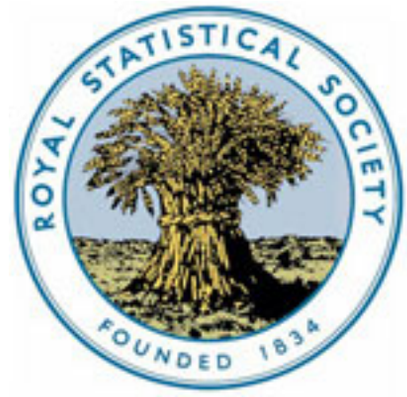

Mapping Disease and Mortality Rates Using Empirical Bayes Estimators
Author(s): Roger J. Marshall
Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 40, No. 2 (1991), pp. 283-294
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/2347593
Accessed: 11/12/2014 01:58

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Wiley and *Royal Statistical Society* are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society. Series C (Applied Statistics).*

http://www.jstor.org

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

Appl. Statist. (1991)
40, *No.* 2, *pp.* 283–294

Mapping Disease and Mortality Rates using Empirical Bayes Estimators

By ROGER J. MARSHALL†

University of Auckland, New Zealand

[Received October 1988. Final revision February 1990]

SUMMARY

Methods for estimating regional mortality and disease rates, with a view to mapping disease, are discussed. A new empirical Bayes estimator, with parameters simply estimated by moments, is proposed and compared with iterative alternatives suggested by Clayton and Kaldor. These methods are shrinkage estimators, in which the crude disease rate is shrunk towards an overall regional rate, and are in this sense 'global' and invariant to spatial configuration. However, it seems unjustifiable, in effect, to ignore the spatial aspect of the problem. A 'local' shrinkage estimator is therefore also suggested in which the crude rate is shrunk towards a local, neighbourhood, rate. Comparison of the estimators is done by some simulation experiments and an example showing infant mortality in Auckland, New Zealand, is presented. When disease is relatively rare a global estimator gives the lowest total mean-square error, but for diseases that are more common and where the underlying spatial pattern is not uniform the local estimator performs best.

Keywords: Disease and mortality rate; Empirical Bayes and James–Stein estimators; Global and local estimators; Neighbourhood

1. Introduction

Mapping mortality and disease rates to display the geographical variability of disease is an increasingly common epidemiological tool. The idea is not new; Howe (1970) traces its development to the 18th century. The advent of computers and computer graphics has made map construction considerably easier, but the statistical problem of deriving estimates of the quantities to be mapped has only recently received attention.

Developments have centred on the idea of pooling information across areas, through the use of empirical Bayes methods, to reduce the total mean-square error. Efron and Morris (1975) seem to be first to use this approach for disease rates—toxoplasmosis in El Salvador—using James–Stein estimators. Clayton and Kaldor (1987) proposed empirical Bayes estimation procedures using a Poisson likelihood and gamma prior framework, a framework also used by Stone (1988), to adjust significance levels in testing for geographical excess risk, as well as by Manton *et al.* (1981, 1987) and, with an additional random effects component, by Tsutakawa (1988). Tsutakawa *et al.* (1985) had earlier proposed an empirical Bayesian approach that is similar to Leonard's (1972) method for estimating binomial proportions. Clayton and Kaldor also suggested estimating a prior distribution nonpara-

†*Address for correspondence*: Department of Community Health, School of Medicine, University of Auckland, Private Bag, Auckland, New Zealand.

© 1991 Royal Statistical Society 0035–9254/91/40283 $2.00

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

metrically, using a method due to Laird (1978), and a method for modelling prior parameters by a spatial autoregressive scheme. Miyawaki and Chen (1981) have also proposed what is essentially an empirical Bayes procedure to estimate a prior density for a standardized mortality ratio.

This work was motivated by a study of the spatial patterns of mortality in the Auckland region of New Zealand (Marshall, 1989). An example is presented in Section 4 to illustrate the methodology developed in Sections 2 and 3. Section 2 is concerned with 'global' shrinkage estimators, i.e. estimators that are shrunk towards an overall, global, mean. An empirical Bayes estimator is derived using results due to Efron and Morris (1973) for construction of James–Stein estimators. It is a non-iterative moment procedure. Similar work by Clayton and Kaldor (1987) is also discussed. However, none of these procedures deals explicitly with the spatial aspect of the data; they are, in fact, invariant to spatial configuration. The spatial character of the problem is, however, what makes it interesting and to ignore it seems unjustifiable. In Section 3 some alternative 'local' estimators, that are shrunk towards a local mean, are proposed. The methods are illustrated using infant mortality data in Section 4 and their properties compared by simulation experiments in Section 5.

2. Empirical Bayes Estimators

Consider a region partitioned into N areas indexed by i ($i = 1, \ldots, N$). Suppose that events (deaths or disease) are recorded over a period of years. Let θ_i be the annual event rate at area i and assume that r_i, the accumulated events, is distributed as a Poisson random variable, with conditional mean $E(r_i|\theta_i) = n_i\theta_i$, where n_i is the person-years at risk. The maximum likelihood (ML) estimator of θ_i is $x_i = r_i/n_i$. We shall refer to x_i as the 'crude' estimator of θ_i. It has mean, conditional on θ_i, $E(x_i|\theta_i) = \theta_i$ and conditional variance $\text{var}(x_i|\theta_i) = \theta_i/n_i$. In practice n_i usually requires estimation by interpolation of census data, but the error in doing so will be ignored. It will be assumed that, conditional on the θ_i, the r_i are independent so that difficult issues concerning spatial dependence among the observed r_i do not arise.

Adopting a Bayesian framework, suppose that θ_i has a prior density with mean and variance $m_i = E_\theta(\theta_i)$ and $A_i = \text{var}_\theta(\theta_i)$ respectively. Unconditionally the mean of x_i is $E_x(x_i) = E_\theta\{E(x_i|\theta_i)\} = E_\theta(\theta_i) = m_i$ and its unconditional variance is

$$\begin{aligned}\text{var}_x(x_i) &= \text{var}_\theta\{E(x_i|\theta_i)\} + E_\theta\{\text{var}(x_i|\theta_i)\} \\ &= \text{var}_\theta(\theta_i) + E_\theta(\theta_i/n_i) \\ &= A_i + m_i/n_i. \end{aligned} \tag{1}$$

Given m_i and A_i the best linear Bayes estimator of θ_i, in terms of total squared error loss (Efron and Morris, 1973), is the shrinkage estimator

$$\hat{\theta}_i = m_i + C_i(x_i - m_i), \tag{2}$$

where $C_i = A_i/(A_i + m_i/n_i) = \text{var}_\theta(\theta_i)/\text{var}_x(x_i)$, the ratio of the prior variance of θ_i to the unconditional variance of x_i. So-called James–Stein estimators of θ_i are constructed by treating the x_i as a sample from the marginal density of x_i to estimate the m_i and C_i in equation (2). Some simplification is required since the model is saturated in the $2N$ parameters m_i, A_i for $i = 1, \ldots, N$. One obvious reduction is to take $A_i = A$ and $m_i = m$ for all i. To estimate m and A by ML requires full specification of the prior and subsequent derivation of the marginal likelihood.

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

For instance, if a gamma prior density is assumed the unconditional density of x_i is negative binomial and parameters can be estimated iteratively by forming a negative binomial likelihood (Clayton and Kaldor, 1987; Stone, 1988) as described later. However, a distribution-free and non-iterative procedure can be obtained by a method of moments. Since $E_x(x_i) = E_\theta(\theta_i) = m$, any weighted mean of the x_i provides an unbiased estimate of m. A natural quantity to use as a weight is n_i so that m is estimated by the pooled mean $\tilde{m} = \Sigma_i x_i n_i/n = \Sigma_i r_i/n$, where $n = \Sigma_i n_i$ is the total person-years at risk. To estimate A consider the weighted sample variance $s^2 = \Sigma_i n_i(x_i - \tilde{m})^2/n$. Ignoring the error in using $\tilde{m}$ as an estimate of m, we have $E_x(x_i - \tilde{m})^2 \simeq \text{var}_x(x_i) = A + m/n_i$ and so

$$E_x(s^2) = n^{-1} \sum_i n_i(A + m/n_i) = A + m/\bar{n}$$

where $\bar{n} = n/N$ is the mean person-years at risk. Therefore a moment estimate of A is, on replacing m by $\tilde{m}$, $\tilde{A} = s^2 - \tilde{m}/\bar{n}$. Since this quantity can be negative we adopt the convention $\tilde{A} = 0$ when $s^2 < \tilde{m}/\bar{n}$. With m and A replaced by $\tilde{m}$ and $\tilde{A}$ the shrinkage estimator in equation (2) becomes

$$\tilde{\theta}_i = \tilde{m} + \tilde{C}_i(x_i - \tilde{m}) \tag{3}$$

where

$$\tilde{C}_i = \frac{s^2 - \tilde{m}/\bar{n}}{s^2 - \tilde{m}/\bar{n} + \tilde{m}/n_i} \tag{4}$$

and $\tilde{\theta}_i = \tilde{m}$ when $s^2 < \tilde{m}/\bar{n}$. Shrinkage to $\tilde{m}$ is greater as s^2 decreases, i.e. with increasing spatial homogeneity of the crude estimates, and as n_i decreases, i.e. when the crude estimator becomes less reliable.

When each prior has a common gamma density with scale and shape parameters ν and α, so that $m = \nu/\alpha$ and $A = \nu/\alpha^2$, the quantity C_i in equation (2) is $C_i = n_i/(n_i + \alpha)$ and $\hat{\theta}_i$ is the posterior mean $(r_i + \nu)/(n_i + \alpha)$. Clayton and Kaldor (1987) suggested both ML and a mixed method of moments and ML to estimate ν and α. The ML equations are derived from an unconditional negative binomial density. The alternative, mixed, procedure used one likelihood equation, the derivative with respect to α, together with an equation obtained by equating a χ^2-statistic to its mean. One solution to both sets of equations (Clayton and Kaldor's equations (5) and (6), and (5) and (7)) occurs as ν and α both approach infinity while keeping $\nu/\alpha = \tilde{m}$. I have found that iterative procedures sometimes converge on this solution, usually when the underlying pattern is uniform; the estimate of θ_i is then $\tilde{m}$, the overall mean. Also, my experience of Clayton and Kaldor's suggested alternating iterative procedure for the mixed method is that it is often slow to converge and I have had more success with a Newton–Raphson procedure.

Alternative estimates of ν and α can be obtained by transforming the moment estimates $\tilde{m}$ and $\tilde{A}$, i.e. by putting $\tilde{\nu} = \tilde{m}^2/\tilde{A}$ and $\tilde{\alpha} = \tilde{m}/\tilde{A}$, and these make good starting values for iteration. In fact, they turn out to be an exact solution to Clayton and Kaldor's mixed method when the n_i are all equal. However, this case seldom arises in practice.

The approaches outlined in this section are all global in the sense that they are estimators that are shrunk towards a global mean $\tilde{m}$. In the next section, estimators which are 'local' are considered.

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

3. Local Estimators

Global approaches are spatially invariant; any rearrangement of the spatial entities leaves the estimates unchanged. However, it is often reasonable to consider areas that are close together to have similar disease rates by virtue of their being similar in other respects, e.g. if areas are small subdivisions of an urban region. One way to account for this is to define a 'neighbourhood' for the ith area, which incorporates the ith area with its neighbours, and to use the neighbourhood to set prior parameters for θ_i. Then θ_i is estimated by shrinking x_i towards the neighbourhood mean. To adopt an iterative method for each area is computationally prohibitive. However, local shrinkage estimators of the form (3), denoted by $\tilde{\theta}_{(i)}$, are easy to obtain; $\tilde{C}_i$, $\tilde{m}$, s^2 and n in equations (3) and (4) are replaced by $\tilde{C}_{(i)}$, $\tilde{m}_{(i)}$, $s^2_{(i)}$ and $\bar{n}_{(i)}$ respectively, these being computed over the ith neighbourhood, e.g. $\tilde{m}_{(i)} = \Sigma_j r_j / \Sigma_j n_j$ with summation over areas in the neighbourhood. Implicit in the construction of $\tilde{\theta}_{(i)}$ is a loose *a priori* assumption that areas within a neighbourhood are relatively homogeneous. Other *ad hoc* local estimators are easy to construct. One obvious possibility is to take the crude rate obtained by pooling the data in a neighbourhood, i.e. the neighbourhood mean $\tilde{m}_{(i)}$.

4. Infant Mortality in Auckland

The methods were used to estimate infant, under 5 year of age, mortality rates in Auckland, New Zealand, for the period 1977–85. In this period there were 1403 deaths in children under 5 years old and the population of children under 5 years old was 59196 in the 1981 census. The basic spatial entity for the analysis was the census area unit (CAU), the Auckland region being divided into 167 such units covering approximately 5000 km^2. These vary in physical and population size, the mean CAU population size of children under 5 years old being 354 with standard deviation 258 and heavily skewed to the right, the maximum being 1407. It was assumed that the total person-years at risk, n_i, for a CAU was nine times its recorded population in the 1981 census, as mortality data were accumulated over a 9-year period. Auckland's population has grown steadily during this period but, as 1981 is the midpoint of the study period, these n_i are probably quite close to those of the actual person-years at risk. In total there were 532764 person-years data. The overall mean is 2.633 deaths per year per thousand. Fig. 1 shows the crude estimates of annual mortality (as number of deaths per thousand) for the region. It is evidently quite patchy with apparently very high rates in the outer north-west, although these areas are relatively rural and sparsely populated. The central and most densely populated urban area of Auckland is an isthmus bounded by the Waitemata and Manukau harbours, marked W and M in Fig. 1.

Table 1 gives global estimates of m, A, α and ν derived by the moment, the mixed and the ML procedures. There was no difficulty in achieving convergence for the last two procedures. Approximate standard errors for the ML estimates, $\hat{\alpha}$ and $\hat{\nu}$, were derived from the information matrix and those of $\tilde{m}$ and $\tilde{A}$ by direct evaluation by assuming that $\tilde{m}$ and s^2 are independent. The 'delta method' was used to obtain standard errors on $\tilde{\nu} = \tilde{m}^2/\tilde{A}$ and $\tilde{\alpha} = \tilde{m}/\tilde{A}$ and on $\hat{m} = \hat{\nu}/\hat{\alpha}$ and $\hat{A} = \hat{\nu}/\hat{\alpha}^2$. There is no obvious way to obtain standard errors for the mixed method, though it seems reasonable that these will be of the same order of magnitude. The estimated correlation between $\hat{\alpha}$ and $\hat{\nu}$ was high, 0.99, while that of $\tilde{m}$ and $\tilde{A}$ was -0.15.

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

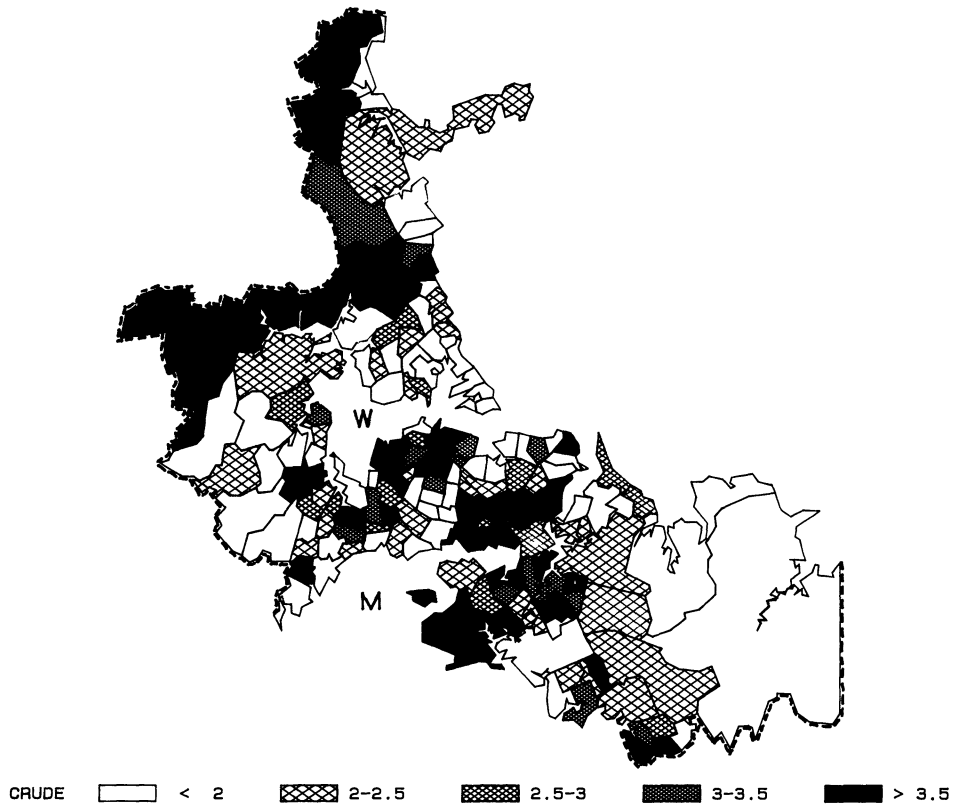

Fig. 1. Crude estimates x_i of mortality in children under 5 years old in Auckland during 1977–85 (expressed as deaths per thousand persons per year): the boundary is coastline except where shown by the broken lines; W and M are the Waitemata and Manukau harbours

The differences in the methods are not great and the resulting estimates of θ_i are similar. Fig. 2 is a map of the moment estimator $\tilde{\theta}_i$. A comparison with Fig. 1 shows how the high rates on the fringes of the region are effectively shrunk towards the overall mean, while the high rates in the main urban areas are shrunk to a much lesser degree. After categorizing by choropleth levels, maps produced by the ML and mixed methods were barely distinguishable from Fig. 2 and have accordingly

TABLE 1
Parameter estimates by three global methods for infant mortality in children under 5 years old in Auckland, New Zealand†

Method	α ($\times 10^{-3}$)	*Estimates of the following parameters:* ν	m ($\times 10^{3}$)	A ($\times 10^{6}$)
Moments	3.615 (1.127)	9.52 (3.13)	2.633 (0.120)	0.728 (0.220)
Mixed	3.022	8.08	2.674	0.884
ML	4.336 (1.197)	11.55 (3.31)	2.664 (0.097)	0.614 (0.175)

†Estimated standard errors are given in parentheses.

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

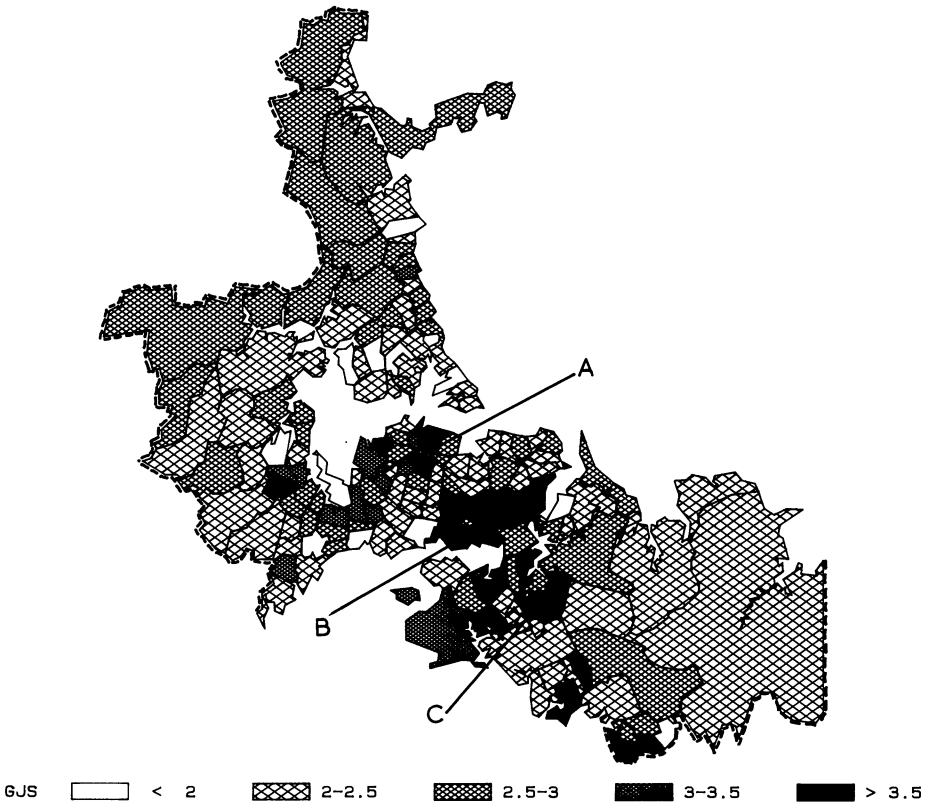

Fig. 2. Global moment estimator $\tilde{\theta}_i$ of infant mortality as a rate per thousand persons per year (see the text for reference to pointers A, B and C)

been omitted. The mortality pattern apparently reflects the socioeconomic gradients within the region; for instance, the areas A, B and C, where mortality seems to be high in Fig. 2, are the less affluent suburbs of the city.

Local estimators were derived by defining the neighbours of each CAU to be those sharing a common boundary. This restriction was relaxed, for two CAUs which only had one such neighbour, to include an additional close CAU. Thus the smallest neighbourhood contained three CAUs; the largest had 13. Fig. 3 is the local moment estimator $\tilde{\theta}_{(i)}$. There is clearly a greater tendency for this estimator to merge areas into homogeneous blocks with high and low rates, a feature not unexpected in view of its construction. The local mean estimator $\tilde{m}_{(i)}$ produced a map very similar to Fig. 3 and its map is omitted.

These maps are an analysis of age-specific rates. To compare rates for particular diseases, across all age groups, requires adjustment for different demographic age structures. This was achieved, in a study of the major causes of death in Auckland (Marshall, 1989), by computing age-specific rates using the local moment method and computing a summary direct age-standardized statistic by using the age structure of the total population of Auckland as a standard. There are, however, other ways to obtain summary age-standardized rates (Clayton and Kaldor, 1987; Manton *et al.*, 1987).

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

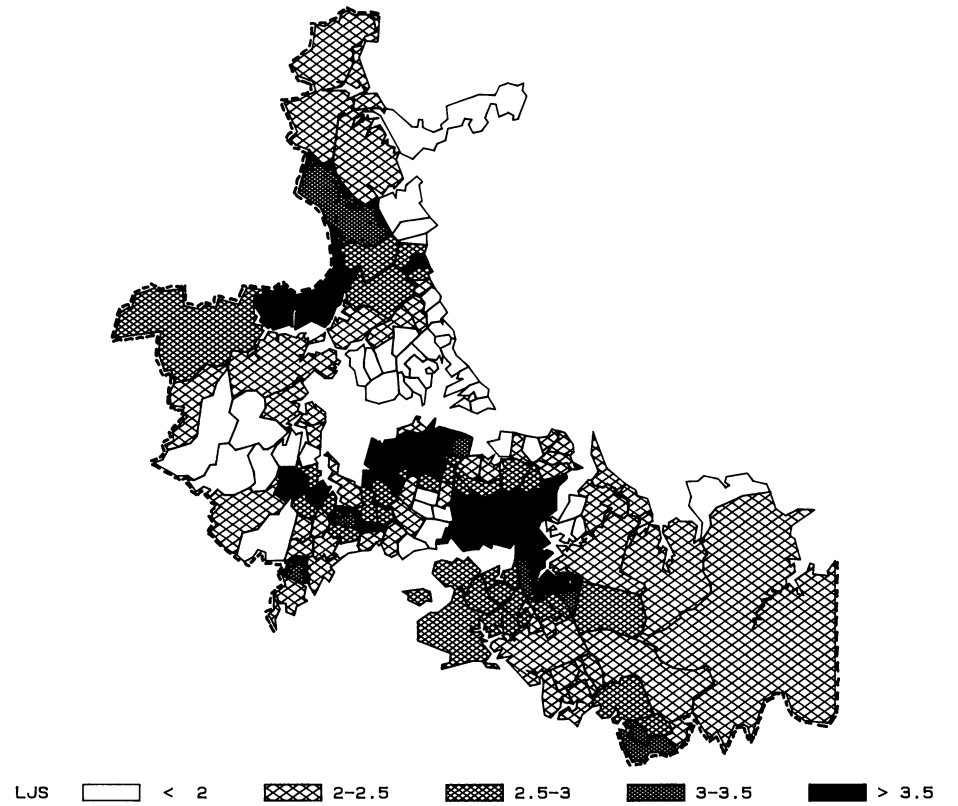

Fig. 3. Local moment estimator $\tilde{\theta}_{(i)}$ of infant mortality as a rate per thousand persons per year

5. Simulation Experiments

A theoretical treatment of the relative merits of the proposed estimators seems to be unattainable but simulation experiments offer some insight. The experiments to be described are not, and cannot, cover *all* possibilities, in terms of geography, population, rates of disease, number of spatial units, and so on. However, the results enable some appreciation of how the estimators behave.

As global estimators are invariant to spatial configuration it is only necessary, for their comparison, to define a set of θ_i and to compare their performance in simulations of sets of Poisson variables r_i with means $n_i\theta_i$. However, for local estimators we need to define a neighbourhood for each i. For this a hypothetical 'region' was divided into 100 areas and each area was indexed by $i=(X, Y)$ for $X=1, \ldots, 10$ and $Y=1, \ldots, 10$. A neighbourhood of (X, Y) was defined as the set of nine areas $\{(X+j, Y+k); j=-1, 0, 1, k=-1, 0, 1\}$ with the rule that on the boundary non-defined areas, e.g. $(X, Y) = (0, 0)$, were discounted. In this way a graph is generated of the neighbour linkages of areas. The graph can be represented as a 10×10 network, though the region itself need not necessarily be constructed as a regular array. Patterns of disease incidence, expressed in terms of (X, Y), are given by each of the spatial arrangements in Table 2. The multiplication factor c,

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

TABLE 2
Key to spatial patterns used in the simulation experiments †

Description	*Pattern*	*Formulation*
Uniform	A	$\theta_i = 5c$
4 sectors, almost uniform	B	$\theta_i = 5c$ if $X \leqslant 5$ and $Y \leqslant 5$, or $X > 5$ and $Y > 5$; $\theta_i = 6c$ otherwise
Random, almost uniform	C	$\theta_i = cU$ where U is a pseudorandom uniform variable between 5 and 7
Linear	D	$\theta_i = c(5 + 0.5X + 0.5Y)$
High cells	E	$\theta_i = 15c$ if $(2, 2) \leqslant (X, Y) \leqslant (4, 4)$; $\theta_i = 10c$ if $(7, 7) \leqslant (X, Y) \leqslant (9, 9)$; $\theta_i = 5c$ otherwise
Dome	F	$\theta_i = c\{20 - 0.4(X - 5.5)^2 - 0.4(Y - 5.5)^2\}$
2 sectors	G	$\theta_i = 5c$ if $X \leqslant 5$; $\theta_i = 10c$ otherwise
4 sectors	H	$\theta_i = 5c$ if $X \leqslant 5$ and $Y \leqslant 5$; $\theta_i = 8c$ if $X \leqslant 5$ and $Y > 5$; $\theta_i = 12c$ if $X > 5$ and $Y \leqslant 5$; $\theta_i = 15c$ if $X > 5$ and $Y > 5$

†$\{X, Y; X = 1, \ldots, 10, Y = 1, \ldots, 10\}$ reference each area. The quantity c is 10^{-3}, 10^{-4} or 10^{-5} for common, less common and rare disease respectively.

equal to 10^{-3}, 10^{-4} or 10^{-5}, was used to evaluate the behaviour of the estimators for 'common', 'less common' and relatively 'rare' disease. The various spatial patterns in Table 2 were arbitrarily defined; they cover the uniform case A, and departures from it in various structured and random ways. Patterns B and C are 'almost' uniform while the remainder have more severe departures. For global estimators only the values of θ_i are relevant, the spatial juxtaposition being immaterial.

The population person-years at risk, n_i for $i = 1, \ldots, 100$, were obtained in one set of simulations as a random sample from a uniform distribution between 1000 and 20000, and in a second set of simulations by a random sample between 10000 and 200000. This allowed comparison of how the estimators behaved for about $n = 1$ million and $n = 10$ million person-years of accumulated data. The disease counts r_i were simulated as independent Poisson counts with mean $n_i\theta_i$. To evaluate precision the total mean-square error (TMSE) of an estimated map was used where

$$\text{TMSE}(\theta_i^*) = \sum_i E(\theta_i^* - \theta_i)^2,$$

θ_i^* being a particular estimator. The TMSE was estimated by the average over 200 map simulations, or 20000 individual θ_i^* calculations. The objective was to compare how θ_i^* behaves relative to the crude rate x_i and $\text{TMSE}(\theta_i^*)$ is expressed as a percentage of $\text{TMSE}(x_i)$ by the quantity $R = 100\ \text{TMSE}(\theta_i^*)/\text{TMSE}(x_i)$. The standard error of the estimated R is of the order of 1% of its value, this being established by some repeated simulations.

Seven estimators of θ_i were calculated in each simulation run: x_i, the crude estimator; the overall mean $\tilde{m}$; the proposed method of moments $\tilde{\theta}_i$; the ML procedure, denoted θ_i^{ml}; Clayton and Kaldor's mixed ML and moment procedure, denoted θ_i^{mx}; the proposed local moment procedure $\tilde{\theta}_{(i)}$; the neighbourhood mean $\tilde{m}_{(i)}$. For the iterative estimates, θ_i^{ml} and θ_i^{mx}, a Newton–Raphson procedure was employed with the moment-derived estimates $\tilde{\nu}$ and $\tilde{\alpha}$ as starting values. Up to 10 iterations were allowed to achieve convergence. If convergence was not then fully achieved, but the

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

incremental change was diminishing, the estimates at the 10th iteration were used. If, after 10 iterations, the increments were diverging the limiting solution $v=\infty$ and $\alpha=\infty$ with $v/\alpha=\tilde{m}$ was assumed and the estimate of θ_i set to $\tilde{m}$.

These criteria were introduced to limit computer time. By experience it was found that if convergence was not achieved in 10 iterations, but the increments were diminishing, further cycles generally made minimal change to the values at the 10th cycle. In contrast, diverging estimates at the 10th cycle always failed to resolve themselves into converged estimates by further iteration, the process continuing indefinitely. As already mentioned, convergence problems were found mainly to occur when the underlying pattern was uniform, or nearly so. For example, when the underlying pattern was uniform the iterative cycle diverged, in the above sense, about 95% of the time for the ML estimator and for about 75% of the time for the mixed estimator. The moment estimator required setting $\tilde{\theta}_i=m$, when $\tilde{A}=0$, about 60% of the time. The more severe the departure from uniformity of the underlying pattern the less often these fixes were required. For example, full convergence was always achieved for patterns D–H in Table 2.

The series graphs in Fig. 4 show the quantity R plotted against rarity of disease for each estimator and for 1 million total person-years of data. Each graph in the series represents the behaviour for the pattern types in Table 2. Consider the uniform pattern A. Each of the global estimators has a similar TMSE, about 1% of the crude estimator and changing little with disease rarity. Although not drawn the best estimator is not surprisingly the overall mean $\tilde{m}$ with a value of $R=0.5\%$. Neither local estimator performs as well, but $\tilde{\theta}_{(i)}$ has a TMSE that is about 10% of the crude estimator and that of $\tilde{m}_{(i)}$ is marginally better. Once there are departures from uniformity R depends on the nature of the pattern and the rarity of the disease. The first point to note is that the R values for the global estimators $\tilde{\theta}_i$, θ_i^{ml} and θ_i^{mx} are all about the same, suggesting little benefit from an iterative procedure over the simple moment method. In particular, R for $\tilde{\theta}_i$ and θ_i^{mx} is often superimposed and indistinguishable. Secondly R decreases with the rarity of the disease once the pattern is anything but uniform, quite sharply for the global estimators but less so for the local estimator $\tilde{\theta}_{(i)}$. $\tilde{\theta}_{(i)}$ is therefore less sensitive to rarity of disease and generally performs better than a global method for common diseases. An exception to this is for the random pattern C where the local estimators fare badly, presumably because of the lack of any spatial structure to θ_i. A final point is that the local estimators, and in particular $\tilde{\theta}_{(i)}$, show less variation between patterns, so that, as well as being less sensitive to rarity, $\tilde{\theta}_{(i)}$ is also less sensitive to the underlying spatial pattern.

When the total population was increased to 10 million person-years and the simulations repeated, the conclusions remained essentially the same. There are some differences, however. Generally R is larger, especially for global estimators and common disease, but $\tilde{\theta}_{(i)}$ is the least sensitive to the change in population size. Adverse behaviour of the local mean $\tilde{m}_{(i)}$ may also occur as it does not weight x_i in relation to n_i and, when the disease is common, may lead to a TMSE that is greater than that of x_i itself. In contrast $\tilde{\theta}_{(i)}$ remains relatively stable with respect to rarity and population size.

An issue not directly addressed in these simulations is the way that the estimators behave as N, the number of area units, varies. Two conceptual mechanisms by which N can increase are disaggregation of area units, within a given region, and by expansion of a region, taking in more area units. Some of the experiments outlined

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

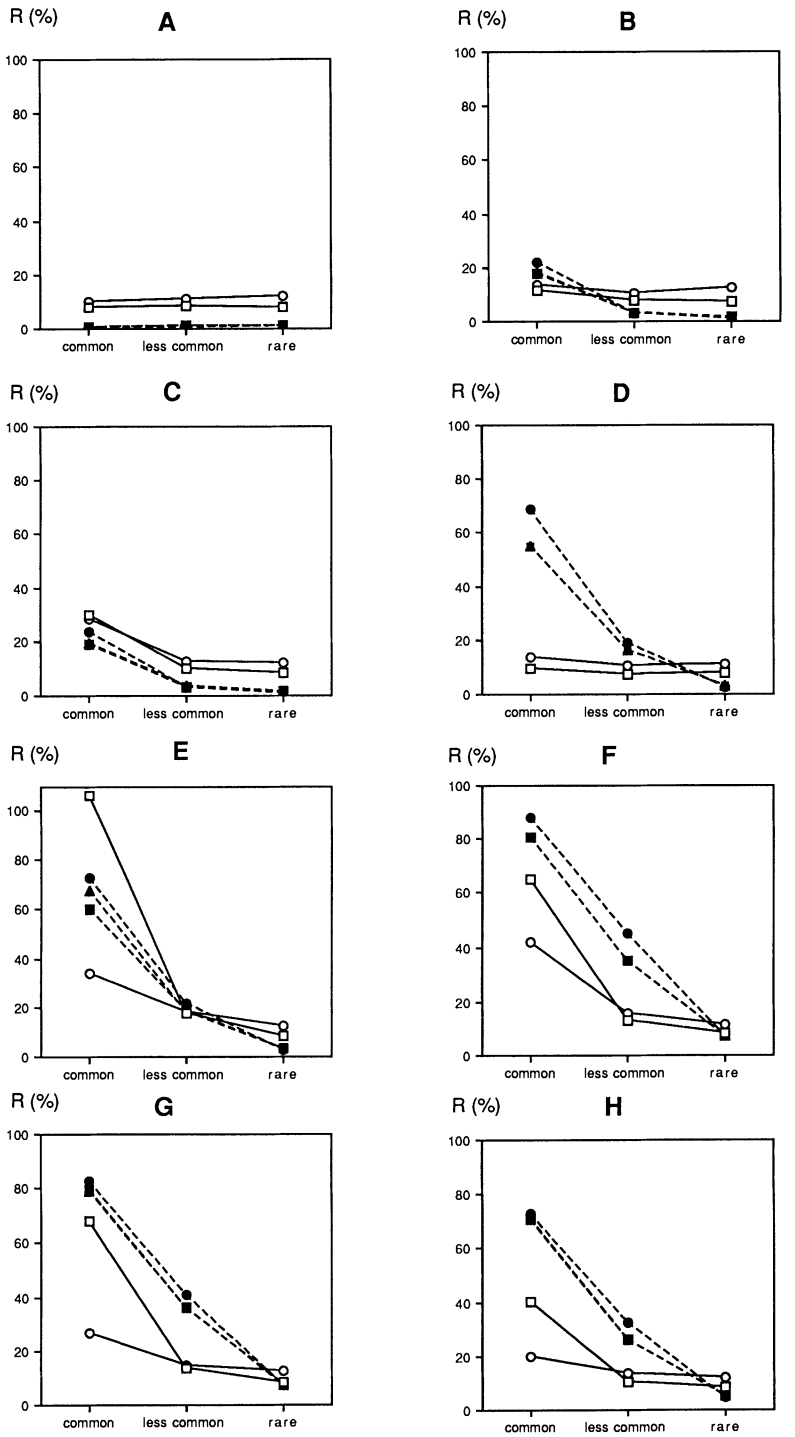

Fig. 4. Percentage reduction in TMSE (R) for different spatial patterns A–H against rarity of disease: global estimators --■-- ($\tilde{\theta}_i$), --●-- (θ_i^{ml}), ----▲---- (θ_i^{mx}); local estimators ——○—— ($\tilde{\theta}_{(i)}$), ——□—— ($\tilde{m}_{(i)}$)

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

were repeated for a larger area formed by increasing the grid size to $N = 20 \times 20$, i.e. by allowing outward expansion. The conclusions are broadly similar to those reported for the 10×10 grid. The magnitude of R for local estimators generally remains little changed, nor is that of global estimators much altered, with the notable exception that, when the underlying spatial pattern is uniform, R is about 0.2% which is substantially smaller than the 1% reported earlier for $N = 10 \times 10$. This behaviour is anticipated in view of the reduction in TMSE in moving from a local to a global method (Fig. 4, pattern A); a 20×20 grid is just a further expansion of the global environment. Examining the behaviour of both local and global estimators to changes to N is less straightforward under a disaggregation mechanism, as the problem of how to disaggregate the population arises. Further, for local estimators, we need to deal with neighbourhoods either remaining unchanged in extent, though encompassing more disaggregated units, or strictly localized to only adjacent units. A detailed examination of the estimators' behaviour with changes to N, and to neighbourhood definition, is a topic for future work.

6. Conclusion

The choice of an appropriate estimator will depend on the true, but unknown, pattern of disease. We ideally require a method which provides good estimates in a variety of situations. However, none of those considered fulfils this requirement entirely, although the local estimator $\tilde{\theta}_{(i)}$ emerges as the least sensitive to rarity of the disease and to pattern type. It is also conceptually appealing in that it employs spatial configuration. Nevertheless its performance is not as good as global methods when θ_i approaches uniformity and for diseases that are rare. Usually the overall incidence of disease will be known so that a judgment about whether to use a local or a global method can be made. For example, in Section 4 the overall total infant mortality rate is 2.63 per thousand per year which is relatively common; the local estimator may therefore be best in this case. In terms of the TMSE there is little difference between the three global estimates $\tilde{\theta}_i$, θ_i^{ml} and θ_i^{mx} so that in choosing a global estimator it is simplest to avoid an iterative method and to select the moment estimator $\tilde{\theta}_i$ which is easy to compute, requiring only a weighted mean and variance calculation. Further, it makes no assumption about the form of an underlying prior density.

The local moment method is essentially an attempt to place different parameters on the marginal prior at each area, but the θ_i are held to be independent *a priori*. An alternative approach is to introduce spatial dependence among the θ_i while retaining common prior marginal parameters. For example, Clayton and Kaldor (1987) suggested a scheme in which spatial dependence among the θ_i is modelled by a spatial autonormal process (Besag, 1974). A more elaborate scheme is to model θ_i as the sum of a spatially structured variable, described by a spatial *intrinsic* autoregressive process, and a local unstructured variable (Besag, 1989). The performance of these methods remains to be evaluated. A perceived drawback of local approaches is that they depend on how neighbourhoods are chosen and therefore introduce some subjectivity. Neighbourhoods are a feature of various methods of analysis of spatial data (e.g. Cliff and Ord (1975), Ord (1975) and Besag (1974, 1975)); they implicitly introduce an *a priori* belief about the local similarity of disease rates which is, in spirit, Bayesian and which seems justifiable for the analysis of small areas. There is, however, a need for guidelines on how to define a neighbourhood. A global method

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions

makes an entire region the neighbourhood, rather than each spatial unit having its own unique neighbourhood. A compromise is to group contiguous areas into neighbourhoods, or zones, if there is *a priori* justification for zoning, for example, in terms of homogeneous geographical or demographic characteristics.

Although this paper has dealt ostensibly with the mapping problem, the methods could also be used to assess the significance of local rates of disease, for example, of apparently raised leukaemia rates near nuclear power installations, by considering empirical posterior probabilities. Various other researchers have made suggestions along these lines (Thomas, 1985; Thomas *et al.*, 1985; Stone, 1988; Hills and Alexander, 1989).

References

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.

——— (1975) Statistical analysis of non-lattice data. *Statistician*, **24**, 179–195.

——— (1989) Towards Bayesian image analysis. *J. Appl. Statist.*, **16**, 395–407.

Clayton, D. and Kaldor, J. (1987) Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671–681.

Cliff, A. D. and Ord, J. K. (1975) Model building and the analysis of spatial pattern in human geography (with discussion). *J. R. Statist. Soc.* B, **37**, 297–328.

Efron, B. and Morris, C. (1973) Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Am. Statist. Ass.*, **68**, 117–130.

——— (1975) Data analysis using Stein's estimator and its generalisations. *J. Am. Statist. Ass.*, **70**, 311–319.

Hills, M. and Alexander, F. (1989) Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review. *J. R. Statist. Soc.* A, **152**, 353–363.

Howe, G. M. (1970) *National Atlas of Disease Mortality in the United Kingdom*, revised and enlarged edn. London: Nelson.

Laird, N. (1978) Non-parametric maximum likelihood estimation of a mixing distribution. *J. Am. Statist. Ass.*, **73**, 805–811.

Leonard, T. (1972) Bayesian methods for binomial data. *Biometrika*, **59**, 581–589.

Manton, K. G., Stallard, E., Woodbury, M. A., Riggan, W. B., Creason, J. P. and Mason, T. J. (1987) Statistically adjusted estimates of geographic mortality profiles. *J. Natn Cancer Inst.*, **78**, 805–815.

Manton, K. G., Woodbury, M. A. and Stallard, E. (1981) A variance components approach to categorical data models with heterogeneous cell population: analysis of spatial gradients in lung cancer rates in North Carolina counties. *Biometrics*, **37**, 259–269.

Marshall, R. J. (1989) An area analysis of major causes of death among under 65 year olds in Auckland—1977–85. *N. Z. Med. J.*, **102**, 67–70.

Miyawaki, N. and Chen, S. C. (1981) A statistical consideration on the mapping of disease. *Soc. Sci. Med.* D, **15**, 93–101.

Ord, J. K. (1975) Estimation methods for models of spatial interaction. *J. Am. Statist. Ass.*, **70**, 120–126.

Stone, R. A. (1988) Investigations of excess environmental risks around putative sources: statistical problems and a proposed test. *Statist. Med.*, **7**, 649–660.

Thomas, D. C. (1985) The problem of multiple inference in identifying point-source environmental hazards. *Environ. Hlth Perspct.*, **62**, 407–414.

Thomas, D. C., Siemiatycki, J., Dewar, R., Robins, J., Goldberg, M. and Armstrong, B. G. (1985) The problem of multiple inference in studies designed to generate hypotheses. *Am. J. Epidem.*, **122**, 1080–1095.

Tsutakawa, R. K. (1988) Mixed model for analyzing geographic variability in disease rates. *J. Am. Statist. Ass.*, **83**, 37–42.

Tsutakawa, R. K., Shoop, G. L. and Marienfeld, C. J. (1985) Empirical Bayes estimation of cancer disease rates. *Statist. Med.*, **4**, 201–212.

This content downloaded from 169.230.243.252 on Thu, 11 Dec 2014 01:58:11 AM
All use subject to JSTOR Terms and Conditions