Disease Mapping of Stage-Specific Cancer Incidence Data

Leonhard Knorr-Held,[1,*] Günter Raßer,[2,**] and Nikolaus Becker[3,***]

[1]Medical Statistics Unit, Department of Mathematics and Statistics,
Lancaster University, Lancaster LA1 4YF, U.K.
[2]Department of Statistics, Ludwig-Maximilians-University Munich,
Ludwigstrasse 33, 80539 Munich, Germany
[3]German Cancer Research Center, Department of Biostatistics,
Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
*email: l.knorr-held@lancaster.ac.uk
**email: rasser@stat.uni-muenchen.de
***email: n.becker@dkfz.de

SUMMARY. We propose two approaches for the spatial analysis of cancer incidence data with additional information on the stage of the disease at time of diagnosis. The two formulations are extensions of commonly used models for multicategorical response data on an ordinal scale. We include spatial and age-group effects in both formulations, which we estimate in a nonparametric smooth way. More specifically, we adopt a fully Bayesian approach based on Gaussian pairwise difference priors where additional smoothing parameters are treated as unknown as well. We argue that the methods are useful in monitoring the effectiveness of mass cancer screening and illustrate this through an application to data on cervical cancer in the former German Democratic Republic. The results suggest that there are large spatial differences in the stage proportions, which indicate spatial variability with respect to the introduction and effectiveness of Pap smear screening programs.

KEY WORDS: Cancer screening; Cervical cancer; Cumulative model; Disease mapping; Ordered categorical response; Pairwise difference prior; Sequential model; Stage-specific cancer incidence data.

1. Introduction

There has been much development of the spatial analysis of observational disease data within the last 10 years. The work can be categorized into two groups, methodology for data where the exact location of each case is known and methodology for aggregated data, where the total number of cases is given in predefined administrative areas (for a review, see Diggle (1996)). Bayesian approaches for the second type of data include the seminal work by Besag, York, and Mollié (1991), who propose a Markov random field model for the spatial smoothing of disease rates. This model is nowadays widely used for disease mapping, the study of spatial variation in disease risk (for reviews, cf., Clayton and Bernardinelli (1992), Knorr-Held and Becker (2000), or Wakefield, Best, and Waller (2000)).

Probably the most prominent application is the statistical analysis of (age-standardized) cancer mortality rates, as such data are routinely collected throughout the world. A spatial analysis may help identify a spatial signal, which is particularly important for rare diseases, where the raw rates exhibit too much variation and are not particularly helpful in order to judge the variation of the underlying disease risk. The estimated spatial pattern may give hints about relevant unobserved risk factors, although some general problems of interpretation can remain due to the observational type of the data.

In this article, we extend the methodology to the analysis of cancer incidence data with additional knowledge on the stage of disease at time of diagnosis. Our aims are (a) to adjust the crude observed data for effects that can be attributed to age and (b) to assess whether there is any spatial variation left in the (adjusted) stage proportions. This is of clear public health importance for diseases for which screening programs have been implemented and spatial variation in stage proportions might indicate heterogeneity in the effectiveness of cancer screening.

We propose two formulations based on regression models for categorical data on an ordered scale (for a recent review, see Fahrmeir and Tutz (2001, Chapter 3)). In the first approach, we model cumulative probabilities of disease risk, whereas in the second, we model conditional probabilities. More specifically, in the latter approach, we consider the probability that a person is diagnosed with the disease in a specific stage given that she is diagnosed in this or in a higher stage. In each formulation, the log odds of these (cumulative or conditional) probabilities are decomposed additively into age-group and spatial effects.

We work directly on data stratified by age, which is in con-

trast with ordinary disease-mapping methods (without stage stratification), where the data are typically standardized by age in advance. Such a two-stage estimation procedure allows one to calculate the expected number of cases, which is subsequently used as an offset in a Poisson regression approach. However, a simultaneous estimation of age and spatial effects should, in general, be preferred because the uncertainty in the age estimates is then automatically incorporated. Furthermore, it is not obvious how to calculate expected cases in our multicategorical setting.

In Section 2, we outline the two different formulations for ordinal disease risk data, and Section 3 illustrates the two approaches in an application to incidence data on cervical cancer in the former German Democratic Republic (GDR) in 1975. We compare our estimates with those obtained from a corresponding maximum likelihood approach with unrestricted age-group and spatial effects. This corresponds to the common comparison of standardized mortality or morbidity ratios with Bayesian relative risk estimates. The results suggest that there are large spatial differences in the (age-adjusted) stage proportions, which indicates spatial variability in the time of introduction and effectiveness of prevention programs. We close with some comments and possible extensions in Section 4.

2. Model

Let n_{ij} denote the number of person-years (or simply people) at risk in district $i = 1, \ldots, I$ and age group $j = 1, \ldots, J$. For each cell (i, j), let y_{ijs} denote the number of diagnosed cases of disease in stage $s = 1, \ldots, S$. We assume that the stages are ordered by severity of the disease, with stage S being the most severe. Finally, let $y_{ij0} = n_{ij} - \Sigma_{s=1}^{S} y_{ijs}$ be the number of all person-years at risk that have not being diagnosed with the disease (stage 0). We now assume that $y_{ij} = (y_{ij0}, y_{ij1}, \ldots, y_{ijS})'$ follows a multinomial distribution with parameters n_{ij} and probability vector $\pi_{ij} = (\pi_{ij0}, \pi_{ij1}, \ldots, \pi_{ijS})'$, where $\Sigma_{s=0}^{S} \pi_{ijs} = 1$.

2.1 The Cumulative Model

In the cumulative model (McCullagh, 1980), we factorize the log odds of the cumulative probabilities $p_{ijs} = \pi_{ij0} + \cdots + \pi_{ijs}$ into an intercept term μ_s, a spatial effect θ_{si}, and an age-group effect φ_{sj}, i.e.,

$$\text{logit}(p_{ijs}) = \log \left(\sum_{t=0}^{s} \pi_{ijt} \Big/ \sum_{t=s+1}^{S} \pi_{ijt} \right)$$
$$= \mu_s + \theta_{si} + \varphi_{sj} \qquad (s = 0, \ldots, S-1). \quad (1)$$

Equivalently, this model can be formulated in terms of descending cumulative probabilities $1 - p_{ijs}$; the corresponding log odds are simply $-(\mu_s + \theta_{si} + \varphi_{sj})$. Hence, the estimates from model (1) can easily be transformed to those corresponding to an analysis of the data with the category order reversed.

The probabilities π_{ijs} entering the multinomial likelihood can be derived from (1) as

$$\pi_{ijs} = \begin{cases} \text{logit}^{-1}(\mu_0 + \theta_{0i} + \varphi_{0j}) & (s = 0) \\ \text{logit}^{-1}(\mu_s + \theta_{si} + \varphi_{sj}) & \\ \quad -\text{logit}^{-1}(\mu_{s-1} + \theta_{s-1,i} + \varphi_{s-1,j}) & \\ & (s = 1, \ldots, S-1) \\ 1 - \text{logit}^{-1}(\mu_{S-1} + \theta_{S-1,i} + \varphi_{S-1,j}) & (s = S), \end{cases}$$
$$(2)$$

where $\text{logit}^{-1}(x) = 1/(1 + \exp(-x))$. To ensure that all these probabilities are positive, the unknown parameters μ_s, θ_{si}, and φ_{sj} have to fulfill the constraints

$$\mu_{s-1} + \theta_{s-1,i} + \varphi_{s-1,j} < \mu_s + \theta_{si} + \varphi_{sj} \quad (3)$$

for all $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $s = 1, \ldots, S-1$.

2.2 The Sequential Model

The rationale for the sequential model is that a categorical response variable can take a specific value s only after the levels $0, 1, \ldots, s-1$ have been reached. This is the case in our application, where cancer diagnosis at a specific stage implies that the cancer has passed undetected through all stages below. A version of the sequential model where covariate effects do not depend on the response value is also known as the continuation ratio model (Agresti, 1984).

The sequential approach to ordinal data thus models the conditional probability that an individual in cell (i, j) gets diagnosed with the disease in stage s, assuming that she gets diagnosed of the disease in stage s or higher, i.e. $q_{ijs} = \pi_{ijs}/(\pi_{ijs} + \cdots + \pi_{ijS})$. Now we decompose the log odds of these conditional probabilities into an intercept term ν_s, a spatial effect η_{si}, and an age-group effect ψ_{sj} as follows:

$$\text{logit}(q_{ijs}) = \log \left(\pi_{ijs} \Big/ \sum_{t=s+1}^{S} \pi_{ijt} \right)$$
$$= \nu_s + \eta_{si} + \psi_{sj} \qquad (s = 0, \ldots, S-1). \quad (4)$$

Note that, formally, the only difference from the cumulative model (1) is that π_{ijs} replaces the cumulative probability $\pi_{ij0} + \cdots + \pi_{ijs}$ in the numerator of the ratio within the logarithm. For $s = 0$, the cumulative and the sequential models are apparently identical (assuming compatible priors). We will comment on this further in Section 2.5.

The probabilities π_{ijs} can now be derived as

$$\pi_{ijs} = \begin{cases} \text{logit}^{-1}(\nu_0 + \eta_{0i} + \psi_{0j}) & (s = 0) \\ \text{logit}^{-1}(\nu_s + \eta_{si} + \psi_{sj}) & \\ \quad \times \prod_{t=0}^{s-1} \left\{ 1 - \text{logit}^{-1}(\nu_t + \eta_{ti} + \psi_{tj}) \right\} & \\ & (s = 1, \ldots, S-1) \\ \prod_{t=0}^{S-1} \left\{ 1 - \text{logit}^{-1}(\nu_t + \eta_{ti} + \psi_{tj}) \right\} & (s = S), \end{cases}$$
$$(5)$$

(see Fahrmeir and Tutz, 2001, p. 94). Note that here the π_{ijs} are defined through products of probabilities, not through differences of probabilities as in the cumulative model. Therefore, no further constraints have to be imposed on the parameters ν_s, η_{si}, and ψ_{sj}. A further difference from the cumulative model is that a sequential model applied to the data but with the category order reversed is not equivalent to model (4) except for the noninteresting binomial case $S = 1$. This is a consequence of the rationale underlying the sequential model, where categories can be reached successively but only in one specific direction.

2.3 *Prior Assumptions*

The two alternative models proposed above are now completed by assigning prior distributions to all unknown parameters. For both the spatial and the age-group parameters, we will use Gaussian pairwise difference priors (Besag et al., 1995), which favor a nearly constant pattern, implied by a high prior mass on very small values of the corresponding variance parameter. However, the priors we use for these variance parameters are highly dispersed; hence, the formulation will be flexible enough to capture spatial or temporal gradients or trends if there is evidence in the data for it. For the spatial effects, this corresponds to the common choice of Markov random field models, while for the age-group parameters, this class reduces to the so-called random walk priors. Exactly the same priors have been used for disease mapping (Besag et al., 1991; Best et al., 1999), for space–time modeling of disease risk (Knorr-Held and Besag, 1998), and in many other areas of application (e.g., Fahrmeir and Lang, 2001). These models neither impose stationarity nor assume a specific parametric form; in fact, they are closely related to non- and semiparametric smoothing methods (see Fahrmeir and Knorr-Held, 2000; Hastie and Tibshirani, 2000).

In the cumulative model, we separate the spatial parameters into independent sets $\theta_0, \ldots, \theta_{S-1}$, $s = 0, \ldots, S-1$, where $\theta_s = (\theta_{s1}, \ldots, \theta_{sI})$, and assume that, for each category s, θ_s follows a Gaussian Markov random field (MRF) model (Besag et al., 1991),

$$p(\theta_s \mid \lambda_{\theta_s}) \propto \lambda_{\theta_s}^{(I-1)/2} \exp\left\{ -\frac{\lambda_{\theta_s}}{2} \sum_{i_1 \sim i_2} (\theta_{s,i_1} - \theta_{s,i_2})^2 \right\},$$
$$(6)$$

where the sum in the exponent goes over all pairs of adjacent areas i_1 and i_2. For some motivation for $I - 1$ instead of I degrees of freedom for the precision (the inverse variance) λ_{θ_s} in (6), see Knorr-Held (2002).

For each unknown precision parameter λ_{θ_s}, $s = 0, \ldots, S-1$, we adopt a gamma prior,

$$p(\lambda_{\theta_s}) \propto \lambda_{\theta_s}^{a-1} \exp(-b\lambda_{\theta_s}),$$

with suitably chosen constants a and b. The S sets of Markov random fields $\theta_0, \ldots, \theta_{S-1}$ are assumed to be independent. Alternatively, one could specify a multivariate MRF model,

$$p(\theta \mid \Lambda_\theta) \propto |\Lambda_\theta|^{(I-1)/2}$$
$$\times \exp\left\{ -\frac{1}{2} \sum_{i_1 \sim i_2} (\theta_{i_1} - \theta_{i_2})' \Lambda_\theta (\theta_{i_1} - \theta_{i_2}) \right\}, \quad (7)$$

where $\theta_i = (\theta_{0i}, \ldots, \theta_{Si})'$. A Wishart prior would be the common choice for the precision matrix Λ_θ, i.e.,

$$p(\Lambda_\theta) \propto |\Lambda|^{a-(S+1)/2} \exp\left\{-\mathrm{tr}(B\Lambda)\right\},$$

again with suitably chosen constants a and B, where a is a scalar and B is an $S \times S$ matrix. Such a multivariate MRF model might be appropriate if the MRFs $\theta_0, \ldots, \theta_{S-1}$ are expected to be correlated. However, note that *a priori* independent fields $\theta_0, \ldots, \theta_{S-1}$ can still be dependent in the posterior if they are dependent in the likelihood. So, without strong prior knowledge about the correlation, we do not expect any major differences between the two formulations

and stick to the simpler form with *a priori* independent MRFs.

The formulation proposed in Besag et al. (1991) is more elaborate, with additional parameters for unstructured spatial heterogeneity. It is computationally convenient to employ a reparameterized version (e.g., Carlin and Louis, 1996, p. 308), where θ_{si} is independent Gaussian with mean $\tilde{\theta}_{si}$ and precision τ_s, say, and a GMRF prior is now placed on the latent vectors $\tilde{\theta}_s$, just like in (6) for θ_s. In our application, we have tested both models with and without the additional unstructured parameters.

For the age–group-specific parameters, we assume along similar lines that, for each category s, the parameters $\varphi_s = (\varphi_{s1}, \ldots, \varphi_{sJ})$ follow a simple Gaussian random walk in time with variance $\lambda_{\varphi_s}^{-1}$, with a flat prior for the initial value φ_{s1}. Such a formulation is the exact temporal analogue of model (6) because the prior can be written again in the pairwise difference form,

$$p(\varphi_s \mid \lambda_{\varphi_s}) \propto \lambda_{\varphi_s}^{(J-1)/2} \exp\left\{ -\frac{\lambda_{\varphi_s}}{2} \sum_{j=2}^{J} (\varphi_{s,j} - \varphi_{s,j-1})^2 \right\}.$$
$$(8)$$

We assume prior independence for the sets of parameters $\varphi_0, \ldots, \varphi_{S-1}$, which again can easily be relaxed by adopting a multivariate Gaussian random walk model. Also, we use again gamma hyperpriors for the precision parameter λ_{φ_s}, $s = 0, \ldots, S - 1$. Finally, for each intercept parameter $\mu_0, \ldots, \mu_{S-1}$, we adopt a flat, locally uniform prior.

Similarly, for the sequential model, we use MRF priors for $\eta_s = (\eta_{s1}, \ldots, \eta_{sI})$, random walk priors for $\psi_s = (\psi_{s1}, \ldots, \psi_{sJ})$, and a flat prior for ν_s, $s = 0, \ldots, S - 1$. The exact forms of the prior densities can easily be obtained by replacing μ_s by ν_s, θ_s by η_s, and φ_s by ψ_s in the above description of the priors in the cumulative model.

2.4 *Model Choice and Parameter Interpretation*

At this point, it might be worth noting that the posterior distribution of the conditional probabilities q_{ijs} can, of course, easily be derived from the cumulative model as well because they are just simple functions of the posterior distribution of the π_{ijs}'s. Similarly, the posterior distribution of the cumulative probabilities p_{ijs} could be calculated from the sequential model. Indeed, both formulations allow the exploration of every functional of the posterior distribution of the π_{ijs}'s. The difference between the two formulations is the different parameterization of the π_{ijs}'s, with different quantities being the focus for smoothing, either the cumulative or the sequential conditional probabilities. Preferences for one or the other model can either be based on interpretation issues or on more formal model-choice criteria.

Regarding parameter interpretation, we are particularly interested in spatial disease risk estimates adjusted for age. The (age-adjusted) overall relative risk (regardless of the stage of the disease) in district i can be obtained from the quantities $\exp(-\theta_{0i})$ and $\exp(-\eta_{0i})$. Similarly, in the sequential model, we can interpret $\exp(-\theta_{si})$, $s = 1, \ldots, S-1$, as the cumulative adjusted relative risk in district i. In the sequential model, $\exp(-\eta_{si})$, $s = 1, \ldots, S - 1$, can be interpreted as the (age-adjusted) odds ratio for the conditional probability of being diagnosed in stage $s + 1$ or higher, given diagnosis in stage s or higher. For the age-group effects, we also prefer to display

$-\varphi_{sj}$ and $-\psi_{sj}$ (rather than φ_{sj} and ψ_{sj}), the age-group effects on the cumulative probabilities $1 - p_{ijs}$ and on the conditional probabilities $1 - q_{ijs}$, respectively. This has the advantage that higher values in the figures displaying age effects and darker colors in the spatial maps can be associated with a higher (cumulative or conditional) risk of a more severe stage of the disease at diagnosis.

For assessment of the model fit, we routinely monitor the posterior distribution of the saturated deviance (Spiegelhalter et al., 2002),

$$D = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij}^2, \tag{9}$$

with the multinomial squared deviance residual

$$d_{ij}^2 = 2 \sum_{s=0}^{S} y_{ijs} \log \left(\frac{y_{ijs}}{n_{ijs}\pi_{ijs}} \right)$$

(using the convention that $0 \log 0 = 0$). Each deviance residual $d_{ij} = (d_{ij}^2)^{1/2}$ can be seen as a (standardized) measure of fit, comparing the observed number of cases y_{ijs} with the fitted number of cases $n_{ijs}\pi_{ijs}$ for all stages $s = 0, \ldots, S$. Note that this is well defined in both models, as only the multinomial cell probabilities enter. For a well-fitting model, D should be asymptotically (with increasing data in each cell (i,j)) around $I \cdot J \cdot S$ (the factor S appears here due to the multinomial response with S "free" categories). The mean posterior deviance $\bar{D}$ can be used as an overall measure of model fit and can be combined with a term p_D called the effective number of parameters to give a deviance information criterion (DIC) for model choice (see Spiegelhalter et al. (2002) for further details)).

2.5 A Comparison of the Two Models

As an illustration, we now consider a simple example with $S = 2$ categories and no further stratification with respect to age or space (i.e., $I = J = 1$).

The difference between the two models is a different parameterization of the multinomial probabilities $\pi = (\pi_0, \pi_1, \pi_2)'$. The cumulative model parameterizes the model with respect to cumulative probabilities $p_0 = \pi_0$ and $p_1 = \pi_0 + \pi_1$ with $p_0 < p_1$. The sequential model uses $q_0 = \pi_0$ and the conditional probability $q_1 = \pi_1/(1 - \pi_0)$. Suppose now we use independent flat beta$(1,1)$ priors for p_0 and p_1 in model 1 or q_0 and q_1 in model 2, respectively. No attempt is made here to choose compatible priors (Dawid and Laurizen, 2001); the following discussion holds for any prior choice as long as the priors are assumed to be independent (note, however, that the order restriction $p_0 < p_1$ already implies a dependence between p_0 and p_1 in the cumulative model).

In the sequential model, it can now easily be seen that, conditional on the data, q_0 and q_1 are still independent because the posterior is proportional to the multinomial likelihood

$$p(q_0, q_1 \mid y) \propto q_0^{y_0}(q_1(1 - q_0))^{y_1}\{(1 - q_0)(1 - q_1)\}^{y_2}$$
$$= q_0^{y_0}(1 - q_0)^{y_1 + y_2} q_1^{y_1}(1 - q_1)^{y_2}, \tag{10}$$

which can be factorized into independent beta terms. Therefore, q_0 and q_1 are independent in the posterior with marginal distribution

$$q_0 \mid y \sim \text{beta}(y_0 + 1, y_1 + y_2 + 1)$$

and

$$q_1 \mid y \sim \text{beta}(y_1 + 1, y_2 + 1).$$

In the cumulative model, however, the posterior

$$p(p_0, p_1 \mid y) \propto p_0^{y_0}(p_1 - p_0)^{y_1}(1 - p_1)^{y_2}$$
$$\text{for } p_0 < p_1 \text{ and } 0 \text{ elsewhere}$$

cannot be factorized and p_0 and p_1 will be dependent. Furthermore, although $p_0 = q_0 = \pi_0$, the marginal posterior distribution of p_0

$$p_0 \mid y \propto p_0^{y_0} \int_{p_0}^1 (p_1 - p_0)^{y_1}(1 - p_1)^{y_2} dp_1$$

is different from the posterior for q_0 and does not seem to be analytically tractable. We can easily sample from the posterior $p(p_0, p_1 \mid y)$, e.g., by Markov chain Monte Carlo, and compare the corresponding multinomial probabilities π with the one obtained from the sequential model. In some empirical comparisons, we have found slight differences for the posterior distribution of π_0 and stronger discrepancies for π_1 and π_2.

This simple example transfers to the general case. Although both the cumulative and the sequential model specify the same model for the probability π_{ij0} of not developing the disease, the posterior distributions of μ_0 and ν_0, θ_{0i} and η_{0i}, and φ_{0j} and ψ_{0j} are not necessarily exactly the same (except for the binomial case $S = 1$) because of the different parameterization of the remaining multinomial probabilities $(\pi_{ij1}, \ldots, \pi_{ijS})$.

Incidentally, the maximum likelihood estimates will be the same in both models due to the invariance property of such estimates with respect to reparameterization (e.g., Cox and Hinkley, 1974). For example, in the above example, the ML estimate for p_0 and q_0 is $y_0/(y_0+y_1+y_2)$ while p_1 is estimated by $(y_0+y_1)/(y_0+y_1+y_2))$ and q_1 is estimated by $y_1/(y_1+y_2)$.

Returning to the factorization (10), we note that the same independence structure holds also in the general sequential model and implies that we could—equivalently to the joint multinomial approach defined by (4) and (5)—estimate S binomial regression models,

$$y_{ij0} \sim \text{B}\left(n_{ij}, \text{logit}^{-1}(\nu_0 + \eta_{0i} + \psi_{0j})\right)$$
$$y_{ij1} \sim \text{B}\left(y_{ij1} + \cdots + y_{ijS}, \text{logit}^{-1}(\nu_1 + \eta_{1i} + \psi_{1j})\right)$$
$$\vdots$$
$$y_{ij,S-1} \sim \text{B}\left(y_{ij,S-1} + y_{ijS}, \right.$$
$$\left. \text{logit}^{-1}(\nu_{S-1} + \eta_{S-1,i} + \psi_{S-1,j})\right),$$

completely separately. This factorization in fact reflects explicitly the conditional definition of the model. Hence, there will be no information in the likelihood about correlation between parameters for different stages. In particular, the extension to multivariate MRF and random walk priors as discussed earlier for the cumulative model does not seem to

be useful here. A separate modeling approach might be advantageous if one is mainly interested in the variation of the stage-specific proportions but not in the overall disease rate. Note that then the actual number of person-years n_{ij} is not even needed for such an analysis. This in fact opens up the possibility for continuous spatial modeling of the risk surface (for a similar non-Bayesian approach for spatial case–control data, see Kelsall and Diggle (1998)) if the exact locations of disease cases are known.

Finally, the factorization (10) implies that the posterior distribution of ν_0, η_{0i}, and ψ_{0j} will be the same whether or not we further stratify by the cancer stages. This would not be exactly the case in the cumulative model.

2.6 Computational Issues

Inference has been carried out using C++ routines developed by the first author. We have used Markov chain Monte Carlo (MCMC) to sample from the relevant posterior distributions, applying univariate Gaussian Metropolis random walk proposals for all components of θ_s (η_s) and φ_s (ψ_s), $s = 1, \ldots, S - 1$, while Gibbs steps have been used for the remaining precision parameter. The spread of each Metropolis proposal was tuned in an automatic fashion, prior to the collection of the posterior samples, so that the corresponding acceptance rate for each parameter was between 35 and 45%. Note that, in the cumulative model, one needs to check the additional restriction (3). If the Metropolis proposal did not fulfill the restriction, it was simply rejected (formally due to a zero prior term in the numerator of the acceptance ratio).

Both formulations impose an identifiability problem on the overall risk parameter μ_s (ν_s), as those can also be absorbed by both age-group and spatial effects. We have recentered both θ_s (η_s) and φ_s (ψ_s) after each iteration, with a corresponding adjustment to μ_s (ν_s) for $s = 0, \ldots, S - 1$. This is a valid approach as long as we assume a locally uniform prior for μ_s (ν_s) because it neither changes the value of the likelihood nor of the prior (all pairwise difference priors have an implicit flat prior on the overall level), hence not of the posterior. Furthermore, it enables us to explore the posterior distribution of the age and spatial effects. Alternatively, one could impose a sum-to-zero restriction directly in the prior for each age-group and spatial parameter block. However, one would need to implement a block updating algorithm, as suggested in Rue (2001) because single-site updating would be impossible due to degenerate full conditionals. Block updating would also be helpful for sparse data, where similar models are known to have convergence and mixing problems (Knorr-Held and Rue, 2002). However, the data we considered in our application are not particularly sparse and MCMC mixing was fine for the single-site scheme we have implemented.

We finally note that Albert and Chib (1993, 2001) suggested a latent variable approach for Bayesian inference by MCMC both in the cumulative and sequential model. This can be advantageous in applications where the number of observations is small or moderate. However, in the current context, the number of latent variables will be equal (in the cumulative model) or even a multiple (in the sequential model) of the number of person-years at risk. This seems to be prohibitive; e.g., in our application, the number of person-years, which is here simply the population number, exceeds seven million.
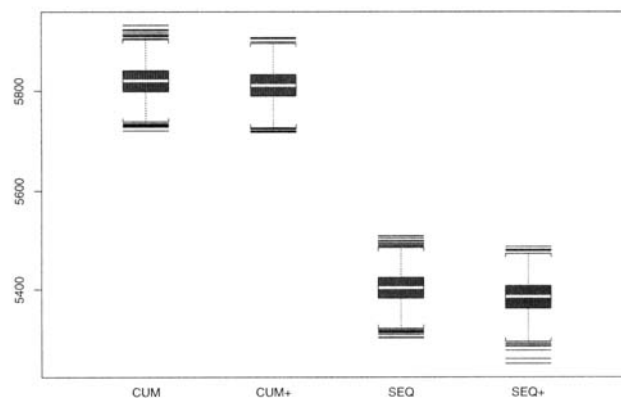

Figure 1. Boxplots of posterior samples from the deviance for the four different models.

3. Application

We now describe an application of the methodology described above to incidence data on cervical cancer in the former German Democratic Republic (GDR). The data are available on a yearly basis; here we present results for the year 1975, shortly after the introduction of Pap smear screening programs. We have used the values $a = 1.0$ and $b = 0.001$ as the default choice for the gamma hyperprior of all precision parameters, which corresponds to an extremely dispersed distribution for the (inverse gamma distributed) variances with infinite mean and variance and a prior mode at 0.0005.

The data are stratified by $I = 216$ administrative districts and $J = 14$ age groups (20–24, 25–29, ..., 80–84, and 85+). There were no cases below age 20. The original records give information on the stage of the detected lesion in six categories: (I) dysplasia, (II) carcinoma in situ (both premalignant), and (III–VI) malignant cancer of increasing severity. Effective screening shifts (a) the stage of the detected lesion toward earlier stages, preferentially to a premalignant condition, and (b) the time of detection toward younger age groups. Here we focus on the effect of stage shift and combine for simplicity the premalignant categories I and II into stage $s = 1$. Similarly, we aggregate the malignant categories III–VI into stage $s = S = 2$. We have deleted 35 cases (0.5%) with missing information on the stage of the disease. The total number of cases sum to 3466 in stage 1 and 3540 in stage 2; the corresponding total female population in the 14 age groups is 7,262,311. The median number of cases per district (regardless of the stage) is 20.5 (range 3–759). Stage-specific medians are 9 (0–433) for stage 1 and 11 (1–326) for stage 2.

In a first assessment of the model fit, Figure 1 compares the posterior distribution of the deviance (9) of the cumulative and the sequential model; both of them either without (denoted by CUM and SEQ) or with (denoted by CUM+ and SEQ+) additional unstructured random effects. Among the simpler formulations without additional unstructured random effects, the sequential model fits the data better than the cumulative model because the mean posterior deviance is smaller (5403 compared with 5820) and the support of the posterior deviance samples of the two models is well separated. Compared with the actual number of cells times the number of stages ($I \cdot J \cdot S = 216 \cdot 14 \cdot 2 = 6048$), this

Table 1

Deviance summaries

Model	$\bar{D}$	p_D	DIC
CUM	5820	130	5950
CUM+	5812	129	5941
SEQ	5403	245	5649
SEQ+	5384	259	5644

seems to be a decent fit to the data and indicates that neither interactions of age with space nor additional unstructured parameters are needed in both formulations. Indeed, the more complex formulations with additional parameters for unstructured heterogeneity give only a minor improvement in model fit, with a slightly smaller mean posterior deviance of 5812 for the cumulative and 5384 for the sequential model.

In the following, we therefore restrict our attention to the formulations without the additional unstructured parameters. We note, however, that the DIC criterion has a slight preference for the sequential model with additional unstructured effects (see Table 1). It is unclear if such a small difference in DIC really matters. One would also like to ensure that this difference is not due to Monte Carlo error; in particular, the assessment of the Monte Carlo error of p_D is difficult (Spiegelhalter et al., 2002). Fortunately, in our application, all maps and figures are virtually indistinguishable so our conclusions are the same with or without the additional unstructured effects.

First we compared the mean deviance residuals d_{ij}. Overall, 69% of the residuals from the sequential model are smaller than the corresponding ones from the cumulative model (see also Figure 2 for a graphical comparison), but no general pattern could be observed that would indicate the lack of fit of the cumulative model in particular age groups or districts, say. One is tempted to study the deviance residuals further stratified by stage, but this does not prove useful because stage-specific contributions $y_{ijs} \log(y_{ijs}/(n_{ijs}\pi_{ijs}))$ can be large in absolute size although their sum d_{ij}^2 may still be small.

Turning now to the estimated age effects, Figure 3 displays posterior median estimates within 90% pointwise credible intervals of $-\varphi_0$ and $-\varphi_1$ from the cumulative model. One can see a fairly similar inverse "bathtub" pattern of the two curves. The second curve, which describes the age pattern relevant for being diagnosed with a malignant form of the disease, has a nearly constant slope for age between 30 and 70, whereas the slope of the first curve, representing the log relative risk for both the premalignant and malignant stage, is already negative in that age range. This reflects the fact that the malignant stage of cervical cancer is more likely to be diagnosed in older age groups, as the cancer needs time to progress (undetected) through the premalignant stage.

The estimates of $-\psi_0$ from the sequential model (Figure 4, left plot) are directly comparable with $-\varphi_0$ (Figure 3, left plot) because both correspond to the overall log relative disease risk (keep in mind, however, that the estimates do not have to be exactly identical, as commented on earlier). Here there is virtually no difference to see. Finally, the right plot in Figure 4 displays the age effect on the conditional risk of the malignant disease stage 2, given a diagnosis in stage 1 or
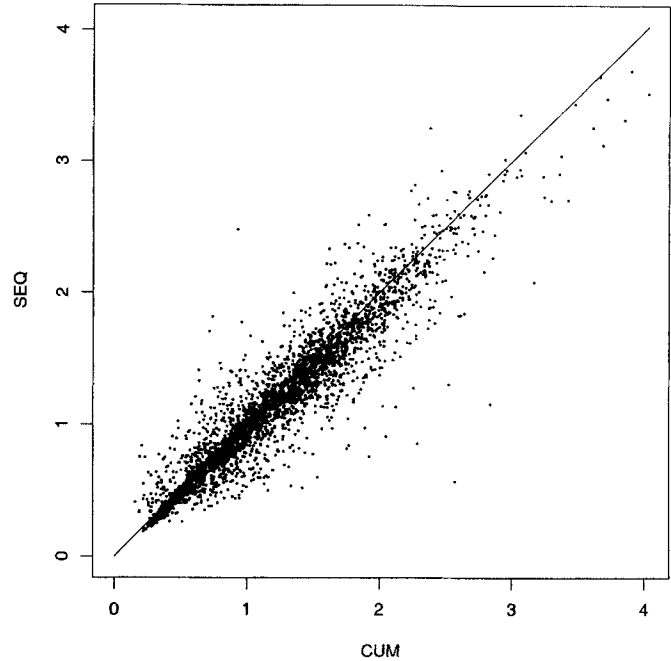

Figure 2. Estimated mean deviance residuals from the sequential model (y-axis) plotted against the corresponding ones from the cumulative model (x-axis).

2. As expected, an increasing conditional risk with increasing age can be seen, which is remarkably linear on the logit scale.

Figure 5 now displays the estimated spatial incidence pattern regardless of the stage. The first map shows standardized morbidity ratios (SMRs) calculated by internal standardization through joint maximum likelihood (ML) estimation (see Breslow and Day, 1987, Chapter 4). More specifically, we obtained the SMRs by applying a standard logistic regression procedure to the aggregated cases in stages 1 and 2 as responses, using age group and district as factors (each of them restricted to sum to zero). Displayed is the exponential of the estimated spatial parameters, which can hence be interpreted as (age-adjusted) relative risk estimates. The other two maps display the corresponding (posterior median) relative risk estimates $\exp(-\theta_0)$ and $\exp(-\eta_0)$ from the cumulative and sequential models, respectively. One can see a fairly similar pattern with the expected smoothing effect, slightly more pronounced for the cumulative model. This might be caused by the additional order restrictions (3). Note that we have used the same scale from 0.4 to 2.5 in all maps, which covers the estimates from the cumulative model (range 0.64–2.39) but not all of the SMRs (0.35–3.19) nor all of the estimates obtained from the sequential model (0.49–2.68). The range was chosen in order to make the spatial pattern in the smoothed maps more visible.

Figure 6 now displays, on the same scale as Figure 5, estimates of the relative risk of a tumor diagnosis in the malignant stage 2 of the disease. The left map gives ML estimates, calculated just as in Figure 5, but only with the cases in stage 2 as responses. The other map displays the median relative risk estimates $\exp(-\theta_1)$ from the cumulative
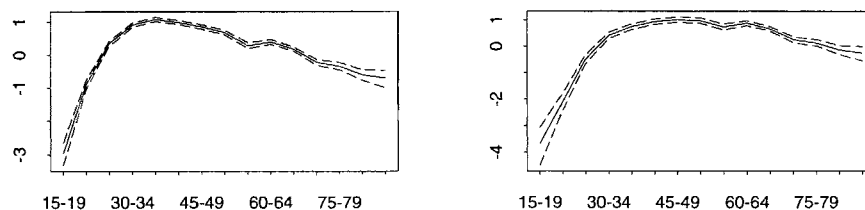
Figure 3.   Estimated median age effects of $-\varphi_0$ (left plot) and $-\varphi_1$ (right plot) within 90% pointwise credible intervals from the cumulative model.
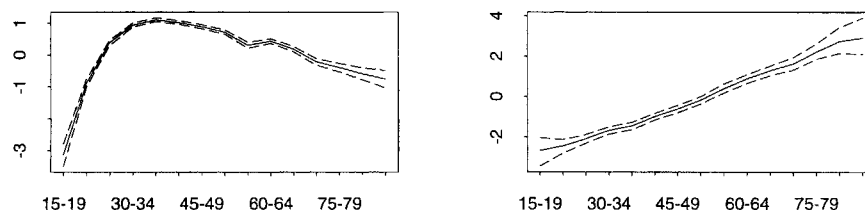

Figure 4.   Estimated median age effects of $-\psi_0$ (left plot) and $-\psi_1$ (right plot) within 90% pointwise credible intervals from the sequential model.

model. There is less spatial variation than for the overall risk $\exp(-\theta_0)$ (Figure 5, middle map), with slightly higher values east of West Berlin (the hatched region in the middle of the map).

Finally, Figure 7 (right map) gives the estimated odds ratio $\exp(-\eta_1)$ from the sequential model for the probability of a diagnosis in a malignant stage of the disease, conditional on a diagnosis in stage 1 or 2. For comparison, the left map displays the corresponding ML estimates. These have consid-

erably more variation; in fact, the district-specific ML estimates did not even exist for 7 out of the 216 districts due to no observations in stage 1. The smoothed map shows higher conditional risk of stage 2 in the southwest and lower conditional risk in the northeast and some other parts of the country. This corresponds roughly to what is known about the local introduction of cervical cancer screening programs. Cervical cancer screening by Pap smear was first introduced in the former GDR as a pilot project in two specific regions
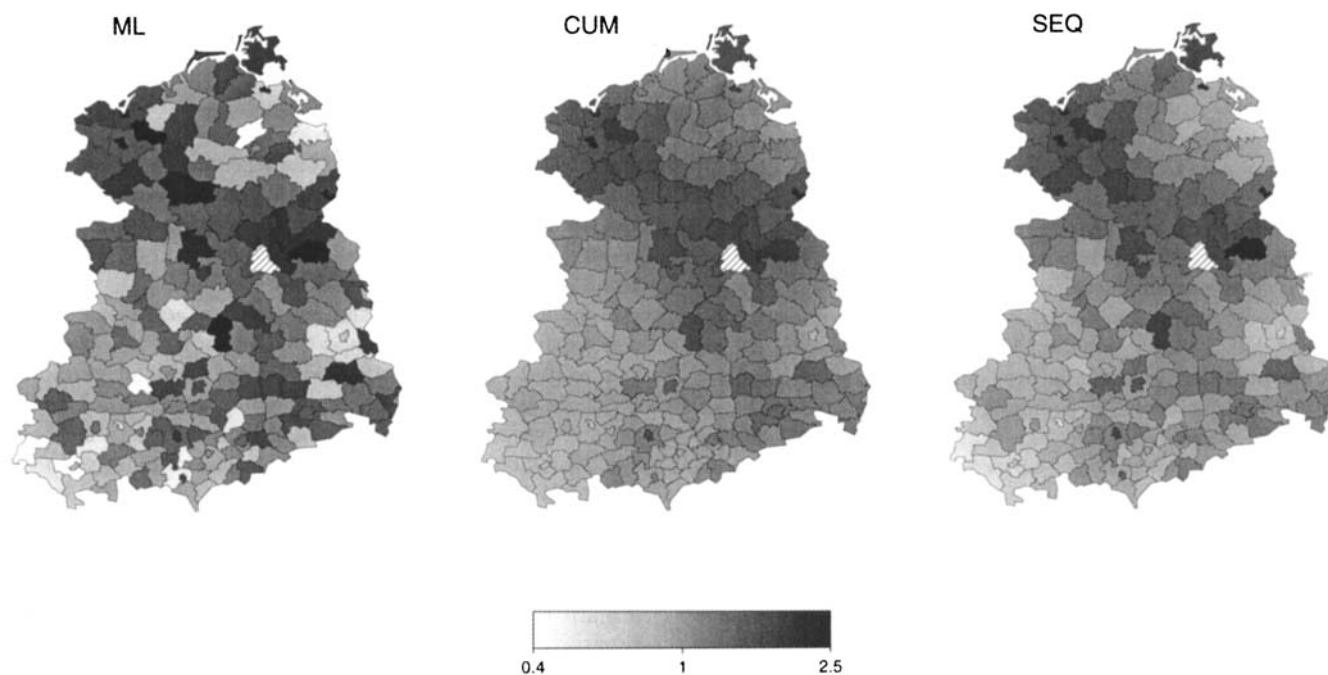

Figure 5.   Relative risk estimates for diagnosis of the disease regardless of the stage by ML (left map), the cumulative (middle map), and the sequential model (right map).
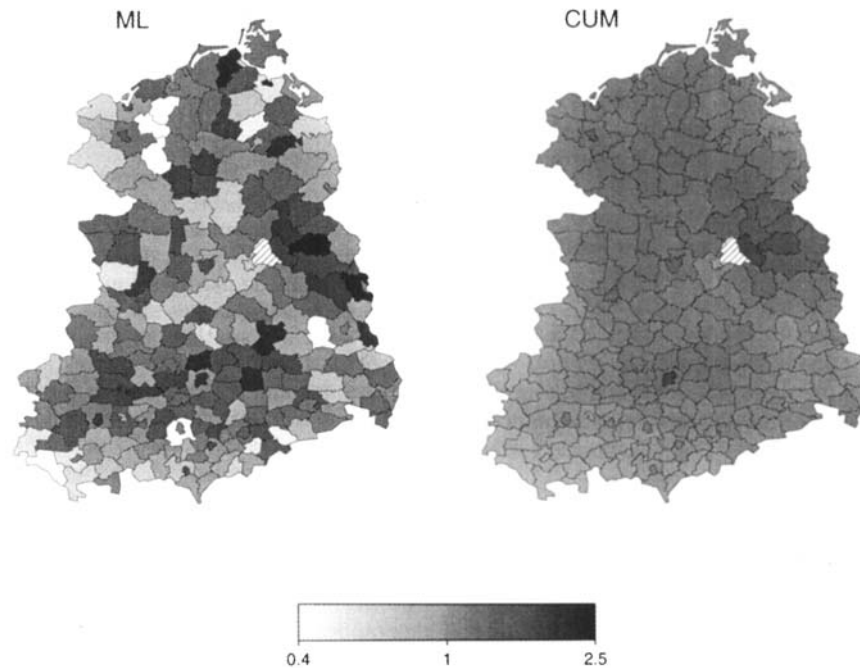
Figure 6. Relative risk estimates for diagnosis of the disease in stage 2 by ML (left map) and the cumulative model (right map).

in 1974, East Berlin and Mecklenburg-Vorpommern (northern coastal region). Available information on the number of lab tests indicates that, in the 1970s, the highest number of tests were carried out in these two areas, while in Saxony-Anhalt and Thuringia (the southeast of the GDR), the lowest numbers were observed (Quaas and Heinrich, 1998).

The maps fit roughly into this pattern: In the northwest (initially high number of tests), they show totally a higher proportion of identified premalignant and malignant cancers (Figure 5), but among them low proportions of malignant cancers (Figure 7). In the southwest of the country (initially low numbers of tests), totally a lower proportion of identi-
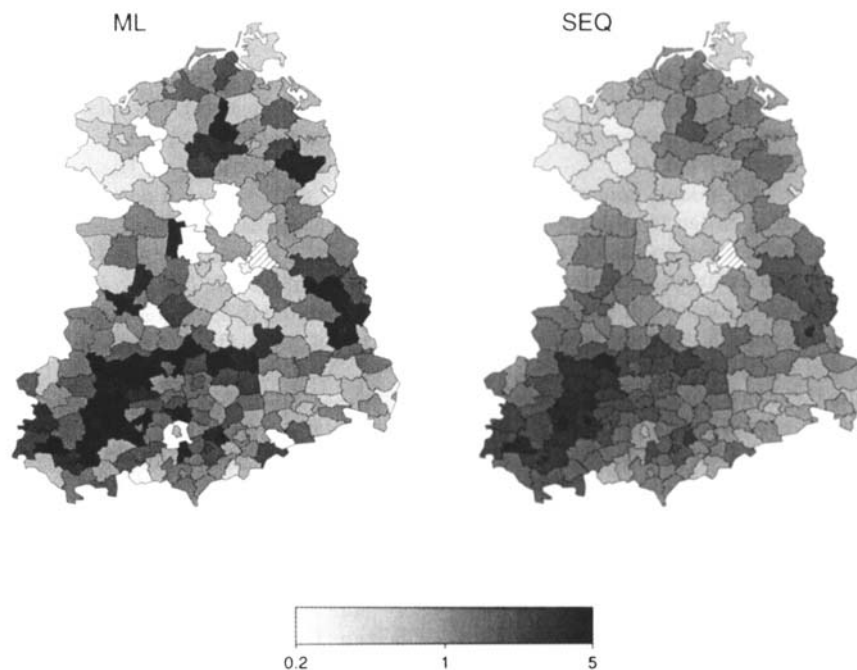

Figure 7. Conditional odds ratio estimates for diagnosis in the malignant stage 2, given diagnosis in stage 1 or 2 by ML (left map) and the sequential model (right map).

fied premalignant and malignant cancers can be seen, but among them high proportions of malignant cancers. In detail, the pattern is more complicated. Not the entire area of Mecklenburg-Vorpommern shows the low proportion of cancers, and areas with initially low frequencies of testing show nevertheless low proportions of malignant cancers (e.g., Saxony in the southeast). These findings may be due to the fact that several factors influence the effectiveness of a screening program: (a) availability of the program, (b) quality of the program, (c) attendance of the eligible population, and (d) quality of outcome reports to the cancer registry. These factors may affect the outcome differently in the different regions of the country. The maps show only the overall effect of these factors. Thus, the method might be valuable in providing indicators of areas with unsatisfactory performance of the screening, whatever the reasons. Their elucidation would need more detailed epidemiological investigation.

4. Discussion

In this article, we have proposed methods for the spatial analysis of cancer incidence data with additional knowledge of the stage of the disease. Throughout, we have used Markov random field models in order to acknowledge the spatial structure of the data. Of course, other models for spatial correlation can be used as well, e.g., the recently developed adaptive smoothing methods based on partition (Knorr-Held and Raßer, 2000; Denison and Holmes, 2001) or mixture models (Green and Richardson, 2000). We are currently investigating the applicability of partition models to such data.

In terms of comparing the two proposed models, it seems that most arguments are in favor of the sequential model: (a) This model is easier to implement because no order constraints are necessary. (b) We can even separate the analysis and fit S binomial regression models separately. (c) The conditional interpretation of the parameters is more useful in order to judge the effectiveness of cancer screening and shows connections to the statistical analysis of spatial case–control studies (Kelsall and Diggle, 1998). (d) In our application, the sequential model provided a substantially better model fit. Only if the interest lies in estimating cumulative relative risks should the cumulative model be preferred.

An obvious extension of the two models considered is the inclusion of relevant covariates in order to reduce (explain) the observed spatial pattern. Depending on the covariate and on the model, the effect could be assumed to be independent of the stage or stage-specific. For example, if the number of lab tests would be available on a district-specific level, it could be included in the sequential model (4) for $s = 1$.

Finally, we note that the incidence data from the GDR cancer registry is actually available for all years between 1961 and 1989. An interesting problem would be to construct a space–time model that captures the increasing number of cases in the premalignant stage and their temporal effect on the number of diagnosed malignant cases sometime later. Here the specification of the time lag between the premalignant and malignant stage is not obvious and could possibly even be estimated from such data as well.

Résumé

Nous proposons deux approches dans l'analyse spatiale des incidences de cancer, conçues pour le cas où l'on dispose également d'informations sur le degré de la maladie au moment du diagnostic. Les deux modèles présentés sont en fait des extensions de modèles couramment utilisés dans l'analyse de données catégorielles sur une échelle ordonnée; nous y incluons des effets spatiaux et des effets liés à l'âge, que nous estimons par lissage non paramétrique. Plus précisément, nous empruntons une voie tout à fait bayésienne en posant, sur les différences liées aux comparaisons par paires que nous voulons explorer, des distributions a priori de type gaussien. Quant aux paramètres de lissage, ils sont, eux aussi, supposés inconnus. Il y a beaucoup d'arguments en faveur de ces méthodes, tout à fait utiles pour contrôler l'efficacité des screenings pratiqués dans les cancers les plus fréquents. Nous illustrons ce fait par une analyse de données portant sur le cancer du col de l'utérus en ex-RDA. Les résultats de cette analyse suggèrent de grandes différences dans les proportions relatives des degrés de la maladie selon les régions, en fonction de la présence et de l'intensité des programmes de dépistage du cancer du col par cytologie cervico-vaginale.

References

Agresti, A. (1984). *Analysis of Ordinal Categorical Data.* New York: Wiley.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88,** 669–679.

Albert, J. and Chib, S. (2001). Sequential ordinal modeling with applications to survival data. *Biometrics* **57,** 829–836.

Besag, J. E., York, J. C., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43,** 1–59.

Besag, J. E., Green, P. J., Higdon, D. M., and Mengersen, K. L. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10,** 3–66.

Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A., and Conlon, E. M. (1999). Bayesian methods for spatially correlated disease and exposure data. In *Bayesian Statistics 6*, J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith (eds), 131–156. Oxford: Oxford University Press.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research,* Volume 2, *The Design and Analysis of Cohort Studies.* Lyon: International Agency for Research on Cancer.

Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis.* London: Chapman and Hall.

Clayton, D. G. and Bernardinelli, L. (1992). Bayesian methods for mapping disease risks. In *Small Area Studies in Geographical and Environmental Epidemiology,* J. Cuzick and P. Elliot (eds), 205–220. Oxford: Oxford University Press.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Dawid, A. P. and Lauritzen, S. L. (2001). Compatible prior distributions. In *Bayesian Methods with Applications to Science, Policy and Official Statistics,* E. I. George (ed), 109–118. Luxembourg: Office for Official Publications of the European Communities.

Denison, D. and Holmes, C. (2001). Bayesian partitioning for estimating disease risk. *Biometrics* **57**, 143–149.

Diggle, P. J. (1996). Spatial analysis in biometry. In *Advances in Biometry,* P. Armitage and H. A. David (eds), 363–384. New York: Wiley.

Fahrmeir, L. and Knorr-Held, L. (2000). Dynamic and semiparametric models. In *Smoothing and Regression: Approaches, Computation and Applications,* M. Schimek (ed.), 513–544. New York: Wiley.

Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society, Series C* **50**, 201–220.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models,* 2nd edition. New York: Springer-Verlag.

Green, P. J. and Richardson, S. (2000). *Spatially correlated allocation models for count data.* Technical Report. University of Bristol, Bristol.

Hastie, T. and Tibshirani, R. (2000). Bayesian backfitting (with discussion). *Statistical Science* **15**, 196–223.

Kelsall, J. E. and Diggle, P. J. (1998). Spatial variation in risk of disease: A nonparametric binary regression approach. *Journal of the Royal Statistical Society, Series C* **47**, 559–573.

Knorr-Held, L. (2002). Some remarks on Gaussian Markov random field models for disease mapping. In *Highly Structured Stochastic Systems,* N. Hjort, P. Green, and S. Richardson (eds). Oxford: Oxford University Press. In press.

Knorr-Held, L. and Becker, N. (2000). Bayesian modelling of spatial heterogeneity in disease maps with application to German cancer mortality data. *Allgemeines Statistisches Archiv (Journal of the German Statistical Society)* **84**, 121–140.

Knorr-Held, L. and Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine* **17**, 2045–2060.

Knorr-Held, L. and Rue, H. (2002). On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics,* in press.

Knorr-Held, L. and Raßer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics* **56**, 13–21.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B* **42**, 109–127.

Quaas, J. and Heinrich, J. (1998). Cervical cancer screening— A retrospective comparison between the old and new German federal states (in German). *Zentralblatt für Gynäkologie* **120**, 13–19.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B* **63**, 325–338.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, in press.

Wakefield, J. C., Best, N. G. and Waller, L. A. (2000). Bayesian approaches to disease mapping. In *Spatial Epidemiology: Methods and Applications,* P. Elliot, J. C. Wakefield, N. G. Best, and D. J. Briggs (eds), 104–127. Oxford: Oxford University Press.