Knorr-Held, Raßer, Becker:

Disease Mapping of Stage-specific Cancer Incidence Data

Sonderforschungsbereich 386, Paper 249 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

Projektpartner

Disease mapping of stage-specific cancer incidence data

Leonhard Knorr-Held, Günter Raßer

Department of Statistics

Ludwig-Maximilians-University Munich

Ludwigstrasse 33

80539 Munich

Germany

leo@stat.uni-muenchen.de

rasser@stat.uni-muenchen.de

Nikolaus Becker

German Cancer Research Center

Department of Biostatistics

Im Neuenheimer Feld 280

69120 Heidelberg

Germany

n.becker@dkfz.de

First version: June 2001

Abstract

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 apply our methods 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 indicates spatial variability with respect to the introduction and effectiveness of 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 for the spatial analysis of observational disease data within the last ten 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. Bayesian approaches for the second type of data includes 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 see for example Clayton and Bernardinelli (1992), Knorr-Held and Becker (2000) or Wakefield *et al.* (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 to 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 to relevant unobserved risk factors, although some general problems of interpretation can remain due to the observational type of the data.

In this paper we extend the methodology to the analysis of cancer incidence data with additional knowledge on the stage of disease at time of diagnosis. Our aim can be described as (a) to adjust the crude observed data for effects which 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 review see Fahrmeir and Tutz, 2001, Ch. 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. Note that we work directly on data stratified by age, which is in contrast to ordinary disease mapping methods (without stage-stratification), where the data are typically standardized by age in advance in order to calculate the expected number of cases, which are then used as an offset in a Poisson regression approach.

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. Here we compare our estimates with those obtained from a corresponding Maximum Likelihood approach with unrestricted age group and spatial effects. 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 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 $y_{ij0} = n_{ij} - \sum_{s=1}^{S} y_{ijs}$ be the number of all person-years at risk, which 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 $\sum_{s=0}^{S} \pi_{ijs} = 1$.

2.1 The cumulative model

In the cumulative model we factorize the log odds of the *cumulative probabilities* $p_{ijs} = \pi_{ij0} + \ldots + \pi_{ijs}$ into an intercept term μ_s, a spatial effect θ_s, and an age group effect φ_s, that is

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

Equivalently this model can be formulated in terms of *ascending* cumulative probabilites $1 - p_{ijs}$; 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

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

where $\text{logit}^{-1}(u) = 1/(1+\exp(-u))$. To ensure that all these probabilities are positive, the parameters μ, θ and φ have to fulfill the constraints

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

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

2.2 The sequential model

The sequential approach to ordinal data considers the *conditional* probability that an individual in cell (i, j) gets diagnosed of the disease in stage s, *assuming* that she gets diagnosed of the disease in stage s or higher, i.e. $q_{ijs} = \pi_{ijs}/(\pi_{ijs} + \ldots + \pi_{ijS})$. Again we decompose the log-odds of these conditional probabilities into an intercept term μ_s, a spatial effect θ_s, and an age group effect φ_s, which can be written as

$$\log\left(\frac{q_{ijs}}{1-q_{ijs}}\right) = \log\left(\frac{\pi_{ijs}}{\sum_{t=s+1}^{S}\pi_{ijt}}\right) = \mu_s + \theta_{si} + \varphi_{sj} \qquad (s = 0, \ldots, S-1). \tag{4}$$

Note that, formally, the only difference to the cumulative model (1) is that π_{ijs} replaces the cumulative probability $\pi_{ij0} + \ldots + \pi_{ijs}$ in the nominator.

The probabilities π_{ijs} can now be derived as

$$\pi_{ijs} = \begin{cases} \text{logit}^{-1}(\mu_0 + \theta_{0i} + \varphi_{0j}) & (s = 0) \\ \text{logit}^{-1}(\mu_s + \theta_{si} + \varphi_{sj}) \cdot \prod_{t=0}^{s-1}\{1 - \text{logit}^{-1}(\mu_t + \theta_{ti} + \varphi_{tj})\} & (s = 1, \ldots, S-1) \\ \prod_{t=0}^{S-1}\{1 - \text{logit}^{-1}(\mu_t + \theta_{ti} + \varphi_{tj})\} & (s = S) \end{cases}, \tag{5}$$

e.g. 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 constraints have to be imposed on the parameters μ, θ and φ. A further difference to 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 degenerate binomial case $S = 1$. Indeed, the sequential model can be derived from an underlying mechanism where categories can be reached successively, but only in one specific direction, see Fahrmeir and Tutz (2001).

2.3 Prior assumptions

The two alternative models proposed above are now completed by assinging prior distributions to all unknown parameters. For both the spatial and the age group parameters we

will use priors which favour 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.

More specifically we use Gaussian pairwise difference priors (Besag *et al.,* 1995) for the district and age group-specific parameters. 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) and Hastie and Tibshirani (2000).

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

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

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 (2001).

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

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

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

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

with a Wishart prior on its precision matrix Λ_θ, i.e.

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

again with suitably chosen constants a and B, where a is a scalar and B is a $S \times S$-matrix. This might be appropriate if the MRF's $\theta_0, \ldots, \theta_{S-1}$ are expected to be correlated a priori. In particular in the cumulative model, one might want to assume that the spatial pattern are similar. However, we do not expect any major differences between the two formulations and therefore stick to the simpler form with a priori independent MRF's.

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 reparametrized version (e.g. Carlin and Louis, 1996, p. 308), where θ_{si} is independent Gaussian with mean u_{si} and precision ν_s, say, and a GMRF prior is now placed on the latent vectors u_s, just like in (6) for θ_s. Of course, this model could also be formulated in a multivariate way with possibly dependent parameters u_i or θ_i. We have tested in our application both models with and without the additional unstructured parameters.

For the age-group specific parameters, we assume in similar lines that, for each category s, the parameters φ_s follow a simple Gaussian random walk in time with variance $\lambda_{\varphi_s}^{-1}$, with a flat prior for the initial value $\varphi_{s,1}$. Such a formulation is the exact temporal analogue of model (6) as the prior can be written as

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

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.

2.4 Model choice and parameter interpretation

At this point it might be worth noting, that the posterior distribution of the *conditional* probabilities can of course easily be derived from the *cumulative* model as well, as they are just simple functions of the posterior distribution of the π_{ijs}'s. Similarly, the posterior distribution of the cumulative probabilities 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 parametrization 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.

We are particularly interested in spatial disease risk estimates, adjusted for age. In both models, $\exp(-\theta_{0i})$ corresponds to the *adjusted relative risk* in district i, regardless of the stage of the disease. In model 1, it is convenient to interpret $\exp(-\theta_{si})$ as the *adjusted cumulative relative risk*. In model 2, $\exp(-\theta_{si})$, $s = 1, \ldots, S$ can be interpreted as the (adjusted) *odds ratio* with respect to the corresponding conditional probabilities. Similarly we prefer to display $-\varphi_s$ (rather than φ_s), the age group effects on the cumulative probabilities $1 - p_{ijs}$ in model 1 and on the (conditional) probabilities $1 - q_{ijs}$ in model 2. This has the advantage that higher values in the figures displaying age effects, and darker colours 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*

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

with the multinomial squared deviance residual

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

(using the convention that $0 \log 0 = 0$). Each squared deviance residual can be seen as a (standardized) measure of fit, comparing the observed number of cases y_{ijs} with fitted number of cases $n_{ijs}\pi_{ijs}$ for all stages $s = 0, \ldots, S$. For a well fitting model, D should be asymptotically (with increasing data in each cell (i, j)) around $I \cdot J \cdot S$, see Spiegelhalter, Best and Carlin (1998) (the factor S appears here due to the multinomial response with S "free" categories).

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 parametrization of the multinomial probabilities $\pi = (\pi_0, \pi_1, \pi_2)'$: The cumulative model parametrizes 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 (similar results can be obtained for other priors, e.g. flat priors on the logit scale). In principle these priors will be assumed to be independent, but note that the order restriction (3), which reduces here to $p_0 < p_1$, already implies a dependence between p_0 and p_1 in the cumulative 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

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

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

$$q_0|y \sim Beta(y_0 + 1, y_1 + y_2 + 1) \text{ and } q_1|y \sim Beta(y_1 + 1, y_2 + 1).$$

In the cumulative model, however, the posterior

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

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

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

is different from the posterior for q_0 and cannot even be calculated analytically. However, we can easily sample from the posterior $p(p_0, p_1|y)$, for example 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 examples 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, we will not get exactly the same posterior distribution of μ_0, θ_{0i} and φ_{0j} (except for the binomial case $S = 1$), even if we have apparently the same prior models for the parameters, because of the different parametrization of the remaining multinomial probabilities $(\pi_{ij1}, \ldots, \pi_{ijS})$.

Note however that the maximum likelihood estimates of μ_0, θ_{0i} and φ_{0j} will be the same in both models due to the invariance property of ML estimates with respect to reparametrization (e.g. Cox and Hinkley, 1974). For example, in the above example the (obvious) 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 $y_1/(y_1 + y_2)$.

As a final point we note that the factorization (10) 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

$$\begin{aligned} y_{ij0} &\sim B(n_{ij}, \text{logit}^{-1}(\mu_0 + \theta_{0i} + \varphi_{0j})) \\ y_{ij1} &\sim B(y_{ij1} + \ldots + y_{ijS}, \text{logit}^{-1}(\mu_1 + \theta_{1i} + \varphi_{1j})) \\ &\vdots \\ y_{ij,S-1} &\sim B(y_{ij,S-1} + y_{ijS}, \text{logit}^{-1}(\mu_{S-1} + \theta_{S-1,i} + \varphi_{S-1,j})) \end{aligned}$$

completely separately. Of course, this is only valid because all parameters are stage-specific and all pairwise difference priors are assumed to be independent. A separate modelling 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 the analysis. On the other hand the factorization (10) of the posterior implies that the posterior of μ_0, θ_{0i} and φ_{0j} will be exactly the same as the one obtained in the usual disease mapping approach, where the disease cases are aggregated over all different stages (using a binomial instead of the Poisson model). Note that 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 and φ, 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%. Samples from u_{is} can be generated by simple Gibbs steps, due to a Gaussian full conditional. 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 nominator of the acceptance ratio).

Both formulations impose an identifiability problem on the overall mean parameter μ_s, as those can also be absorbed by both θ_s and φ_s. We have recentered both θ_s and φ_s after each iteration with a corresponding adjustment to μ_s for $s = 0, \ldots, S-1$. This is a valid approach as long as we assume a locally uniform prior for μ_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 block θ_s and φ_s, $s = 0, \ldots, S-1$. However, one would need to implement a block updating algorithm, as for example 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, 2001). However, the data we considered in our application is 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) have suggested to use a latent variable appraoch for Bayesian inference by MCMC both in the cumulative and sequential model. This can be advantageous in many applications when 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; for example, in our application the number of person-years exceeds seven millions.

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 is available on a yearly basis; here we present results for the year 1975, shortly after the introduction of Pap smear screening programs. We will denote the cumulative model by model 1 and the sequential model by model 2. The extended formulations with additional unstructured spatial effects are called model 1b and 2b respectively. We have used the values $a = 1.0$ and $b = 0.001$ as a default choice for the gamma hyperprior of all precision parameters, which corresponds to an extremely dispersed distribution for the variances with infinite mean and variance and a prior mode at 0.0005.

The data is 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 6 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 towards earlier stages, preferentially to a premalignant condition, and (b) the time of detection towards 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 up to 3,466 in stage 1 and 3,540 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, the posterior median deviance turns out to be 5,820

(5,770-5,875) and 5,403 (5,353-5,457) for model 1 and 2 respectively (90% credible intervals in brackets). This is a remarkably good fit of both models to the data - compared to the actual number of cells times the number of stages ($I \cdot J \cdot S = 216 \cdot 14 \cdot 2$=6,048) - 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 (model 1b and 2b) gives only a minor improvement in model fit, as can be seen from Figure 1, which compares the posterior distribution of the deviance of the four models. There seems to be evidence, however, that model 2 fits the data better than model 1 as the median posterior deviance is smaller and the ranges of the posterior deviance samples of the two models are well separated.

In the following we therefore restrict our attention to model 1 and 2 without additional parameters for unstructured heterogeneity. We first discuss the estimated age effects in both models. Figure 2 displays posterior median estimates within 90% pointwise credible intervals of $-\varphi_s$, $s = 0, 1$, obtained from model 1 (top row) and model 2 (bottom row). There is a decreasing precision of the estimates (with increasing s) due to the decrasing number of cases in the relevant likelihood in both models. For model 1, one can see a fairly similar inverse "bathtub" pattern of the two curves. Note that the slope of the curves is slightly increasing with s increasing. This reflects the fact that a higher stage of cervical cancer is more likely to be diagnosed in older age groups, as the cancer needs time to progress (undetected) through stage 1. 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 and lower risk outside, especially for age below 30.

The estimates of $-\varphi_0$ in model 2 (bottom left plot) are directly comparable to the corresponding ones obtained with model 1 (top left plot), as both correspond to the overall log relative disease risk (keep in mind, however, that the estimates don't have to be exactly

identical, as commented earlier). Here there is virtually no difference to see. The bottom right plot in Figure 2 displays the age effect on the conditional risk of the malignant disease stage 2, given a diagnosis in stage 1 or 2. As expected, an increasing conditional risk with increasing age can be seen, which is remarkably linear on the logit scale.

Figure 3 now displays the observed and estimated spatial incidence pattern of the disease, regardless of the stage. The first map shows Standardized Morbidity Ratios (SMR's) calculated by internal standardization through *joint* maximum likelihood (ML) estimation; see Breslow and Day (1987), chapter 4. More specifically, we obtained the SMR's by applying a standard logistic regression procedure to the aggregated cases in stage 1 and 2 as responses, using age group and district-specific parameters (each of them restricted to sum up to zero). Displayed are 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 $\exp(-\theta_{0i})$ from model 1 and 2. Compared to the ML estimates, one can see a fairly similar pattern with the expected smoothing effect, slightly more pronounced for model 1. This might be caused by the additional order restrictions (3) on the parameters in the cumulative model. As in the usual disease mapping context this indicates that the ML estimates have too much variation to be reliable estimates of the overall disease risk. Note that we have used the same scale from 0.4 to 2.5 in all maps, which covers the estimates from model 1 (range 0.64-2.39), but not all of the SMR's (0.35-3.19) nor of the estimates obtained from model 2 (0.49-2.68). The range was chosen in order to make the spatial pattern in the smoothed maps more visible.

Figure 4 now displays - on the same scale as Figure 3 - estimates of the relative risk of a tumour diagnosis in the malignant stage 2 of the disease. The left map gives ML estimates, calculated just as in Figure 3, but only with the cases in stage 2 as responses. The other

map displays the posterior median estimates from $\exp(-\theta_{1i})$ in model 1. There is less spatial structure compared to the cumulative relative risk estimates from $\exp(-\theta_{0i})$, with a slightly higher risk east of West-Berlin (the hatched region in the middle of the map).

Finally, Figure 5 gives the estimated odds ratio for the probability of a diagnosis in a malignant stage of the disease, conditional on a diagnosis in stage 1 or 2. The left map displays ML estimates, which shows considerably more variation than those obtained by model 2 (right map). 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 south-west, and lower conditional risk in the north-east 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 has been first introduced in the former GDR as a pilot project in two specific regions 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 have been carried out in these two areas, while in Saxony-Anhalt and Thuringia (the south-east of the GDR) the lowest numbers were observed (Quaas and Heinrich, 1998).

The maps fit roughly into this pattern: in the north-west (initially high number of tests) they show totally a higher proportion of identified premalignant and malignant cancers (Figure 3), but among them low proportions of malignant cancers (Figure 5). In the south-west of the country (initially low numbers of tests) totally a lower proportion of identified 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 malignant cancers, and areas with initially low frequencies of testing show nevertheless low proportions of malignant cancers (e.g. Saxony in the south-east). These findings may be due to the fact that *several* factors influence the

effectiveness of a screening programme: (a) availability of the programme, (b) quality of the programme, (c) attendance of the eligible population and (d) quality of outcome report 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 to provide indicators to areas with unsatisfactory performance of the screening whatsoever the reasons are. Their elucidation would need more detailed epidemiological investigation.

4 Discussion

In this paper we have demonstrated the application of Bayesian disease mapping methods for cancer incidence data with additional knowledge on 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, for example the recently developed adaptive smoothing methods based on partition (Knorr-Held and Raßer, 2000, Dension and Holmes, 2001) or mixture models (Green and Richardson, 2000). We are currently investigating the applicability of partition models to such data.

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 for stage 2.

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 space-time models that will capture the increasing number of cases in the pre-malignant stage and

their temporal effect on the number of diagnosed malignant cases some time later. Here the specification of the time lag between the pre-malignant and malignant stage is not obvious and could possibly even be estimated from such data as well.

Acknowledgments

This research was funded by the German Science Foundation (DFG), SFB 386. The authors thank R. Stabenow from the joint cancer registry (GKR) in Berlin for providing the dataset on cervical cancer in the GDR. Some of the work was done while the first author was at Imperial College London.

References

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**, to appear.

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.

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.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research*, vol. 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*: J. Cuzick and P. Elliot (eds), *Small Area Studies in Geographical and Environmental Epidemiology*, pp. 205-220. Oxford University Press, Oxford.

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

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

Fahrmeir, L. and Knorr-Held, L. (2000). Dynamic and semiparametric models. *in*: M. Schimek (ed.), *Smoothing and Regression: Approaches, Computation and Applications*, Ch. 18, pp. 513-544, Wiley & Sons, New York.

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

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

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

Knorr-Held, L. (2001). Some remarks on Gaussian Markov random field models for disease mapping. In N. Hjort, P. Green and S. Richardson (eds.), *Highly Structured Stochastic Systems*. Oxford: Oxford University Press. forthcoming.

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 Rue, H. (2001). On block updating in Markov random field models for disease mapping. Revised for *Scandinavian Journal of Statistics*.

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

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. and Carlin, B. P. (1998). Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models, Technical Report, MRC Biostatistics Unit, Cambridge, UK.

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

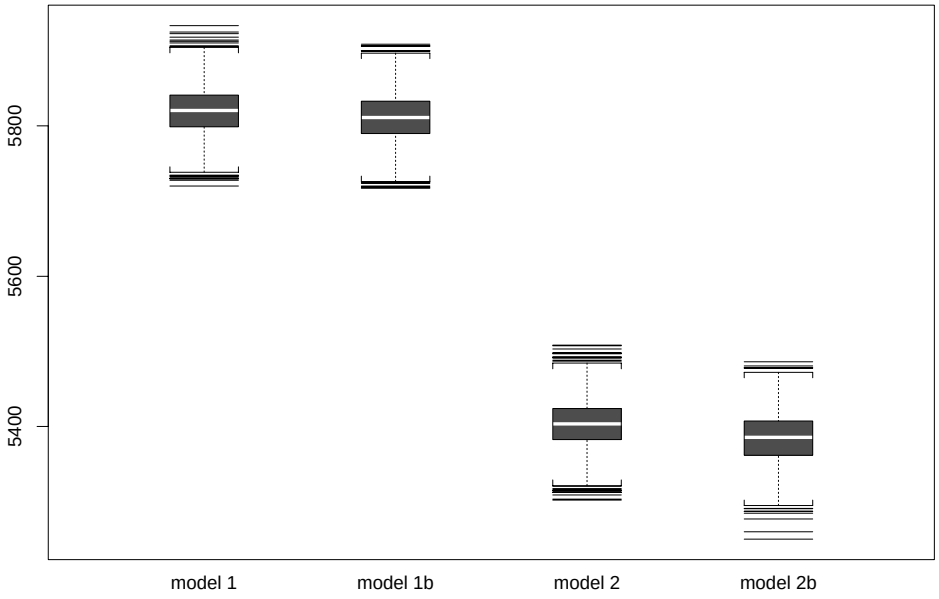

Figure 1: Boxplots of posterior samples from the deviance for the four different models

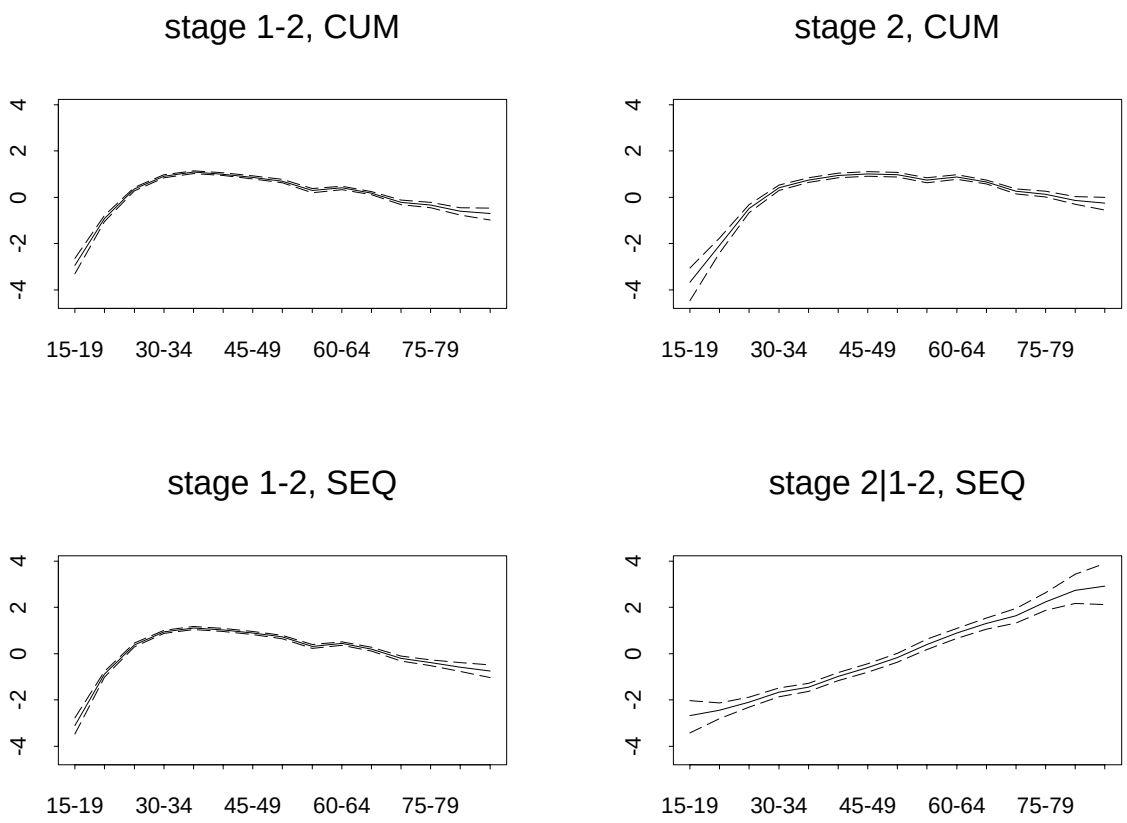

Figure 2: Estimated median age effects within 90% pointwise credible intervals from the cumulative (top row) and the sequential model (bottom row)

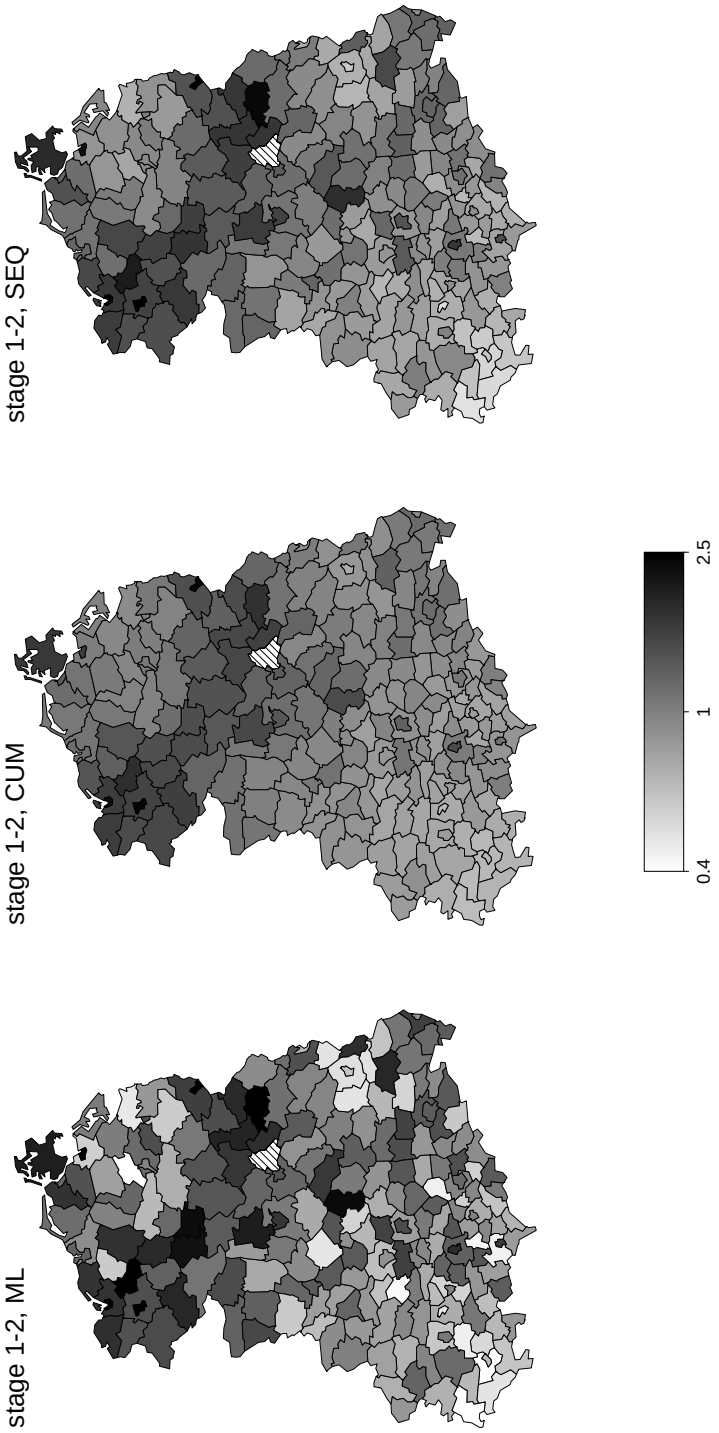

Figure 3: Relative risk estimates for diagnosis of the disease regardless of the stage by ML (left map), model 1 (middle map) and model 2 (right map).

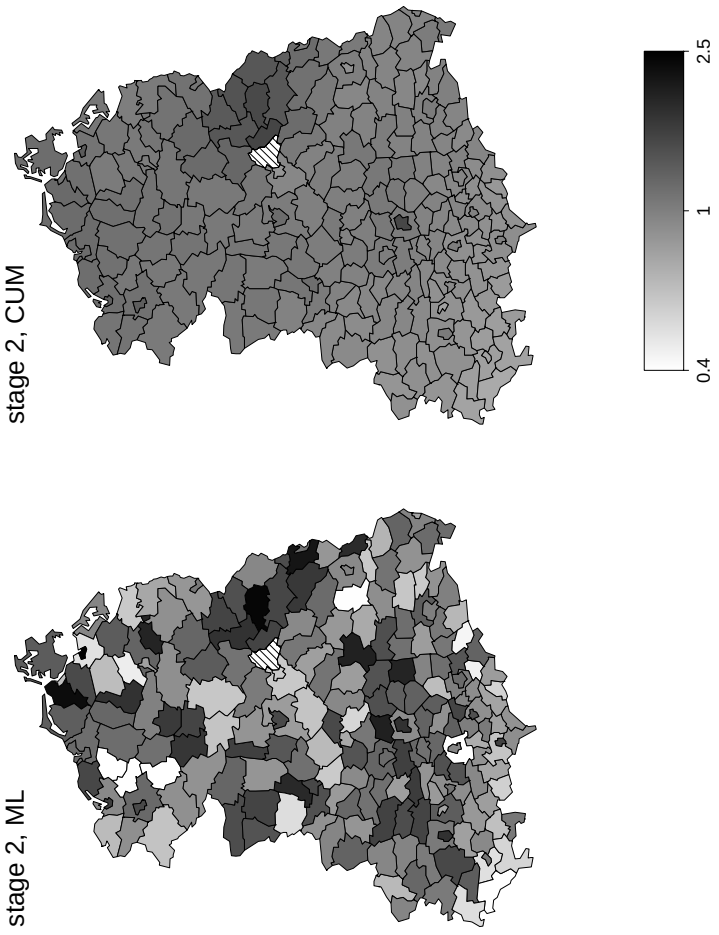

Figure 4: Relative risk estimates for diagnosis of the disease in stage 2 by ML (left map) and model 1 (right map).

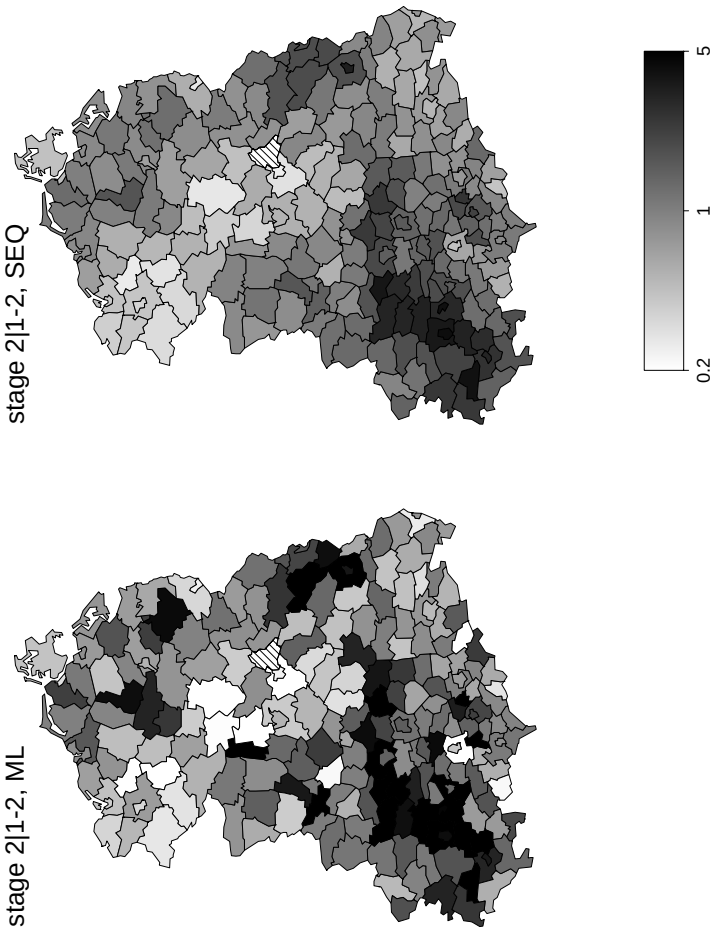

Figure 5: Conditional odds ratio estimates for diagnosis in the malignant stage 2, given diagnosis in stage 1 or 2 by ML (left map) and model 2 (right map).