16

Multilevel Modelling of Area-Based Health Data

Ian H. Langford
University of East Anglia and University College, London

Alastair H. Leyland
University of Glasgow

Jon Rasbash
University of London

Harvey Goldstein
University of London

Rosemary J. Day
University of East Anglia and University College, London

Anné-Lise McDonald
University of East Anglia and University College, London

16.1 INTRODUCTION

Multilevel modelling is a form of random coefficient modelling which was first used in an educational setting (Goldstein, 1995). It is appropriate for data that have a natural hierarchy, such as educational data, where pupils are nested within classes within schools. However, it is also appropriate for geographically distributed data, where we may have individual cases of diseases nested within households, within postcode sectors, and so on. At larger scale, we may have ecological models where data are collected

Disease Mapping and Risk Assessment for Public Health. Edited by A.B. Lawson *et al.*

in larger administrative units, such as local authority districts, which are in turn nested within regions and nations. There are several examples of multilevel modelling being used in environmental and geographical epidemiology now appearing in the literature (see Langford *et al.*, 1998). Epidemiologists are increasingly using more complex methods of statistical analysis to investigate the distribution of diseases (Elliott *et al.*, 1992b, 1995), and the motivation behind this chapter is to introduce a multilevel modelling framework which also allows for the analysis of complex spatial processes such as:

(i) spatial autocorrelation in a Poisson generalised least squares model, and
(ii) simultaneous modelling of spatial effects that occur at different scales in a geographical hierarchy.

In the example given in this chapter we concentrate on investigating data that consist of observed and expected counts of disease occurring in discrete spatial units. Hence, for a sample of geographical areas we have a number of cases occurring within a distinct population at risk in each area. Whether we are embarking on an exploratory analysis, where we are simply interested in producing a map of the relative risks of disease, or an inferential analysis, where we are interested in investigating potential causal factors, it is useful to break down the likely effects on the distribution of a disease into three separate categories:

(i) Within-area effects, such as social characteristics of the population at risk. Since we are modelling aggregated data, we do not have information on individual cases, although we may have aggregate information on the mean and variance of social indicators such as income, employment and so on for each area. However, we can at least model these unmeasured variables by allowing for extra-Poisson variation in our model.
(ii) Hierarchical effects. These are due to the fact that small areas are grouped into larger areas, for administrative purposes, or for cultural and geographical reasons. For instance, in the example we present on mortality from prostate cancer in Scotland, local authority districts are grouped into Health Boards, which have different methods of treatment or classification of a disease, or different ascertainment rates. Again, if we have accurate information on these factors, then we could include them directly in the model, but we can allow for random variation between Health Boards even if we do not know the direct causes of this variation.
(iii) Neighbourhood effects. Areas that are close to each other in geographical space may share common environmental or demographic factors which influence the incidence or outcome of disease, but have a smoother distribution than that of the disease. For example, climatic factors such as temperature may vary between different part of a nation, but not at the smaller scale of a local authority district. In addition, as areas are usually formed from geopolitical boundaries that have nothing to do with the disease we are interested in, we may wish to spatially smooth the distribution or relative risks to remove any artifactual variation brought into the data by the method of aggregating the data.

The use of empirical Bayes and fully Bayesian techniques has allowed for alternative models of spatial and environmental processes affecting the distribution of a disease which rely on different underlying beliefs or assumptions about aetiology (Bernardinelli *et al.*, 1995a; Bernardinelli and Montomoli, 1992; Cisaghli *et al.*, 1995; Clayton and Kaldor, 1987; Langford *et al.*, 1998; Langford, 1994; 1995; Lawson, 1994; Lawson and Williams, 1994; Mollié and Richardson, 1991; Schlattmann and Böhning, 1993). Two main statistical techniques have been used to model geographically distributed health data in this way. The first is Markov chain Monte Carlo (MCMC) methods, using Gibbs sampling (Gilks *et al.*, 1993) often fitted using the BUGS software (Spiegelhalter *et al.*, 1995), which can be used to fit fully Bayesian or emprical Bayesian models. The second set of methods is multilevel modelling techniques based on iterative generalised least squares procedures (IGLS) and are the focus of this chapter.

In the following section we discuss the basic multilevel Poisson model, and develop a computational method for modelling spatial processes within the software package MLn. MLn (for MS-DOS and Windows 3.1) and its successor MLwiN (for Windows 97 and NT) are widely used tools for multilevel modelling, and information about them can be found from a number of websites worldwide; for information, see `http://www.ioe.ac.uk/multilevel/`. We then present an example of how our model can be used using morbidity data for prostate cancer in Scottish local authority districts, and comment on how the results may be interpreted. The discussion section then focuses on methodological and substantive issues in a more general setting, and discusses work in progress to generalise the procedures we have developed.

16.2 DEVELOPING A POISSON SPATIAL MULTILEVEL MODEL

The basic model of fixed and random effects described by Goldstein (1995) and Breslow and Clayton (1993) is

$$Y = X\beta + Z\theta, \tag{16.1}$$

with a vector of observations Y being modelled by explanatory variables X and associated fixed parameters β, and explanatory variables Z with random coefficients θ. The fixed and random part design matrices X and Z need not be the same. θ is assumed to contain a set of random error terms in addition to other random effects. Goldstein (1995) describes a two-stage process for estimating the fixed and random parameters (the variances and covariances of the random coefficients) in successive iterations using IGLS. A summary of this process follows.

First, we estimate the fixed parameters in an initial ordinary least squares regression, assuming the variance at higher levels on the model to be zero. From the vector of residuals from this model we can construct initial values for the dispersion matrix V. Then, we iterate the following procedure, first estimating fixed parameters in a generalised least squares regression as

$$\hat{\beta} = (X^{\mathrm{T}}V^{-1}X)^{-1}X^{\mathrm{T}}V^{-1}Y \tag{16.2}$$

and again calculating residuals $\tilde{Y} = Y - X\hat{\beta}$. By forming the matrix product of these residuals, and stacking them into a vector, i.e. $Y^* = \mathrm{vec}(\tilde{Y}\tilde{Y}^{\mathrm{T}})$, we can estimate the variance of the random coefficients θ, $\gamma = \mathrm{cov}(\theta)$, as,

$$\hat{\gamma} = (Z^{*\mathrm{T}}V^{*-1}Z^*)^{-1}Z^{*\mathrm{T}}V^{*-1}Y^*, \tag{16.3}$$

where V^* is the Kronecker product of V, namely $V^* = V \otimes V$, noting that $V = \mathrm{E}(\tilde{Y}\tilde{Y}^{\mathrm{T}})$, Z^*

is the appropriate design matrix for the random parameters. Assuming multivariate Normality, the estimated covariance matrix for the fixed parameters is

$$\text{cov}(\hat{\beta}) = (X^{\mathrm{T}}V^{-1}X)^{-1}, \tag{16.4}$$

and for the random parameters, Goldstein and Rasbash (1992) show that

$$\text{cov}(\hat{\gamma}) = 2(Z^{*\mathrm{T}}V^{*-1}Z^{*})^{-1}. \tag{16.5}$$

We can therefore estimate random parameters, and their variances, in the same way as we estimate fixed parameters and their variances from the model. To compare what we are doing with ordinary least squares resgression, we are extend the process by modelling the random part of the model with respect to the structure of our data, estimating a set of parameters rather than simply having a residual error term.

However, we now need to develop a model for the relative risks of a disease. If we consider a population of areas with O_i observed cases and E_i expected cases and relative risk $\theta_i = O_i/E_i$, where E_i may be calculated from the incidence in the population N_i for each area as

$$E_i = N_i \cdot \frac{\sum O_i}{\sum N_i} \tag{16.6}$$

and may be additionally divided into different age and sex bands, then we can write the basic Poisson model as

$$\begin{aligned} O_i &\sim \text{Poisson}(\mu_i), \\ \log(\mu_i) &= \log(E_i) + \alpha + x_i\beta + u_i + v_i, \end{aligned} \tag{16.7}$$

where $\log(E_i)$is treated as an offset, α is a constant and x_i is an explanatory variable with coefficient β (this may be generalised to a number of explanatory variables). We take account of the distribution of cases *within* each area by assuming that the cases have a Poisson distribution. In contrast, the μ_i represent heterogeneity effects between areas (Clayton and Kaldor, 1987; Langford, 1994), which may be viewed as constituting extra-Poisson variation caused by the variation among underlying populations at risk in the areas considered. The v_i are spatially dependent random effects, and may have any one of a number of structures describing adjacency or nearness in space (Besag *et al.*, 1991). Hence, we have a hierarchical model where within-area effects are modelled with a Poisson distribution (the first line of (16.7)) and relative risks between areas are considered as having a lognormal distribution (the second line of (16.7)). Other formulations for spatial effects are possible using normal approximations with covariance priors (see,for example, Besag *et al.*, 1991; Bailey and Gatrell, 1995; Lawson *et al.*, 1996).

Before discussing the structure of these spatial effects, we must first account for the fact that we have a non-linear (logarithmic) relationship between the outcome variable and the predictor part of the model. There are two options:

(i) If the cases in each area are sufficiently large, say $O_i > 10$, then it may be reasonable to model the logarithm of the relative risks directly (Clayton and Hills, 1993), assuming these follow a Normal distribution. In this case, heterogeneity effects can be accommodated by weighting the random part of the model by some function of the population at risk in each area.

(ii) When the Normal approximation is inappropriate, we can make a linearising approximation to estimate the random parameters. If we take the case of having heterogeneity effects only for the sake of simplicity, we can estimate the residuals $\hat{u}_i$ from the model using penalised quasi-likelihood (PQL) estimation with a second-order Taylor series approximation (Breslow and Clayton, 1993; Goldstein, 1995; Goldstein and Rasbash, 1992). After each iteration t we make predictions H_i from the model, where $H_t = X_i\hat{\beta}_t + \hat{u}_i$, and hence use these to calculate new predictions for iteration $t+1$, so that

$$\begin{aligned} f(H_{t+1}) = f(H_t) &+ x_i(\hat{\beta}_{t+1} - \hat{\beta}_t)f'(H_t) \\ &+ \hat{u}_i f'(H_t) + \hat{u}_i^2 f''(H_t)/2, \end{aligned} \tag{16.8}$$

where the first two terms on the right-hand side of (16.8) provide the updating function for the fixed part of the model, and $f(\cdot)$ is a link function. The third term comprises a linear random component created by multiplying the first differential of the predictions by the random part of the model, and the fourth term is the next term in the Taylor expansion about H_t. For the Poisson distribution:

$$f(H) = f'(H) = f''(H) = \exp(X_i\hat{\beta}_t + \hat{u}_i). \tag{16.9}$$

Hence, at each iteration we estimate about the fixed part of the model plus the residuals. A full description of this procedure can be found in Goldstein (1995) and Goldstein and Rasbash (1992). This can lead to problems with convergence, or with the model 'blowing up' if some of the residuals are particularly large. In these cases, the second-order term in (16.8) can be omitted, or, in extreme cases, estimates can be based on the fixed part of the model only. This latter case is called marginal quasi-likelihood (MQL: Breslow and Clayton, 1993; Goldstein, 1995), but may lead to biased parameter estimates. However, bootstrap procedures can potentially be used to correct for these biases (Goldstein, 1996a,b; Kuk, 1995). For (16.7) we substitute $\hat{u}_i + \hat{v}_i$ for $\hat{u}_i$, in (16.8) and (16.9)

There are several possibilities for specifying the structure of the random effects in the model (see, for example, Besag *et al.*, 1991, and Bailey and Gatrell, 1995). These models assume two components, namely a random effects or 'heterogeneity' term and a term representing the spatial contribution of neighbouring areas as in (16.7).

We adopt a somewhat different approach, which allows a more direct interpretation of the model parameters and can be fitted in a computationally efficient manner within a multilevel model. For the heterogeneity effects, this is not a problem, because we simply have a variance–covariance matrix with 1 or other specified values on the diagonal, and the model is analogous to fitting an iteratively weighted least squares model (McCullagh and Nelder, 1989). However, the case of the spatial effects is more complex, because we require off-diagonal terms in the variance–covariance matrix. This can be achieved through careful consideration of the structure of the spatial part. Our formulation of the spatial model is to consider the spatial effects v_i to be the weighted sum of a set of independent random effects v_i^* such that

$$v_i = \sum_{j \neq i} z_{ij} v_j^*. \tag{16.10}$$

The v_i^* can be considered to be the effect of area upon other areas, moderated by a

measure of proximity of each pair of areas z_{ij}. The v_i^* can be estimated directly from the model—these are the residuals—due to their independence. Returning to the matrix notation used in (16.1), we can rewrite (16.7) as

$$Y = \{\log(E_i)\ 1\ x_i\}\begin{bmatrix}1\\ \alpha\\ \beta\end{bmatrix} + [Z_u Z_v^*]\begin{bmatrix}\theta_u\\ \theta_v^*\end{bmatrix}, \tag{16.11}$$

where Z_u is the identity matrix and $Z_v^* = \{z_{ij}\}$. With a variance structure such as

$$\text{var}\left(\begin{bmatrix}\theta_u\\ \theta_v^*\end{bmatrix}\right) = \begin{bmatrix}\sigma_u^2 I & \sigma_{uv} I\\ \sigma_{uv} I & \sigma_v^2 I\end{bmatrix}, \tag{16.12}$$

which is equivalent to

$$\text{var}\left(\begin{bmatrix}u_i\\ v_i^*\end{bmatrix}\right) = \begin{bmatrix}\sigma_u^2 & \sigma_{uv}\\ \sigma_{uv} & \sigma_v^2\end{bmatrix},$$

the overall variance from (16.1), conditional on the fixed parameters, is given by

$$\text{var}(Y|X\beta) = Z\sum\nolimits_\theta Z^{\mathrm{T}}, \tag{16.13}$$

where $\sum_\theta$ is the variance of the random terms in θ. The structure of $\sum_\theta$ will often lead to simplifications; for example, in a random effects model when $\theta = \{u_i\}$ and $\text{var}(u_i) = \sigma_u^2$, $\text{cov}(u_i, u_j) = 0$ then $\sum_\theta = \sigma_u^2 I$ and so $\text{var}(Y|X\beta) = \sigma_u^2 ZZ^{\mathrm{T}}$. Similarly, in the spatial model defined by the partitions in θ and Z given by (16.11) and the variance structure of (16.12), we can see that

$$\text{var}(Y|X\beta) = \sigma_u^2 Z_u Z_u^{\mathrm{T}} + \sigma_{uv}(Z_u Z_v^{*\mathrm{T}} + Z_v^* Z_u^{\mathrm{T}}) + \sigma_v^2 Z_v^* Z_v^{*\mathrm{T}}. \tag{16.14}$$

There are many ways in which the z_{ij} can be formulated; in general we can write

$$z_{ij} = w_{ij}/w_{i+}. \tag{16.15}$$

The w_{ij} can either 1's and 0's representing an adjancency matrix, or be functions of the distance between areas (see Section 16.3). Common choices for the w_{i+} would be $w_{i+} = (\sum_{j\neq i} w_{ij})^{0.5}$, which ensures that the variance contribution is the same for all areas, or $w_{i+} = \sum_{j\neq i} w_{ij}$, in which case the variance of an area decreases as the information about that area (in terms of, for example, the number of neighbours in an adjacency model) increases.

Finally, there is the problem of specifying the random effects for heterogeneity and spatial effects within a generalised linear modelling framework, in this case using IGLS estimation within the MLn software. We do this by constructing weights matrices associated with the random effects and fit these directly into the model. The variance of the data conditional on the fixed part of the model, as given in (16.14), is expressed in terms of three matrices: $Z_u Z_u^{\mathrm{T}}(Z_u Z_v^{*\mathrm{T}} + Z_v^* Z_u^{\mathrm{T}})$, and $Z_v^* Z_v^{*\mathrm{T}}$. Expressing the model is terms of these design matrices overcomes the need to place multiple equality constraints upon the random parameters. This is generalisable to the non-linear model expressed in (16.7)–(16.9).

16.3 INCIDENCE OF PROSTATE CANCER IN SCOTTISH LOCAL AUTHORITY DISTRICTS

In this example we wish to investigate the hypothesis that the relative risk of prostate cancer is higher in rural than urban areas, as previous research has indicated an association between agricultural employment and the incidence of prostate cancer (Key, 1995). The data cover six years, from 1975 to 1980, of the incidence of prostate cancer in 56 districts in Scotland (Kemp *et al.*, 1985). Table 16.1 shows the observed and expected cases, plus the relative risks for incidence of prostate cancer.

To examine the effect of rural location, we use a variable measuring the percentage of the male workforce employed in agriculture, fishing and forestry industries as a surrogate measure of the rurality of an area. However, we have to look not only at the incidence of prostate cancer within districts, but account for a potential artifactual effect caused by differential diagnosis rates between Health Board areas in Scotland. Hence, we are modelling spatial effects caused by different processes at two different scales, namely:

(i) a spatial autocorrelation model at district scale, where we are accounting for the possibility that areas closer in geographical space have similar incidence of prostate cancer; and
(ii) a variance components model at Health Board scale, where we investigate the possibility that different Health Boards have different relative risks of prostate cancer, potentially because diagnostic criteria are variable.

Hence, we can extend (16.1) and (16.11) so that

$$\log(\mu) = \log(E) + X\beta + [Z_u Z_v^*]\begin{bmatrix}\theta_u\\ \theta_v^*\end{bmatrix} + Z_{\text{hb}}\theta_{\text{hb}}. \tag{16.16}$$

In this case the expected cases, E, have been calculated from national incidence rates for Scotland for discrete age bands. We use three explanatory variables in the fixed part of the model ($X\beta$) in addition to the intercept term (*CONS*), namely the proportion of the population in higher social classes (*SC12*); the estimated incidence to ultraviolet light at the earth's surface *(UVBI)*; and the percentage of the male employment in agriculture, fishing and forestry *(AGRI)*. Social class and ultraviolet light exposure have been included as these have previously been postulated as risk factors for prostate cancer.

The Z_v^* are calculated using distances between district centroids. The choice of distance decay function is largely user-dependent, and should ideally be based on some prior hypothesis about the data (Bailey and Gatrell, 1995). Here we have used a simple exponential decay model where we define the w_{ij} as

$$w_{ij} = \exp(-\lambda d_{ij}), \tag{16.17}$$

where d_{ij} are the Euclidian distances between the centroids of areas i and j, and λ is a constant to be estimated from the data. The estimation of λ is problematical because it is non-linear in the random part of the model. Goldstein *et al.* (1994) show that maximum likelihood estimates can be obtained using a Taylor series expansion for the Normal model. However, things become more complicated for a Poisson model, and in general

Table 16.1 Observed and expected cases, and relative risks for the incidence of prostate cancer in Scottish districts, 1975–1980

District	Health Board	Observed	Expected	SMR
Caithness	Highland	15	25.587	0.58625
Sutherland	Highland	18	12.319	1.46110
Ross–Cromarty	Highland	42	42.644	0.98489
Skye–Lochalsh	Highland	10	9.477	1.05520
Lochaber	Highland	22	18.005	1.22190
Inverness	Highland	51	51.173	0.99662
Badenoch	Highland	15	8.529	1.75870
Nairn	Highland	10	9.477	1.05520
Moray	Grampian	107	75.812	1.41140
Banff–Buchan	Grampian	95	74.920	1.25310
Gordon	Grampian	70	58.754	1.19140
Aberdeen	Grampian	249	189.530	1.31380
Kincardine	Grampian	52	38.854	1.33840
Angus	Tayside	104	86.236	1.20600
Dundee	Tayside	176	168.680	1.04340
Perth–Kinross	Tayside	148	108.030	1.37000
Kirkcaldy	Fife	145	135.510	1.07000
NE–Fife	Fife	91	56.859	1.60050
Dunfermline	Fife	117	115.610	1.01200
West-Lothian	Lothian	106	129.830	0.81646
Edinburgh	Lothian	538	402.750	1.33580
Midlothian	Lothian	87	77.707	1.11960
East-Lothian	Lothian	77	74.864	1.02850
Tweeddale	Borders	27	13.267	2.03510
Ettrick	Borders	38	29.377	1.29350
Roxburgh	Borders	44	33.168	1.32660
Berwickshire	Borders	23	17.058	1.34840
Clackmannan	Forth Valley	37	44.540	0.83072
Stirling	Forth Valley	100	72.969	1.37040
Falkirk	Forth Valley	149	136.460	1.09190
Argyll–Bute	Argyll & Clyde	56	60.650	0.92334
Dumbarton	Argyll & Clyde	80	72.969	1.09640
Renfrew	Argyll & Clyde	118	194.270	0.60741
Inverclyde	Argyll & Clyde	84	94.765	0.88640
Glasgow	Greater Glasgow	627	721.160	0.86943
Clydebank	Greater Glasgow	31	49.278	0.62909
Bersden	Greater Glasgow	31	36.958	0.83878
Strathkelvin	Greater Glasgow	57	82.446	0.69137
Eastwood	Greater Glasgow	43	50.225	0.85614
Cumbernauld	Lanarkshire	24	58.754	0.40848
Monklands	Lanarkshire	58	104.240	0.55640
Motherwell	Lanarkshire	100	141.200	0.70822
Hamilton	Lanarkshire	58	102.350	0.56670
East-Kilbride	Lanarkshire	40	78.655	0.50855
Clydesdale	Lanarkshire	47	54.0.16	0.87011
Cunninghame	Ayrshire & Arran	103	128.880	0.79919
Kilmarnock	Ayrshire & Arran	66	77.707	0.84934

Table 16.1 (*continued*)

District	Health Board	Observed	Expected	SMR
Kyle–Carrick	Ayrshire & Arran	108	106.140	1.01750
Cumnock–Doon	Ayrshire & Arran	29	42.644	0.68004
Wigtown	Dumfries & Galloway	28	28.430	0.98489
Stewartry	Dumfries & Galloway	40	20.848	1.91860
Nithsdale	Dumfries & Galloway	56	52.121	1.07440
Annandale	Dumfries & Galloway	48	33.168	1.44720
Orkney	Orkney	22	17.058	1.28970
Shetland	Shetland	17	21.796	0.77996
Western Isles	Western Isles	45	29.377	1.53180

an alternative is to fit a series of models with different values of λ_k and determine the residual deviance from each model, D_k. We can then regress the deviance against the distance decay parameter so that

$$D_k = a + b\lambda_k + c\lambda_k^2 + e_k. \tag{16.18}$$

Differentiating, the approximate solution will be where $\lambda = -b/2c$. Successive approximations then converge towards the estimated value. However, care must be taken when estimating λ, as the likelihood function may be multimodal (Ripley, 1988).

Z_{hb} is a vector of 1 which allows for a variance component for each Health Board to be estimated, and hence a measure of the variance at this scale, σ^2_{hb}. Table 16.2 presents the results for four different models, representing:

model A: a simple, single-level model with no spatial effects.
model B: a model with district scale spatial effects, but no Health Board effects.
model C: a model with only Health Board effects.
model D: a model with both district and Health Board effects as given in (16.16).

The simple model (model A) presented in Table 16.2 seems to indicate a strong and significant effect of rurality, as measured by percentage male agricultural employment (*AGRI*). However, this is weakened by fitting a spatial autocorrelation parameter in model B, which suggests that the effect of *AGRI* may be due to adjacent areas having similar mortality. The change in deviance between the two models is 14.89 on two degrees of freedom ($p < 0.001$: we have fitted a covariance parameter as well as a variance term). The third model (model C), using Health Boards as a level with no spatial autocorrelation between districts, shows how ignoring autocorrelation between residuals at a lower level of a multilevel model (in this case districts) could lead to misleading results at higher levels (in this case, Health Boards), as the parameter for the variance between Health Boards is statistically significant at $p < 0.05$, but the deviance statistic suggests that the model is not as good a fit to the data as model B.

Unexplained random variation at the district level can appear spuriously at the Health Board level, and the final model, with both Health Board effects and spatial effects between districts, suggests that this may be the case. The parameter estimate for *AGRI* becomes smaller in models B, C and D, although it is still significant at the

Table 16.2 Parameter estimates and standard errors for the prostate cancer models

	(A) Simple model		(B) Spatial effects		(C) Health Board effects		(D) Both effects	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Fixed part								
Intercept	−0.0257	0.584	−0.513	0.605	−0.0108	0.636	−0.321	0.670
SC12	−0.000645	0.00524	0.00145	0.00389	−0.00339	0.00477	0.000825	−0.000184
UVBI	−0.0141	0.0635	0.0565	0.0704	−0.00112	0.0705	0.0279	0.0764
AGRI	0.0272	0.00603	0.0163	0.00636	0.0180	0.00634	0.0167	0.00663
Random part								
σ^2_{hb}					0.327	0.0183	0.0225	0.00550
σ^2_u	0.0822	0.0155	0.0665	0.0141	0.0530	0.0117	0.0657	0.0126
σ_{uv}			0.000805	0.000414			0.000682	0.000130
σ^2_v			0.000159	0.0000167			0.000	0.000
λ			7.23				7.00	
Residual deviance	18.98		4.09		10.93		−4.97	

0.05 level. Hence, misspecification of the random part of a model can noticeably affect the fixed as well as the random parameters. Further work needs to be done on the analysis of residuals in these complex models: Langford and Lewis (1998) details some procedures for the general analysis of outliers in multilevel models.

However, we must be careful in drawing conclusions about the size of the parameter estimate for *AGRI*, because we are not postulating that there is some genuine spatial correlation between cases of prostate cancer, for example if the disease had an infectious aetiology and was transmitted between individuals by contact. The spatial effects are not, therefore, in this case an alternative causal factor, but merely a statistical manipulation to account for correlation amongst the residuals in our model. One problem is that the values of the variable *AGRI* are also spatially correlated, because rural districts tend to be adjacent to each other, as do urban ones. Hence, we must be cautious in making inferences from our models without corroborative evidence from elsewhere, although it is interesting to note that the parameter estimate for *AGRI* remains significant in all four model formulations.

16.4 DISCUSSION

We have attempted to demonstrate a general method for modelling geographical data which is distributed in hierarchical administrative units, but which also displays spatial autocorrelation. The models can be implemented within a widely available software package called MLn/MLwiN. However, there are several issues that still need to be addressed, both methodologically and substantively:

(i) We are extending the basic method to model multiple causes of disease simultaneously. Hence, we could model the joint distribution of prostate cancer and another cancer simultaneously. This is the equivalent of adding in another level to the model, so that we have diseases nested within districts within Health Board areas. A further extension to the model can be where areas, such as districts, are not discretely nested within higher level units, such as Health Boards. In this case, a multiple membership model (Goldstein, 1995) may be used, where weights are attached to allocate portions of districts to different Health Boards.

(ii) Space–time models are also possible, as time is simply an extra dimension that requires a variance parameter in the random part of the model, and covariance terms with any spatial parameters

(iii) The main problem in fitting the models is poor convergence properties, usually caused by a high correlation between the heterogeneity and spatial components of the model. One of the authors (AHL) is developing an orthogonalisation procedure to overcome this problem

(iv) The residuals from the model are measured with error, but the IGLS procedure used will tend to underestimate the variance of the residuals. To overcome this, MLwiN has the capability of using the IGLS convergence of the model as the starting point for either a Gibbs sampling or Metropolis–Hastings run of simulations which will provide for better estimates of, for example, the confidence intervals around the posterior relative risks of disease for each district or Health Board. These techniques could also be used to provide better estimates of the standard errors for fixed para-

meters in the model, rather than relying on those estimated from the model to judge statistical significance.

Substantively, multilevel spatial models suffer from similar problems of interpretation as single–level spatial models. It is often difficult to know whether one has modelled a genuine spatial pattern or merely accounted for unmeasured explanatory variables, and fitting a spatial smoothing parameter masks a genuine relationships with an explanatory variable which has its own distinct spatial distribution. However, we believe that the use of a multilevel model can shed light on different processes which may be operating at different spatial scales, and hence provide a valuable tool for the analysis of geographically distributed health data.

ACKNOWLEDGEMENTS

This work was supported by the Economic and Social Research Council, UK. The Public Health Research Unit is financially supported by the Chief Scientist Office of the Scottish Office Department of Health. The opinions expressed in this chapter are not necessarily those of the Chief Scientist Office.