Knorr-Held, Best:

Shared component models for detecting joint and selective clustering of two diseases

Shared component models for detecting joint and selective clustering of two diseases

Leonhard Knorr-Held

Institut für Statistik, Ludwig-Maximilians-Universität München,
Ludwigstr. 33, 80539 München, Germany.

and

Nicola G. Best

Department of Epidemiology and Public Health, Imperial College School of Medicine,
Norfolk Place, London W2 1PG, UK.

Abstract

The study of spatial variations in disease rates is a common epidemiological approach used to describe geographical clustering of disease and to generate hypotheses about the possible 'causes' which could explain apparent differences in risk. Recent statistical and computational developments have led to the use of realistically complex models to account for overdispersion and spatial correlation. However, these developments have focused almost exclusively on spatial modelling of a *single* disease. Many diseases share common risk factors (smoking being an obvious example) and if similar patterns of geographical variation of related diseases can be identified, this may provide more convincing evidence of real clustering in the underlying risk surface. In this paper, we propose *shared component models* for the joint spatial analysis of two diseases. The key idea is to identify shared and disease-specific spatially-varying latent risk factors by appropriate partitioning of the underlying risk surface for each disease. The various components of this partition are modelled simulataneously using nonparametric cluster models implemented via reversible jump Markov chain Monte Carlo methods. We illustrate the methodology through an analysis of oral and oesophageal cancer mortality in the 544 districts of Germany, 1986-1990.

Keywords: cluster models; joint disease mapping; latent variables; reversible jump Markov chain Monte Carlo; shared component models.

1 Introduction

The study of spatial variations in disease rates (disease mapping) is a classic epidemiological technique, where location is used as a surrogate for the mix of lifestyle, environmental and possibly genetic factors that may underly geographical differences in risk (Elliott and Best, 1998). The purpose is both to describe such variations and to generate hypotheses about the possible 'causes' which could explain them. The last decade has seen rapid development in the statistical and computational methods available to carry out such analyses, including the use of realistically complex models to account for overdispersion and spatial correlation, as well as to study the association between disease incidence and spatially varying covariates, such as deprivation, urbanization or environmental pollution (Besag et al., 1991; Clayton and Bernardinelli, 1992; Cressie, 1993; Best et al., 1998; Lawson and Clark, 1999; Langford et al., 1999; Knorr-Held and Raßer, 2000). These developments have focused almost exclusively on spatial modelling of a *single* disease. However, many diseases share common risk factors (smoking being an obvious example); if similar patterns of geographical variation of related diseases can be identified in a joint analysis, this may provide more convincing evidence of real clustering in the underlying risk surface than would be available from the analysis of a single disease. Some authors have suggested using incidence rates from other diseases as surrogate exposure measures, possibly allowing for measurement error, in a non-symmetric regression fashion (see, for example, Clayton et al., 1993; Bernardinelli et al., 1997). However a *joint* formulation which simultaneously models spatial variations in the risk of two or more related diseases seems a more natural and powerful design for detecting geographical patterns in the underlying risk surface.

In this paper, we propose so-called *shared component models* for the joint spatial analysis of two or more diseases. The key idea of the formulation is to separate the underlying risk surface for each disease into a shared component, common to both diseases, and a disease-specific component. Identifiability of the different components is achieved through spatial cluster models for each of them, similar to the ones proposed in Knorr-Held and Raßer (2000) for the spatial analysis of a single disease. The shared component can be interpreted as a surrogate for unobserved covariates that display spatial structure and are common to both diseases. Similarly, each disease-specific component represents only those spatially-varying risk factors which are specific to the respective disease.

This paper is organized as follows. Section 2 gives a short review of the spatial cluster model proposed by Knorr-Held and Raßer (2000), which forms the basis of the shared component models that are introduced in Section 3. The methodology will be illustrated in Section 4 through a joint analysis of oral and oesophageal cancer mortality in the 544 districts of Germany, 1986-1990. Section 5 contains a discussion.

2 Cluster models for disease mapping

We now review a model described in detail by Knorr-Held and Raßer (2000) for the detection of clusters and discontinuities in disease maps. The approach taken differs from the Markov random fields which are widely used to model spatial correlation in disease maps (Besag et al., 1991) in that the underlying risk surface is modelled nonparametrically using a spatial mixture distribution with an unknown number of components. Technically our method is based on reversible jump MCMC (Green, 1995), an extension of the Metropolis-Hastings algorithm to simulate from distributions of variable dimension.

Let n be the number of districts in the study region and y_i and e_i, $i = 1, \dots, n$, denote the observed and expected number of cases respectively. The basic idea is to assume that the study region can be partitioned into k clusters $C_j \subset \{1, \dots, n\}$, $j = 1, \dots, k$, i.e. sets of contiguous regions, where each cluster has constant relative risk λ_j. The clusters cover the whole study region, i.e. $\cup_{j=1}^{k} C_j = \{1, \dots, n\}$, and do not overlap. Disease counts y_i, $i \in C_j$, $j = 1, \dots, k$, are assumed to be conditionally independent Poisson random variables with mean $e_i \cdot \lambda_j$ so the likelihood function can be written as

$$L(y|\lambda) = \prod_{j=1}^{k} \prod_{i \in C_j} \frac{(e_i \lambda_j)^{y_i}}{y_i!} \exp(-e_i \lambda_j).$$

For a detailed discussion of the assumptions and properties underlying the Poisson model see Wakefield et al. (2000). We note that an alternative way of formalizing the partitioning of the study region into clusters is through a function C, which maps each region i into the relevant cluster j if and only if $i \in C_j$. The formulation can now simply be written as conditional independent Poisson responses with mean $e_i \cdot \lambda_{C(i)}$. With a slight misuse of notation we will sometimes denote $\lambda_{C(i)}$ simply by λ_i. A similar notation will be used in Section 3.

The number k, and shape and size of the clusters, as well as the risk of disease within each cluster are now treated as unknown variables. As the dimension of the problem depends on the number of clusters, reversible jump MCMC is the appropriate inference technique. It is important to understand that the final risk estimates are not based on a specific cluster configuration but are an *average* over a large number of cluster configurations, weighted by the corresponding posterior probabilities. These estimates thus incorporate all the uncertainty about the number, location and level of risk of the clusters. The formulation can therefore be considered as nonparametric (Arjas, 1996; Heikkinen and Arjas, 1998).

To define the clusters, Knorr-Held and Raßer propose a construction where k regions are marked as so-called cluster centres, each of them defining a separate cluster. Each of the remaining regions is assigned to the cluster which is closest in terms of the minimal number of boundaries that have to be crossed to move from that region to the cluster centre. The cluster centres are collected in a vector, say G_k. Regions, which have the same "distance" (in terms of the minimal number of boundaries to be crossed) to two or more cluster centres are assigned to the center with the smallest index position in G_k. Hence G_k is kept unordered to avoid any unjustifiable preference for centers with smaller indices.

The model now assumes a truncated geometric prior on $\{1, \ldots, n\}$ for k, i.e. $Pr(k) \propto (1-c)^k$, $k = 1, \ldots, n$, where $c \in [0, 1)$ is a suitably chosen constant. Conditional on the number of clusters, a uniform prior is specified for the $(n-k)!/n!$ possible choices for the (unordered) vector of cluster centres $G_k|k$. Finally, the logarithms of the relative risk parameters $\log \lambda_j$, $j = 1, \ldots, k$, are assumed to be independent realizations from a normal distribution with unknown mean μ and variance σ^2. A flat prior is chosen for μ, i.e. uniform on the whole real line; an inverse gamma prior with fixed parameters a and b is adopted for σ^2. Hence the prior mode for σ^2 is at $b/(a+1)$.

Due to independence of the risk parameters λ_j, the model is able to detect spatial discontinuities if adjacent districts are assigned to different clusters. In particular, clusters of size one are not excluded from the model, which implies that the formulation does not necessarily smooth the data. In practice, some smoothing is inevitable, since there will always be some uncertainty about whether or not a region forms a cluster by itself. However, the sizes of the clusters, which determine the local degree of smoothing, are variable, hence the smoothing is adaptive. This is in contrast to Markov random field models, where the smoothing parameter is constant across the study region and so smoothing is non-adaptive.

3 Shared component models

We now consider joint modelling of two diseases with disease counts y_{1i} and y_{2i}, $i = 1, \ldots, n$, for diseases 1 and 2 respectively. Similarly, expected counts are denoted by e_{1i} and e_{2i}. We assume that the overall relative risk $\sum_i y_{di} / \sum_i e_{di}$, $d = 1, 2$, is the same for both diseases. We first outline our proposed model formulation and then present some motivating arguments.

3.1 Model notation and formulation

For each disease, the relative risk in district i is modelled as the product of a shared component λ_i and a disease-specific component $\phi_{di}, d = 1, 2$. Responses y_{1i} and y_{2i} are thus assumed to be conditionally independent Poisson random variables

$$y_{1i} \sim Po(e_{1i} \cdot \lambda_i^{\delta} \cdot \phi_{1i}) \quad \text{and} \quad y_{2i} \sim Po(e_{2i} \cdot \lambda_i^{1/\delta} \cdot \phi_{2i}). \tag{1}$$

The contribution of the shared component to the overall relative risk is weighted by the scaling parameter δ to allow a different "risk gradient" (on the log scale) to be associated with this component for each disease. The three components λ_i, ϕ_{1i} and ϕ_{2i} are assumed to be independent, with each one following a cluster model as described in Section 2.

In the limiting case of a shared component without any spatial structure, that is $\lambda_i =$ constant for all i, y_{1i} and y_{2i} will be independent Poisson with the spatial variation in relative risk determined only through the specific components ϕ_1 and ϕ_2, respectively. On the other hand, if both specific components are constant, both diseases will have a common relative risk pattern determined through the shared component λ. The overall relative risk level will be the same for both diseases but the magnitude of the area-specific relative risks may differ — hence the need for the scaling parameteter δ. In practice, it may sometimes be the case that one or two of the components λ, ϕ_1 and ϕ_2 dominate the risk surfaces of the two diseases, but which one will not usually be known in advance. The general formulation (1) is a flexible modelling framework which includes the whole range between those extreme cases and lets the data decide about the strength of each component.

Note that the number of different levels in the risk surface of each disease, determined through two overlaid cluster models, is much larger than the number of clusters in each model. We therefore penalize large values of the number of components k_λ, k_{ϕ_1} and k_{ϕ_2} heavily a priori and take larger values for c_λ, c_{ϕ_1} and c_{ϕ_2} than in an analysis of a single disease.

As in Section 2, a flat prior will be used for the parameter μ_λ of the lognormal prior on λ. However, for identifiability reasons, we fix the means of the log relative risk parameters of the two specific components to zero. Inverse gamma priors are assumed for the variances on the three sets of log relative risks as before. Finally, we assume that the logarithm of the scaling parameter δ has a normal prior with mean zero and variance τ^2. Since the prior for δ is symmetric around zero on a log-scale, any value δ_0 is as "equally likely" as the reciprocal value $1/\delta_0$ a priori. More precisely,

$$P(\delta_l \leq \delta \leq \delta_u) = P(1/\delta_u \leq 1/\delta \leq 1/\delta_l)$$

for any positive values $\delta_l < \delta_u$. Consequently, the formulation (1), where both δ and the reciprocal value $1/\delta$ enter, has an attractive invariance feature: if we switch the indices of the two diseases, we will get exactly the same posterior distribution for the joint and specific components and the posterior for δ will change to the reciprocal distribution. Therefore, the posterior distribution of the relative risk for each disease will be exactly the same.

3.2 Some motivation for shared component models

We now provide both formal and heuristic arguments for our model. Consider first the (unrealistic) case where there is only one “true” but unobserved (continuous) covariate Z_i, common to both diseases, and assume that the relationship to the relative risk is of the usual log-linear type. Then the true log relative risks η_{1i} and η_{2i} for diseases 1 and 2 are given by

$$\begin{aligned} \eta_{1i} &= \alpha_1 + \beta_1 \cdot Z_i \qquad (2) \\ \eta_{2i} &= \alpha_2 + \beta_2 \cdot Z_i, \qquad (3) \end{aligned}$$

where β_1 and β_2 are the different risk gradients associated with the covariate for the two diseases. (Note that, for technical reasons, we assume that β_1 and β_2 have the same sign.) Now suppose we specify the following model for the log relative risk

$$\begin{aligned} \eta_{1i} &= \log \lambda_i \cdot \delta \qquad (4) \\ \eta_{2i} &= \log \lambda_i / \delta. \qquad (5) \end{aligned}$$

where the $\log \lambda_i$ follow a cluster model as described above, with marginal mean and variance denoted by μ_λ and σ_λ^2 respectively. This is simply a special case of our shared component model (1) without the disease-specific components ϕ_1 and ϕ_2. If we now assume that the distribution of the true covariate Z_i across the study region has some arbitrary (possibly spatially correlated) form with marginal mean μ_z and variance σ_z^2, then it is straightforward (Wakefield et al., 2000) to show that

$$\begin{aligned} (\log \lambda_i - \mu_\lambda) \cdot \delta &= \beta_1 \cdot (Z_i - \mu_z) \qquad (6) \\ (\log \lambda_i - \mu_\lambda) / \delta &= \beta_2 \cdot (Z_i - \mu_z). \qquad (7) \end{aligned}$$

Dividing Eqn. (6) by Eqn. (7) gives $\delta^2 = \beta_1/\beta_2$, i.e. the squared scaling parameter can be interpreted as the ratio of the two risk gradients. Similar arguments lead to the identities $\mu_\lambda^2 = (\alpha_1 + \beta_1 \mu_z) \cdot (\alpha_2 + \beta_2 \mu_z)$ and $\sigma_\lambda^2 = \beta_1 \cdot \beta_2 \cdot \sigma_z^2$.

The above equations have been derived under the assumption of only one true covariate, common to both diseases, whose between-area distribution exhibits spatial variation that may be modelled by the nonparametric cluster model described in Section 2. In a more general scenario, the "true" model might look like

$$\begin{aligned}\eta_{1i} &= \alpha_1 + \beta_1 \cdot Z_i + \gamma_1 \cdot V_i \\ \eta_{2i} &= \alpha_2 + \beta_2 \cdot Z_i + \gamma_2 \cdot W_i,\end{aligned}$$

where Z_i is a shared risk factor as before, and V_i and W_i are disease-specific risk factors relevant to one or other of the diseases only. The idea of our general formulation Eqn. (1) is that the distributions of the true exposures Z, V and W will each display different patterns of spatial variation across the study region, so that the cluster model for the shared component λ will capture the spatial distribution of the Z_i's (possibly shifted and on a different scale), while the cluster models for ϕ_1 and ϕ_2 will account for the underlying distributions of the V_i's and W_i's respectively. Of course, there is an apparent lack of identifiability in the model since the spatial risk surface of one disease can, in principle, be represented by either the joint or the specific component. However, in all applications we have tried so far, the estimated spatial patterns of the three of components are rather distinct and there was no sign of severe confounding, as long as we penalize larger values of k_λ, k_{ϕ_1} and k_{ϕ_1} a priori. Values of k_λ are typically much larger than those of k_{ϕ_1} and k_{ϕ_2} in the posterior. Hence it seems that the shared component captures finer differences in the risk surface with larger number of components k_λ, probably because both datasets contribute directly to the likelihood. The disease-specific components have a coarser resolution with much smaller number of components k_{ϕ_1} and k_{ϕ_2}, as there is less direct information in the likelihood.

3.3 Implementation

The method has been implemented with reversible jump MCMC moves similar to those used in Knorr-Held and Raßer (2000) and Giudici et al. (2000) for each spatial component. Proposals for each accept-reject step have been constructed in order to achieve high acceptance rates and to avoid the need for tuning parameters. The only additional parameter is the scaling parameter δ for which a simple Metropolis update step has been used.

4 Application to German cancer mortality data

We now describe an application of our shared component models to the joint analysis of both oral cavity and oesophageal cancer mortality among males in Germany. These two cancer sites are closely related anatomically and are often studied as a single diagnostic group (e.g. Kjaerheim et al. (1998); Gronbaek et al. (1998)). Tobacco smoking and alcohol abuse are the two most important established risk factors for both diseases (Blot et al., 1994; Schottenfeld and Fraumeni, 1996). However, there may also be differences in the pathogenesis of oral cavity and oesophageal cancers and some have argued that it is misleading to treat these and other cancers of the upper digestive tract as a single group (Fitzgerald and Caygill, 1999). The shared component models proposed here allow us to exploit the aetiologicial similarities between the two diseases, yet still identify any differences in their respective patterns of risk.

Our analysis considers mortality from oral cavity and oesophageal cancer in males in the 544 districts in Germany, 1986-1990. The two datasets have been internally standardized by maximum likelihood under the restriction that the sum of observed cases equals the sum of expected cases for each of the diseases. Fig. 1 displays the standard mortality ratios ($\text{SMR}_i = y_i/e_i$) for each cancer site, while Fig. 2 shows the relative risk estimates obtained from separate applications of the Knorr-Held and Raßer cluster model (with $a = 1$, $b = 0.01$ and $c = 0.02$) to each dataset. The spatial structure of the estimates resemble each other quite closely with high values in the north-east and in large parts of the south-west. Hence a joint analysis of the two diseases seems appropriate. Note that a different scale with a wider range has been used for the maps in Fig. 1, to accomodate the more extreme SMRs. The change in scale for the relative risk maps shown in Fig. 2 and the remaining figures in this section was chosen in order to show more clearly the different spatial patterns identified by the separate cluster models and by the various components of the joint cluster models. For oral cavity cancer, maps showing the SMRs and cluster model relative risk estimates on the same scale, together with a more detailed discussion of the results, can be found in Knorr-Held and Raßer (2000).

For the joint analysis, we have chosen $c_\lambda = 0.1$ and $c_{\phi_1} = c_{\phi_2} = 0.2$ and all other values as above. These choices have been made in order to approximately match those made in the separate analyses with a sufficient amount of penalization. Finally, the parameter τ^2 (the variance of the log scaling parameter) has been set to 0.17, which corresponds to a prior belief that the ratio of relative risks associated with the shared latent covariates for each disease (*i.e.* β_1/β_2) is between $1/5$ and 5 with 95% probability.

Fig. 3 displays the estimated shared component λ. Most striking are two large clusters, one in the north–east in Mecklenburg–West Pomerania and one in the south–west covering the whole of Saarland and parts of Rhineland–Palatinate and Baden–Württemberg along the border to France. Both clusters are consistent with what is known about the distribution of established risk factors in Germany, since they coincide with regions where alcohol (Mecklenburg–West Pomerania) or smoking (Saarland, Rhineland–Palatinate) consumption are known to be high, e.g. Becker and Wahrendorf (1997). Also, many urbanized areas can be identified in Fig. 3, especially in the north with higher estimates of the shared component in Bremen, Hamburg, Kiel, Berlin and parts of the Ruhr area.

Prior and posterior distribution of the scaling parameter δ can be seen in Figure 4. The posterior median is estimated as 1.10 with a 90% credible interval of (0.99,1.25). Hence the effect of the shared component on cancer of the oral cavity (median estimates of λ_i^δ in the range of 0.70 to 1.29) is slightly larger than on cancer of the oesophagus (median estimates of $\lambda_i^{1/\delta}$ in the range of 0.75 to 1.23). This indicates that the unobserved risk factors common to both diseases are associated with a slightly higher risk of oral cavity cancer than oesophageal cancer. If one assumes that the joint component mainly reflects spatial variation in alcohol and tobacco consumption, than our findings are in accordance with Baron et al. (1993), who find that the combined alcohol and smoking risks for oral cancer are significantly greater than those for oesophageal cancer in a unified analysis of data from a case-control study.

Fig. 5 displays the two disease-specific components ϕ_1 and ϕ_2. Interestingly, the specific component for oral cavity cancer has a distinct spatial pattern with higher values (around 1.15) in the south and lower values (below 1.0) in the north. This indicates the existence of additional risk factors relevant only to oral but not to oesophageal cancer. A possible explanation might be the higher consumption or different preferences for alcohol products in the south, where all the main wine-growing areas are. Of course, this is rather speculative but supported by Leclerc et al. (1987), who analyse case-control data on cancer of the upper respiratory and digestive tract, and find more wine consumers among mouth cancer cases than cancer on other locations, controlling for the total amount of alcohol consumed. The oesophageal specific component shows a different spatial pattern with less variation and slightly higher values in the West and North of Germany. Note that there is a “negative” cluster in the south of former East Germany in *both* specific components, a bit more pronounced for oesophageal cancer. This is somewhat surprising as one would think that such a common cluster would be captured mainly by the shared component. The shared component also has low estimates in the same area but on a finer resolution (actually there seem to be two distinct clusters). This may be indicative of artefacts in the data due to different data collection procedures in the former East Germany (in particular, in the ICD coding process) possibly resulting in apparent under-ascertainment of mortality from certain cancers relative to West Germany.

We have therefore studied sensitivity of the results to the choice of hyperparameters, especially to different values for c_λ. As one would expect, for $c_\lambda = 0.2$, the shared component captures more of the joint cluster in the south of East Germany, while the specific components have estimates closer to one. The other noticeable difference to the analysis with $c_\lambda = 0.1$ is that the shared component now seems to link the two clusters in West Berlin and Mecklenburg–West Pomerania to a larger one. Apart from these two changes, however, there is surprisingly little sensitivity in the estimates of the shared and specific components and the relative risk surfaces.

Finally Fig. 6 displays the overall posterior median relative risk surface for each disease. The estimates are quite similar to those obtained by separate analyses (Fig. 2), although they show some differences, especially in more sparsely populated areas. The differences are caused by "borrowing strength" from the spatial pattern of the other disease through the shared component. For example, in the joint analysis estimates for oesophageal cancer are slightly lower in east Brandenburg (south of Mecklenburg–West Pomerania) on the border to Poland.

5 Discussion

The shared component models proposed in this paper offer a straightforward approach to the joint spatial analysis of two related diseases. They represent a natural extension of the cluster models proposed by Knorr-Held and Raßer (2000), and have the attractive feature that the various components have a direct interpretation in terms of latent covariates which are either shared by both diseases or are selectively associated with only one or other of the outcomes. The example illustrates some of the advantages to be gained by simulataneous analysis of the spatial variation in two diseases. For example, not only were we able to clearly identify joint clusters of oral cavity and oesophageal cancers associated with common risk factors (alcohol and smoking consumption), but our analysis also revealed a north-south trend specific to oral cavity cancer which was less apparent from the separate analysis of this disease alone. Another interesting application would be the joint analysis of a single disease, split by gender.

Our model makes a number of strong assumptions and so some cautionary comments must be made in order to avoid overinterpretation of the results. In particular, we assume that the net effect of all shared covariates may be modelled by a single latent variable (λ) with a different risk gradient for each disease. This may be reasonable when there is one dominant, common risk factor such as smoking or a genetic effect, but may be less satisfactory if there are a number of shared risk factors, each with different risk gradient ratios for the two diseases. Likewise, we assume that the spatial structure of the disease-specific covariates may be adequately captured by a single latent variable for each disease. Our models also assume that the shared and specific components are independent, which ignores the possibility of interaction between the "true" covariates. Nevertheless, we think that these models provide a useful approximation to the underlying risk surface and may help gain further insight into the true pattern of exposures relevant to each disease.

5.1 Extensions and future work

We note that the shared component models proposed here are easily extended to the joint analysis of three or more diseases, although the number of possible permutations of shared and specific components may rapidly become prohibitive. Observed area-level covariates may also be combined with shared component cluster models in the same way as for a single disease (Giudici et al., 2000). Another variation on the above models would be to assume a Markov random field prior rather than the Knorr-Held and Raßer cluster model for each component.

Mollié (1990) adopts an alternative approach to the joint analysis of areal counts of two diseases, based on a bivariate normal prior for the corresponding log relative risk parameters in each region. However, this model ignores possible spatial correlation in the relative risks across regions. Recently, Assunção et al. (1999) use multivariate Markov random field models in a linear regression context with spatially varying coefficents. We are currently working on a comparison of shared component models versus multivariate spatial models for joint disease mapping.

Acknowledgements

The first author was supported by the German Science Foundation (DFG), SFB 386. The second author was partially supported by a Medical Research Council Career Establishment Grant (G9803841). The authors express thanks to Nikolaus Becker, German cancer research center, for access to the data, and to Annibale Biggeri for helpful epidemiological comments and references.

References

Arjas, E. (1996), "Discussion of paper by Hartigan," in "Bayesian Statistics 5," eds. J. M. Bernardo, J. O. Berger, A. P.David, and A. F. M. Smith, Oxford: Oxford University Press, pp. 221–222.

Assunção, J. J., Gamerman, D., and Assunção, R. M. (1999), "Regional differences in factor productivities of Brazilian agriculture: a space varying parameter approach," Technical report, Statistical Laboratory, Universidade Federal do Rio de Janeiro.

Baron, A. E., Franceschi, S., Barra, S., Talamini, R., and La Vecchia, C. (1993), "A comparison of the joint effects of alcohol and smoking on the risk of cancer across sites in the upper aerodigestive tract," *Cancer Epidemiol Biomarkers Prev*, 2(6), 519–523.

Becker, N. and Wahrendorf, J. (1997), *Atlas of Cancer Mortality in the Federal Republic of Germany 1981-1990*, Berlin: Springer Verlag.

Bernardinelli, L., Pascutto, C., Best, N. G., and Gilks, W. R. (1997), "Disease mapping with errors in covariates," *Statistics in Medicine*, 16, 741–752.

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

Best, N. G., Ickstadt, K., and Wolpert, R. L. (1998), "Spatial Poisson regression for health and exposure data measured at disparate resolutions," Discussion Paper 98-36, Duke University ISDS, USA.

Blot, W. J., Devesa, S. S., McLaughlin, J. K., and Fraumeni, J. F. (1994), "Oral and pharyngeal cancers," in "Cancer Surveys: Trends in Cancer Incidence and Mortality, Vol. 19/20," eds. R. Doll, J. F. Fraumeni, and C. S. Muir, New York: Cold Spring Harbor Laboratory Press, pp. 23–42.

Clayton, D. G. and Bernardinelli, L. (1992), "Bayesian methods for mapping disease risk," in "Geographical and Environmental Epidemiology: Methods for Small-Area Studies," eds. P. Elliott, J. Cuzick, D. English, and R. Stern, Oxford: Oxford University Press, chapter 18, pp. 205–20.

Clayton, D. G., Bernardinelli, L., and Montomoli, C. (1993), "Spatial correlation in ecological analysis," *International Journal of Epidemiology*, 22, 1193–1202.

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, New York, NY, USA: John Wiley & Sons.

Elliott, P. and Best, N. G. (1998), "Geographical patterns of disease," in "Encyclopaedia of Biostatistics," eds. P. Armitage and T. Colton, London: J. Wiley & Sons.

Fitzgerald, R. and Caygill, C. (1999), "Treating upper digestive tract cancers as a single entity may be misleading," *British Medical Journal*, 318, 1289.

Giudici, P., Knorr-Held, L., and Raßer, G. (2000), "Modelling categorical covariates in Bayesian disease mapping by partition structures," *Statistics in Medicine*, (to appear).

Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

Gronbaek, M., Becker, U., Johansen, D., Tonnesen, H., Jensen, G., and Sorensen, T. I. (1998), "Population based cohort study of the association between alcohol intake cancer of the upper digestive tract," *British Medical Journal*, 317, 844–7.

Heikkinen, J. and Arjas, E. (1998), "Nonparametric Bayesian estimation of a spatial Poisson intensity," *Scandinavian Journal of Statistics*, 25, 435–50.

Kjaerheim, K., Gaard, M., and Andersen, A. (1998), "The role of alcohol, tobacco and dietary factors in upper aerogastric tract cancers: a prospective study of 10,900 Norweigan men," *Cancer Causes Control*, 9(1), 99–108.

Knorr-Held, L. and Raßer, G. (2000), "Bayesian detection of clusters and discontinuities in disease maps," *Biometrics*, (to appear).

Langford, I. H., Leyland, A. H., Rasbash, J., and Goldstein, H. (1999), "Multilevel modelling of the geographical distributions of disease," *Applied Statistics*, 48, 253–68.

Lawson, A. B. and Clark, A. (1999), "Markov chain Monte Carlo methods for putative sources of hazard and general clustering," in "Advanced Methods of Disease Mapping and Risk Assesment for Public Health Decision Making," eds. A. B. Lawson, D. Boehning, E. Lessafre, A. Biggeri, J. Viel, and R. Bertollini, Chichester, UK: John Wiley & Sons, pp. 119–42.

Leclerc, A., Brugere, J., Luce, D., Point, D., and Guenel, P. (1987), "Type of alcoholic beverage and cancer of the upper respiratory and digestive tract," *Eur J Cancer Clin Oncol*, 23(5), 529–534.

Mollié, A. (1990), *Representation Geographique des Taux de Mortalire: Modelisation Spatiale et Methodes Bayesiennes*, Ph.D. thesis, Universite Paris 6.

Schottenfeld, D. and Fraumeni, J. F. J., eds. (1996), *Cancer Epidemiology and Prevention*, New York, NY, USA: Oxford University Press, 2nd Edition.

Wakefield, J. C., Best, N. G., and Waller, L. A. (2000), "Bayesian Approaches to Disease Mapping," in "Spatial Epidemiology: Methods and Applications," eds. P. Elliott, J. C. Wakefield, N. G. Best, and D. J. Briggs, Oxford: Oxford University Press, (in press).

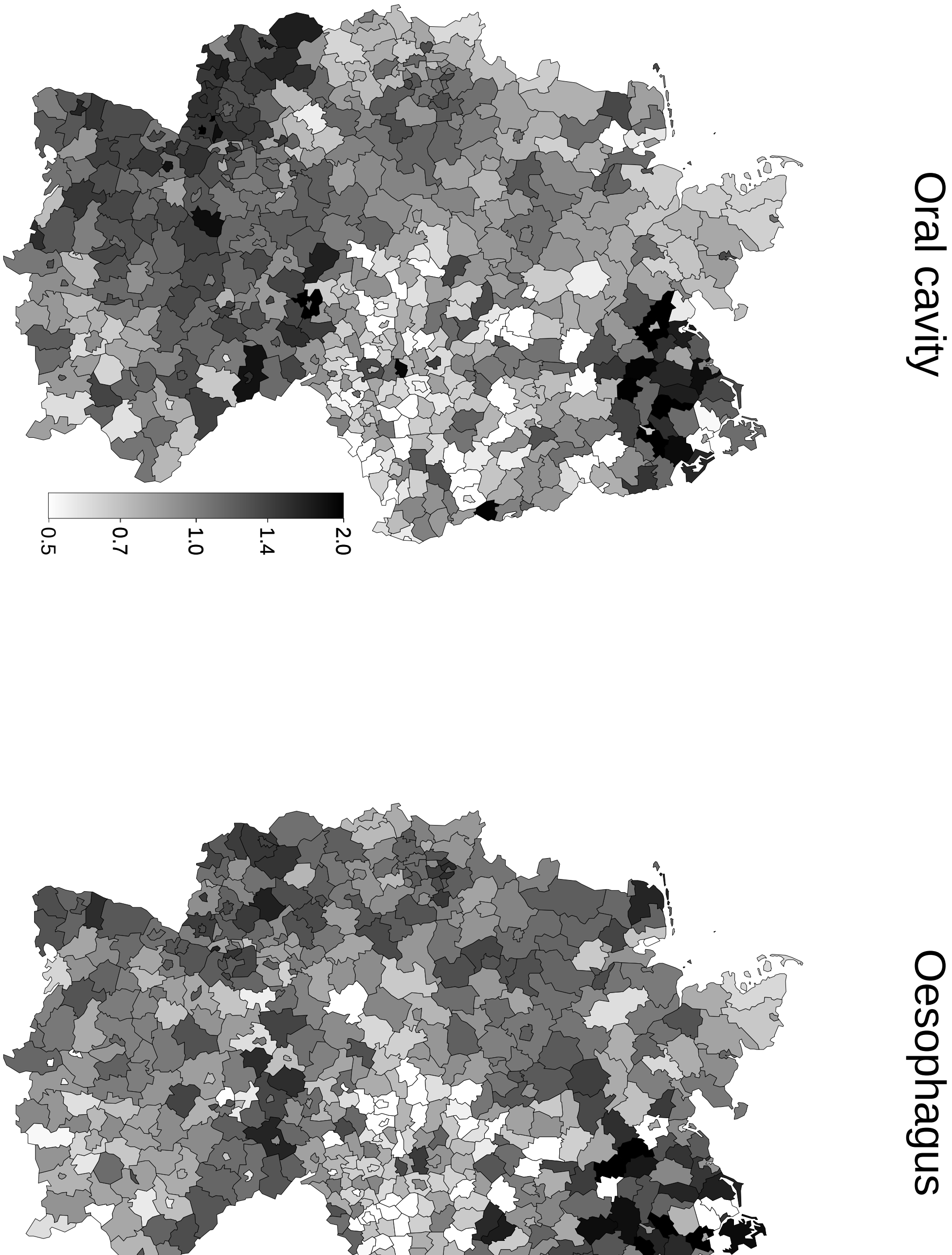

Figure 1: Standard mortality ratios for oral cavity and oesophageal cancer for males in Germany (note that these maps are shown on a different scale to the other maps in this section).

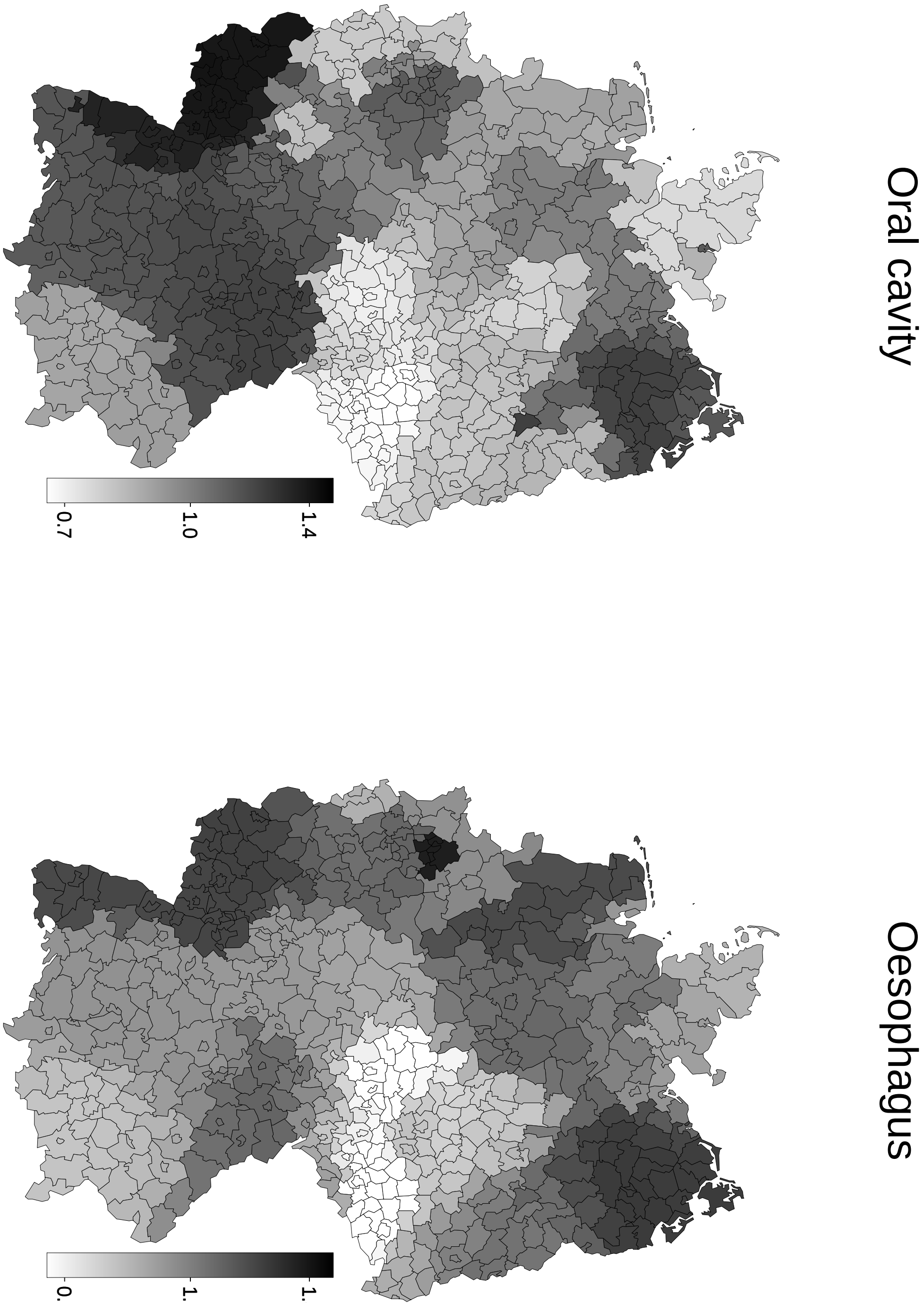

Figure 2: Posterior median relative risks for oral cavity and oesophageal cancer for males in Germany, estimated using separate cluster models for each disease.

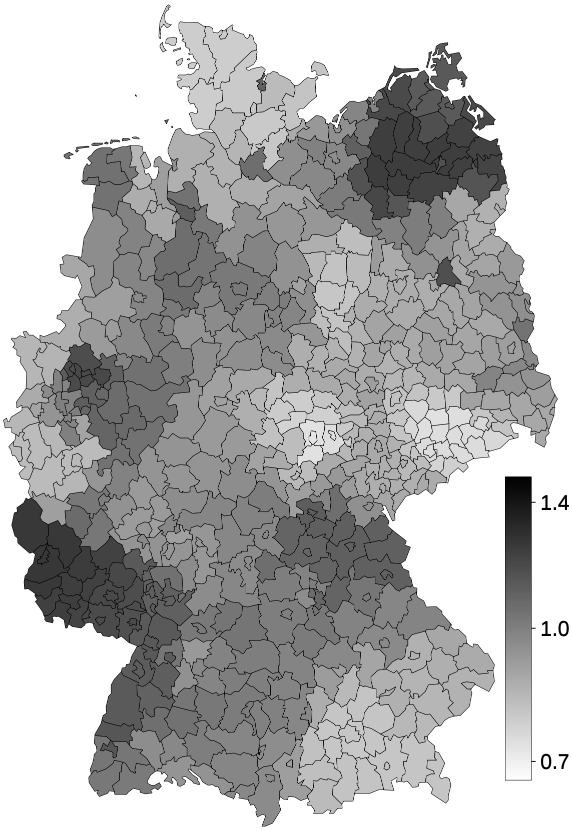

Figure 3: Estimated posterior median for the shared component λ in the joint cluster model.

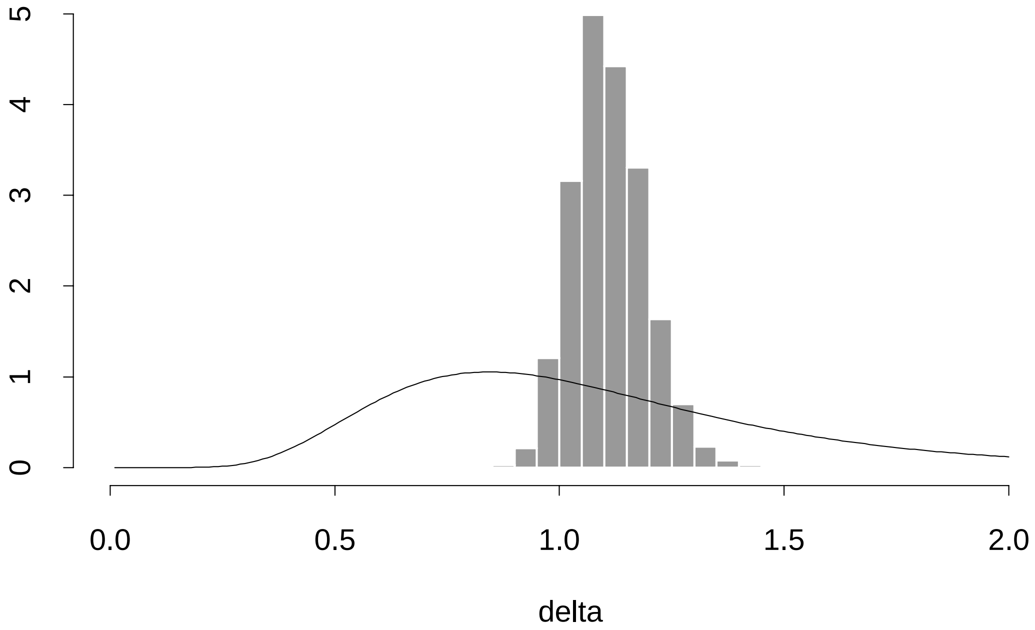

Figure 4: Prior (solid line) and posterior (histogram) distribution of the scaling parameter δ.

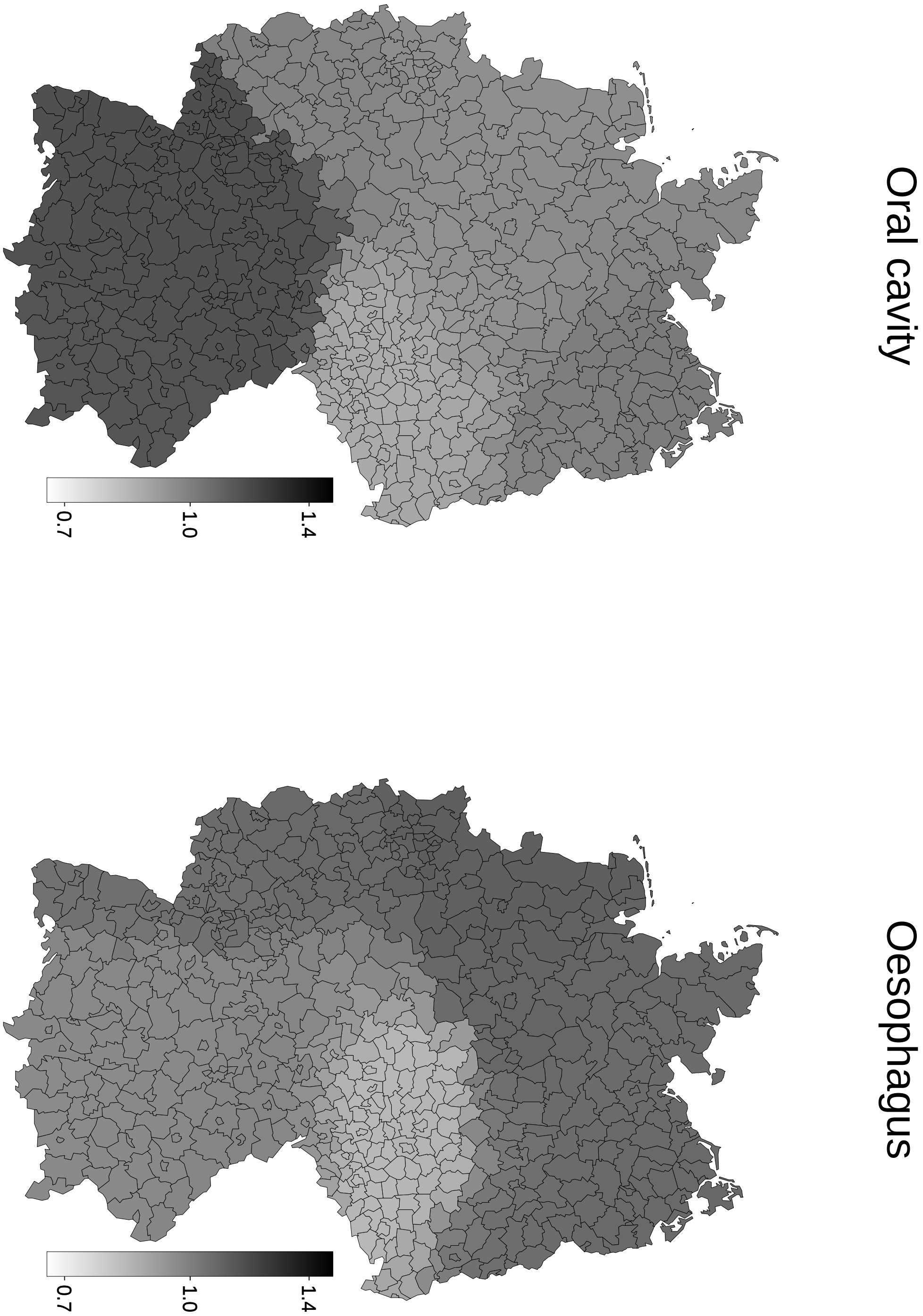

Figure 5: Estimated posterior median for the disease specific components ϕ_1 (oral cavity cancer) and ϕ_2 (oesophageal cancer).

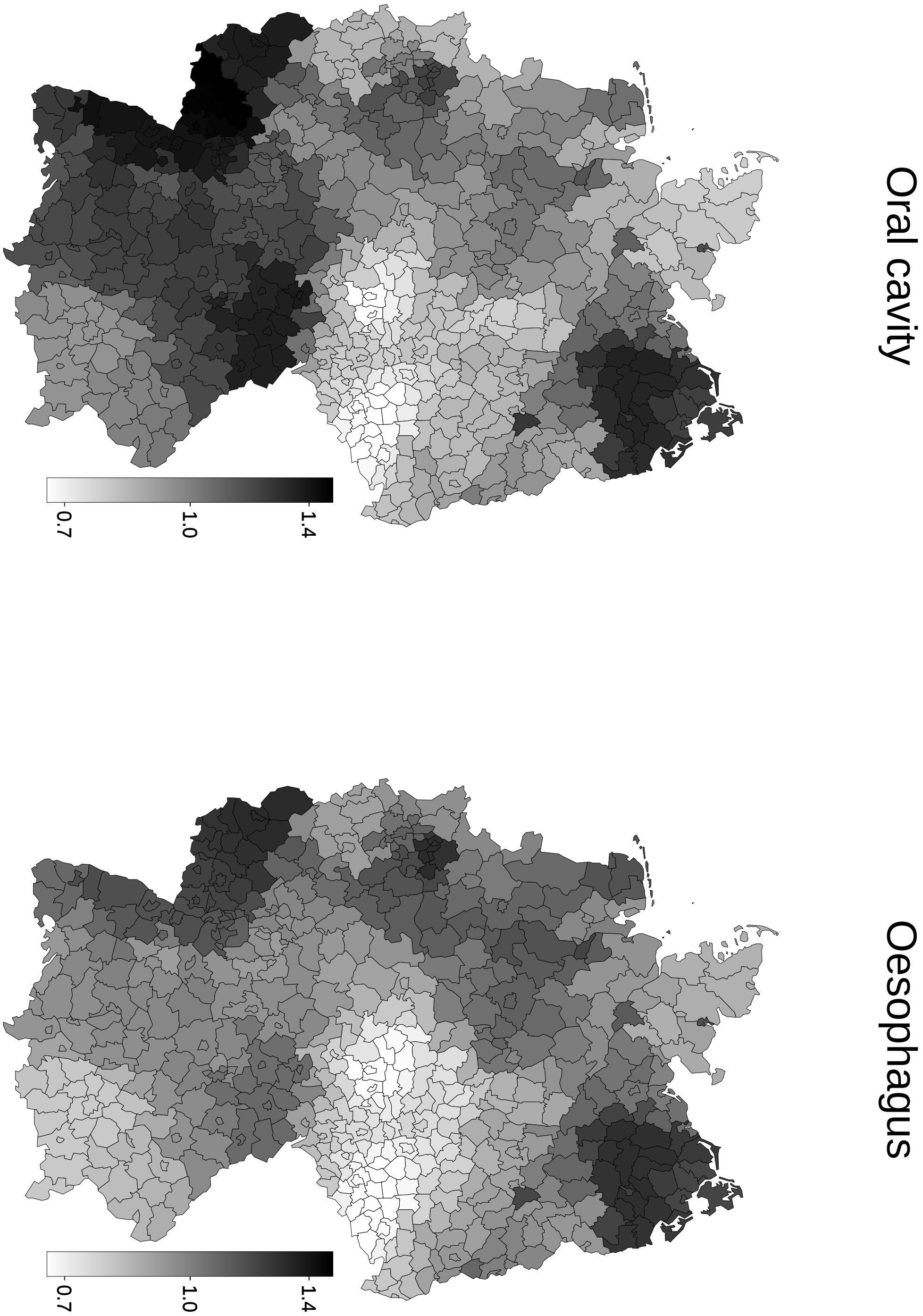

Figure 6: Posterior median relative risks for oral cavity and oesophageal cancer for males in Germany, estimated using the joint cluster model.