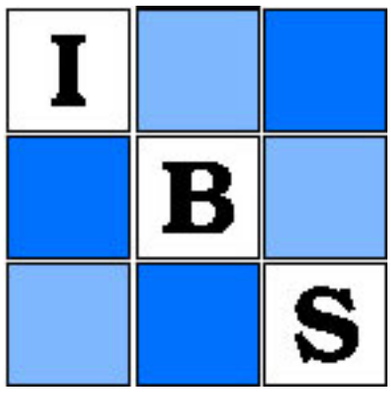

BIOMETRICS 43, 671–681
September 1987

Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping

David Clayton

Department of Community Health,
Leicester Royal Infirmary, Leicester, United Kingdom

and

John Kaldor

Unit of Biostatistics and Field Studies, International Agency for Research on Cancer,
150 cours Albert Thomas, 69372 Lyon Cedex 08, France

SUMMARY

There have been many attempts in recent years to map incidence and mortality from diseases such as cancer. Such maps usually display either relative rates in each district, as measured by a standardized mortality ratio (SMR) or some similar index, or the statistical significance level for a test of the difference between the rates in that district and elsewhere. Neither of these approaches is fully satisfactory and we propose a new approach using empirical Bayes estimation. The resulting estimators represent a weighted compromise between the SMR, the overall mean relative rate, and a local mean of the relative rate in nearby areas. The compromise solution depends on the reliability of each individual SMR and on estimates of the overall amount of dispersion of relative rates over different districts.

1. Introduction

In recent years, a considerable amount of effort has been put into the production of cancer atlases—that is, collections of maps in which the geographical distribution of cancer is presented by site, for a specific country or region. The International Agency for Research on Cancer (1982) provides a summary of work in this field, and discusses the issues involved. One of the main problems has been the choice of the appropriate measure of cancer incidence or mortality to map. Some atlases have presented measures of relative risk, usually standardized mortality ratios (SMRs), while others display the statistical significance of local deviations of risk from the overall rates on the map.

Both these approaches can badly misrepresent the geographical distribution of cancer incidence. In the former case, no account is taken of varying population size over the map, so that imprecisely estimated SMRs, based on only a few cases, may be the extremes of the map, and hence dominate its pattern. On the other hand, mapping significance alone totally ignores the size of the corresponding effect, so that on the map, two areas with identical SMRs may be indicated quite differently if they are of unequal population size, and the most extreme areas may simply be those with the largest populations. Not only are these approaches unsatisfactory, but the lack of a common format of presentation frustrates the comparison of cancer risk across atlases.

Key words: Cancer atlas; Empirical Bayes estimate; Gaussian spatial process; James–Stein estimate; Relative risk; Standardized mortality ratio.

Such considerations have led us to seek an alternative solution to the mapping problem. Clearly it is generally preferable to estimate and present a measure of risk rather than significance, but the variation in estimation precision across the map must also be taken into account. We suppose that the country or region to be mapped is divided into N mutually exclusive districts. Under the conventional approach to mapping risk, it is implicit that the relative risk in the ith district, θ_i, is an unknown parameter to be estimated. If O_i and E_i are, respectively, the observed and expected numbers of deaths in the ith district, the standard maximum likelihood estimate (MLE) of θ_i under Poisson assumptions on O_i is the SMR, $\hat{\theta}_i = O_i/E_i$. However, taken together, $\{\hat{\theta}_i, i = 1, \ldots, N\}$ are not necessarily the best estimates of $\{\theta_i\}$.

For the analogous problem of the simultaneous estimation of several normal means, James and Stein (1961) showed the existence of estimators with lower total squared error loss than the MLEs. Efron and Morris (1973) have further justified these estimators using an empirical Bayes approach, effectively introducing a mixture distribution for the unknown means. A fully Bayesian derivation of these estimates was given by Lindley and Smith (1972), and an application of the Bayesian analysis to the case of observations from different binomial distributions was given by Leonard (1972), who also indicated a possible approach for approximate Bayes estimators for multinomial observations (Leonard, 1975). Louis (1984) has derived similar estimates for the situation where there is an ensemble of parameters to be estimated, by requiring that their dispersion be that of the estimated mixture distribution.

In this paper we derive empirical Bayes estimates (Robbins, 1964) of the relative risks $\{\theta_i\}$. We adopt a random-effects (or mixture) model that assumes a parametric probability density function (pdf), $f(\boldsymbol{\theta})$, for the distribution of relative risks between districts. We further assume that, conditional on θ_i, O_i is a Poisson variable with expectation $\theta_i E_i$. The marginal distribution of the $\{O_i\}$ allows estimation of the parameters of the relative risk distribution, $f(\boldsymbol{\theta})$. The posterior expectations of $\{\theta_i\}$ given $\{O_i\}$ may then be estimated and provide the empirical Bayes estimates of the relative risks.

In the next section, we consider three types of mixture model for $f(\boldsymbol{\theta})$, the distribution of relative risk, and describe how parameter estimation may be carried out under each one. In Section 4, the theory is extended to allow the estimation of age effect when valid $\{E_i\}$ are unavailable, and the methods are illustrated in Section 5, using data on lip cancer in Scotland. In a recent paper, Tsutakawa, Shoop, and Marienfeld (1985) have derived empirical Bayes estimates of mortality rates. Our work differs from theirs in a number of important respects, which are pointed out in the body of the paper.

2. Mixture Models for the Distribution of Relative Risks

In this section, three models are considered for $f(\boldsymbol{\theta})$. The first, the gamma distribution, produces observed counts $\{\theta_i\}$ that are negative binomial, and the empirical Bayes estimates are of a natural form and straightforwardly computable. In Section 2.2, we introduce a log-normal model, which requires some approximations but allows for the possibility of spatial correlation. There does not appear to be a multivariate form of the gamma distribution suitable for this purpose (Besag, 1974). Finally, we outline a nonparametric method of estimating $f(\boldsymbol{\theta})$, when the $\{\theta_i\}$ are independent and identically distributed (iid).

Tsutakawa et al. (1985) use a normal distribution for the logit of the probability of disease in each geographical unit, but do not consider spatial correlation.

2.1 *The Gamma Model*

We assume that the relative risks $\{\theta_i\}$ are iid, following a gamma distribution with scale parameter α, and shape parameter ν, i.e., with mean ν/α and variance ν/α^2. Since

conditional on θ_i, the observed deaths O_i are Poisson variates with expectation $\theta_i E_i$, it follows that unconditionally, O_i are negative binomial with

$$\mathrm{E}(O_i) = E_i\nu/\alpha; \tag{1}$$

$$\mathrm{var}(O_i) = E_i\nu/\alpha + E_i^2\nu/\alpha^2. \tag{2}$$

It is easily shown that, conditional on O_i, the distribution of θ_i is also of the gamma form with scale parameter $(E_i + \alpha)$ and shape parameter $(O_i + \nu)$. Thus, the posterior expectation of θ_i conditional on O_i is

$$\mathrm{E}(\theta_i \mid O_i;\ \alpha,\ \nu) = \frac{O_i + \nu}{E_i + \alpha}. \tag{3}$$

Empirical Bayes estimates $\hat{\theta}_i$ of θ_i are provided by (3) with estimates $\hat{\nu}$, $\hat{\alpha}$ substituted for ν and α. We note that the empirical Bayes estimates are less dispersed than the SMRs: each estimate is a compromise between the SMR, O_i/E_i, and $\hat{\nu}/\hat{\alpha}$, the estimated mean of the distribution of $\{\theta_i\}$. Estimates based on large numbers of observed deaths are close to the corresponding SMR, while those based on small numbers are close to the overall mean relative risk.

Estimation of α and ν in the negative binomial distribution can be accomplished via maximum likelihood. The log-likelihood based on the distributions of $\{O_i\}$ is

$$L(\alpha,\ \nu) = \sum_i \left[\log \frac{\Gamma(O_i + \nu)}{\Gamma(\nu)} + \nu \log(\alpha) - (O_i + \nu)\log(E_i + \alpha)\right]. \tag{4}$$

Setting the first derivative of (4) with respect to α and ν equal to 0 yields the equations

$$\frac{\hat{\nu}}{\hat{\alpha}} = \frac{1}{N}\sum_i \frac{O_i + \hat{\nu}}{E_i + \hat{\alpha}} = \frac{1}{N}\sum_i \hat{\theta}_i \tag{5}$$

and

$$\sum_{i=1}^{N} \sum_{j=0}^{O_i - 1} \frac{1}{\nu + j} + N \log(\alpha) - \sum_{i=1}^{N} \log(E_i + \alpha) = 0, \tag{6}$$

where we adopt the convention that $\sum_{j=0}^{-1} 1/(\nu + j) = 0$.

From (5), the MLE of the mean of the distribution of $\{\theta_i\}$ is equal to the arithmetic mean of the empirical Bayes estimates. Equations (5) and (6) can be solved simultaneously for α and ν, by standard iterative procedures. A computationally simpler alternative to maximum likelihood estimation involves using a moment estimator for the variance of the mixing distribution, ν/α^2 (McCullagh and Nelder, 1983). We set the Pearsonian chi-square based on (1) and (2) to $N - 1$, its asymptotic expectation, giving

$$\frac{\hat{\nu}}{\hat{\alpha}^2} = \frac{1}{N - 1}\sum_i \left(1 + \frac{\hat{\alpha}}{E_i}\right)(\hat{\theta}_i - \hat{\nu}/\hat{\alpha})^2, \tag{7}$$

where $\{\hat{\theta}_i\}$ are the empirical Bayes estimates. Together, (5) and (7) can be used recursively to compute $\hat{\nu}$ and $\hat{\alpha}$. At each stage of the iteration, we calculate empirical Bayes estimates of $\{\theta_i\}$ from (3) based on current estimates of ν and α, and then use the right-hand sides of (5) and (7) to provide new estimates of ν and α.

The model for the distribution of $\{\theta_i\}$ can be extended to allow for covariates, $\{\mathbf{z}_i\}$. For example, if the degree of urbanisation of districts is strongly predictive of mortality, then it seems logical to use this relationship to improve the estimates of relative risk. Thus, the estimate for a district with relatively few observed and expected cases should be drawn not towards the overall mean relative mortality, but towards an estimated value consistent with

its degree of urbanisation. Such an extension may be incorporated by allowing distinct values, $\{\alpha_i\}$, for the scale parameters of the distributions of each θ_i, and adopting the log-linear model

$$\mathrm{E}(\theta_i) = \frac{\nu}{\alpha_i} = \exp(\mathbf{z}_i^{\mathrm{T}}\boldsymbol{\phi}).$$

The problem of estimation of the parameters $\boldsymbol{\phi}$ and ν of the log-linear model in the presence of overdispersion has been discussed by McCullagh and Nelder (1983) and Breslow (1984).

2.2 *Log-normal Model*

In this section, we consider a multivariate log-normal distribution (Johnson and Kotz, 1973, pp. 17–20) for $\boldsymbol{\theta}$. Thus, we suppose that $\boldsymbol{\beta}$, the vector of log relative risks [i.e., $\theta_i = \exp(\beta_i)$], has a multivariate normal distribution, with mean $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$.

For the empirical Bayes estimates, we require the expected values of $\{\theta_i\}$ conditional on $\{O_i\}$. However, under the log-normal model, these cannot be obtained in closed form. We therefore adopt an approximation, in which $\psi(\boldsymbol{\beta})$, the Poisson likelihood for $\{\beta_i\}$ given the data $\{O_i\}$, is assumed to be quadratic. With this assumption the posterior density of $\boldsymbol{\beta}$ conditional on $\mathbf{O}$ can be shown to be multivariate normal (Leonard, 1975), with mean $\mathbf{b}$ and variance $\mathbf{S}$, which are given in Appendix 1.

The quadratic approximation for ψ requires that $\psi(\boldsymbol{\beta})$ be expanded about a suitable $\tilde{\boldsymbol{\beta}}$. An obvious choice for $\tilde{\boldsymbol{\beta}}$ is the vector whose ith element is $\log(O_i/E_i)$, the conventional MLE of the log relative risk. However, this choice suffers from the fact that it is undefined when $O_i = 0$, so we use its bias-corrected version (Snedecor and Cochran, 1967), with ith element

$$\tilde{\beta}_i = \log\left(\frac{O_i + \frac{1}{2}}{E_i}\right). \tag{8}$$

Although the quadratic approximation for $\psi(\boldsymbol{\beta})$ provides an explicit formula for the empirical Bayes estimate of θ_i, it does not result in an explicit form of the distribution of $\{O_i\}$. Nevertheless, maximum likelihood estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be accomplished using the EM algorithm (Dempster, Laird, and Rubin, 1977), since only the marginal distribution of $\boldsymbol{\beta}$ and the conditional distribution of $\boldsymbol{\beta}$ given $\mathbf{O}$ are required, both of which we have. Details are given in Appendix 1.

In the simplest case, where the β_i's are iid, their distribution is defined by two parameters: $\phi = \mu_i$ and σ^2, where $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$. The resulting EM algorithm involves successive updating of the estimates

$$\hat{\phi} = \frac{1}{n}\sum_i b_i = \bar{b}, \tag{9}$$

$$\hat{\sigma}^2 = \frac{1}{n}\left\{(\hat{\sigma}^2)'\sum_i [1 + (\hat{\sigma}^2)'(O_i + \tfrac{1}{2})]^{-1} + \sum_i (b_i - \hat{\phi}')^2\right\}, \tag{10}$$

which are in turn used to update the empirical Bayes estimate $\mathbf{b}$. In this case, this has the form

$$b_i = \frac{\hat{\phi}' + (O_i + \frac{1}{2})(\hat{\sigma}^2)'\log[(O_i + \frac{1}{2})/E_i] - (\hat{\sigma}^2)'/2}{1 + (O_i + \frac{1}{2})(\hat{\sigma}^2)'}. \tag{11}$$

It can be seen that, like (3), (11) is a compromise between $\tilde{\beta}_i$, the "local" estimate of log relative risk, and ϕ, the overall mean of the distribution of the β_i's; again, log-SMRs based on large numbers are affected less than those based on small numbers. When O_i is very

small, the quadratic approximation is clearly suspect, but should still serve the purpose of weighting the corresponding relative risk estimate b_i towards $\hat{\mu}$.

We now suppose that the log relative risks are correlated. In particular, we focus on models where correlation is dependent on geographical proximity. Models of this kind for Gaussian variates on a map have been considered by a number of authors, including Ord (1975), Besag (1974), and Cook and Pocock (1983). All of the proposed models have in common that $\boldsymbol{\Sigma}$ may be expressed as a function of a small number of parameters. For the numerical example in Section 5 of this paper, we have used the conditional autoregression (CAR) defined by

$$\mathrm{E}(\beta_i \mid \beta_j, j \neq i) = \mu_i + \rho \sum_j W_{ij}(\beta_j - \mu_j), \tag{12}$$

$$\mathrm{var}(\beta_i \mid \beta_j, j \neq i) = \sigma^2, \tag{13}$$

where $\mathbf{W}$ is the adjacency matrix of the map, defined by

$$W_{ij} = \begin{cases} 1 & i \text{ and } j \text{ are adjacent districts} \\ 0 & \text{otherwise} \end{cases}$$

It can be shown (Besag, 1974) that under (12) and (13), $\mathrm{E}(\boldsymbol{\beta}) = \boldsymbol{\mu}$ and the variance–covariance matrix of $\boldsymbol{\beta}$ is $\boldsymbol{\Sigma} = \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}$.

In the discussion in Section 5, we mention some other possible models for spatial correlation. Specific aspects of the estimation of parameters under the CAR are given in Appendix 2.

2.3 *Nonparametric Model*

A third alternative we have considered is to ignore any spatial correlation and assume that $\{\theta_i\}$ are iid random variables with density, $f(\theta)$, of unspecified parametric form. Laird (1978) has shown that $f(\theta)$ may be estimated nonparametrically using an EM algorithm and that the MLE, $\hat{f}$, has its support concentrated onto a finite number of points. The number of support points for $\hat{f}$ cannot be larger than N, the number of Poisson observations.

3. Estimation of Age Effects

The empirical Bayes procedures were derived under the implicit assumption that a valid set of external age-specific rates, $\{\xi_j, j = 1, \ldots, p\}$, say, is available, and that O_{ij}, the observed number of cases in the jth age group in the ith district, has expectation $\mathrm{E}(O_{ij}) = \theta_i\xi_j Y_{ij}$, where Y_{ij} is the total person–years of observation giving rise to O_{ij}. By the conventional approach, the expected number of cases in the ith district is

$$\theta_i E_i = \theta_i\left(\sum_j Y_{ij}\xi_j\right),$$

and the SMR, O_i/E_i, is the MLE of θ_i. When $\{\xi_j\}$ are unknown, they can be estimated simultaneously with the $\{\theta_i\}$ by maximum likelihood, most conveniently using a recursive algorithm first proposed by Mantel and Stark (1968). We now show how an analogous procedure can be used in the context of empirical Bayes estimation of $\{\theta_i\}$.

Since $f(\boldsymbol{\theta})$ does not depend on the parameters $\{\xi_j\}$, it may be shown that, under regularity conditions,

$$\frac{\delta}{\delta\boldsymbol{\xi}} \log f(\mathbf{O}) = \mathrm{E}\left[\frac{\delta}{\delta\boldsymbol{\xi}} \log f(\mathbf{O} \mid \boldsymbol{\theta}) \mid \mathbf{O}\right],$$

where $\mathbf{O}$ now refers to $\{O_{ij}, i = 1, \ldots, N; j = 1, \ldots, p\}$, and the expectation is taken over the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{O}$.

The inner part of the right-hand side of this equation is the conventional log-likelihood for simultaneous maximum likelihood estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\xi}$, i.e.,

$$\log f(\mathbf{O} \mid \boldsymbol{\theta}; \boldsymbol{\xi}) = \sum_{ij} [O_{ij}\log(\theta_i\xi_j) - Y_{ij}\theta_i\xi_j]$$
$$= \sum_i O_{i.}\log(\theta_i) + \sum_j O_{.j}\log(\xi_j) - \sum_i \sum_j Y_{ij}\theta_i\xi_j.$$

The first term is irrelevant for estimating ξ_j, and the second term does not involve θ_i. However, the last term is linear in θ_i, so that taking its expectation over the posterior distribution of θ_i is equivalent to replacing the $\{\theta_i\}$ by empirical Bayes estimates. Thus, the MLE of $\boldsymbol{\xi}$ obeys

$$\hat{\xi}_j = \frac{O_{.j}}{\sum_i Y_{ij}\hat{\theta}_i}.$$

Estimation of ξ_j and $\boldsymbol{\theta}$ may be achieved simultaneously, by updating the estimates of $\{\xi_j\}$ at each step of the recursion used to estimate $\boldsymbol{\theta}$.

As in the case of conventional maximum likelihood estimation of θ_i and ξ_j, a linear constraint is required to obtain a unique set of estimates of the age effects. It is natural to constrain the mean of the age effects to be unity.

It should be noted that the results of this section lead to the empirical Bayes equivalent of indirectly standardized rate estimates, as proposed by Mantel and Stark (1968) and Breslow and Day (1975). These authors used a model in which the joint effect of area and

Table 1
Lip cancer incidence in Scotland by county: Observed numbers (O), SMRs, and empirical Bayes estimates of the relative risk

			Empirical Bayes estimate				
County	*O*	SMR	I	II	III	IV	Adjacent counties
1	9	652.2	421.9	495.5	453.4	345.0	5, 9, 11, 19
2	39	450.3	414.6	424.5	415.7	367.2	7, 10
3	11	361.8	302.2	310.6	313.0	362.2	6, 12
4	9	355.7	289.7	298.1	271.8	320.5	18, 20, 28
5	15	352.1	308.0	313.9	328.7	320.7	1, 11, 12, 13, 19
6	8	333.3	272.1	277.9	290.3	310.5	3, 8
7	26	320.6	298.7	300.6	307.5	320.9	2, 10, 13, 16, 17
8	7	304.3	251.0	253.3	266.5	291.9	6
9	6	303.0	244.6	247.1	250.0	281.2	1, 11, 17, 19, 23, 29
10	20	301.7	278.4	279.5	285.7	318.3	2, 7, 16, 22
11	13	295.5	264.1	265.0	290.9	314.3	1, 5, 9, 12
12	5	279.3	226.4	226.2	269.0	254.9	3, 5, 11
13	3	277.8	208.7	208.8	254.1	224.9	5, 7, 17, 19
14	8	241.7	216.5	213.3	197.2	248.0	31, 32, 35
15	17	216.8	207.5	204.8	195.3	242.0	25, 29, 50
16	9	197.8	186.9	182.3	193.1	175.9	7, 10, 17, 21, 22, 29
17	2	186.9	164.4	156.2	201.7	166.2	7, 9, 13, 16, 19, 29
18	7	167.5	162.3	156.7	146.7	137.8	4, 20, 28, 33, 55, 56
19	9	162.7	159.4	154.8	198.1	128.1	1, 5, 9, 13, 17
20	7	157.7	154.7	149.3	166.8	129.7	4, 18, 55
21	16	153.0	152.0	49.2	152.5	117.4	16, 29, 50
22	31	136.7	137.1	135.7	146.9	116.5	10, 16
23	11	125.4	127.5	124.5	116.8	116.5	9, 29, 34, 36, 37, 39
24	7	124.6	127.7	123.6	83.5	116.8	27, 30, 31, 44, 47, 48, 55, 56
25	19	122.8	124.2	122.5	131.6	116.5	15, 26, 29
26	15	120.1	122.0	120.0	119.6	116.5	25, 29, 42, 43
27	7	115.9	120.3	116.6	118.9	115.5	24, 31, 32, 55
28	10	111.6	115.2	112.7	121.6	116.0	4, 18, 33, 45

age on risk is multiplicative. In our case, a random component is estimated to represent the effect of area on risk, but it still acts multiplicatively with the age effect. This contrasts with the approach of Tsutakawa et al. (1985), who obtain empirical Bayes estimates for a model in which the random components due to area are independent, and hence different, for each age–sex group within area. As well as multiplying the number of parameters to be estimated, their model is not consistent with the empirical finding that area effects on relative risk are roughly constant across age groups.

4. An Example: Lip Cancer in Scotland

Table 1 sets out observed and "expected" cases of lip cancer registered during the 6 years from 1975 to 1980 in each of the 56 counties of Scotland, and gives four different sets of estimates of relative incidence, starting with the SMR. The "expected" numbers of cases, $\{E_i\}$, were actually calculated by the method of Mantel and Stark (1968), i.e., they are based on MLEs of the age effects $\{\xi_j\}$ in the simple multiplicative model. However, for purpose of the example we regard them as known constants. In the table, all four estimates of relative incidence have been multiplied by 100, in keeping with the convention established for SMRs.

The data are given for 56 counties, and these are arranged in descending order of incidence as measured by SMRs, which vary between 0 and 652. The columns labelled I and II show the empirical Bayes estimates based on the assumption that $\{\theta_i\}$ are iid gamma and log-normal variates, respectively. There is little difference between the two sets of

Table 1—*Continued*

			Empirical Bayes estimate				
County	*O*	SMR	I	II	III	IV	Adjacent counties
29	16	111.3	113.7	112.1	104.3	116.4	9, 15, 16, 17, 21, 23, 25, 26, 34, 43, 50
30	11	107.8	111.4	109.3	89.1	116.1	24, 38, 42, 44, 45, 56
31	5	105.3	112.6	108.8	90.3	111.9	14, 24, 27, 32, 35, 46, 47
32	3	104.2	115.3	109.7	126.5	113.1	14, 27, 31, 35
33	7	99.6	105.7	103.2	106.3	110.0	18, 28, 45, 56
34	8	93.8	99.6	97.9	68.7	112.9	23, 29, 39, 40, 42, 43, 51, 52, 54
35	11	89.3	93.9	92.9	95.8	114.2	14, 31, 32, 37, 46
36	9	89.1	94.6	93.4	96.2	112.5	23, 37, 39, 41
37	11	86.8	91.4	90.6	87.4	113.4	23, 35, 36, 41, 46
38	8	85.6	91.8	90.8	70.7	109.8	30, 42, 44, 49, 51, 54
39	6	83.3	91.5	90.3	85.7	105.0	23, 34, 36, 40, 41
40	4	75.9	87.9	86.7	71.3	95.0	34, 39, 41, 49, 52
41	10	53.3	58.5	60.1	51.9	40.7	36, 37, 39, 40, 46, 49, 53
42	8	50.7	56.9	59.0	60.6	41.0	26, 30, 34, 38, 43, 51
43	2	46.3	66.6	69.8	85.9	65.1	26, 29, 34, 42
44	6	41.0	48.4	51.9	53.5	37.4	24, 30, 38, 48, 49
45	19	37.5	39.8	41.4	48.8	36.2	28, 30, 33, 56
46	3	36.6	49.6	55.2	52.3	42.2	31, 35, 37, 41, 47, 53
47	2	35.8	54.0	60.2	48.0	49.7	24, 31, 46, 48, 49, 53
48	3	32.1	44.2	50.7	55.7	38.7	24, 44, 47, 49
49	28	31.6	33.0	34.2	30.8	36.2	38, 40, 41, 44, 47, 48, 52, 53, 54
50	6	30.6	36.8	41.3	68.7	36.2	15, 21, 29
51	1	29.1	57.5	65.0	71.9	57.3	34, 38, 42, 54
52	1	27.6	55.4	63.5	64.9	55.1	34, 40, 49, 54
53	1	17.4	38.3	51.6	53.9	40.4	41, 46, 47, 49
54	1	14.2	32.3	47.1	50.3	37.8	34, 38, 49, 51, 52
55	0	0.0	30.9	58.5	92.7	61.2	18, 20, 24, 27, 56
56	0	0.0	56.4	70.4	72.6	40.9	18, 24, 30, 33, 45, 55

estimates. Both show considerably less dispersion than the SMRs, their range being from 31 to 422 for the gamma model, and from 34 to 495 for the log-normal model. However, the ranks of the empirical Bayes estimates differ little from the corresponding ranks of the SMRs for most counties. The exceptions are counties such as (55) and (56), where SMRs are based on such small numbers of cases as to be highly unreliable. The empirical Bayes estimates for these counties are drawn towards the overall mean. Conversely, certain counties, notably (45) and (49), change in rank order because they are based on so much data that the empirical Bayes estimates are virtually unchanged from the SMRs while others are modified. The estimates of coefficient of variation of the relative risk distribution agreed fairly closely—.78 for the gamma model and .85 for the log-normal model.

The column labelled III of Table 1 gives the estimates based on the spatial autocorrelation model (CAR), and the final column of the table gives the numbers of adjacent counties. There is a high degree of spatial relationship in this data set: the MLE of ρ is .174 as compared with the maximum value of .175 that ρ can take in a valid CAR process, for this pattern of adjacencies. Most of the empirical Bayes estimates are little different from their equivalents in the uncorrelated log-normal model. The only counties affected substantially are those with few cases and SMRs that differ appreciably from surrounding counties. A clear example is (24), in which only 7 cases were observed, but which is adjacent to several low-risk areas; the spatially correlated estimate is substantially reduced.

Finally, column IV gives the nonparametric empirical Bayes estimates. The MLE of $f(\theta)$ is concentrated in four masses at $\theta = .362$, 1.16, 3.08, and 3.89 (SMRs of 36.2, 116, 308, and 389, respectively) with probabilities .275, .478, .186, and .061, respectively. This suggests that approximately one in four of the areas have a very low risk of lip cancer (36% of expectation), another one in four have a very high risk, and the remaining 50% are intermediate. The nonparametric empirical Bayes estimates differ little in rank order from their two parametric equivalents (columns I and II). Here it is very clear that the distribution of the empirical Bayes estimates does not mirror $\hat{f}(\theta)$.

5. Discussion

In this paper we have presented empirical Bayes procedures for the simultaneous estimation of a large number of relative risks, with a view to their visual presentation on a map. The method involves estimation of the parameters of a random process that is assumed to generate the relative risks, and calculation of posterior expectations as the estimates of individual relative risks. For each district, these may be viewed as smoothed SMR estimates, where the extent of smoothing is determined by the size of the crude SMR, its precision, and the estimated underlying relative risk distribution. Thus, in contrast to other possible methods of smoothing such as spline fitting, the extent of smoothing is totally determined by the data.

The empirical Bayes estimates can be expressed in a simple, closed form, in contrast to those of Tsutakawa et al. (1985), which must be obtained by numerical integration.

A number of other Gaussian spatial autoregression processes have been suggested in the statistical literature, and could be used in the same way as the CAR. Besag (1974) discusses the simultaneous autoregression (SAR), in which the effect of neighbouring districts on each other is defined simultaneously, rather than conditionally as in the CAR process. Ord (1975) defines a similar process. In all of these cases, weighting schemes other than the **W** matrix defined above can be employed to account for such factors as relative populations and sizes of common borders, although certain restrictions must be satisfied in order to produce valid multivariate normal distribution functions. Cook and Pocock (1983) parametrized the covariance between mortality in pairs of cities by an inverse distance function. As we did, these authors viewed the spatial process as generating unobservable log relative

risks, but their primary interest was in the systematic part of the model ($\mathbf{z}^T\boldsymbol{\phi}$ in our notation), rather than improved estimation of the relative risks themselves. The EM approach to the maximum likelihood estimation of model parameters provides a very simple algorithm, which could readily be modified for use with any of these spatial processes.

However, in view of the inherently descriptive nature of disease mapping, it may be felt that the spatially correlated approach implies the adoption of hypotheses that are too specific. After all, there is no a priori reason why geographic proximity should be reflected in correlated cancer rates. The uncorrelated mixture models, on the other hand, do no more than provide an ensemble of risk estimates in which relative precision has been taken into account. The nonparametric version is based on the fewest assumptions, and in addition provides information of a qualitative as well as quantitative nature, in that it has the possibility of classifying areas into one of a small number of categories, based on their posterior probabilities of belonging to each category.

Our preference for mapping data such as has been presented here is, therefore, for the nonparametric empirical Bayes estimates which seem to overcome the main shortcomings of the SMRs with minimal assumptions. However, for mapping diseases in very small areas when the expected case frequencies are small, clearly it will be necessary to allow for spatial autocorrelation and work is in progress to evaluate such methods.

Acknowledgements

We would like to thank M. Smans for bringing the mapping problem to our attention, and for computational assistance.

Résumé

Beaucoup de tentatives ont eu lieu ces dernières années pour présenter des cartes géographiques d'incidence ou de mortalité liées à des maladies telles que le cancer. Ces cartes présentent en général soit des taux relatifs dans chaque unité géographique, comme on peut les mesurer par les rapports de mortalité standardisés (SMR) ou des index similaires, soit le degré de signification d'un test comparant la différence des taux entre une unité géographique et le reste. Aucune de ces approches n'est pleinement satisfaisante et nous proposons une nouvelle approche utilisant l'estimation Bayesienne empirique. Les estimations fournies représentent une pondération du SMR, du taux moyen relatif général, et du taux moyen relatif "local," dans les unités géographiques voisines. Le compromis obtenu dépend de la fiabilité de chaque SMR individuel, et de la dispersion d'ensemble des taux relatifs dans les différentes unités.

References

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36,** 192–236.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33,** 38–44.

Breslow, N. E. and Day, N. E. (1975). Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases* **28,** 289–303.

Cook, D. G. and Pocock, S. J. (1983). Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. *Biometrics* **39,** 361–371.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm (with Discussion). *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association* **68,** 117–130.

International Agency for Research on Cancer (1982). *Report on the Workshop on Mapping of Cancer, Lyon, 10–11 December* 1981. IARC Internal Technical Report 82/003. Lyon: IARC.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the 4th Berkeley Symposium* **1,** 361–379. Berkeley, California: University of California Press.

Johnson, N. L. and Kotz, S. (1973). *Distributions in Statistics: Continuous Multivariate Distributions.* New York: Wiley.
Laird, N. M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73,** 805–811.
Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika* **59,** 581–589.
Leonard, T. (1975). Bayesian estimation methods for two-way contingency tables. *Journal of the Royal Statistical Society, Series B* **37,** 23–37.
Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with Discussion). *Journal of the Royal Statistical Society, Series B* **34,** 1–41.
Louis, T. A. (1984). Estimating a population of parameter values using Bayes and empirical Bayes methods. *Journal of the American Statistical Association* **79,** 393–398.
Mantel, N. and Stark, C. R. (1968). Computation of indirect-adjusted rates in the presence of confounding. *Biometrics* **24,** 997–1005.
McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.
Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70,** 120–126.
Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Annals of Mathematical Statistics* **35,** 1–20.
Snedecor, G. W. and Cochran, W. (1967). *Statistical Methods,* 6th edition. Ames, Iowa: Iowa State University Press.
Tsutakawa, R. K., Shoop, G. L., and Marienfeld, C. J. (1985). Empirical Bayes estimation of cancer mortality rates. *Statistics in Medicine* **4,** 201–212.

Received September 1984; *revised February* 1986 *and March* 1987.

Appendix 1

Estimation Under the Log-normal Model

The mean of the posterior density of $\boldsymbol{\beta}$ conditional on $\mathbf{O}$ (and hence the empirical Bayes estimate of $\boldsymbol{\beta}$) is

$$\mathbf{b} = [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\psi}''(\tilde{\boldsymbol{\beta}})]^{-1}[\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \boldsymbol{\psi}''(\tilde{\boldsymbol{\beta}})\tilde{\boldsymbol{\beta}} + \boldsymbol{\psi}'(\tilde{\boldsymbol{\beta}})], \tag{A.1}$$

and the variance–covariance matrix is

$$\mathbf{S} = [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\psi}''(\tilde{\boldsymbol{\beta}})]^{-1}, \tag{A.2}$$

where $\boldsymbol{\psi}'$ and $\boldsymbol{\psi}''$ are, respectively, the vector of first derivatives and (diagonal) matrix of second derivatives of $\boldsymbol{\psi}$ with respect to $\boldsymbol{\beta}$. With $\tilde{\boldsymbol{\beta}}$ given by (8), these are

$$\psi_i'(\tilde{\boldsymbol{\beta}}) = -\tfrac{1}{2} \quad \text{and} \quad \psi_{ii}''(\tilde{\boldsymbol{\beta}}) = -(O_i + \tfrac{1}{2}).$$

The E-step in the EM algorithm (Dempster et al., 1977) consists of calculating

$$Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathrm{E}_{\boldsymbol{\mu}',\boldsymbol{\Sigma}'}[l(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\beta}) \mid \mathbf{O}],$$

where l is the log-likelihood of the unknown parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ based only on the missing data $\boldsymbol{\beta}$, and the expectation $\mathrm{E}_{\boldsymbol{\mu}',\boldsymbol{\Sigma}'}$ of $\boldsymbol{\beta}$ is taken conditional on the observed data $\mathbf{O}$ using current estimates $\boldsymbol{\mu}'$ and $\boldsymbol{\Sigma}'$ of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The M-step consists of maximizing Q with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, to produce updated estimates, and the process is repeated until convergence. Now

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \boldsymbol{\beta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu})$$

so that

$$Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{E}_{\boldsymbol{\mu}',\boldsymbol{\Sigma}'}[(\boldsymbol{\beta} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}) \mid \mathbf{O}]. \tag{A.3}$$

The expectation of the quadratic form in (A.3) can be evaluated using the mean and covariance matrix of the conditional distribution of $\boldsymbol{\beta}$ given $\mathbf{O}$ defined by (A.1) and (A.2) with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ replaced

by $\boldsymbol{\mu}'$ and $\boldsymbol{\Sigma}'$, to give

$$Q(\mu, \Sigma) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log|\mathbf{D}| - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\left[\sum_i \sum_j D_{ij}S_{ij} + (\mathbf{b}-\mu)^{\mathrm{T}}\mathbf{D}(\mathbf{b}-\mu)\right], \quad \text{(A.4)}$$

where $\mathbf{D}$ is defined by $\boldsymbol{\Sigma} = \sigma^2\mathbf{D}^{-1}$. We also suppose for generality that $\boldsymbol{\mu} = \mathbf{z}\boldsymbol{\phi}$, where $\mathbf{z}$ is a known matrix of covariates, and $\boldsymbol{\phi}$ is a parameter vector to be estimated.

Maximization of Q with respect to σ^2, $\mathbf{D}$, and $\boldsymbol{\phi}$ for the M-step proceeds via differentiation of Q, and equating the results to 0. Differentiation with respect to $\boldsymbol{\phi}$ and σ^2 gives

$$\hat{\boldsymbol{\phi}} = (\mathbf{z}^{\mathrm{T}}\mathbf{D}\mathbf{z})^{-1}\mathbf{z}^{\mathrm{T}}\mathbf{D}\mathbf{b}, \quad \text{(A.5)}$$

$$\hat{\sigma}^2 = \frac{1}{n}\left[\sum_i \sum_j S_{ij}D_{ij} + (\mathbf{b}-\mathbf{z}\boldsymbol{\phi})^{\mathrm{T}}\mathbf{D}(\mathbf{b}-\mathbf{z}\boldsymbol{\phi})\right], \quad \text{(A.6)}$$

and substitution back into (11) leaves

$$g(\mathbf{D}) = \log|\mathbf{D}| - n\log(\hat{\sigma}^2)$$

to be maximized over $\mathbf{D}$.

Maximization of $g(\mathbf{D})$ with respect to the parameters of $\mathbf{D}$ can be accomplished either by direct search or Newton–Raphson solution of $g' = 0$. The resulting parameter values are then used to estimate $\mathbf{D}$, and $\boldsymbol{\phi}$ and σ^2 from (A.5) and (A.6) to complete the M-step, and the EM iteration continued by recalculation of $\mathbf{b}$ and $\mathbf{S}$ from (A.1) and (A.2), using the new parameter estimates. The empirical Bayes estimates of $\boldsymbol{\beta}$ are obtained from (A.1) using the final estimates $\hat{\boldsymbol{\mu}} = \mathbf{z}\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\Sigma}}^{-1} = \hat{\mathbf{D}}/\hat{\sigma}^2$.

Appendix 2

Maximization of $g(D)$ for the CAR in the Absence of Covariates

Applying the formula (12) yields

$$\hat{\phi} = \frac{1}{\sum_i \sum_j D_{ij}} \sum_i \sum_j D_{ij}b_j$$

$$= \frac{1}{N - \rho W_{\cdot\cdot}}(b_{\cdot} - \rho\mathbf{b}^{\mathrm{T}}\mathbf{W}_{\cdot}),$$

where $\mathbf{W}_{\cdot}$ is the vector of row sums of the weight matrix $\mathbf{W}$, and $W_{\cdot\cdot}$ is the total of all its elements. Then formula (13) gives

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i S_{ii} + \mathbf{b}^{\mathrm{T}}\mathbf{b} - \rho\left(\frac{1}{n}\sum_i \sum_j W_{ij}S_{ij} + \mathbf{b}^{\mathrm{T}}\mathbf{W}\mathbf{b}\right) - \frac{1}{n - \rho W_{\cdot\cdot}}(\rho\mathbf{b}^{\mathrm{T}}\mathbf{W}_{\cdot} - b_{\cdot})^2.$$

Substitution in (A.1) leaves

$$g(\mathbf{D}) = g(\rho) = \sum_i \log(1 - \rho\lambda_i) - n\log\left[K + \rho L - \frac{1}{N - \rho W_{\cdot\cdot}}(\rho\mathbf{b}^{\mathrm{T}}\mathbf{W}_{\cdot} - b_{\cdot})^2\right]$$

to be maximized over ρ. We used a NAG direct search subroutine for ρ on the interval $I = (0, 1/\lambda_{\max})$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{W}$. This maximum permissible value of ρ is defined by the requirement that $\mathbf{D} = \mathbf{I} - \rho\mathbf{W}$ be positive definite in order for $\boldsymbol{\Sigma}$ to be a covariance matrix. Thus, the solutions λ^* of $|\mathbf{D} - \lambda^*\mathbf{I}| = 0$ must all be positive, or equivalently, the solutions λ^* of

$$\left|\mathbf{W} - \left(\frac{1 - \lambda^*}{\rho}\right)\mathbf{I}\right| = 0.$$

Now the eigenvalues λ of $\mathbf{W}$ satisfy $|\mathbf{W} - \lambda\mathbf{I}| = 0$, so that ρ must be such that $\lambda^* = 1 - \rho\lambda > 0$ for all the eigenvalues of $\mathbf{W}$; this requirement implies that $\rho < 1/\lambda_{\max}$, as stated above. It may be noted that the results of this Appendix are applicable for any symmetric weight matrix $\mathbf{W}$, not just the adjacency matrix.