_i from the Bayesian model (7)–(10). Since the number of malaria cases is quite high in all the Sardinian communes, the two maps are almost identical. That is, there appears to be little sampling error in the malaria data. The communes for which Figures 3(*a*) and (*b*) do differ are generally quite small and have extreme values of prevalence.

Figures 2(*b*) and 3(*b*) do not exhibit a clear association between malaria prevalence and IDDM incidence. To examine this more formally, we introduced malaria prevalence as a covariate in the model. To begin with, we ignored sampling error in the observed proportions z_i/n_i and simply substituted

$$x_i = \log\left(\frac{z_i}{n_i - z_i}\right) \tag{11}$$

directly into (3). We then calculated Bayesian estimates for model (1), (3)–(6). This resulted in a posterior mean for β of $-0{\cdot}036$, with 95 per cent Bayesian credible interval $[-0{\cdot}066, -0{\cdot}007]$. This indicates that communes with historically high malaria prevalence currently have relatively low incidence of IDDM and *vice versa*. Our next step was to fit model (1), (3)–(10) which allows for sampling error and spatial correlation in the covariate. This gave a posterior mean of $-0{\cdot}039$ for β, with 95 per cent credible interval $[-0{\cdot}071, -0{\cdot}008]$.

The negligible increase in the estimated magnitude of β after accounting for covariate sampling error supports our earlier suggestion of quite accurate measurements of malaria prevalence in 1938–1940. Recall, however, that the hypothesis of interest concerns how genetic adaptation in areas of endemic malaria affects susceptibility to IDDM. Thus, the true covariate is the long-term malaria endemicity averged over may centuries in each commune.

Consider the following model for the observed covariate z_i/n_i (malaria prevalence between 1938–1940):

$$z_i \sim \text{binomial}(n_i, \theta_i)$$
$$\log\left(\frac{\theta_i}{1 - \theta_i}\right) \sim \text{N}(x_i, \omega). \tag{12}$$

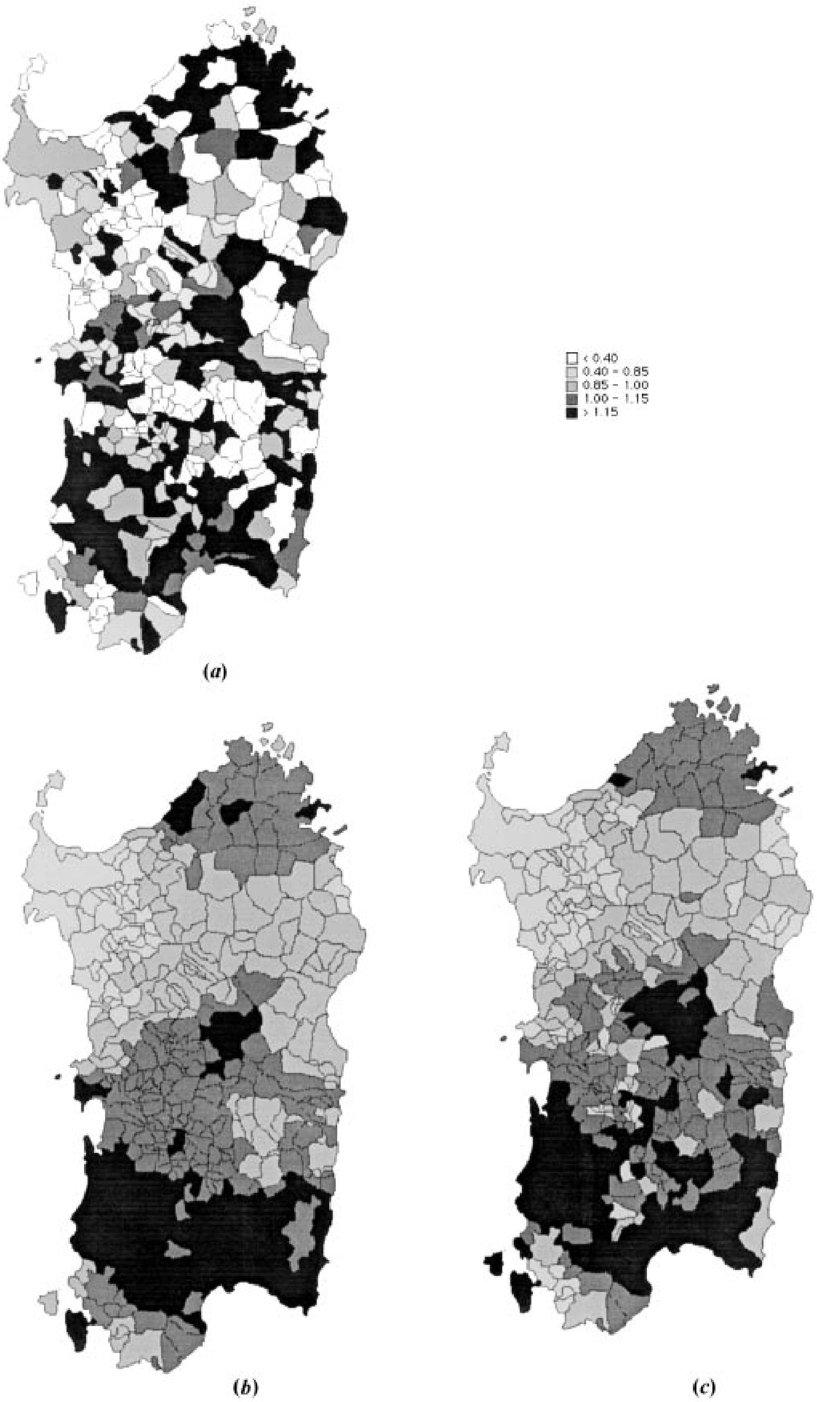

Figure 2. (*a*) Diabetes SMRs ($\hat{\rho}_i$) calculated from Sardinian regional IDDM rates, estimated by equation (2). (*b*) Bayesian estimates of the relative risk of IDDM (ρ_i) in model (1), (3)–(6), but without covariates ($\beta = 0$). (*c*) Bayesian estimates of the relative risk of IDDM (ρ_i) in model (1), (3)–(10), (12), allowing for spatial correlation and long-term error with fixed variance $\omega = 2.25$ in the covariate

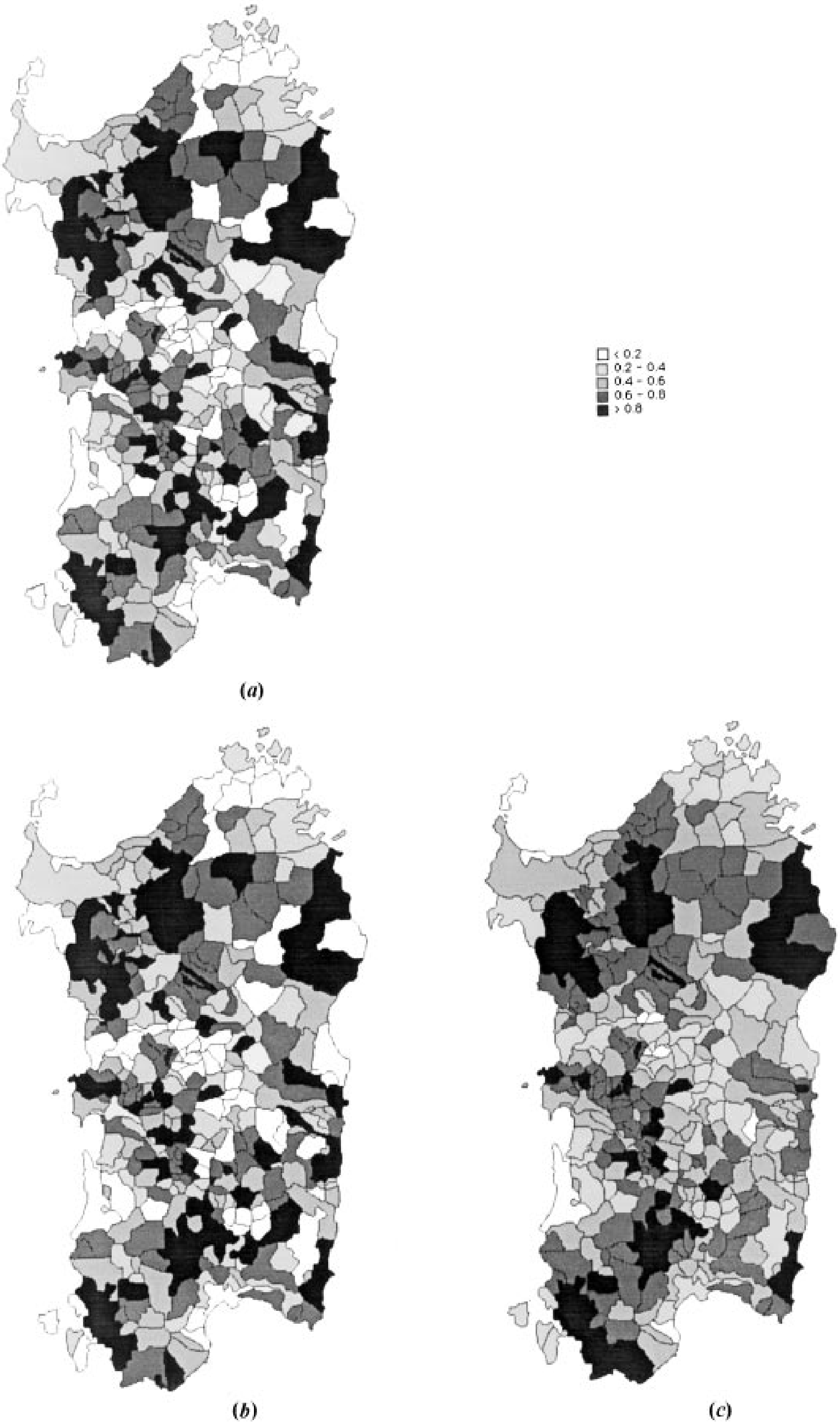

Figure 3. (*a*) Malaria prevalence in Sardinia in the period 1938–1940: proportion of the population affected (z_i/n_i). (*b*) Bayesian estimates of malaria prevalence θ_i in model (7)–(10). (*c*) Bayesian estimates of malaria prevalence θ_i in model (9), (10), (12), obtained by setting the long-term variance $\omega = 2.25$

Table I. Relative risk of IDDM for selected communes estimated according to various covariate models of malaria prevalence. MLE refers to the maximum likelihood estimate given by equation (2); model A refers to the disease model (1), (4)–(6) with $\log \rho_i = \alpha_i$; model B refers to model (1), (3)–(6) which allows dependence of ρ_i on the *observed* malaria covariate; model C refers to model (1), (3)–(10) which allows for spatial correlation in the covariance; model D refers to model (1), (3)–(6), (9), (10), (12) which incorporates spatial correlation and long-term error in the covariate

Commune	Observed malaria prevalence	Model	RR of IDDM ρ_i	(95% CI)	RR associated with malaria $e^{\beta x_i}$
Low-lying areas					
33	77·0%	MLE	$\frac{4}{5.54} = 0{\cdot}72$	—	—
		A	0·80	(0·53, 1·08)	—
		B	0·76	(0·53, 1·03)	0·96
		C	0·77	(0·57, 0·99)	0·95
		D	0·75	(0·54, 0·98)	0·89
42	96·1%	MLE	$\frac{0}{0.69} = 0{\cdot}00$	—	—
		A	0·86	(0·54, 1·22)	—
		B	0·77	(0·59, 1·09)	0·89
		C	0·77	(0·55, 1·04)	0·88
		D	0·76	(0·53, 1·02)	0·84
Mountainous/hilly areas					
90	12·0%	MLE	$\frac{0}{0.39} = 0{\cdot}00$	—	—
		A	1·13	(0·71, 1·78)	—
		B	1·19	(0·78, 1·79)	1·07
		C	1·18	(0·83, 1·66)	1·08
		D	1·21	(0·84, 1·72)	1·11
103	3·0%	MLE	$\frac{1}{0.27} = 3{\cdot}70$	—	—
		A	0·97	(0·62, 1·46)	—
		B	1·09	(0·70, 1·63)	1·13
		C	1·12	(0·79, 1·56)	1·14
		D	1·15	(0·80, 1·83)	1·15

Comparison of (12) with (8) shows that we have replaced a deterministic relationship with a stochastic relationship, and have thus introduced an extra layer of uncertainty into the model. This corresponds to replacement of the dashed arrow between x_i and θ_i in Figure 1 with a solid arrow, and inclusion of an extra node that represents ω. We continue to assume the spatial smoothing prior (9)–(10) for x_i.

We can interpret model (12) as follows: the true log odds of malaria in commune i in 1938–1940 (that is, $\log(\theta_i/(1-\theta_i))$) represents a single realization from a latent Normal distribution with mean x_i (that is, the long-term average endemicity of malaria in commune i) and unknown long-term variance, ω. Since the data contain no information by which to estimate ω, we must fix its value *a priori*. We carried out exploratory analyses to select a suitable value as follows. Using only the malaria prevalence data, we repeatedly fitted model (9)–(10), (12) using different values for ω in the range 0·1–10. We produced maps of the estimated long-term prevalence $\psi_i = e^{x_i}/(1+e^{x_i})$ for each value of ω. We compared these subjectively 'by eye' to identify one that showed a 'reasonable' amount of smoothing, that is, a map neither under-smoothed (neighbouring regions showing sharp changes in colour, giving the map a 'speckled' appearance) or over-smoothed (virtually all regions mapped with the same colour). Figure 3(*c*) shows the selected map, which corresponded to $\omega = 2{\cdot}25$. We then substituted this value into the full model (1), (3)–(6), (9), (10), (12), to estimate the relationship between long-term malaria endemicity and susceptibilty to IDDM. Figure 2 (*c*) shows the estimated relative risk of IDDM obtained for the above model with allowance for spatial correlation and long-term error in the covariate. The posterior mean of β is now $-0{\cdot}060$ with 95 per cent Bayesian credible interval $[-0{\cdot}112, -0{\cdot}012]$. In addition, we estimated the correlation coefficient between relative risk of IDDM (ρ_i) and long-term malaria prevalence (ψ_i) in each region as $-0{\cdot}568$, with 95 per cent credible interval $[-0{\cdot}812, -0{\cdot}182]$.

Some low-lying areas near the north-west coast of Sardinia are characterized by a historically high prevalence of malaria, whilst prevalence has tended to be low in the mountainous central regions (Figure 3(*a*)). Table I gives the overall relative risk of IDDM in a selection of these communes, estimated according to each model described above. To illustrate the influence of malaria prevalence on the estimated risk of diabetes, we have also calculated $e^{\beta x_i}$, the component of the overall relative risk attributable specifically to the covariate. This appears in the final column of Table I.

6. CONCLUSIONS

6.1. Substantive conclusions

Sardinians are known for their susceptibility to autoimmune diseases. The significant negative association that emerged between long-term malaria endemicity and diabetes indicates that people who live in areas where malaria has been particularly frequent have a lesser risk of IDDM than those who live in areas with a low prevalence of malaria as observed in 1938. This is illustrated by the relative risk estimates in Table I: diabetes risk is consistently lower in the low-lying regions than in the hills and mountains. A possible interpretation of this finding is that, since malaria has been endemic in the plains of Sardinia for centuries, places with high prevalance of malaria in 1938 are those in which a stronger selection process took place, both providing resistance to malaria and preventing the onset of autoimmune conditions.[24]

The estimated correlation coefficient of nearly $-0{\cdot}6$ represents an alternative means to quantify the association between incidence of IDDM and possible genetic selection due to past prevalance of malaria. Although the 95 per cent credible interval of $[-0{\cdot}812, -0{\cdot}182]$ is rather wide, and suggests uncertainty about the true strength of this relationship, it does tend towards scientifically significant values.

6.2. Methodological issues

We have shown that one can construct disease maps that take account of covariates measured with error using Bayesian hierarchical-spatial models, where one posits spatial smoothing priors for both disease relative risks and underlying covariates. We consider that such models have particular value when the covariates are themselves incidence or prevalence data for other diseases.

Our choice of prior for malaria prevalence is particularly suitable since it varies between areas in a spatially structured way. Malaria prevalence tends to be higher in low lying and damp regions and lower in the mountains and hills. The spatial prior enables us to obtain a map of this geographical variation in malaria prevalence in which we have filtered out the random variation.

Specification of a spatial smoothing prior for the disease relative risks represents a way to allow for unmeasured risk factors (other than malaria) that vary smoothly with location. If the pattern of variation in such covariates is similar to that of disease risk, location may act as a confounder. Of course, the location effect is only a surrogate for other confounding factors. Introduction of a spatial prior to model the effect of location thus causes the estimate of the regression coefficient β to be controlled for these factors.

One potentially controversial aspect of our analysis concerns the subjective choice of ω, the long-term variation parameter used in the final model. Future studies could plan to obtain repeated measurements of the covariate to estimate the value of ω statistically. Although such data were unavailable for the present application, we did examine the sensitivity of the β estimate to different values of ω. For $\omega > 2{\cdot}25$ (that is, increased long-term variation) the posterior mean of β became slightly more negative, with a wider credible interval. Values of ω much less than 0·5 yielded estimates of β similar to those obtained using the model (1), (3)–(10), which does not account for long-term sampling error.

The results of the present analysis are typical of many studies that involve covariate measurement error: when not properly accounted for, such errors tend to disguise real associations. By fully acknowledging all potential sources of error in our final model, we could clarify considerably the relationship between malaria prevalence and incidence of IDDM – the estimate of β almost doubled in magnitude after accounting for both sampling and measurement error. In addition, Table I illustrates how an increase in the sophistication of the model, although producing quite small numerical changes in each relative risk estimate, does reduce the credible interval and produce greater distinction between low-lying and hilly areas. Interpretation of any model requires care, however, and ecological regression models are no exception. There is a danger of misinterpreting association as causation. With models as complex as those considered here, it is particularly important to investigate the sensitivity of any conclusions to changes in model specification. In this regard our analysis of the IDDM/malaria data is continuing. We also plan to apply our model to epidemiological data on other diseases.

ACKNOWLEDGEMENTS

The support of the Economic and Social Research Council (UK) is gratefully acknowledged. The work was funded in part by ESRC (UK) Award Number H519 25 5023 and MURST 60% (I).

REFERENCES

1. Clayton, D. G., Bernardinelli, L. and Montomoli, C. 'Spatial correlation in ecological analysis', *International Journal of Epidemiology*, **22**, 1193–1202 (1994).
2. Kemp, I., Boyle, P., Smans, M. and Muir, C. *Atlas of Cancer in Scotland*, 1975–1980. *Incidence and Epidemiologic Perspective*, IARC Scientific Publication No. 72, International Agency for Research on Cancer, Lyon, 1985.
3. Morgenstein, H. 'Uses of ecologic analysis in epidemiologic research', *American Journal of Public Health*, **72**, 1336–1344 (1982).
4. Walter, S.D. 'The ecologic method in the study of environmental health. I. Overview of the method', *Environmental Health Perspective*, **94**, 61–65 (1991).
5. Walter, S. D. 'The ecologic method in the study of environmental health. II. Methodologic issues and feasibility', *Environmental Health Perspective*, **94**, 67–73 (1991).
6. Bernardinelli, L. and Montomoli, C. 'Empirical Bayes versus Fully Bayesian analysis of geographical variation in disease risk', *Statistics in Medicine*, **11**, 983–1007 (1992).
7. Besag, J., York, J. and Mollié, A. 'Bayesian image restoration, with applications in spatial statistics (with discussion), *Annals of the Institute of Statistics and Mathematics*, **43**, 1–59 (1991).
8. Clayton, D. G. 'Hierarchical model in descriptive epidemiology', *Proceeding of the XIVth International Biometrics Conference*, 201–213 (1989).
9. Clayton, D. G. and Bernardinelli, L. 'Bayesian methods for mapping disease risk', *in* Elliot, P., Cuzick, J., English, D. and Stern, R. (eds), *Geographical and Environmental Epidemiology: Methods for Small Area Studies*, Oxford University Press, New York (1992).

10. Richardson, S. and Gilks, W. R. 'Conditional independence models for epidemiological studies with covariate measurement error', *Statistics in Medicine*, **12**, 1703–1722 (1993).
11. Geman, S. and Geman, D. 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741 (1984).
12. Gelfand, A. E. and Smith, A. F. M. 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1990).
13. Spiegelhalter, D. J., Thomas, A., Best, N. G. and Gilks, W. R. *BUGS*: *Bayesian inference Using Gibbs Sampling*, Version 0·50. Medical Research Council Biostatistics Unit, Cambridge, 1995.
14. Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. and Songini, M. 'Bayesian analysis of space-time variation in disease risk', *Statistics in Medicine*, **14**, 2433–2443 (1995).
15. Bernardinelli, L., Clayton, D. and Montomoli, C. 'Bayesian estimates of disease maps: how important are priors?', *Statistics in Medicine*, **14**, 2411–2431 (1995).
16. Ebert, D. and Lorenzi, R. 'Parasites and polymorphisms', *Nature*, **369**, (1994).
17. Todd, J. A., Bell, J. I. and McDevitt, H. O. 'A molecular basis for genetic susceptibility in insulin dependent diabetes mellitus', *Trends in Genetics*, **4**, 129–134 (1988).
18. Piazza, A., Mayr, W. R., Contu, L., Amoroso, A., *et al.* 'Genetic and population structure of four Sardinian villages', *Annals of Human Genetics*, **4**, 47–63 (1985).
19. Muntoni, S. and Songini, M. 'High incidence rate of IDDM in Sardinia', *Diabetes Care*, **15**, 1317–1322 (1992).
20. Songini, M., Loche, M. and Muntoni, S. 'Increasing prevalence of juvenile onset type-1 (insulin dependent) diabetes mellitus in Sardinia: the military service approach', *Diabetologia*, **36**, 457–552 (1993).
21. Fermi, C. 'Provincia di Nuoro. Malaria, danni economici. Risanamento e proposte per il suo risorgimento', *Gallizzi*, **2**, 1–311 (1938).
22. Fermi, C. 'Provincia di Cagliari, Malaria, danni economici. Risanamento e proposte per il suo risorgimento', *Gallizzi*, **3**, 1–610 (1940).
23. Best, N. G., Cowles, M. K. and Vines, S. K. *CODA*: *Convergence Diagnosis and Output Analysis for Gibbs sampling output*, Version 0·30. Medical Research Council Biostatistics Unit, Cambridge, 1995.
24. Jacob, C. O. 'Tumor necrosis factor α in autoimmunity: pretty girl or old witch?', *Immunology Today*, **13**, 122–125 (1992).