ENVIRONMETRICS
*Environmetrics* 2003; **14**: 431–451 (DOI: 10.1002/env.598)

# Empirical Bayes estimators and non-parametric mixture models for space and time–space disease mapping and surveillance

Dankmar Böhning*,†

*Institute for International Health, Epidemiology and Biometry, Joint Centre for Humanities and Health Sciences, Free University Berlin/Humboldt University at Berlin, Fabeckstr. 60-62, 14195 Berlin, Germany*

SUMMARY

The analysis of the geographic variation of disease and its representation on a map is an important topic in epidemiological research and in public health in general. Identification of spatial heterogeneity of relative risk using morbidity and mortality data is required. Frequently, interest is also in the analysis of space data with respect to time, where typically data are used which are aggregated in certain time windows like 5 or 10 years. The occurrence measure of interest is usually the *standardized mortality (morbidity) ratio* (SMR). It is well known that disease maps in space or in space and time should not solely be based upon the crude SMR but rather some smoothed version of it. This fact has led to a tremendous amount of theoretical developments in spatial methodology, in particular in the area of hierarchical modeling in connection with fully Bayesian estimation techniques like Markov chain Monte Carlo. It seems, however, that at the same time, where these theoretical developments took place, on the practical side only very few of these developments have found their way into daily practice of epidemiological work and surveillance routines. In this article we focus on developments that avoid the pitfalls of the crude SMR and simultaneously retain a simplicity and, at least approximately, the validity of more complex models. After an illustration of the typical pitfalls of the crude SMR the article is centered around three issues: (a) the separation of spatial random variation from spatial structural variation; (b) a simple mixture model for capturing spatial heterogeneity; (c) an extension of this model for capturing temporal information. The techniques are illustrated by numerous examples. Public domain software like Dismap is mentioned that enables easy mixture modeling in the context of disease mapping. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: descriptive measure for spatial heterogeneity; simple estimators for heterogeneity variance; mixture models for space-time data

## 1. INTRODUCTION

Environmental justice and equity are emerging concepts in the development of environmental health policy. These concepts are related to questions on the spatial distribution of environmental contaminants in the population leading to the potential occurrence of certain diseases in different parts of the population. Disease mapping can be defined as a method for displaying the spatial distribution of disease occurrence (or exposure occurrence), the most prominent forms being the

*Correspondence to: D. Böhning, Institute for International Health, Epidemiology and Biometry, Joint Centre for Humanities and Health Sciences, Free University Berlin/Humboldt University at Berlin, Fabeckstr. 60-62, 14195 Berlin, Germany.
†E-mail: boehning@zedat.fu-berlin.de


*Received 16 March 2002*
*Accepted 30 April 2002*

variety of existing cancer atlases (Holland, 1991; 1992; Becker *et al.*, 1984; Becker and Wahrendorf, 1997; Cartwright *et al.*, 1990). A frequent objective in geographic epidemiology is to present that part of the spatial variation of a disease occurrence distribution, which cannot be explained by the different distribution of known factors in the various regions nor is it due to random variation. Often, interest is in the hope to find hints to unknown risk factors. Typically, the occurrence measure used in epidemiology and public health institutions is the standardized mortality ratio (SMR), computed only for some aggregated unit such as an area (county, municipality, etc.). It is defined as

$$SMR_i = O_i/E_i$$

in the $i$th region, where $O_i$ are the observed death (mortality) or disease (morbidity) cases, and $E_{\mathrm{i}}$ are the expected cases computed from a reference population, for each of the $i = 1, \ldots, n$ areas. A typical layout of this situation is displayed in Figure 1 for the states of Austria.

Two conventional methods are used to construct the disease map. The first one uses a classification based upon a certain percentile of the empirical SMR-distribution. An example is given in Figure 2 (top) for childhood leukemia in the new states of Germany for the period of 1980–1989, using quartiles. The second conventional method used frequently is based upon the Poisson distribution

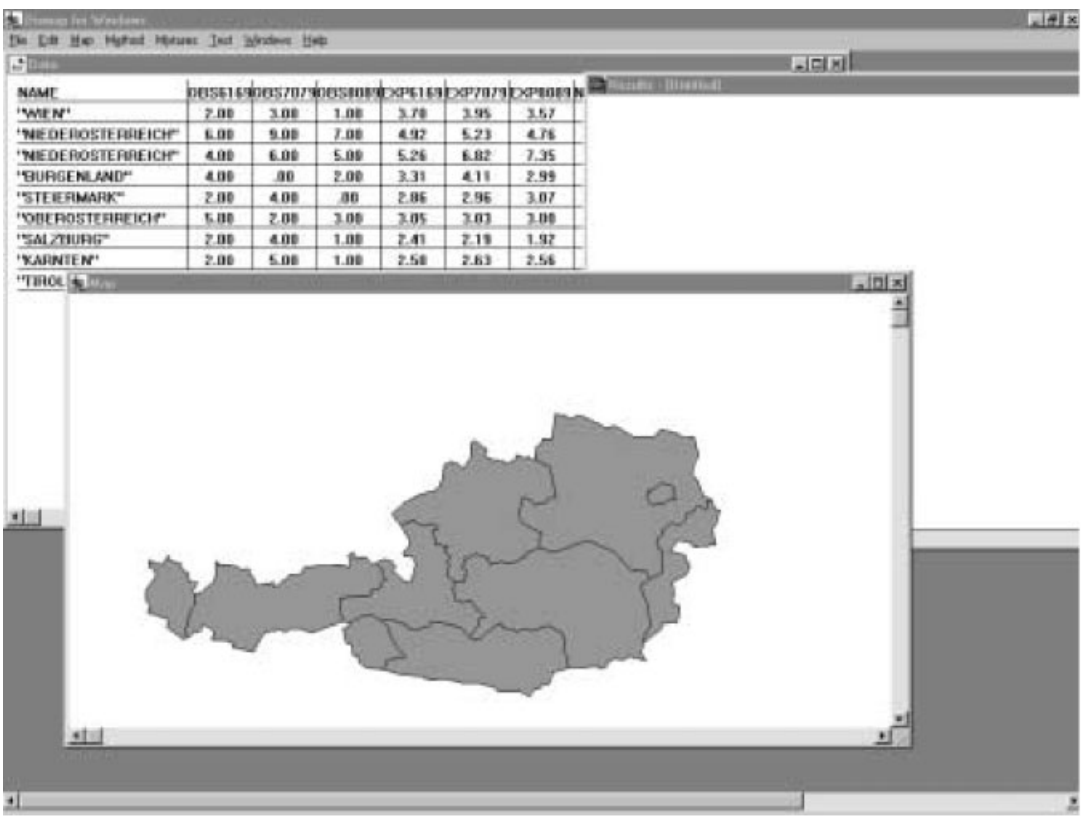


Figure 1. Typical layout for a disease mapping situation here exemplified at the states of Austria using the package dismap

 

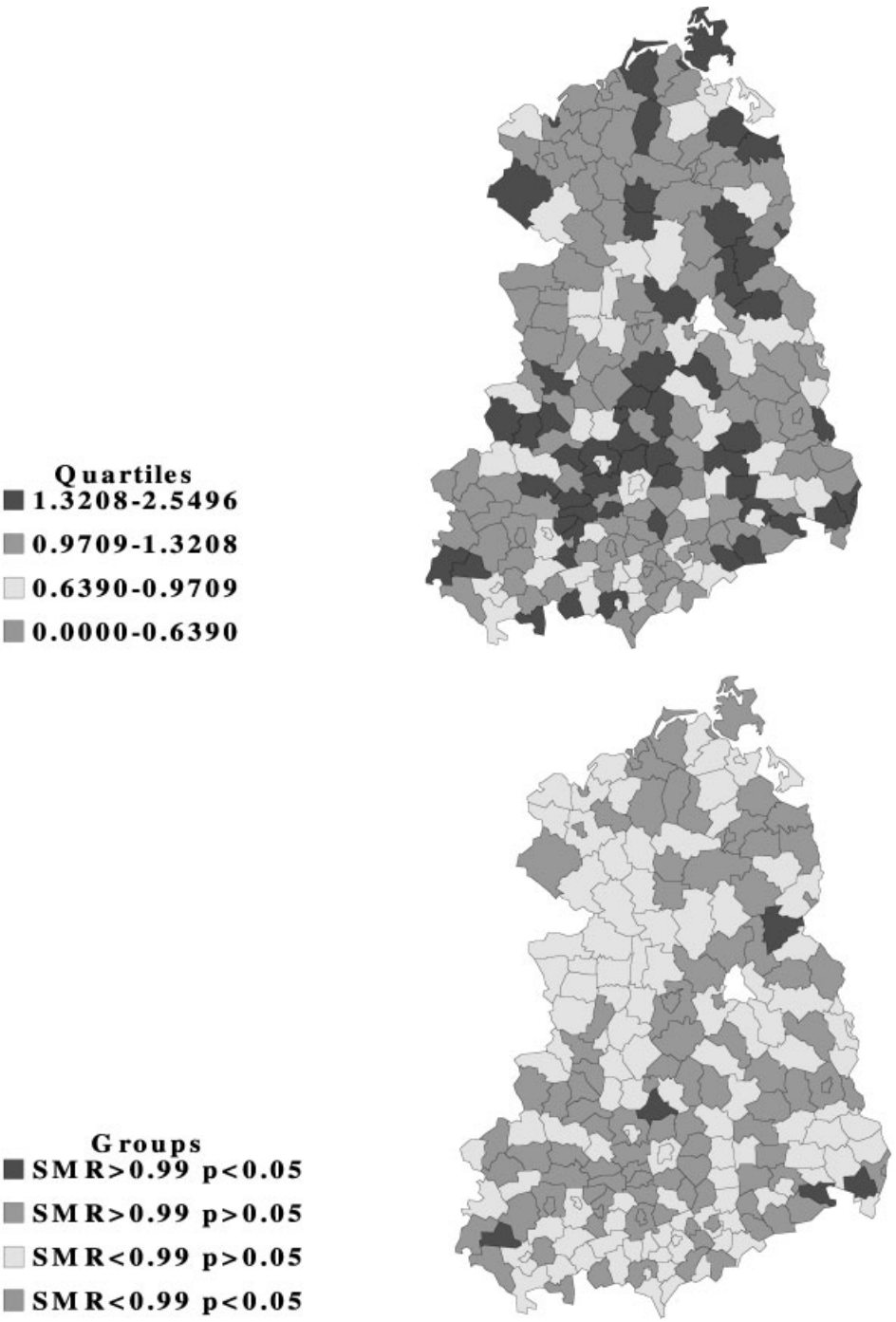


Figure 2. Classification based upon quartiles (top) and the *P*-value of the Poisson distribution (bottom) for the example of childhood leukemia, new states of Germany, 1980–89

 

$\mathrm{Po}(o_i \mid \lambda E_i) = \exp(-\lambda E_i)(\lambda E_i)^{o_i}/o_i!$; in particular, it is classification based on the $P$-value under the Poisson distribution,

$$P(O_i \geq o_i) = \mathrm{Po}(o_i \mid \lambda E_i) + \mathrm{Po}(o_i + 1 \mid \lambda E_i) + \cdots$$

This method is illustrated in Figure 2 (bottom) for the same data.
One of the disadvantages of the conventional methods can be seen in the fact that they can represent considerable random variation on the map. Clayton and Kaldor (1987) formulate the critical issues:

> *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 across atlases.*

For further discussion see also Böhning (2000, Chap. 7). In the forgoing example of childhood leukemia both methods represent only (Poisson) random variation. If the random variation is removed from the map there is no variation left, as can be seen in Figure 3. How this was found out will be revealed in the following section.

## 2. SOME SIMPLE DESCRIPTIVE MEASURES FOR SPATIAL HETEROGENEITY

We describe the usual, basic assumptions. Conditional on the values of $\lambda$ the observed number of cases is assumed to be Poisson: $O_i \sim \mathrm{Po}(o_i \mid \lambda E_i)$. Furthermore, we allow for heterogeneity in $\lambda$. It is assumed that $\lambda$ has some (not further specified) distribution $P$ with existing mean $\mu$ and variance $\tau^2$ ($P$ will be denoting the heterogeneity distribution of $\lambda$). Taking both distributional concepts together we are led to the marginal distribution:

$$\int_0^\infty \left[\exp(-\lambda E_i)(\lambda E_i)^{o_i}/o_i!\right] p(\lambda)\,\mathrm{d}\lambda$$

Thus we have three distributions:

- the conditional distribution $O_i \mid \lambda \sim \mathrm{Po}(o_i \mid \lambda E_i)$ (or Level I distribution)
- the heterogeneity distribution $\lambda \sim P$, having density $p(\lambda)$ (or Level II distribution)
- and the marginal distribution

$$O_i \sim \mathrm{f}(o_i \mid E_i, P) = \int_0^\infty \mathrm{Po}(o_i \mid \lambda E_i) p(\lambda)\,\mathrm{d}\lambda$$

 

Figure 3. True map for the example of childhood leukemia, new states of Germany, 1980–89: homogenous spatial structure

Now, the mean and the variance of $O_i$ with respect to the marginal distribution is given as $E_i$ for the mean and $\mathrm{Var}(O_i) = \mu E_i + \tau^2 E_i^2$ for the variance, which can be written alternatively as

$$\mathrm{Var}(\mathrm{SMR_i}) = \mu/\mathrm{E_i} + \tau^2 \tag{1}$$

Note that for a homogeneous Poisson case ($\tau^2 = 0$) we yield the conventional Poisson variance $\mathrm{Var}(SMR_i) = \mu/E_i$. Solving Equation (1) for $\tau^2$ and taking averages we achieve a simple estimate for $\tau^2$:

$$\hat{\tau}^2 = \frac{1}{n}\sum_i (SMR_i - \mu)^2 - \frac{1}{n}\mu \sum_i E_i^{-1} \tag{2}$$

which is of the form:

$$\text{estimate of } \tau^2 = (\text{sample variance of } SMR) - (\text{predicted variance under homogeneity})$$

 

Equation (2) estimates a residual variance which is not due to Poisson random variation, but expresses some form of extra-spatial variation in risk. Furthermore, we are able to give some simple proportional measures, such as proportion of spatial heterogeneity (PSH),

$$PSH = \hat{\tau}^2 \Big/ \frac{1}{n}\sum_i (SMR_i - \mu)^2$$

and the proportion of spatial random variation (PSRV),

$$PSRV = 1 - PSH$$

We summarize some properties of the PSH measure.

***Properties***. Let PSH be as defined above: $\text{PSH} = \hat{\tau}^2 / \frac{1}{n}\sum_i (SMR_i - \mu)^2$. Then,

(i) $0 \leq PSH \leq 1$
(ii) E (numerator of PSH) $= \tau^2$
E (denominator of PSH) $= \tau^2 + \frac{1}{n}\mu \sum_i E_i^{-1}$
E (PSH) $\approx \tau^2 / [\tau^2 + \frac{1}{n}\mu \sum_i E_i^{-1}]$

*Examples*:
(a) Childhood leukemia in the new states of Germany, 1980–89. We find that the proportion of spatial heterogeneity is PSH = 0, and consequently all variation seen in Figure 1 is due to Poisson noise, or PSRV = 1.
(b) Lip cancer in males in the new states of Germany, 1980–89. Here we find a considerable proportion of spatial heterogeneity, namely $PSH = 0.6260$, $PSRV = 0.3740$. This means that here, 63% of the total spatial variation is explained by spatial heterogeneity. In this case the true map of spatial heterogeneity is provided in Figure 4, indicating a remarkable north–south gradient in East Germany. (Figure 4 was constructed using the mixture model, Equation (10), which will be introduced in Section 5; the reader should be able possibility to visualize the meaning of the high value of PSH in this application already at this stage.)

## 3. ESTIMATION OF HETEROGENEITY VARIANCE

We are interested in estimating the variance $\tau^2$ of the heterogeneity distribution $P$, the distribution describing the variation in $\lambda$. Sometimes this variance is also called prior variance. We go back to Equation (1), where the marginal variance was provided as

$$\text{Var}(O_i) = \mu E_i + \tau^2 E_i^2$$

Consider the random variables

$$W_i = \frac{\{O_i - \mu E_i\}^2 - \mu E_i}{E_i^2} = (SMR_i - \mu)^2 - \mu / E_i \tag{3}$$

 

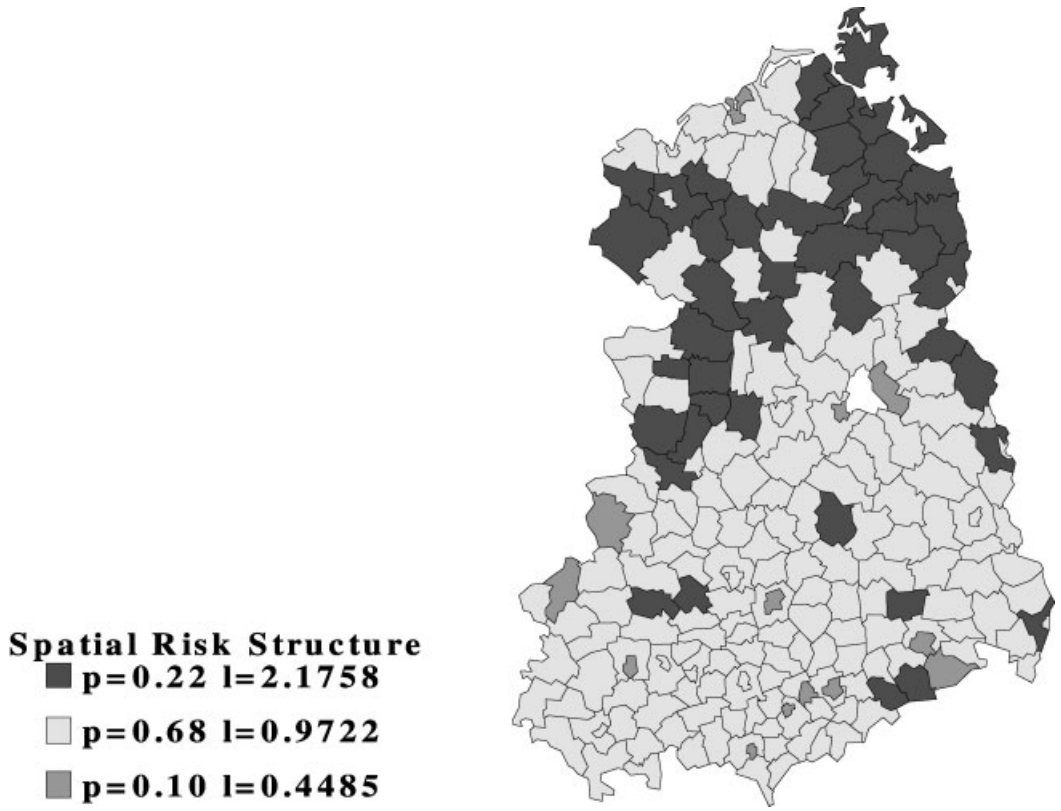


Figure 4. True spatial structure of relative risk for lip cancer in the new states of Germany, 1980–89 (true spatial risk structure was found using the mixture model discussed in Section 5)

We note that, for all $i = 1, \ldots, n$, the $W_i$ are unbiased: $\mathrm{E}(W_i) = \tau^2$. It is clear that some form of pooling is necessary. We consider the linear family

$$T_\alpha(\mathbf{W}) = \sum_i \alpha_i W_i \Big/ \sum_i \alpha_i \tag{4}$$

with positive, but otherwise arbitrary, $\alpha_i$. This is a rather rich family of (quite simple) estimators for $\tau^2$. Specific values of $\alpha_i$ lead to estimators suggested in the literature such as: $\alpha_i = 1/n$ (Böhning and Sarol, 2000), $\alpha_i = E_i$ (Marshall, 1991), and $\alpha_i = E_i^2$ (Bautista, 1997). Clearly, the best choice for $\alpha_i$ is given as the inverse variance of $W_i$ : $1/\mathrm{Var}(W_i)$. However, $\mathrm{Var}(W_i)$ depends on the first four moments and involves also the unknown parameters $\tau^2$ and (potentially) $\mu$. Therefore, the question of which $\alpha_i$ to be used is still open. Some progress has been achieved in Böhning *et al.* (2001, 2002), providing some advantageous evidence for the Marshall estimator; the estimator using equal weights is very appropriate if there is large heterogeneity. To use any estimator based upon Equation (3), knowledge about the mean parameter $\mu$ needs to be incorporated. For example, if internal indirect standardization is used, then the mean $\mu$ is necessarily estimated by $\mu = 1$. Recall that internal indirect standardization characterizes the fact that the same data set is used to provided the $E_i$. To be more specific, let $O_{ij}$ be

 

the observed number of cases and $n_{ij}$ the number at risk in the $i$th area and $j$th age group. Then, age-specific reference rates are constructed as $\lambda_j = \sum_i O_{ij} / \sum_i n_{ij}$ for each age group $j$. Then, define $E_i$ as $\sum_j \lambda_j n_{ij,}$, and it follows that $\sum_i E_i = \sum_i \sum_j [\sum_i O_{ij} / \sum_i n_{ij}] n_{ij} = \sum_i O_i$. However, in full generality it is required that $\mu$ is left unspecified, in which case it needs to be replaced by an estimate. There are two natural estimates of $\mu$: the maximum likelihood estimate $\sum_i O_i / \sum_i E_i$ and the simple mean $1/n \sum_i O_i/E_i$. If the simple mean is used, it is shown in Böhning and Sarol (2000) that $\sum_i W_i/n$ is unbiased for $\tau^2$. In all other cases, Equation (3) will no longer be unbiased. However, it has been shown recently (Böhning *et al.*, 2002) that this bias is contributing rather little to the MSE of estimating $\tau^2$ (when compared to the case when the true value of $\mu$ is used).

## 4. DISEASE MAPPING WITH EMPIRICAL BAYES

Let us consider the heterogeneity distribution $p(\lambda)$ of $\lambda$ as prior and compute the posterior distribution (simply using Bayes theorem),

$$p(\lambda \mid o_i, E_i) = \frac{\mathrm{Po}(o_i \mid \lambda E_i) p(\lambda)}{\mathrm{f}(o_i \mid E_i, P)} = \frac{\mathrm{Po}(o_i \mid \lambda E_i) p(\lambda)}{\int_0^\infty \mathrm{Po}(o_i \mid \lambda E_i) p(\lambda)\, \mathrm{d}\lambda} \tag{5}$$

and the associated posterior expectation,

$$\mathrm{E}(\lambda \mid o_i, E_i) = \int_0^\infty \lambda p(\lambda \mid o_i, E_i)\, \mathrm{d}\lambda \tag{6}$$

Sometimes $\mathrm{E}(\lambda \mid o_i, E_i)$ takes on a particular simple form as in the following example.

*Example*: Let $p(\lambda)$ be the Gamma distribution $\Gamma(\mu, \tau^2)$ with a specific parameterization in which $\mu$ denotes its mean and $\tau^2$ its variance. Then we have that

$$\mathrm{E}(\lambda \mid o_i, E_i) = \frac{o_i + \mu^2/\tau^2}{E_i + \mu/\tau^2} = \frac{\tau^2 o_i + \mu^2}{\tau^2 E_i + \mu}$$

Note that in the case of no heterogeneity $(\tau^2 = 0), \mathrm{E}(\lambda \mid o_i, E_i) = \mu$, and in the case of dominating heterogeneity $(\tau^2 \text{ large}), \mathrm{E}(\lambda \mid o_i, E_i) = o_i/E_i$.

For the posterior expectations to be useful we need to replace the prior (or the parameters therein) by estimates, which leads to empirical Bayes estimates for the SMR. In the example of the $\Gamma$-prior simply replace $\mu$ and $\tau^2$ by one of the estimates of Section 3; $\mu$ can be estimated by the simple average of the $\mathrm{SMR}_i$ or by the pooled mean $\sum_i o_i / \sum_i E_i$,

$$SMR_i^{\mathrm{EB}} = \frac{o_i + \hat{\mu}^2/\hat{\tau}^2}{E_i + \hat{\mu}/\hat{\tau}^2} \tag{7}$$

The empirical Bayes estimators (Equation 7) were applied to the example of leukemia in the New States of Germany, 1980–89. Figure 5 shows the original quartile map based upon the crude SMR-values (top) as well as the one based upon Equation (7) (bottom), and redrawn on the same scale as the quartile map for the crude SMRs (bottom left). The result is rather clear: all spatial heterogeneity has

 

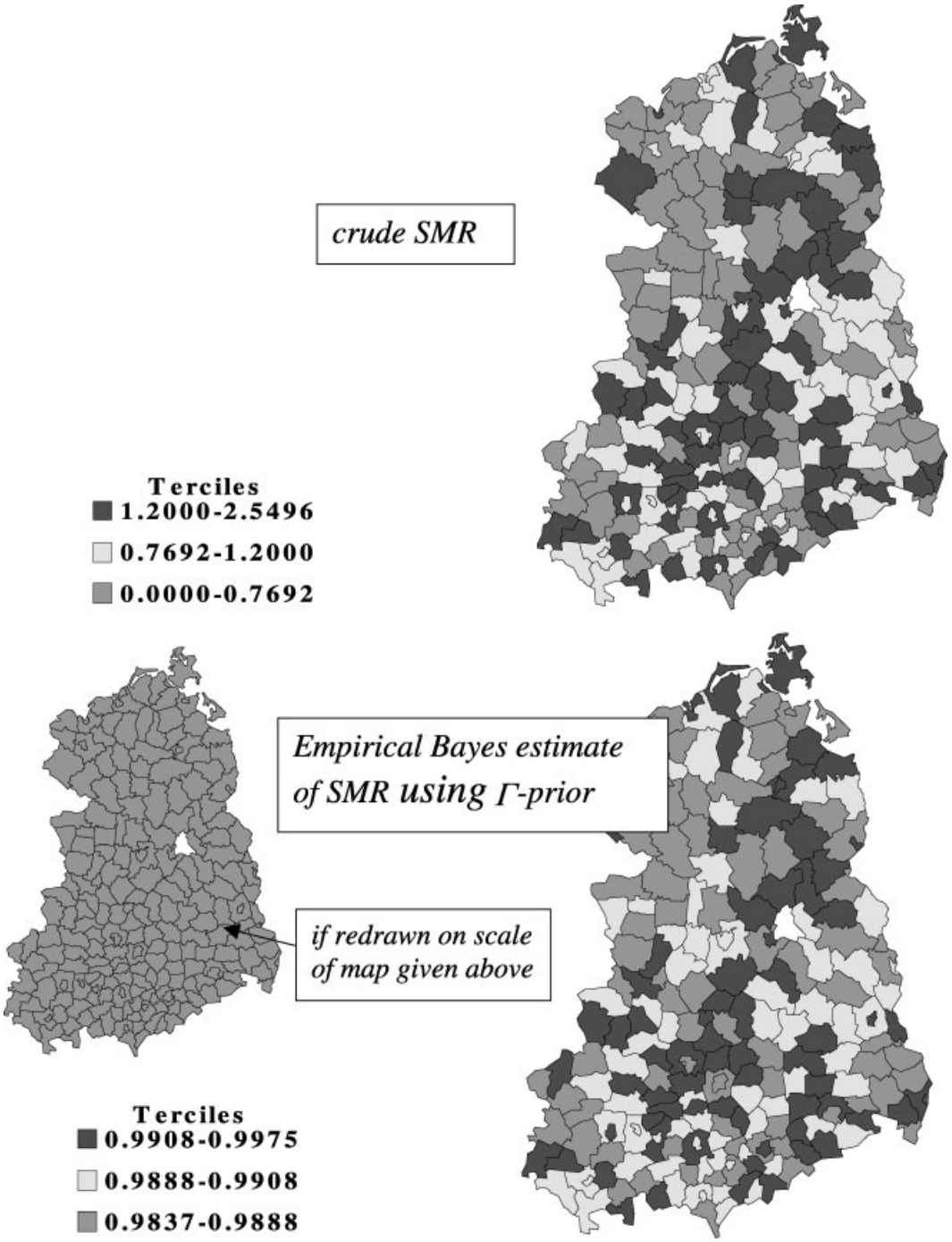


Figure 5. Childhood leukemia in the new states of Germany, 1980–89: based upon crude SMR (top) and $SMR_{EB}$ using a Gamma prior (bottom)

 

disappeared. This is due to the fact that PSH = 0 or, equivalently, $\tau^2$ is estimated to be 0. This means that all variation observed is random variation, and the empirical Bayes estimator (Equation 7) adjusts for this in a quite natural way.

The empirical Bayes estimators (Equation 7) were also applied to the example of lip cancer in the new states of Germany, 1980–89. Figure 6 shows the original quartile map based upon the crude SMR-values (top) as well as the one based upon Equation (7) (bottom). The result is quite clear: spatial heterogeneity is dominating, so there is not much difference in the two maps. This is due to the fact that PSH is more than 50 per cent or, equivalently, $\tau^2$ is estimated to be much larger than 0. This means that most of the variation observed is due to spatial variation, and the empirical Bayes estimator (Equation 7) becomes close to the crude SMR. Theses examples illustrate nicely how the empirical Bayes estimator accounts for the amount of random variation in a given map.

For further discussion on empirical Bayes methods the interested reader may look at Lawson *et al.* (1999), Böhning (2000), or Biggeri *et al.* (1993).

In general, estimates of the posterior distribution can be achieved as follows. We replace $p(\lambda)$ in the posterior distribution,

$$p(\lambda \mid o_i, E_i) = \frac{\mathrm{Po}(o_i \mid \lambda E_i)p(\lambda)}{\int_0^\infty \mathrm{Po}(o_i \mid \lambda E_i)p(\lambda)\,\mathrm{d}\lambda}$$

by an estimate $\hat{p}(\lambda)$, where a general way to determine this is to use the marginal likelihood (mixture likelihood),

$$L(P) = \prod_i \mathrm{f}(o_i \mid E_i, P) = \prod_i \int_0^\infty \mathrm{Po}(o_i \mid \lambda E_i)p(\lambda)\,\mathrm{d}\lambda$$

$P$ can be a parametric distribution with parametric density $p(\lambda)$ like Gamma or Normal, but $P$ can also be left as non-parametric distribution on $\lambda$ with discrete $p(\lambda)$, usually written as

$$P = \begin{pmatrix} \lambda_1 \lambda_2 \cdots \lambda_k \\ p_1 p_2 \cdots p_k \end{pmatrix} \tag{8}$$

Then we use the non-parametric maximum likelihood estimate (NPMLE) of $P$. The NPMLE is the discrete distribution which maximizes the likelihood

$$L(P) = \prod_i \mathrm{f}(o_i \mid E_i, P) = \prod_i \sum_j \mathrm{Po}(o_i \mid \lambda_j E_i)p_j$$

over all discrete probability measures $P$. Then, we are able to compute the non-parametric empirical Bayes estimate of the SMR as

$$SMR_i^{\mathrm{EB-NP}} = \underbrace{\sum_j \hat{\lambda}_j}_{\text{value of } \lambda} \times \underbrace{\mathrm{Po}(o_i \mid \hat{\lambda}_j E_i)\hat{p}_j \Big/ \sum_l \mathrm{Po}(o_i, \mid \hat{\lambda}_l E_i)\hat{p}_l}_{\text{posterior density}} \tag{9}$$

These developments in the context of disease mapping in its general form go back to Clayton and Kaldor (1987).

 

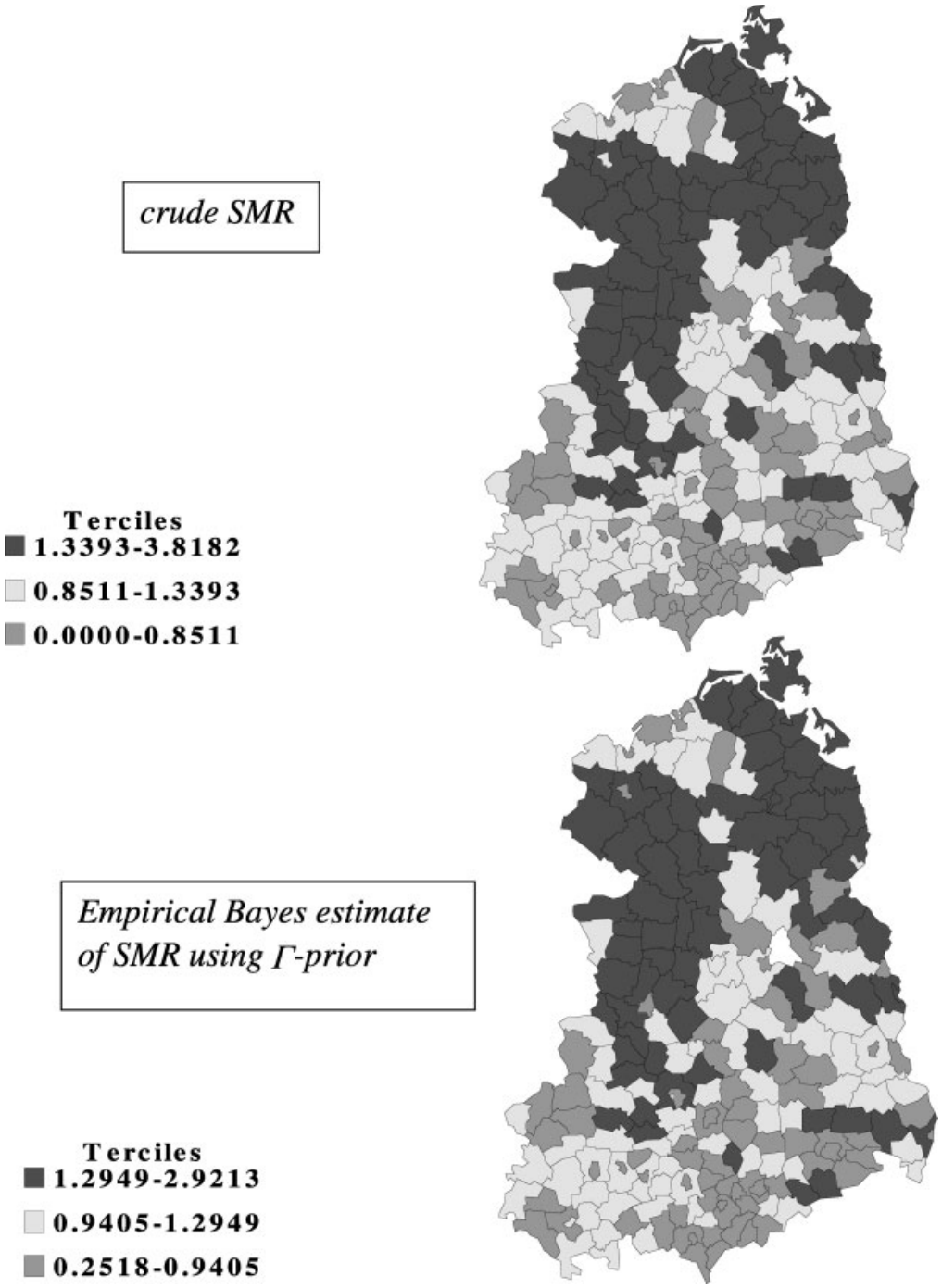


Figure 6. Lip cancer in the new states of Germany, 1980–89: based upon crude SMR (top) and $SMR_{EB}$ using a Gamma prior (bottom)

 

## 5. DISEASE MAPPING IN SPACE USING (NON-PARAMETRIC) MIXTURES

The basic idea is as follows. Suppose that there are $n$ areas given defining the observed risk structure. In addition, there is a second, unobserved map providing the true, but not observable, spatial structure of risk. The situation is illustrated in Figure 7. Furthermore, we have the observed cases $o_1, \ldots, o_n$, the expected cases $E_1, \ldots, E_n$ and non-observed indicators $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, where $Z_{ij} = 1$ means that area $i$ is from component $j$ in the non-observable map with relative risk $\lambda_j$.

This situation is readily modeled as follows. As before, we assume that the $SMR_i = O_i/E_i$ is provided by a Poisson distribution

$$O_i \sim \mathrm{Po}(o_i \mid \lambda_j E_i) \tag{10}$$

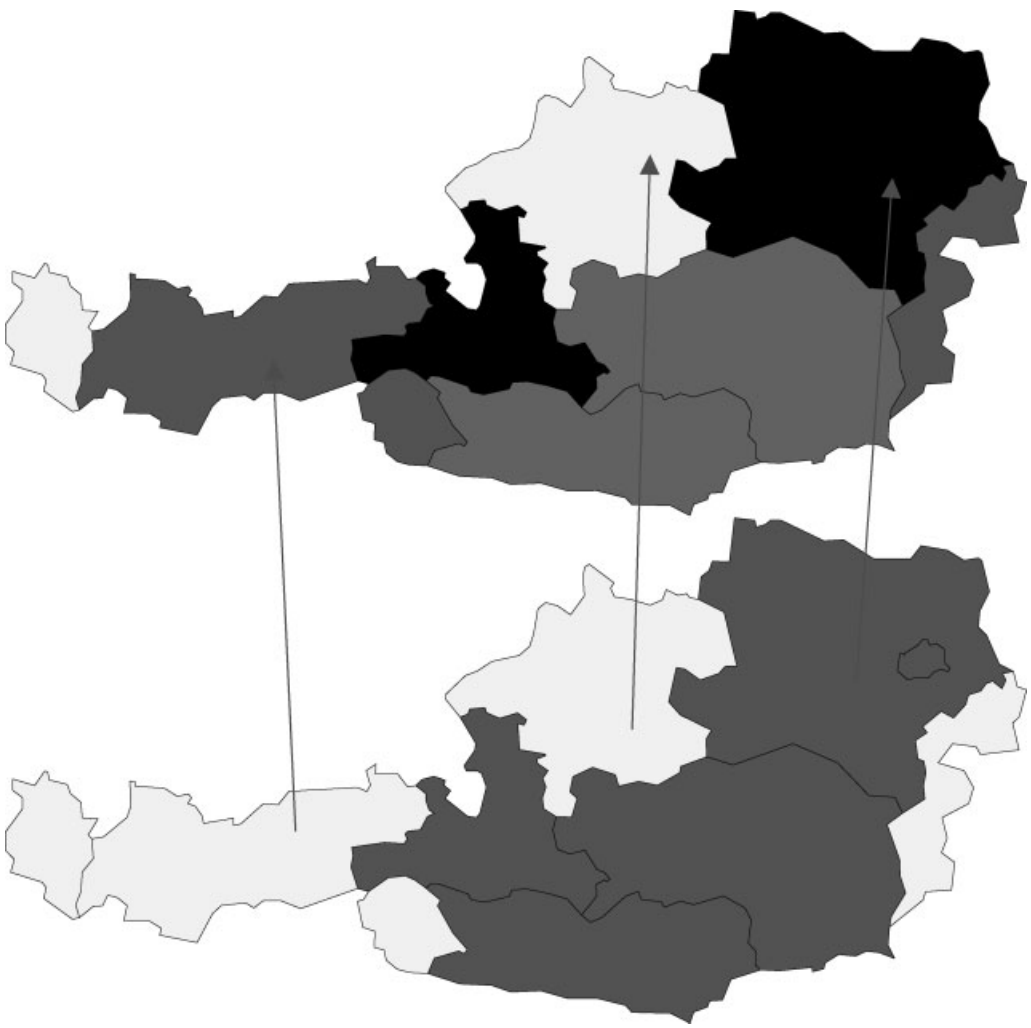

Figure 7. Observable map of relative risk (top) and non-observable map of spatial relative risk structure, here with two components (bottom)

 

conditionally, $O_i$ being from the area with mortality rate $\lambda_j$. Furthermore, there might be $k$ different components, and the proportion of these $k$ components is provided by $p_1, \ldots, p_k$. If we think of an experiment that allocates an area to one of the $k$ components, the associated distribution is provided by the discrete mass distribution, giving mass $p_j$ to component with $\lambda = \lambda_j$, in other words $P = \begin{pmatrix} \lambda_1 \lambda_2 \cdots \lambda_k \\ p_1 p_2 \cdots p_k \end{pmatrix}$. In summary, conditional on component membership a Poisson distribution is assumed, whereas the component distribution is left non-parametric.

***Marginal distribution***. Now, let us consider the unconditional probability $\Pr\{O_i = o_i\}$:

$$\begin{aligned} \Pr\{O_i = o\} &= \sum_j \Pr\{O_i = o, \mathbf{Z}_i = \mathbf{e}_j\} \\ &= \sum_j \Pr\{O_i = o \mid \mathbf{Z}_i = \mathbf{e}_j\} \Pr\{\mathbf{Z}_i = \mathbf{e}_j\} \\ &= \sum_j \mathrm{Po}(o \mid \lambda_j E_i) p_j \end{aligned}$$

where $\mathbf{e}_j$ is the vector of 0s with 1 in the $j$th position. This result can be expressed as

$$O_i \sim p_1 \mathrm{Po}(O_i \mid \lambda_1 E_i) + p_2 \mathrm{Po}(o_i \mid \lambda_2 E_i) + \cdots + p_k \mathrm{Po}(o_i \mid \lambda_k E_i) \tag{11}$$

which is a non-parametric mixture of Poisson distributions. Equation (11) can be nicely interpreted as various forms of spatial heterogeneity. If $k = 1$ there exists a homogeneous risk structure; if $k = 2$ we have a spatial heterogeneity consisting of two different components, etc. Estimation of the parameters $p_1, \lambda_1, \ldots, p_k, \lambda_k$, including the number of components $k$,

$$P = \begin{pmatrix} \lambda_1 \lambda_2 \cdots \lambda_k \\ p_1 p_2 \cdots p_k \end{pmatrix}$$

is usually done by maximum likelihood and leads to the non-parametric maximum likelihood estimate of a mixing distribution (Laird, 1978; Simar, 1976; Lindsay, 1983; Aitkin, 1996, 1999). The NPMLE is well investigated theoretical (Lindsay, 1995) and is algorithmically possible (Böhning, 1995). Frequently, the EM algorithm for mixtures (Dempster *et al.*, 1977) is used.

***Map construction***. The actual disease map construction utilizes the Bayes theorem. If we think of the mixing distribution $P$ as a prior distribution of the Poisson parameter $\lambda$, we are able to achieve the posteriori distribution (actually also the expected values of $z_{ij}$) as

$$\mathrm{f}(\lambda_j \mid o_i, E_i, \hat{P}) = \mathrm{Po}(o_i \mid \hat{\lambda}_j E_i) \hat{p}_j \Big/ \sum_l \mathrm{Po}(o_i \mid \hat{\lambda}_l E_i) \hat{p}_l \tag{12}$$

for each of the $n$ areas. Each area $i$ is classified into component (color) $j$ such that the posterior distribution $\mathrm{f}(\lambda_j \mid o_i, E_i, \hat{P})$ is maximized over all components $1, \ldots, k$.

***Two applications***. As a health region we consider the 219 counties of the former German Democratic Republic (the five new states of Germany). We look at the occurrence distribution of the following two cancer sites:

1. incidence on lung cancer (ICD 162) for women, 1980–89
2. incidence on mamma carcinoma (ICD 174), 1980–89.

 

For lung cancer in females the following estimate of the mixing distribution (NPMLE) was identified:

$$\begin{pmatrix} \lambda_1\, \lambda_2 \cdots \lambda_k \\ p_1\, p_2 \cdots p_k \end{pmatrix} = \begin{pmatrix} 1.33 & 0.98 & 0.78 & 0.56 \\ 0.12 & 0.27 & 0.51 & 0.10 \end{pmatrix}, \qquad \hat{k} = 4$$

The associated posterior classification is provided in Figure 8 (top). About 12 per cent of the region are classified into the elevated relative risk component of 1.33. These are mainly clustered in the State of Brandenburg around Berlin (the white spot on the map).

For the other cancer site (mamma carcinoma), a less pronounced map emerges. The estimate of spatial heterogeneity is again provided via the estimate of the mixing distribution (NPMLE):

$$\begin{pmatrix} \lambda_1\, \lambda_2 \cdots \lambda_k \\ p_1\, p_2 \cdots p_k \end{pmatrix} = \begin{pmatrix} 1.21 & 1.15 & 1.04 & 0.92 & 0.73 \\ 0.02 & 0.06 & 0.20 & 0.68 & 0.04 \end{pmatrix}, \qquad \hat{k} = 5$$

There is a few elevated regions with some clustering of the minorly elevated regions belonging to the component with relative risk 1.15; these are partly concentrated in the south-west of the health region (see Figure 8, bottom).

## 6. DISEASE MAPPING IN SPACE AND TIME

Suppose that again $n$ areas are given. However, in addition we have disease occurrence information in form of SMRs for $T$ time periods:

$$\begin{array}{llll} O_1^{(1)}, O_2^{(1)}, \ldots, O_n^{(1)}; & O_1^{(2)}, O_2^{(2)}, \ldots O_n^{(2)}; & \ldots. & O_1^{(T)}, O_2^{(T)}, \ldots, O_n^{(T)} \\ E_1^{(1)}, E_2^{(1)}, \ldots, E_n^{(1)}; & E_1^{(2)}, E_2^{(2)}, \ldots, E_n^{(2)}; & \ldots. & E_1^{(T)}, E_2^{(2)}, \ldots, E_n^{(T)} \end{array}$$

Typically, data in public health are available for time windows of 5 or 10 years. In the two cancer applications mentioned in the previous section, data for two more time periods ($T = 3$) are available:

1980–89, 1970–79, 1960–69

There are several ways to proceed in this situation.

### 6.1. *Modeling each time period separately*

It is assumed that each period $t$ has its specific latent map; for example, there are $T$ independent mixture models assumed. Area $i$ in period $t$ is from component $j_t$ with risk $\lambda_j^{(t)}, j = 1, \ldots, k$, where the components depend on $t$ and also $k = k_t$ might depend on $t$. The result is not one mixture model but $T$ mixture models—one for each time period:

$$O_i^{(t)} \sim p_1^{(t)} \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_1^{(t)} E_i^{(t)}\right) + p_2^{(t)} \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_2^{(t)} E_i^{(t)}\right) + \cdots + p_k^{(t)} \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_k^{(t)} E_i^{(t)}\right) \qquad (13)$$

 

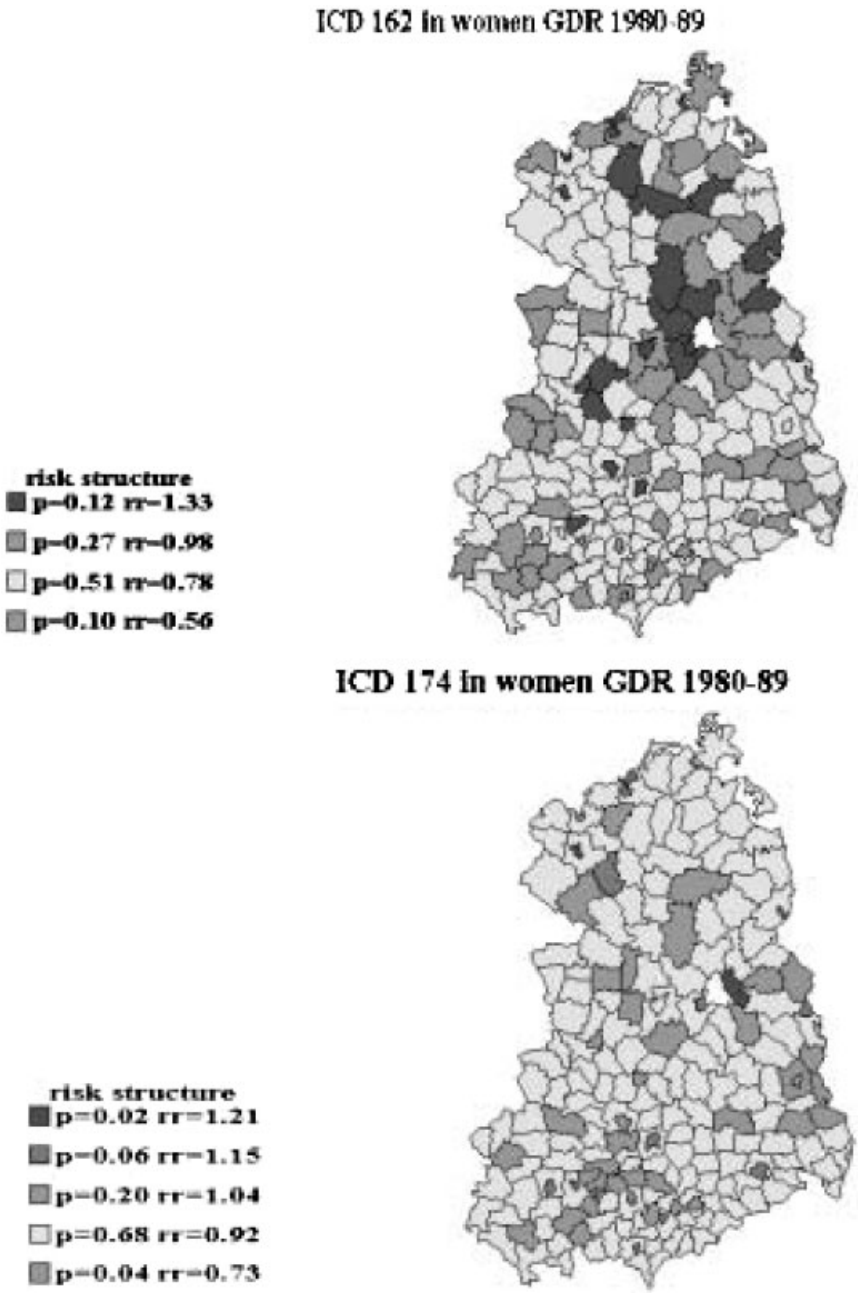


Figure 8. Estimated spatial heterogeneity distribution (mixing distribution) for the new states of Germany, 1980–89, for two cancer sites: lung cancer in females (top) and mamma carcinoma (bottom)

 

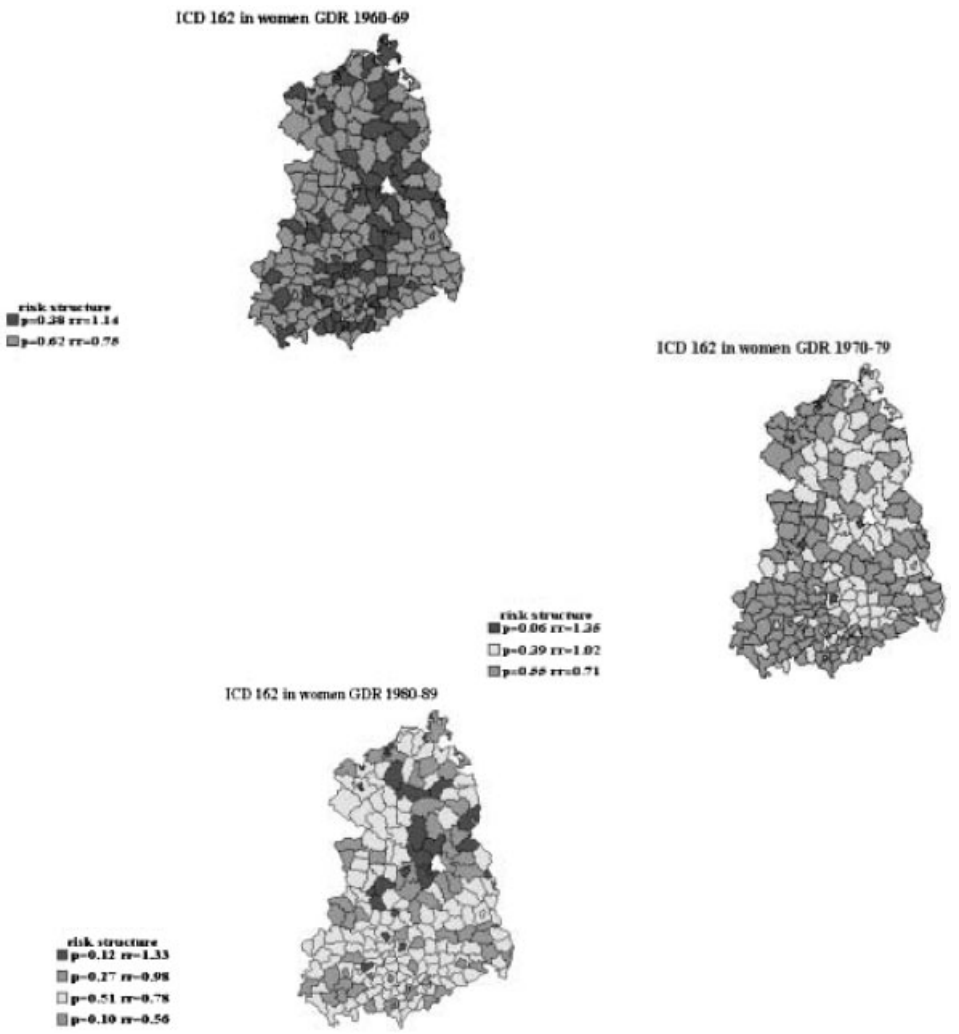


Figure 9. Disease mapping with independent mixture models exemplified for lung cancer in females for the new states of Germany, 1980–89

The likelihood is given as

$$\mathrm{L}(P_1,\ldots,P_T)=\prod_t\prod_i \mathrm{f}\left(o_i^{(t)}\,\middle|\,E_i^{(t)},P_t\right)=\prod_t\prod_i\sum_j \mathrm{Po}\left(o_i^{(t)}\,\middle|\,\lambda_j^{(t)}E_i^{(t)}\right)P_j^{(t)} \tag{14}$$

which can clearly be maximized by separately maximizing $T$ likelihoods $\Pi_i \sum_j \mathrm{Po}(o_i^{(t)} \mid \lambda_j^{(t)} E_i)\, p_j^{(t)}$.

This approach is exemplified for the lung cancer application discussed before. There are three independent mixture models visualized in Figure 9. The three maps are difficult to compare. Not only are the number of components different for each map, but also the mean components, the relative risks, are on a different level.

 

### 6.2. *Modeling in space and time simultaneously*

Instead of developing $T$ models for the $T$ periods independently, one can try to think of one latent, unobservable map of true relative risk which has not only spatial components but also space–time components.

The situation is exemplified in Figure 10, which shows a latent map with two space–time components. Note that here an area can belong to different components at different time periods, as also demonstrated in Figure 10.

It is assumed that an area $i$ in period $t$ is from component $j$ with risk $\lambda_j$, $j = 1, \ldots, k$, where $\lambda_j$ and $k$ does not depend on $t$. Again, the component membership is not observed. Consequently, the marginal model is one mixture model of the form

$$O_i^{(t)} \sim p_1 \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_1 E_i^{(t)}\right) + p_2 \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_2 E_i^{(t)}\right) + \cdots + p_k \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_k E_i^{(t)}\right) \tag{15}$$

leading to the likelihood

$$\mathrm{L}(P) = \prod_t \prod_i \mathrm{f}\left(o_i^{(t)} \,\middle|\, E_i^{(t)}, P\right) = \prod_t \prod_i \sum_j \mathrm{Po}\left(o_i^{(t)} \,\middle|\, \lambda_j E_i^{(t)}\right) p_j \tag{16}$$

Since there are no restrictions on the number of components, Equation (15) again provides a rather flexible tool to capture spatial and temporal heterogeneity, simultaneously. Application to space–time data is straightforward. We consider again the two cancer sites for the new states of Germany, 1980–89. Figure 11 shows the results of the fitted mixture models. In the upper part of Figure 11 we see that a four-component model is found for the cancer site of the lung in females. There is some evidence of space–time clustering: the region of Brandenburg (around the white spot in the map, which is the western part of Berlin) contains most of the elevated components of relative risk 1.31 with increasing tendency towards later time periods. Considering the mamma carcinoma we see a less pronounced picture. A mixture model with three components is found, but the component means are close together. Moreover, there seems to be no consistent pattern if we consider the more elevated group, for example.

### 6.3. *Comparing the two mixture models*

Note that in both cases $T$ maps are drawn; however:

(a) there are $k^{(1)} + \cdots + k^{(T)}$ colors in the first mixture model (Section 6.1)
(b) there are $k$ colors in the second mixture model (Section 6.2)

where $k^{(t)}$ is the number of components in the $t$th time-specific mixture model. The second model appears more attractive since the space–time components are directly comparable for maps different in time periods. Both models are not comparable in a nested sense. An example might illustrate this. Suppose there are three time periods and there is spatial homogeneity in relative risk for each time period, but the mean level of relative risk is different for each time period. The first approach would fit three independent mixture models and would find homogeneity, each period having a different mean. The second approach would find a mixture model giving equal weights to three different component mean levels. Each component would then represent exactly one period map, and the space–time components would decrease to represent only variation in time. However, there are also cases for which both approaches might lead to the same mixture model. To illustrate this, suppose that there is

 

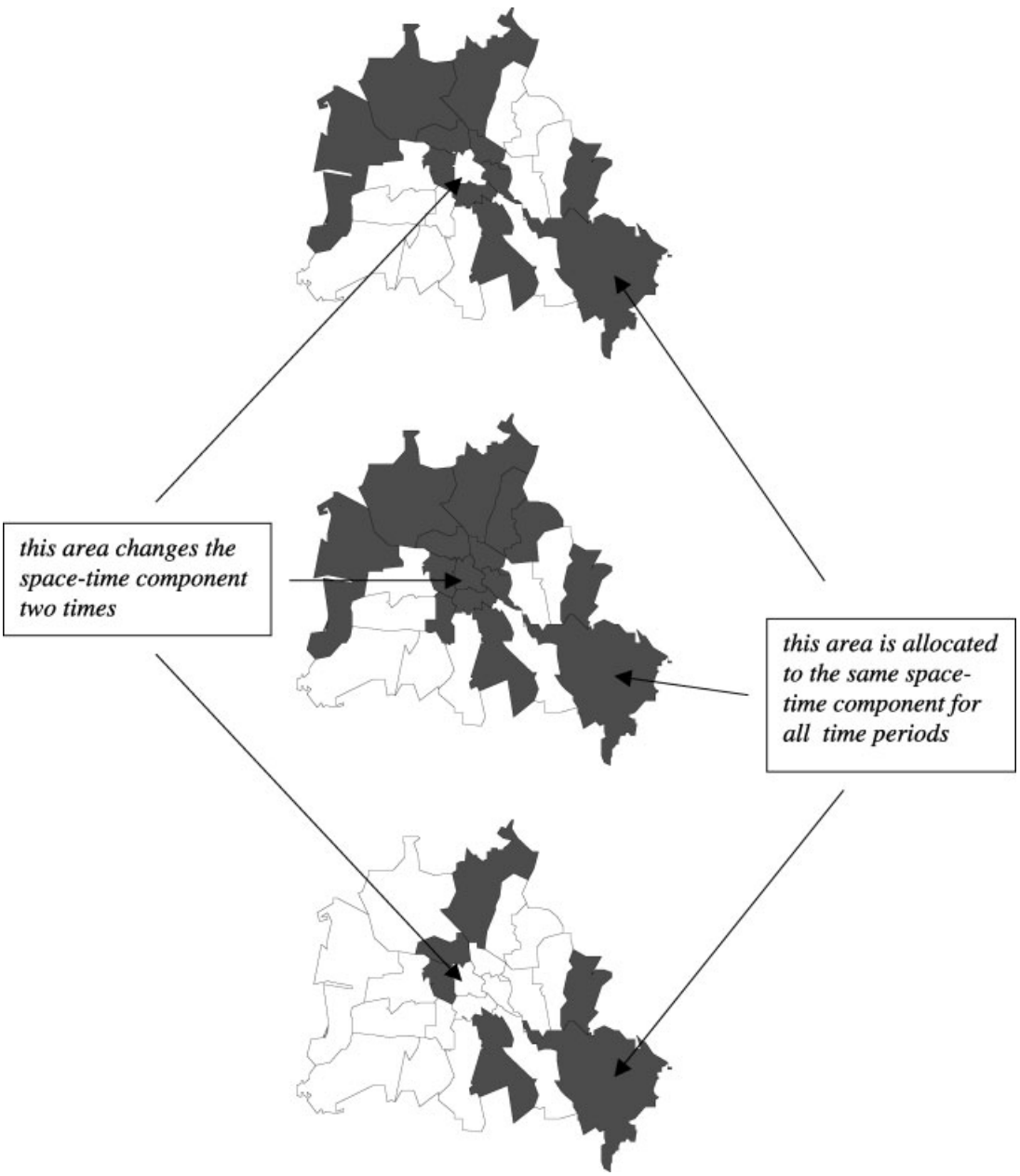


Figure 10. Principal layout of latent space–time map for three time periods and 23 areas and a two-component mixture model

 

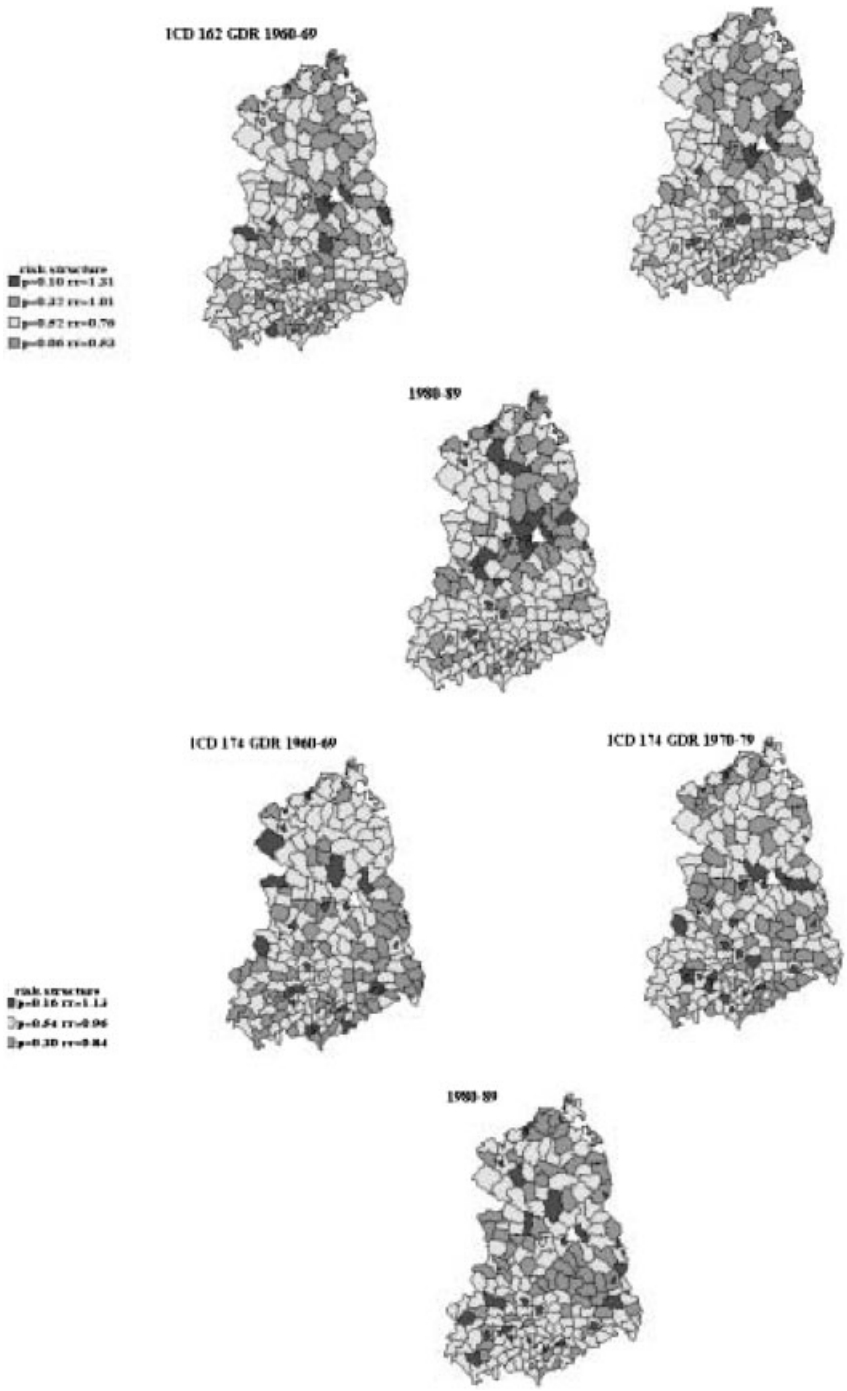


Figure 11. Space–time disease mapping via mixture models with space–time components for the new states of Germany: lung cancer in females (top); mamma carcinoma (bottom)

 

spatial heterogeneity but no temporal change. Then, the first approach would find the same mixture model $T$ times, whereas the second approach would find this mixture model directly. Both approaches have considerable flexibility, though the second offers ease of interpretation and presentation.

As a guideline, it appears that the second approach, the mixture model having space–time components, provides more advantages, in terms of modeling and in terms of interpretation and presentation. However, one should keep in mind that we have with these models public health data in mind where the time windows are rather long—typically 5 or 10 years. Over these periods many things can change, like the boundaries of administrative regions (new ordering of districts) or disease classification systems, where either comparisons in time are no longer possible or make no more sense. Then the first approach is the only of the two which can be used validly.

## 7. DISCUSSION

The article has investigated possibilities of consistently estimating spatial and spatial–temporal heterogeneity via non-parametric mixture models. Of course, these procedures represent only a small part of a whole range of procedures available for disease mapping. For an overview on this issue the interested reader is referred to Lawson *et al.* (1999), specifically part I therein. The suggested models have a close relationship to the class of Bayesian models developed and discussed more recently, specifically to the area of empirical Bayesian models as discussed in Section 4 of this article.

Coming more specifically to mixture models, frequently the question of the number of components is raised. In this approach we have used the non-parametric maximum likelihood proposal. However, critical voices claim that the NPMLE of the number of components might overestimate the true number considerably. For a discussion on this issue see Schlattmann (2002). Alternative proposals favor criteria based on the likelihood but with a penalty for the number of parameters involved, like the Akaike or the Bayesian information criterion (see Celeux, 2001, for details). In our experience, to protect for oversmoothing, the NPMLE should be compared with mixture models having fewer components. This raises the question of valid inference, since the likelihood ratio is no longer following a chi-square distribution (Böhning *et al.*, 1994; Böhning, 2000). Here, the Bootstrap approach might provide reasonable approximations of the true null-distribution. As a by-product of these resample procedures, standard errors of the estimates of the mixture model parameters can be readily obtained.

For the practitioner. the availability of these procedures in easy-to-use software like C.A.MAN (Böhning *et al.*, 1992a; Böhning *et al.*, 1998) or DISMAP (Schlattmann and Böhning, 1993; Schlattmann, 1996) is of great help and facilitates modeling to a large extent. These tools can be obtained for no cost from the website: http://www.medizin.fu-berlin.de/sozmed/bo1.html.

ACKNOWLEDGEMENTS

The work of the author is supported by the German Research Foundation. The author is grateful to two unknown referees for constructive comments. In particular, the author would like to thank Prof. Dr. Annibale Biggeri for the invitation to submit this article to the Special Issue of *Environmetrics*.

REFERENCES

Aitkin MA. 1996. A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* **6**: 251–262.

Aitkin MA. 1999. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**: 117–128.

 

Bautista OM. 1997. *Analysis of Overdispersed Poisson Count.* Doctoral Dissertation, UMI: Ann Arbor.
Becker N, Wahrendorf J. 1997. *Atlas of Cancer Mortality in the Federal Republic of Germany 1981–1990.* Springer-Verlag: Berlin.
Becker N, Frentzel-Beyme R, Wagner G. 1984. *Atlas of Cancer Mortality in the Federal Republic of Germany, Second Completely Revised Edition.* Springer-Verlag: Berlin.
Biggeri A, Braga M, Marchi M. 1993. Empirical Bayes *interval* estimates: an application to geographical epidemiology. *Journal of the Italian Statistical Society* **3**: 251–267.
Böhning D. 1995. A review of reliable maximum likelihood algorithms for the semi-parametric mixture maximum likelihood estimator. *Journal of Statistical Planning and Inference* **47**: 5–28.
Böhning D. 2000. *Computer Assisted Analysis of Mixtures and Applications: Disease Mapping, Meta-Analysis and Others.* Chapman & Hall/CRC: Boca Raton.
Böhning D, Sarol J. 2000. A nonparametric estimator of heterogeneity variance with applications to SMR- and rate data. *Biometrical Journal* **42**: 321–334.
Böhning D, Schlattmann P, Lindsay BG. 1992a. Computer assisted analysis of mixtures (C.A.MAN): Statistical algorithms. *Biometrics* **48**: 283–303.
Böhning D, Dietz E, Schaub R, Schlattmann P, Lindsay BG. 1994. The distribution of the likelihood ratio for mixtures of densities from the one-parametric exponential family. *Annals of the Institute of Statistical Mathematics* **46**: 373–388.
Böhning D, Dietz E, Schlattmann P. 1998. Recent developments in computer assisted analysis of mixtures (C.A.MAN). *Biometrics* **54**: 367–377.
Böhning D, Dietz E, Schlattmann P, Viwatwonkasem C, Biggeri A. 2001. Some general points in estimating heterogeneity variance with the DerSimonian-Laird-Estimator. *Biostatistics* **3**: 1–13.
Böhning D, Malzahn U, Sarol J, Rattanasiri S, Biggeri A. 2002. Efficient non-iterative and nonparametric estimation of heterogeneity variance for the standardized mortality ratio. *Annals of the Institute of Statistical Mathematics* (to appear).
Cartwright RA, Alexander FE, McKinney PA, Ricketts TJ. 1990. *Leukemia and Lymphoma. An Atlas of Distribution within Areas of England and Wales 1984–1988.* Leukemia Research Fund: London.
Celeux G. 2001. Different points of view for choosing the number of components in a mixture model. *Proceedings of the 10th International Symposium on Applied Stochastic Models and Data Analysis*, Vol. 1, Govaert G, Janssen J, Limnios N (eds). 21–28.
Clayton D, Kaldor J. 1987. Empirical Bayes estimates for age-standardized relative risks. *Biometrics* **43**: 671–681.
Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**: 1–38.
Holland WW. 1991. *European Community Atlas of 'Avoidable Death'*, Vol. 1. Oxford University Press: Oxford.
Holland WW. 1992. *European Community Atlas of 'Avoidable Death'*, Vol. 2. Oxford University Press: Oxford.
Laird NM. 1978. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**: 805–811.
Lawson A, Böhning D, Biggeri A, Lesaffre E, Viel J-F, Bertollini R (eds). 1999. *Disease Mapping and Risk Assessment for Public Health.* Wiley: New York.
Lindsay BG. 1983. The geometry of mixture likelihoods, Part I: a general theory. *Annals of Statistics* **11**: 783–792.
Lindsay BG. 1995. *Mixture Models: Theory, Geometry, and Applications*. NSF-CBMS regional conference series in probability and statistics, Vol. 5. Institute of Statistical Mathematics: Hayward.
Marshall RJ. 1991. Mapping Disease and Mortality Rates Using Empirical Bayes Estimators. *Applied Statistics* **40**: 283–294.
Schlattmann P. 1996. The computer package DismapWin. *Statistics in Medicine* **15**: 931.
Schlattmann P. 2002. Estimating the number of components in a finite mixture model: the special case of homogeneity. *Computational Statistics and Data Analysis* (to appear).
Schlattmann P, Böhning D. 1993. Mixture models and disease mapping. *Statistics in Medicine* **12**: 943–50.
Simar L. 1976. Maximum likelihood estimation of a compound Poisson process. *Annals of Statistics* **4**: 1200–1209.

 