



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Knorr-Held, Becker:

Bayesian Modelling of Spatial Heterogeneity in Disease Maps with Application to German Cancer Mortality Data

Sonderforschungsbereich 386, Paper 176 (1999)

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

Projektpartner



Bayesian Modelling of Spatial Heterogeneity in Disease Maps with Application to German Cancer Mortality Data

Leonhard Knorr-Held

Institut für Statistik

Universität München

Ludwigstr. 33, 80539 München

Nikolaus Becker

Deutsches Krebsforschungszentrum

Im Neuenheimer Feld 280

69120 Heidelberg

November 24, 1999

Abstract

This paper starts with a short overview of basic concepts in disease mapping such as relative risk and age-standardization. Then two recent methods for advanced statistical analysis of areal summary measures of health outcomes are reviewed which overcome difficulties in traditional mapping methods. Both methods account for spatial correlation in an hierarchical Bayesian framework and use computer-intensive Markov chain Monte Carlo methods for statistical inference. The methods are compared through analyses of cancer mortality data from Germany, 1986-1990.

1 Introduction

Recently the interest in spatial epidemiology has moved from large to small scale variation of health statistics. This process has been caused by the increasing availability of geographically referenced health and population data, advances in computing power and suitable mapping methods, and, last but not least, public interest in effects of environmental pollution. From

a statistical perspective, this development contains a number of challenges and problems. Statistical methods for the analysis of such data have to account for spatial dependencies in the data, as well as for the fact that rare disease events and the corresponding small area health statistics have inherent large variability.

This paper is concerned with the topic of *disease mapping*, an important subtopic in spatial epidemiology. Here the location of the disease cases is not exactly known, only the number of cases within areal units are given. Also no attempt is made to study the correlation of the risk variation with ecological covariates such as deprivation, urbanization or environmental factors. Disease mapping is carried out to summarize spatial variation in risk. This is typically done in order to generate hypothesis, to provide a context to place individual studies and to highlight clusters of elevated or lowered risk.

The general goal could be described as to identify the extra-sample variation due to unobserved heterogeneity by filtering out the sample variation. For this purpose it is often necessary to adopt models that account for spatial correlation in the data. Bayesian approaches in combination with recent advances in statistical computing have proven to be extremely useful for this purpose. In particular the development of Markov chain Monte Carlo (MCMC) methods in the last decade made it possible to adopt realistically complex model formulations without strong simplifying assumptions. Improved estimates of disease risk and of the uncertainty associated with it can now be calculated quite easily and are not based on asymptotic assumptions.

This paper applies two competing Bayesian models to cancer mortality data from Germany for the period 1986-1990. It will be argued that the resulting estimates from both models might be superior to more common risk estimates such as standardized mortality ratios. Differences between the two methods turn out to be more subtle but still of general interest. In Section 2 we start reviewing the usual binomial and Poisson model setup, com-

mon to most disease mapping approaches, and discuss the process of age-standardization. Section 3 describes the two competing Bayesian models for disease mapping, Markov random field models and cluster models. Both methods are applied in Section 4 to German mortality data for brain and lung cancer of males, 1986-1990. Similarities and differences between the methods are pointed out. We also compare the results with those obtained from a simpler empirical Bayes method which does not incorporate the spatial structure in the data. The last section contains concluding remarks.

2 Relative risk and age-standardization

In its simplest form the data for studies of the geographical variation of disease risk consists of pairs (y_i, n_i) , where n_i is the number of persons under risk and y_i is the number of cases in area $i = 1, \dots, I$. The usual starting point is now to assume that y_i is binomial with parameters n_i and an unknown risk probability π_i . Responses y_1, y_2, \dots, y_I are assumed to be independent. Note that in Bayesian models, where prior distributions are placed on $\pi = (\pi_1, \pi_2, \dots, \pi_I)$, responses are assumed to be *conditionally* independent, given π . Such a two-stage formulation is the simplest instance of an hierarchical model.

As a side comment we remark here that the binomial formulation does not require the risk for each individual in area i to be equal π_i . In fact, π_i can be seen as a weighted average of risk probabilities defined on a finer grid (Knorr-Held and Besag, 1998). For a detailed discussion of this issue in connection with the related problem of overdispersion see Wakefield, Best and Waller (1999).

Under the “rare disease assumption” (n_i large, π_i small), the binomial model can be approximated by the Poisson model, where $y_i \sim Po(n_i\pi_i)$. Using a reference probability $p > 0$, the model can be rewritten as $y_i \sim Po(e_i\lambda_i)$, where $e_i = n_i \cdot p$ is the *expected* number

of cases (if p is the probability of being a case) and $\lambda_i = \pi_i/p$ is the risk of being a case in area i , relative to the reference risk p , known as the *relative risk* with respect to p .

It can easily be seen that the maximum likelihood estimate of λ_i is the ratio of y_i over e_i , known as the standard mortality ratio (SMR). However, interpretation of standard mortality ratios is difficult, since the variance of y_i/e_i is equal to $\lambda_i e_i / (e_i)^2 = \lambda_i / e_i = \lambda_i / (n_i \cdot p)$ and therefore inversely proportional to the population counts n_i . Hence, the most extreme SMR's are typically found in the areas with the smallest populations, a problem which is sometimes called the "small number problem". The non-appropriateness of the SMR as an estimate of relative risk is most obvious for $y_i = 0$ which is quite common for rare diseases. It is therefore necessary to incorporate information ("borrow strength") from other areas $j \neq i$ to improve the risk estimate in area i , see Section 3.

The usual choice for the reference probability p is the overall risk $\sum y_i / \sum n_i$, which implies that $\sum e_i = \sum y_i$. This is the simplest case of a so-called *internal* standardization, as p is derived internally from the data while for *external* standardization reference rates are taken from an external data source.

In the more general case, the data are additionally stratified by confounding factors, usually age. Assume that the data are now given as (y_{ij}, n_{ij}) , where the index $i = 1, \dots, I$ denotes area and $j = 1, \dots, J$ age group. Now $y_{ij} \sim Po(n_{ij}\pi_{ij})$ and under the assumption of no interaction between age and area effects, π_{ij} can be written as the product of a reference probability p_j for age group j times a (age group independent) relative risk λ_i (equal to π_{ij}/p_j for all j) in area i . Under conditional independence of responses y_{ij} , the model can be simplified to $y_i \sim Po(e_i \lambda_i)$ where $y_i = \sum_j y_{ij}$ and $e_i = \sum_j e_{ij} = \sum_j n_{ij} p_j$. The model hence reduces to the simpler case without any stratifying factors and the comments made there apply again to the maximum likelihood estimate y_i/e_i of the relative risk. In passing we note that this collapsing over age group is a great advantage of the Poisson model and would not

be possible in the binomial model except in the uninteresting case $\pi_{ij} = \pi_i$ (Wakefield, Best and Waller 1999).

The use of rate probabilities, such as p_1, \dots, p_J , for standardization is called the *indirect* method (Mantel and Stark, 1968). For the *direct* method, in contrast, one usually employs an external known standard population such as “Segi’s world population” (Segi, 1982). If no external set of standard rates p_j is available, joint maximum likelihood estimation of p_j and λ_i can be performed under a suitable identifiability constraint (Breslow and Day, 1987). It is natural to either restrict the mean of the log relative risks to zero, or to set $\sum_i y_i = \sum_i e_i$. A simpler approach is to estimate the p_j ’s through $\hat{p}_j = \sum_i y_i / \sum_i n_i$ which automatically ensures $\sum_i y_i = \sum_i e_i$.

3 Spatial modelling of relative risk

Estimation of relative risk by SMR’s suffers essentially from two main problems. The first, which has already been mentioned, is the “small number problem”: the dependence of the statistical variation of the SMR on the number of population at risk. The second is that the SMR’s totally ignore the spatial structure of the data. Both problems can be addressed through stochastic modelling of the relative risk parameters. We focus here on models that introduce spatial dependencies between relative risk parameters. However, it should be mentioned that simpler techniques can be used to account for the small number problem alone. Such methods do not incorporate the spatial structure in the data and are typically based on mixture models (Schlattmann and Böhning, 1993) or on generalized linear mixed models with exchangeable random effects (Clayton and Kaldor, 1987, Breslow and Clayton, 1993). For comparison with the more elaborate spatial models in Section 4, we report also results obtained from the Clayton and Kaldor method, where the relative risk in

each area i is assumed to be independently gamma distributed and the parameters of that gamma distribution are estimated iteratively through empirical Bayes techniques (Clayton and Kaldor, pp. 672-673).

3.1 Markov random field models

Since the pioneering work of Besag, York and Mollie (1991), Markov random field models are increasingly used in disease mapping applications. The main assumption is that the log relative risks $\eta_i = \log(\lambda_i)$, $i = 1, \dots, I$, are spatially correlated and follow a Markov random field. However, there are numerous ways to specify such a formulation with different notions of “spatially close” and different probabilistic assumptions for the log relative risk parameters. The most commonly used model is a so-called Gaussian intrinsic or conditional autoregression (CAR), where the joint distribution of η , given an unknown precision parameter κ , is

$$p(\eta|\kappa) \propto \exp\left(-\frac{\kappa}{2} \sum_{i \sim j} (\eta_i - \eta_j)^2\right) \quad (1)$$

and the sum in the exponent goes over all pairs of adjacent areas i and j . Typically those areas are considered as adjacent, that share a common border. Such a prior is nonstationary and improper (i.e. the integral over η is infinite), which can easily be seen as only differences of log relative risk parameters enter in (1). Hence there is an implicit flat prior on the overall level of the η 's. Model (1) originates from statistical image analysis (e.g. Geman and Geman, 1984) and has a spatial smoothing effect on the relative risk estimates where the degree of smoothing is determined through the parameter κ .

The conditional distribution of η_i , given all other parameters η_j , $j \neq i$, turns out to be Gaussian with expectation equal to the mean of the η_j 's in adjacent areas of region i and (conditional) precision equal to κ times the number of neighbours of area i . The *marginal* prior precision (given a fixed overall level, say $\sum \eta_i = 0$) has more direct interpretation as

the *prior amount of smoothing* in region i , but can only be determined through simulation. Results from Bernardinelli, Clayton and Montomoli (1995) suggest that there is also strong dependence of the marginal prior precision on the number of neighbours. This is a rather unattractive feature of model (1) as ideally the prior degree of smoothing should be the same for all areas or should at least be not linked directly to the number of neighbours. However, the dependence of the conditional (and hence the marginal) prior precision on the number of neighbours is necessary for consistency of the Markov random field definition. To overcome this problem, Cressie (1992) and Stern and Cressie (1999) have suggested to use specific functions of the expected cases e_i and e_j as multiplicative weights w_{ij} for each term $(\eta_i - \eta_j)^2$ in (1), which has the advantage that the conditional variance no longer depends on the number of neighbouring areas. However, this choice implies that the conditional expectation is no longer the (weighted) mean of adjacent parameters but now a (weighted) sum without any normalization, which has interpretation difficulties.

Alternatively a number of other weights have also been suggested. For example, the weights could be based on the distance of the centroids of the districts or on the length of the common boundary. However, comparative studies in Best *et al.* (1999) between adjacency and distance-based weights suggest that different choices of weights do not have too much influence on the final estimates.

The approach taken in Besag *et al.* (1991), which we present here in a slightly different but equivalent formulation, is more elaborate but can easily be presented by introducing another hierarchy in the above formulation. The log relative risks η_i , $i = 1, \dots, I$, are now assumed to be *conditionally independent* Gaussian with expectation u_i and precision λ . The spatial dependence is moved from the second to a third level, assuming a Markov random

field (again with precision κ) for the u_i 's. More specifically, the model can be written as

$$p(\eta|u, \lambda) \propto \exp\left(-\frac{\lambda}{2} \sum_i (\eta_i - u_i)^2\right) \text{ and } p(u|\kappa) \propto \exp\left(-\frac{\kappa}{2} \sum_{i \sim j} (u_i - u_j)^2\right). \quad (2)$$

Model (2) boils down to model (1) for $\lambda = \infty$. On the other extreme, for $\kappa = \infty$ the spatially structured component vanishes and the model accounts now only for unstructured heterogeneity.

Finally hyperpriors are assigned to the precision parameters κ and λ . It is not possible to use classical noninformative priors here as they are improper and the resulting posterior distribution will also be improper. This is a property already known from mixed models for Gaussian responses, e.g. O'Hagan (1994, p. 272). The usual choice is therefore a highly dispersed but proper prior from the conjugate gamma family although other choices can be made as well. In all applications we use gamma hyperpriors with parameters 1 and 0.005, which is close to Kelsall and Wakefield (1999), who recommend 0.5 and 0.0005 as a default choice. See also Bernardinelli, Clayton and Montomoli (1995) for a thorough comparison of different hyperprior specifications.

The idea behind the Besag *et al.* model is that, through data-driven estimation of κ and λ , the formulation is able to distinguish if the observed heterogeneity is *spatially structured* or *unstructured*. One of the main disadvantages of the Besag *et al.* model, however, is that both κ and λ are *global* parameters. Hence the amount of spatially structured and unstructured heterogeneity is assumed to be the same over the whole study region which is a rather restrictive assumption. For the spatially structured term, this implies that the degree of smoothing is the same over the whole map, hence the smoothing is *non-adaptive*.

Approximate statistical inference in model (1) can be performed within the generalized linear mixed model framework, now with spatially correlated random effects (Breslow and Clayton, 1993, Breslow, Leroux and Platt, 1998). In the more complex model (2), inference can only be performed through Markov chain Monte Carlo (MCMC) techniques, e.g. Be-

sag *et al.* (1991), Clayton and Bernardinelli (1992), Knorr-Held and Besag (1998) or Best *et al.* (1999). These methods are based on variants of the Metropolis-Hastings algorithm (Hastings, 1970) and are nowadays the common choice in more complex hierarchical models. The WinBUGS software (Spiegelhalter, Thomas and Best, 1998) has incorporated the CAR model in its latest version together with modules for mapping the results.

3.2 Cluster models

Markov random fields models in disease mapping assume that the risks in adjacent regions are somehow similar and that the degree of similarity is the same over the whole study region. We now describe an alternative model, where risk parameters in adjacent areas are assumed to be either exactly the same or independent from each other. Such a formulation has been proposed in Knorr-Held and Raßer (2000) based on a modification of so-called Voronoi tessellations (e.g. Green, 1995). These Voronoi tessellations are piecewise constant model formulations and have been successfully used in various applications in continuous space, e.g. Heikkinen and Arjas (1998, 1999). Technically the methods are based on reversible jump MCMC (Green, 1995), an extension of the Metropolis-Hastings algorithm to simulate from distributions of variable dimension.

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

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

The number k , the shape, the size of the clusters as well as the risk within each cluster

are treated (directly or indirectly) 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 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, hence will incorporate all the uncertainty about the number, the location and the risk level of the clusters. The formulation can therefore be seen as nonparametric (Arjas, 1996, Heikkinen and Arjas, 1998). Typically such a model will have a spatial smoothing effect, just as Markov random field models, but is able to retain discontinuities in the risk surface.

To define the clusters a construction is proposed where k regions are marked as so-called cluster centres, each of them defining a 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 one to the other. The cluster centres are collected in a vector, say G_k . Regions, which have the same “distance” (in terms of the 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 centres with smaller indices.

The model now assumes a uniform or truncated geometric prior on $\{1, \dots, I\}$ for k and a uniform prior for $G_k|k$. That means each specific choice of G_k has a prior probability equal to $(n - k)!/n!$ as the number of all (unordered) G_k 's is $n!/(n - k)!$. Finally, the log relative risk within each cluster is assumed to be independent Gaussian with unknown mean μ and precision ν . In all applications we use similar priors as in the Markov random field approach, i.e. a gamma hyperprior for ν with parameters 1 and 0.005, and a prior for μ uniform on the whole real line. The prior for k is chosen as geometric with parameter 0.02, i.e. $Pr(k) \propto 0.98^k$.

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. Note that clusters of size one are not excluded from the model which implies that the formulation does not necessarily smooth the risk estimates. In practice it will always do so, at least to some extent, since there will always be some uncertainty whether 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 κ is constant and smoothing is non-adaptive. Furthermore, simulations in Knorr-Held and Raßer (1999) have shown that the influence of the number of adjacent regions on the average size of the cluster, to which the corresponding region is assigned to, is minimal a priori. This indicates that, in contrast to the Markov random field models, there is no strong dependence of the prior amount of smoothing on the number of neighbours.

Related formulations based on continuous Voronoi tessellations have recently been proposed in Denison and Holmes (1999) for disease mapping. If applied to the usual areal data, their formulation has the disadvantage that clusters will not necessarily be connected. The discrete Knorr-Held and Raßer (2000) model fully acknowledges the discrete nature of space and automatically ensures that all clusters are connected.

4 Application to German cancer mortality data

We now present results of two analyses of cancer mortality data from Germany, 1986-1990: brain cancer and lung cancer, both for males. For an application to oral cavity cancer in Germany see Knorr-Held and Raßer (2000). The raw data is given in $I = 544$ regions and $J = 16$ age groups. 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. We

compare two models: the Markov random field model proposed by Besag *et al.* (1991) and the cluster model by Knorr-Held and Raßer (2000). The two models are abbreviated by BYM and KHR, respectively. As a side comment we remark here that the 544 districts show large variability with respect to the number of neighbours: 36 districts have only one neighbour while the maximum number of neighbours is 11 (median of 5).

4.1 Brain cancer

A large number of malignant neoplasms of different histological types are classed together as “brain tumors” (ICD category 191). We look only at males where the total number of cases is 9,348 with a median number of 11 cases per district. The corresponding SMR’s vary between 0 and 2.73 and are shown in Figure 1.

The estimates from the BYM model and from the KHR model are in a much smaller range with values between 0.87 and 1.18 (BYM) and 0.90 and 1.19 (KHR), see Figure 2 and 3. For both methods, estimates above 1.0 can be found in the very south-west, in large parts of the west and north-west, and at isolated spots in the east. The estimates show a strong overall agreement with stronger differences only for the extreme values. In particular, the highest risk estimate from the BYM model can be found in the city of Leipzig ($y_i = 84$, $e_i = 59.2$, $SMR_i = 1.42$), a district with just one neighbour. The median relative risk is here 1.18 with a 97% posterior probability of an elevated risk. We are somewhat sceptical with respect to this high risk estimate as it can be largely attributed to a high value of the spatially correlated component u_i . The relatively high estimate might be a consequence of the dependence of the prior variance on the number of neighbours. The KHR model, where the amount of smoothing does not depend so strongly on the number of neighbours, gives a lower estimate of 1.06 for Leipzig with a corresponding posterior probability of 0.80.

Figure 4 displays the posterior probabilities from the KHR model that the relative risk

in a given district is larger than the overall risk $\sum y_i / \sum e_i = 1.0$. The highest values in the KHR model can be found in 10 districts in the very south-west with relative risk estimates between 1.16 and 1.19 and corresponding posterior probabilities in the range of 87 to 99%. The BYM estimates are in the range of 1.03 to 1.16 with posterior probabilities between 0.64 and 0.91. A second cluster can be isolated in the area between Frankfurt, Wiesbaden and Giessen (Taunuskreis) with relative risk estimates between 1.10 to and 1.12 and posterior probabilities between 90 and 99%. Finally, posterior probabilities of 95% and higher can also be found in the very north. However, the real scientific interest lies in the point estimates which do not have much variation. Hence, the main message of both analyses is that there is not much geographical variation in disease risk from brain cancer for males, although a naive inspection of the SMR's would suggest so.

We have also applied the empirical Bayes method by Clayton and Kaldor (1986) to this dataset. The estimates are in a similar range (0.91 to 1.14) with highest values in Leipzig (1.14), Frankfurt (1.11) and Fulda (1.10). However, the approach ignores spatial location so spatial clusters like the one in the very south-west are hardly to identify.

4.2 Lung cancer

Lung cancer is by far the leading cause of cancer death among males in West Germany with a total number of 134,820 cases for the period 1986-1990 (median number of 150 cases per district). The corresponding SMR's vary in the range of 0.54 to 1.58 and are displayed in Figure 5. The issue of random variation of the SMR due to the problem of small numbers is here certainly not as crucial as for brain cancer. Indeed, the variation of the empirical Bayes estimates with Clayton and Kaldor's method show only slightly less variation with values between 0.60 and 1.54 and a standard deviation of the log relative risk estimates of 0.18 (0.21 for the SMR's). However, it is sensible to apply spatial models and search for spatial

patterns in the geographical distribution of lung cancer risk.

Figure 6 and 7 give the risk estimates with the model of Besag *et al.* (1991) and Knorr-Held and Raßer (2000) respectively. The maps closely resemble each other although the estimates from the BYM model seem to be noisier. This is a consequence of the non-adaptivity of the BYM model as the overall variability in the risk surface induces a variability of the BYM estimates in the whole map. In contrast, estimates from the KHR model are locally smoother without oversmoothing important spatial gradients and clusters.

The KHR model identifies three cluster of relative risk above 1.2, one in the west around the Ruhr area, one a bit more south covering most of the state Saarland, and one in the north-east in the state of Mecklenburg–West Pomerania. These findings are consistent with known high smoking prevalence rates in these areas (Becker and Wahrendorf, 1997). Figure 8 maps the posterior probabilities (from the KHR model) that the relative risk is higher than the overall rate which is in overall agreement with Figure 7. The small “significant” cluster in the former East German state of Saxony can be associated with large-scale Uranium mining projects after World War II, exposing thousands of mine workers to extremely high levels of radon emissions. The lower values in most part of the south are interesting from a protective point of view, as they indicate the potential for cancer prevention under similar socioeconomic conditions.

5 Conclusions

This paper has reviewed recent advances in the geographical analysis of disease risk. Our focus has been on Bayesian models, which allow for sound statistical inference in realistically complex models for spatial dependence without any asymptotic assumptions. This techniques overcome difficulties in traditional methods of mapping disease risk. Markov random

field models have already been applied in cancer mapping of Sardinia (Bernardinelli *et al.*, 1994) and France (Rezvani *et al.*, 1997). The KHR method has not yet been applied outside of academic research, as their work is rather new. However, there are some arguments to prefer this method over the BYM approach, the important ones being the issue of adaptive smoothing and the strong dependence of the BYM model on the number of neighbours. Non-Bayesian techniques for smoothing have been applied in the recent atlas of mortality in the United States (Pickle *et al.*, 1996). The algorithm used is based on weighted headbanging (Mungiole *et al.*, 1996), a useful exploratory method.

Of course, the data for such studies are of an observational type and there is a danger of overinterpretation of the results. Observational studies can only suggest the need for further investigation as there is an inherent difficulty to control for confounding factors, such as socio-economic status or deprivation. Also long latencies of many diseases, the issue of migration, and other data anomalies constitute fundamental problems for the interpretation of the results. However, the design, implementation and analysis of controlled studies is often very expensive, both in terms of time and financial costs. It is certainly not an appropriate attitude to ignore the large amount of available observational data, instead proper statistical techniques should be applied to the data and dangers of misinterpretation should be emphasized.

References

- Arjas, E. (1996). Discussion of paper by Hartigan, *in* J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith (eds), *Bayesian Statistics 5*, Oxford: University Press, pp. 221-222.
- Becker, N. and Wahrendorf, J. (1997). *Atlas of Cancer Mortality in the Federal Republic of Germany 1981-1990*, Berlin: Springer Verlag.
- Bernardinelli, L., Clayton, D. and Montomoli, C. (1995). Bayesian estimates of disease maps: How important are priors?, *Statistics in Medicine*, **14**, 2411-2431.

- Bernardinelli, L., Maida, A., Marinoni, A., Clayton, D., Romano, G., Montomoli, C., Fadda, D., Solinas, G., Castiglia, P.L., Cocco, P.L., Ghislandi, M., Berzuini, C., Pascutto, C., Nerini, M., Styles, B., Capocaccia, R., Lispi, L. and Mallardo, E. (1994). *Atlas of cancer mortality in Sardinia*, Rome: Unitá Operativa del P.F. FATMA.
- Best; N.G., Arnold, R.A., Thomas, A., Waller, L.A. and Conlon, E.M. (1999). Bayesian models for spatially correlated disease and exposure data. *in* J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith (eds), *Bayesian Statistics 6*, Oxford: University Press, pp. 131-156.
- Besag, J. E., York, J. C. & 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. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9-25.
- Breslow, N.E. and Day, N. (1987). *Statistical Methods in Cancer Epidemiology. Vol. II - The Design and Analysis of Cohort Studies* IARC Scientific Publications No. 82. International Agency for Research on Cancer, Lyon.
- Breslow, N. and Leroux, B. and Platt, R. (1998). Approximate hierarchical modelling of discrete data in epidemiology. *Statistical Methods in Medical Research*, **7**, 49-62.
- 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*, Oxford University Press, pp. 205-220.
- Clayton, D. G. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics* **43**, 671-681.
- Cressie, N. (1992). Smoothing regional maps using empirical Bayes predictors. *Geographical Analysis* **24**, 75-95.
- Denison, D.G. and Holmes, C.C. (1999). Bayesian partitioning for estimating disease risk. Technical Report, Imperial College London.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721-741.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* **82**, 711-732.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97-109.
- Heikkinen, J. and Arjas, E. (1998). Nonparametric Bayesian estimation of a spatial Poisson intensity, *Scandinavian Journal of Statistics*, **25**, 435-450.
- Heikkinen, J. and Arjas, E. (1999). Modelling a Poisson forest in variable elevations: A nonparametric Bayesian approach, *Biometrics* **55**, 738-745.
- Kelsall, J. E. and Wakefield, J. C. (1999). Discussion of ‘Bayesian models for spatially correlated disease and exposure data’, by Best *et. al.* In J. M. Bernardo, J. O. Berger, A. P. David and A. F. M. Smith (eds), *Bayesian Statistics 6*, Oxford: University Press, p. 151.
- 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 Raßer, G. (1999). Bayesian detection of clusters and discontinuities in disease maps: Simulations, *Discussion paper 142*, SFB 386, University Munich, Available at www.stat.uni-muenchen.de/sfb386/publikationen.html.
- Knorr-Held, L. and Raßer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps, *Biometrics* **56**, ???-???
- Mantel, N. and Stark, C. R. (1968). Computation of indirect-adjusted rates in the presence of confounding, *Biometrics* **24**, 997-1005.
- Mungiole, M., Pickle, L.W., Simonson, K.H., White, A.A. (1996). Application of a weighted head-banging algorithm to mortality data maps. *Proceedings of the Statistical Graphics Section of the 1996 Annual Meeting of the American Statistical Association*, 45-49.
- O’Hagan, T. (1994). *Kendall’s Advanced Theory of Statistics. Volume 2B: Bayesian Inference*. London: Edward Arnold.
- Pickle, L.W., Mungiole, M., Jones G.K., White, A.A. (1996). *Atlas of United States Mortality*, Hyattsville, Maryland: National Center for Health Statistics.
- Rezvani, A., Mollié, A., Doyon, F. and Sancho-Gernier, H. (1997). *Atlas de la mortalité par cancer en France (1986-1993)*. Paris: Les éditions INSERM.
- Schlattmann, P. and Böhning, D. (1993). Mixture models and disease mapping. *Statistics in Medicine* **12**, 1943-1950.

- Segi, M. (1982). *Age-adjusted Death Rates for Cancer for Selected Sites in 43 Countries in 1977*. Segi Institute of Cancer Epidemiology, Nagoya.
- Spiegelhalter, D.J., Thomas, A. and Best, N. G. (1998) *WinBUGS User Manual, version 1.1.1*. Cambridge University. Available at www.stat.mrc-bsu.cam.ac.uk/bugs.
- Stern, H.S., and Cressie, N. (1999). Small-area and point-level Bayesian models for inference on extremes in disease maps. In A.B. Lawson, D. Boehning, E. Lesaffre, A. Biggeri, J.F. Viel and R. Bertollini (eds.), *Disease Mapping and Risk Assessment for Public Health*, pp. 63–84. Wiley, Chichester.
- Wakefield, J.C., Best, N.G. and Waller, L. (1999). Bayesian approaches to Disease Mapping. In: ???, ???.

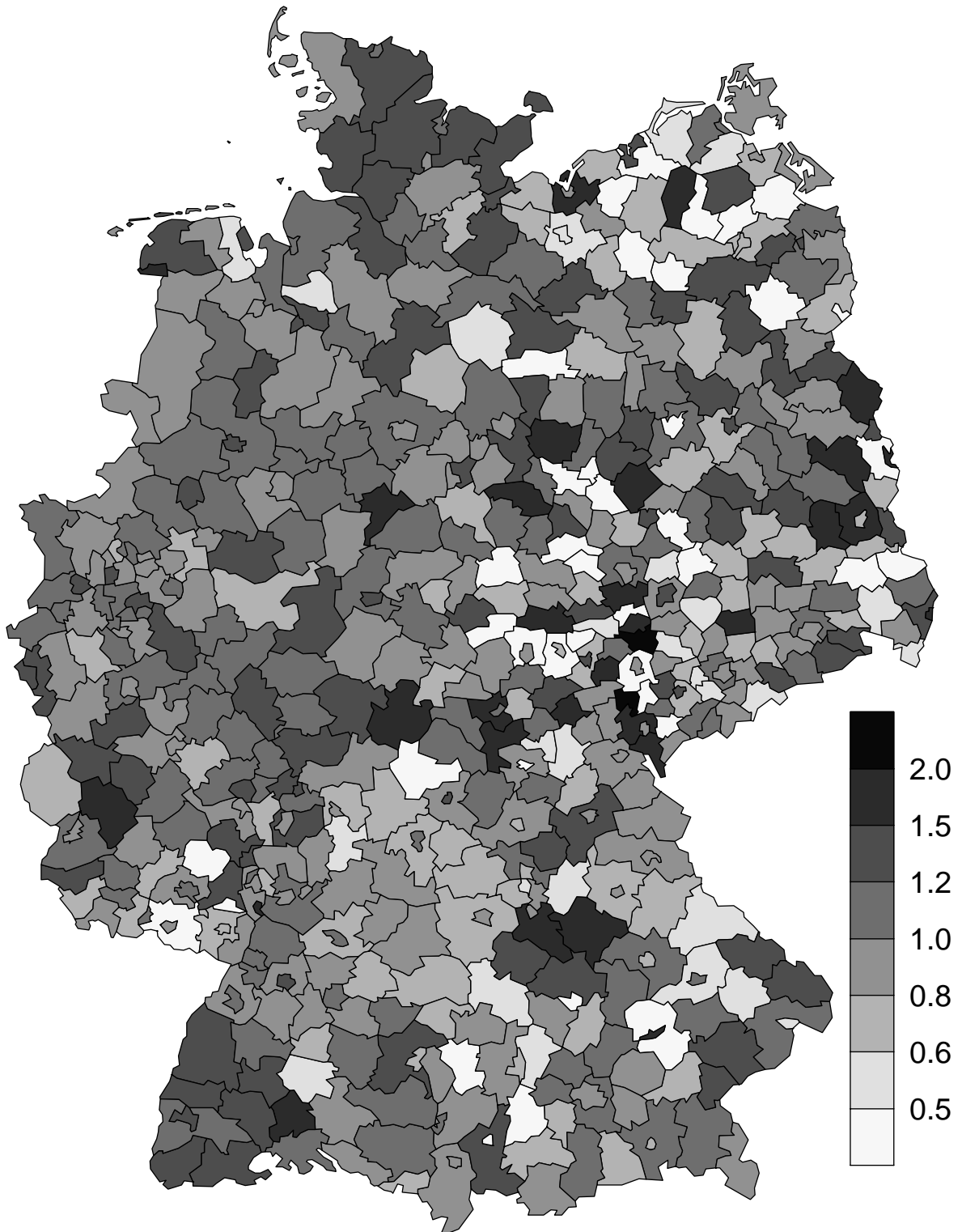


Figure 1: Standard mortality ratios for brain cancer of males in Germany.

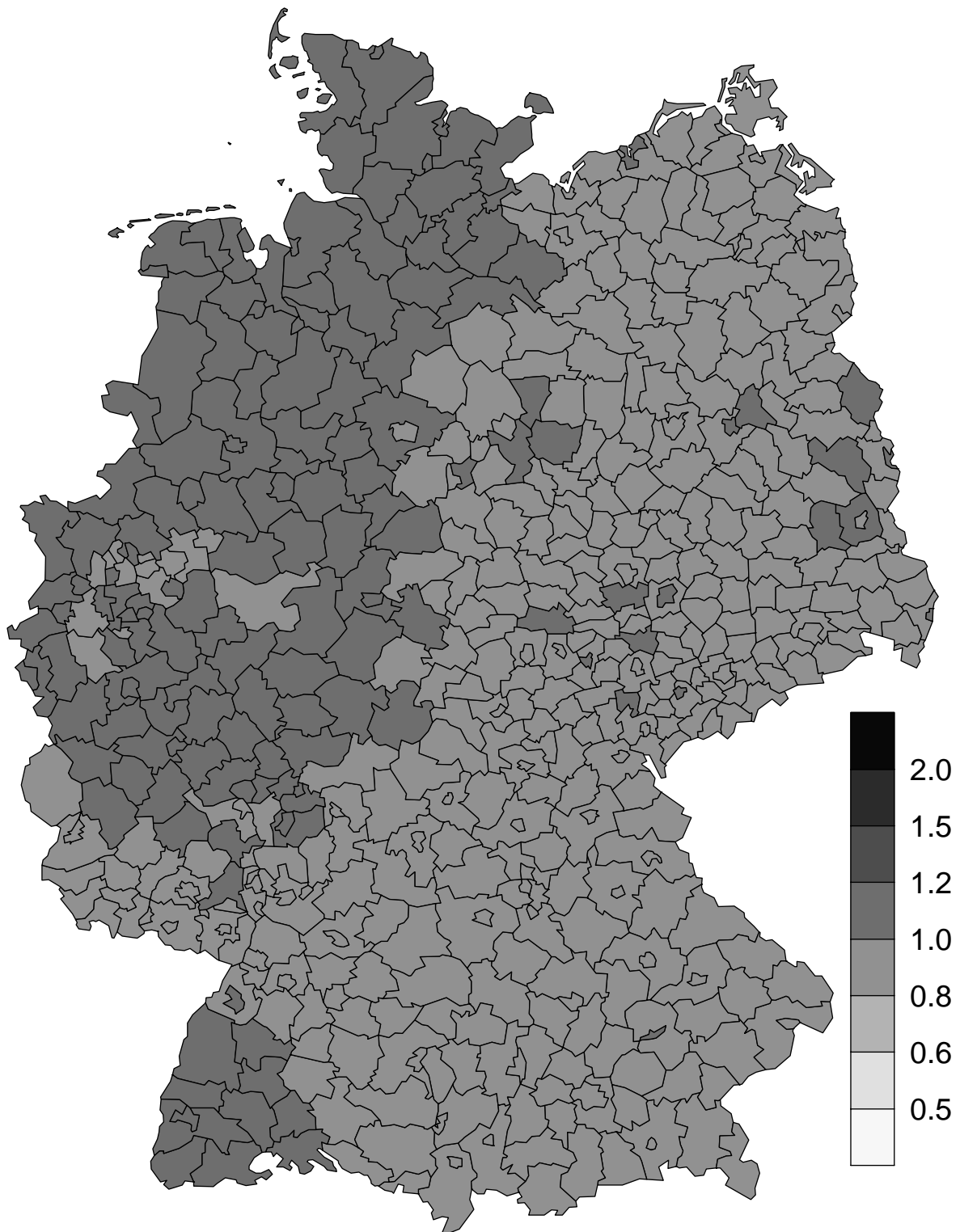


Figure 2: Estimated median relative risks for brain cancer of males in Germany with the method of Besag *et al.*

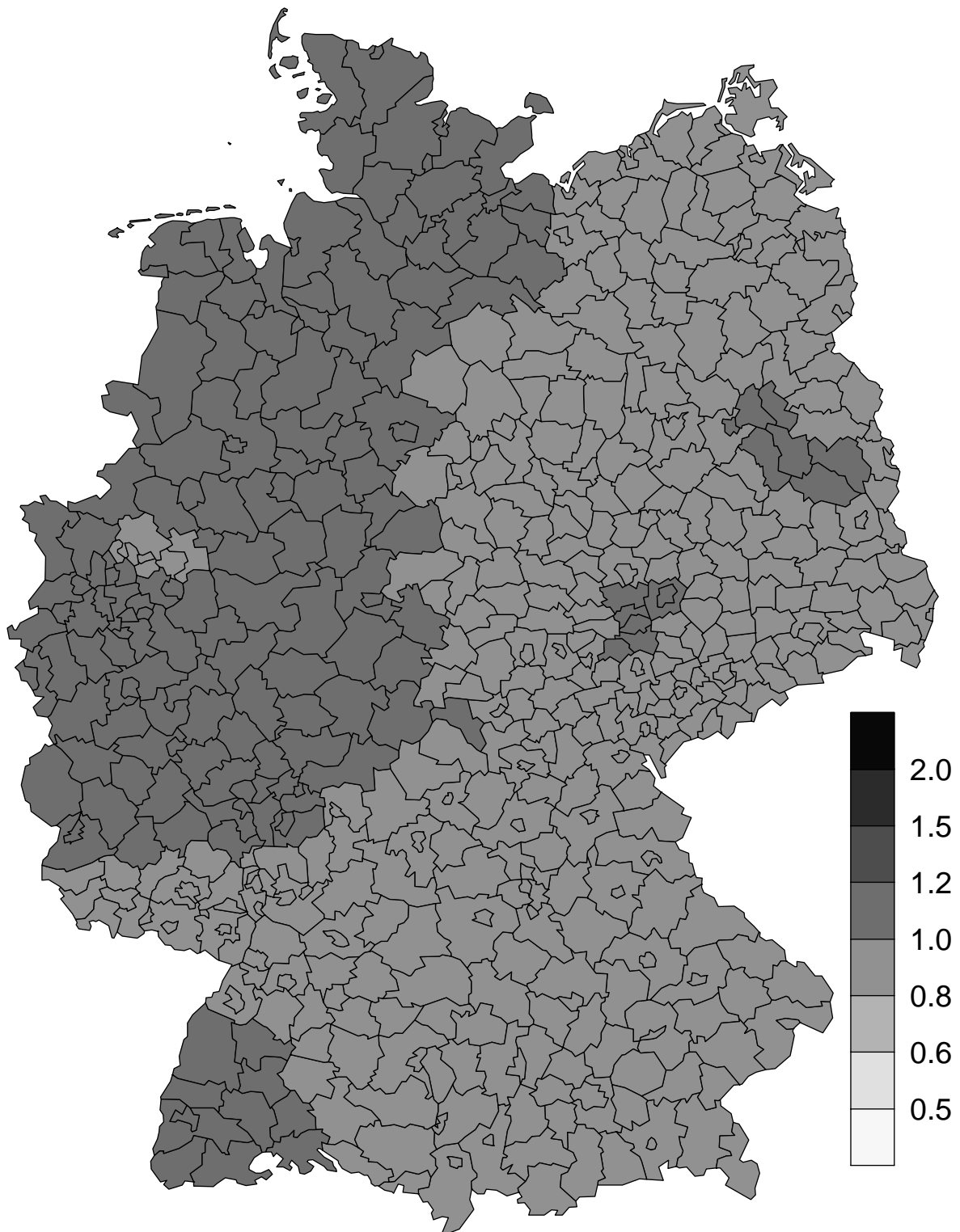


Figure 3: Estimated median relative risks for brain cancer of males in Germany with the method of Knorr-Held and Raßer.

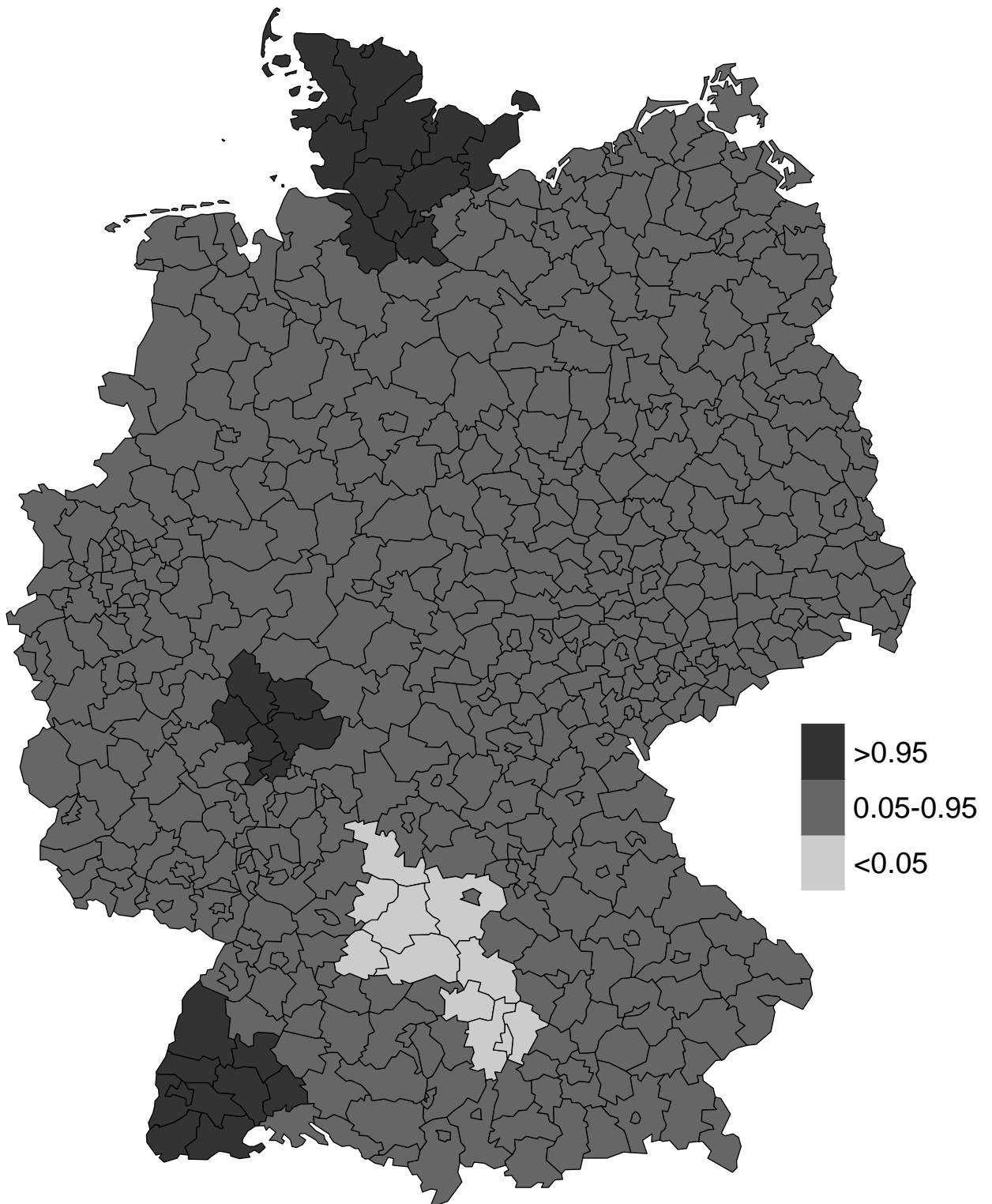


Figure 4: Posterior probabilities of a relative risk above 1.0 for brain cancer of males with the method of Knorr-Held and Raßer.

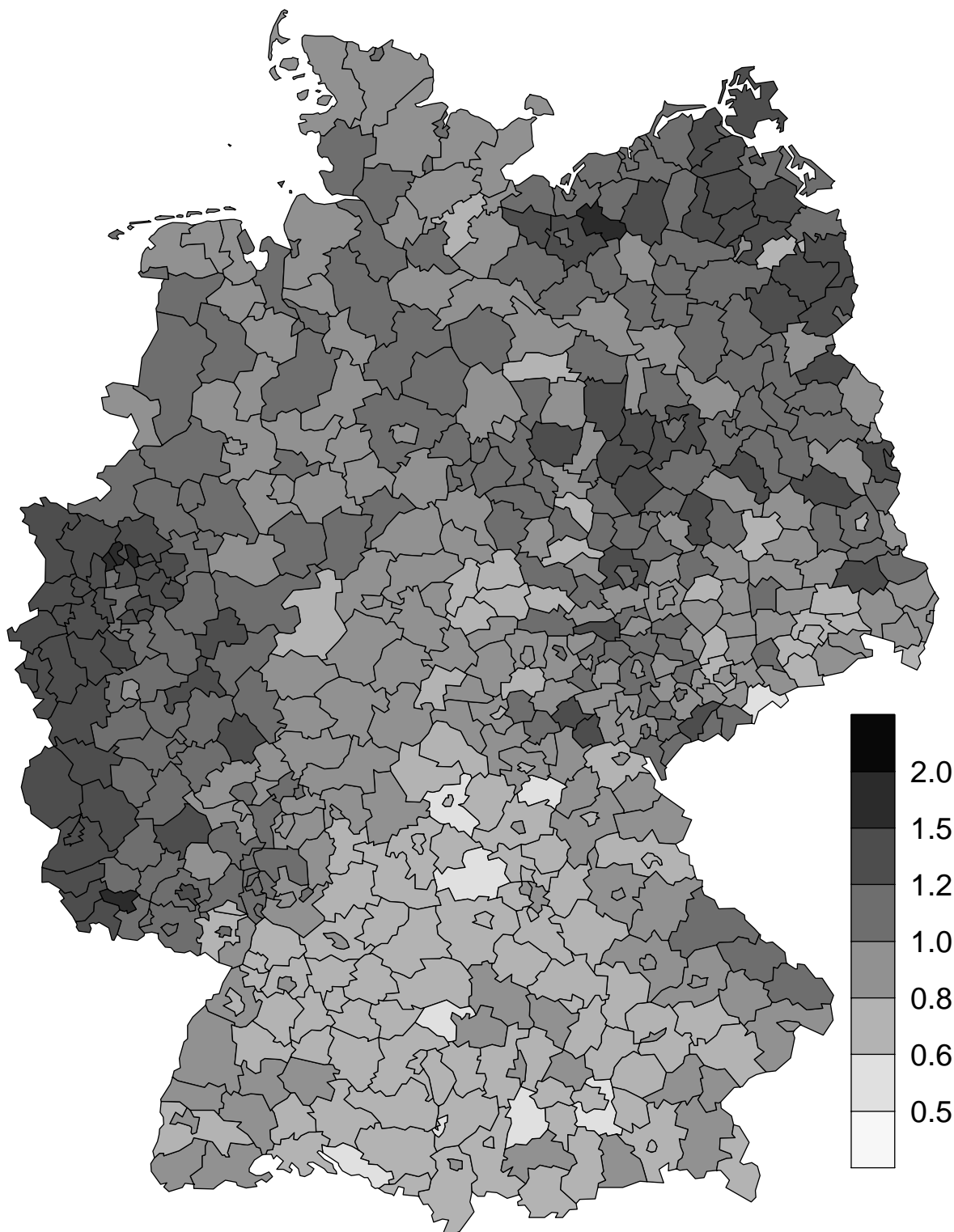


Figure 5: Standard mortality ratios for lung cancer of males in Germany.

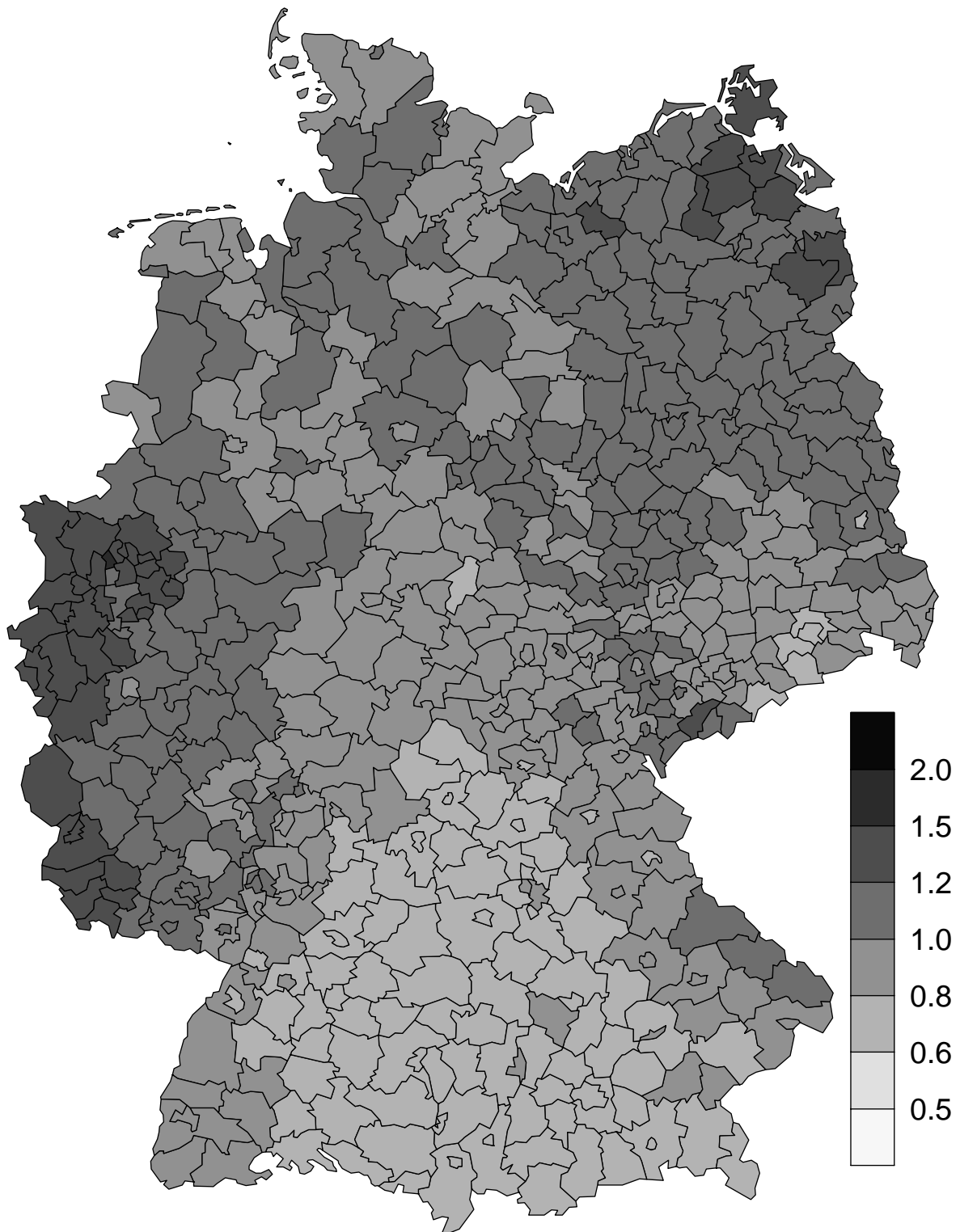


Figure 6: Estimated median relative risks for lung cancer of males in Germany with the method of Besag *et al.*

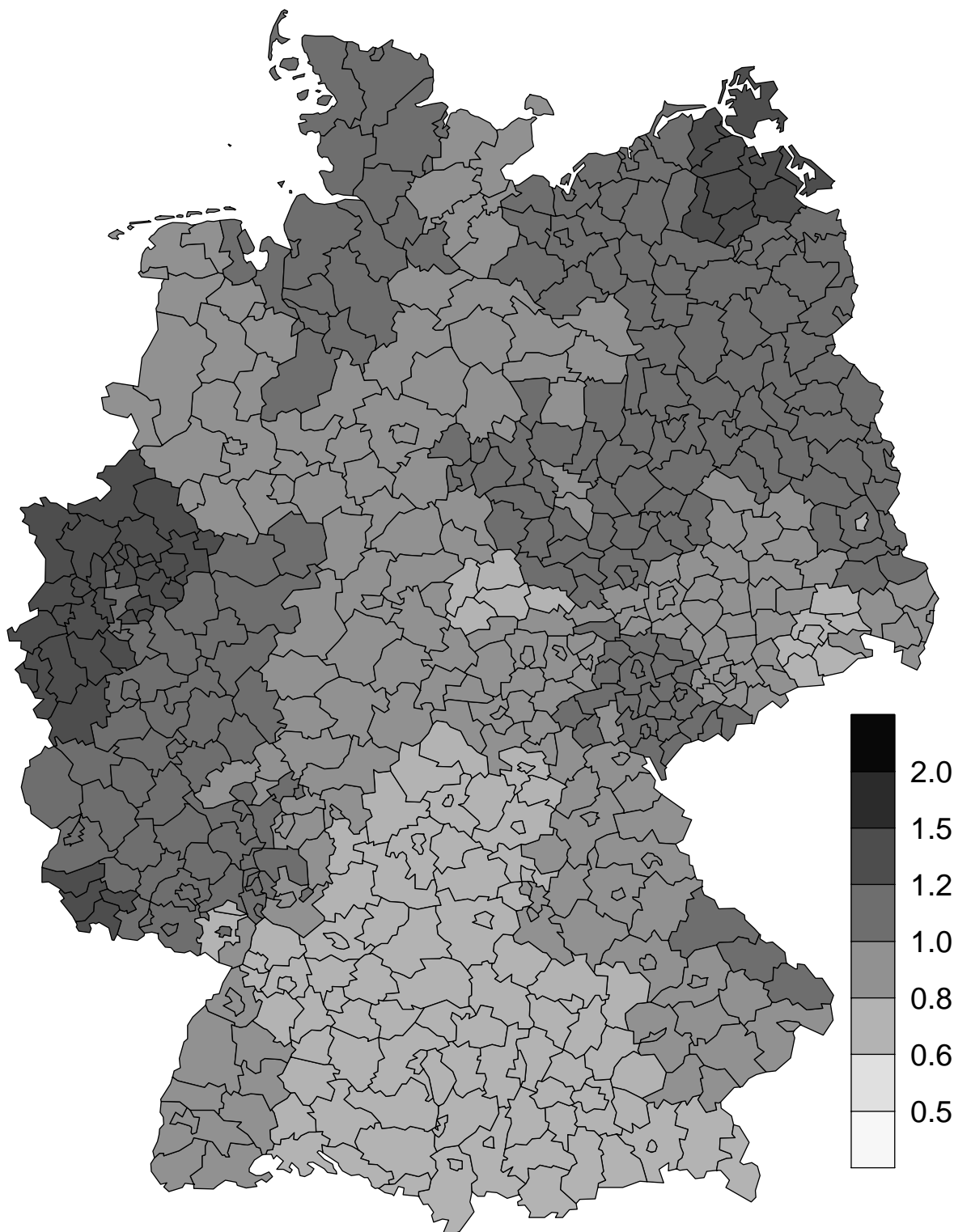


Figure 7: Estimated median relative risks for lung cancer of males in Germany with the method of Knorr-Held and Raßer.

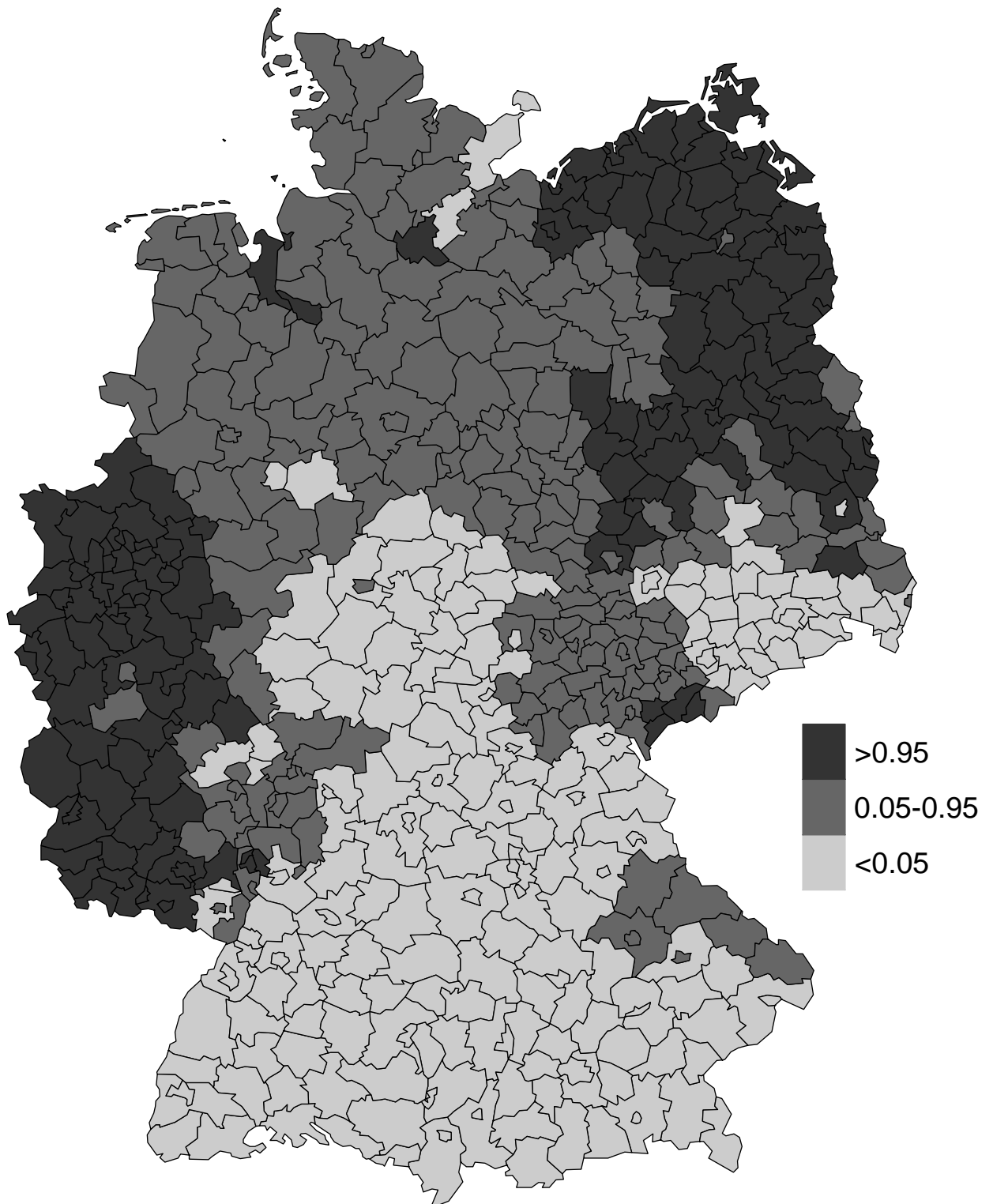


Figure 8: Posterior probabilities of a relative risk above 1.0 for lung cancer of males with the method of Knorr-Held and Raßer.