STATISTICS IN MEDICINE
Statist. Med. 2004; **00**:1–17 *Prepared using* simauth.cls *[Version: 2002/09/18 v1.11]*

A multiscale method for disease mapping in spatial epidemiology

Mary M. Louie[1*†], Eric D. Kolaczyk[2]

[1] *Channing Laboratory, Brigham & Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115*
[2] *Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215*

SUMMARY

The effects of spatial scale in disease mapping are well-recognized, in that the information conveyed by such maps varies with scale. Here we provide an inferential framework, in the context of tract count data, for describing the distribution of relative risk simultaneously across a hierarchy of multiple scales. In particular, we offer a multiscale extension of the canonical standardized mortality ratio (SMR), consisting of Bayesian posterior-based strategies for both estimation and characterization of uncertainty. As a result, a hierarchy of informative disease and confidence maps can be produced, without the need to first try to identify a single appropriate scale of analysis. We explore the behavior of the proposed methodology in a small simulation study, and we illustrate its usage through an application to data on gastric cancer in Tuscany. Copyright © 2004 John Wiley & Sons, Ltd.

1. Introduction

In epidemiology, interest generally centers on understanding how the occurrence of a disease relates to potential explanatory factors. Furthermore, it is not uncommon to include for consideration, when available, the spatial locations at which cases of the disease occur. This area of study and its associated tools are known as 'spatial epidemiology'. And arguably one of the most fundamental tasks in spatial epidemiology is the characterization of the spatial distribution of disease through the creation of a map(s).

Statistical models play an important role in disease mapping, typically being used to 'clean' the map of noise and to help extract structure in the underlying measurements. Yet disease mapping is not a purely statistical exercise – various concepts and principles from Geography are critical as well. Perhaps foremost among these is the concept of *scale*, as the use of a comparatively coarser or finer scale in a map generally entails a correspondingly greater or lesser degree of aggregation of the information underlying the data. The effects of scale on

*Correspondence to: Mary M. Louie, National Center for Health Statistics, 3311 Toledo Road, Room 3215, Hyattsville, MD 20782
†Email: mlouie@cdc.gov

Contract/grant sponsor: National Science Foundation; contract/grant number: BCS0079077

Contract/grant sponsor: Office of Naval Research; contract/grant number: N00014–99–1–0219

Received
Revised

the analysis and inference of geo-spatial data are known to be many and varied, and often are referred to collectively under the label of the so-called modifiable areal unit problem [1]. Disease mapping is by no means immune to such effects, of course. See [2], for example.

The data encountered in spatial epidemiology tend to take either of two forms: (i) the locations at which cases of the disease occurred or (ii) aggregate counts of cases that occurred within geographical subregions of some overall region of study. The latter form is arguably that in which spatial disease data are most readily available, and will be the focus of our attention in this paper. In the event of the former, our methods will of course still apply if the investigator has available and is willing to use some degree of aggregation. The choice of subregions over which aggregation is done typically is influenced by geo-political considerations, and the conventions adopted by relevant government statistical agencies tend to dominate. Census tracts are often used, for example.

Usually there are a number of levels at which the data either could have been aggregated or may be further aggregated. In addition, these levels frequently are nested in a hierarchical fashion. For example, in an analysis of male gastric cancer data, presented in Section 6, we considered aggregations at regional, provincial, and municipal levels, as depicted in Figure 3. Since with each successive degree of aggregation there comes a corresponding coarsening of spatial scale, the issue of scale arises in a natural and explicit manner in such contexts. Note, however, that it is quite possible that the data contain useful information at more than just one level of aggregation. Even in the event that the information at one scale dominates in its relevance, it may not be clear a priori just which is that scale. Therefore, what is needed is a capability to conduct multiscale statistical analyses.

Multiscale methods have been the focus of intense development by researchers in a variety of disciplines over the last 15 years, including mathematics, statistics, signal and image processing, and computer vision. This work is based primarily on the use of wavelets and variations thereof, and pertains largely to data that are regularly spaced over some domain, such as are commonly obtained through instrumental measuring devices. Unfortunately, the spatially irregular and typically aggregated nature of most geographical data sets (such as the tract count data considered in this paper) precludes the application of such methods. The framework of Kolaczyk and Huang [4] seeks to address this shortcoming, using multiscale likelihood factorizations to extend the standard wavelet-based paradigm to a class of polygon-based geographical data structures. The subsequent work of Louie and Kolaczyk [5] further explores this framework, and coins the term multiscale spatial process models (MSSpP). These two papers together explore the design, usage, and properties of such models for modeling and inference with generic Gaussian and Poisson measurements. In this paper we extend the Poisson-based MSSpP model to the context of tract count data for the purpose of creating disease maps.

It is standard in epidemiology to quantify the occurrence of a disease through measures of disease frequency, wherein the observed levels of the disease are related to the 'at-risk' population from which they arose, thereby controling for natural variation in the latter [6]. These quantities are termed 'relative risks'. In disease mapping, one seeks to present a visual representation of the manner in which the disease frequency varies with spatial location. Therefore, in constructing a disease map it is necessary to incorporate some sense of a spatial control distribution, so as to control for spatial variation in the 'at-risk' population. A canonical quantity of this type is the standardized mortality ratio (SMR), defined to be the ratio within each subregion of the observed counts to an estimate of the counts expected based on just the

'at-risk' population. Currently, given disease counts over subregions at some single, fixed level of aggregation, disease maps essentially characterize the spatial distribution of the disease at just that level. We take as our goal in this paper to create a multiscale extension of the SMR for tract count data.

The rest of this paper is organized as follows. In Section 2, we describe a Bayesian framework for modeling disease count data at multiple scales. In Section 3, we present our multiscale extension of the SMR estimator. Two versions are given, reflecting empirical and fully Bayesian implementations of our framework. In Section 4 we provide methods for characterization of posterior uncertainty. Section 5 contains results of a small simulation study, quantifying the mean-squared error performance of our methods in comparison to a standard competitor in the literature. Section 6 gives a summary of an analysis of a well-known dataset on male gastric cancer from the Tuscan region of Italy. Additional discussion may be found in Section 7.

2. Multiscale Models of Relative Risk

The state of the art in wavelet-based statistical inference is primarily Bayesian in nature (particularly empirical Bayesian) from both practical [9] and theoretical [10, 11] perspectives. The standard framework consists of two key components: (i) a wavelet transformation of an unknown function, yielding a set of wavelet coefficients, and (ii) the specification of a prior probability structure on these coefficients. Our multiscale modeling framework for disease count data rests similarly on two key components: (i) a factorization of the data likelihood in the standard Poisson-based sampling model, yielding a multiscale reparameterization of the underlying intensity process, and (ii) a prior probability structure defined with respect to the resulting multiscale parameters.

2.1. Multiscale Representation of Data Likelihood

Let $D \subseteq \mathbb{R}^2$ be an arbitrary spatial region, and define the collection $\mathcal{B}^{(J)} = \{\{B_{j,k}\}_{k=1}^{N_j}\}_{j=0}^{J}$ to be a hierarchy of $J+1$ nested partitions of D. In particular, the $B_{j,k}$ will represent subregions in D at scales $j = 0, 1, \ldots, J$ and positions $k = 1, \ldots, N_j$ within scale, such that

$$\bigcup_{k=1}^{N_j} B_{j,k} \;=\; D \qquad \text{and} \qquad \bigcup_{k' \in ch(k)} B_{j+1,k'} \;=\; B_{j,k}\,, \tag{1}$$

where for a given choice of j and k, $ch(k)$ denotes the set of indices k' for which $B_{j+1,k'} \subseteq B_{j,k}$. For example, in Figure 3 we have $J+1 = 3$, and the $B_{j,k}$ at scales $j = 0, 1$, and 2 correspond to the overall Tuscan region, its nine provinces, and their 287 municipalities, respectively. For convenience, and without loss of generality, we will assume that $N_0 = 1$ in our development of the methodology below.

We will let the occurrences of cases of a given disease be represented by a point process $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$. And as is common [2], we will assume conditional on an intensity process $\{\mu(\mathbf{s}) : \mathbf{s} \in D\}$, that $Y(\cdot)$ is Poisson. We then define $Y_{j,k}$ to be the aggregate count of occurrences within the subregion $B_{j,k}$, for each (j,k). Clearly $Y_{j,k} \sim \text{Poisson}(\mu_{j,k})$, where $\mu_{j,k} \equiv \int_{B_{j,k}} \mu(\mathbf{s}) d\mathbf{s}$. In addition, by (1), we have the relations

$$Y_{j,k} \;=\; \textstyle\sum_{k' \in ch(k)} Y_{j+1,k'} \qquad \text{and} \qquad \mu_{j,k} \;=\; \textstyle\sum_{k' \in ch(k)} \mu_{j+1,k'}\ . \tag{2}$$

What is important to note here is the inter-dependency among the $\{Y_{j,k}\}$ and $\{\mu_{j,k}\}$ among themselves, which serves as an obstacle to an effective and interpretable analysis of the data that incorporates information across multiple scales. However, this obstacle can be surmounted through the use of an alternative representation, obtained by factorizing the data likelihood in a multiscale fashion under the assumed Poisson model.

Specifically, let $\mathbf{Y}_j = (Y_{j,1}, \ldots, Y_{j,N_j})^T$ and let $\boldsymbol{Y}_{j+1,ch(k)}$ be those measurements $Y_{j+1,k'}$ for whom $B_{j+1,k'} \subseteq B_{j,k}$. Define $\boldsymbol{\mu}_j$ and $\boldsymbol{\mu}_{j+1,ch(k)}$ similarly. Then one can write

$$\Pr\left(\boldsymbol{Y}_J \,|\, \boldsymbol{\mu}_J\right) = \Pr\left(Y_{0,1}|\mu_{0,1}\right) \prod_{j=0}^{J-1} \prod_{k=1}^{N_j} \Pr\left(\boldsymbol{Y}_{j+1,ch(k)} \,|\, Y_{j,k}, \boldsymbol{\omega}_{j,k}\right) \quad , \tag{3}$$

where $Y_{0,1}|\mu_{0,1} \sim \text{Poisson}(\mu_{0,1})$ and $\boldsymbol{Y}_{j+1,ch(k)} \,|\, Y_{j,k}, \boldsymbol{\omega}_{j,k} \sim \text{Multinomial}(Y_{j,k}\,;\, \boldsymbol{\omega}_{j,k})$, with

$$\boldsymbol{\omega}_{j,k} = \mu_{j,k}^{-1} \times \boldsymbol{\mu}_{j+1,ch(k)} \quad . \tag{4}$$

The derivation of this result is straightforward, and may be found in [4]. Multiscale representations like that above are analogous to orthonormal wavelet basis decompositions in many ways, and can be used to create methods of inference with analogous statistical risk properties [8]. Note that the expression in (3) illustrates how the information in the original measurements at the finest scale i.e., in $\boldsymbol{Y}_J$, can be effectively separated out into independent components as those measurements are aggregated across each successively coarser scale. Furthermore, each component $\boldsymbol{Y}_{j+1,ch(k)}|Y_{j,k}$ is equipped with its own unique parameter $\boldsymbol{\omega}_{j,k}$, which describes the corresponding local structure in the intensity process $\mu(\cdot)$. Together, this re-expression of data and parameters allows for the simultaneous localization of information in both, relative to each $B_{j,k}$ and its subregions.

2.2. Multiscale Prior Probability Structure

Suppose that the underlying intensity process μ is locally constant across much of the spatial region D. Then the values of many of the aggregate intensity parameters $\mu_{j,k}$ will likely be proportional to the areas $|B_{j,k}|$, and hence in turn many of the multiscale parameters $\boldsymbol{\omega}_{j,k}$ in (4) should be approximately equal to the relative areas of the subregions $B_{j+1,k'}$ in $B_{j,k}$, for $k' \in ch(k)$. In taking a Bayesian approach to inference on μ, this knowledge provides strong motivation to define the associated prior probability structure on the multiscale parameterization, instead of on the original spatial parameterization, as has been done in [4]. A similar reasoning applies when the object of interest is relative risk, and not intensity, and we define an analogous prior accordingly.

First, we introduce a measure of relative risk into the model, at the finest scale J, by re-expressing the elements of the mean vector $\boldsymbol{\mu}_J$ as $\mu_{J,k} = \theta_{J,k}\, e_{J,k}$, where $e_{J,k}$ denotes the expected counts in the 'at-risk' population in $B_{J,k}$. The parameter $\theta_{J,k} = \mu_{J,k}/e_{J,k}$ therefore represents the relative risk in $B_{J,k}$. More generally, we define the relative risk in $B_{j,k}$, for arbitrary scales j, as $\theta_{j,k} = \mu_{j,k}/e_{j,k}$, where the values $e_{j,k}$ are defined recursively through the expression $e_{j,k} = \mathbf{1}^T \boldsymbol{e}_{j+1,ch(k)}$, in analogy to the expressions in (2). We will assume the $e_{J,k}$, and hence the entire collection $\{e_{j,k}\}$, to be known or, more realistically, to have been assessed separately.

Next, we specify independent prior distributions for the parameters $\boldsymbol{\omega}_{j,k}$ in the form

$$\boldsymbol{\omega}_{j,k} \sim Dirichlet(c_j T \boldsymbol{e}_{j+1,ch(k)}). \tag{5}$$

Here T is used to denote the global relative risk for D, and the $c_j > 0$ are scale-dependent hyperparameters. These hyperparameters may be considered either fixed or random, and we explore both cases in this paper. In the latter case, we model the c_j themselves as independent draws from a Uniform$(0, c_{max})$ distribution, for some $c_{max} \gg 0$.

Lastly, it remains to specify a prior distribution for $\mu_{0,1}$. The distribution for such parameters typically is left only vaguely specified or completely unspecified in the multiscale modeling literature, due to the fact that the high degree of aggregation at coarser scales leads the data to dominate posterior inferences under any reasonable choice of prior. Here we will proceed similarly and simply specify that the prior distribution of $\mu_{0,1}$ be diffuse and independent of the $\boldsymbol{\omega}_{j,k}$'s, and that it have mean $E[\mu_{0,1}] = T\, e_{0,1}$.

Our choice of the Dirichlet distribution for the $\boldsymbol{\omega}_{j,k}$ is motivated primarily by its conjugacy to the multinomial distribution governing the components $\boldsymbol{Y}_{j+1,ch(k)}|Y_{j,k}$. The motivation for our choice of parameterization can be seen by noting that, due to the independence specified between the $\boldsymbol{\omega}_{j,k}$ (conditionally, given the c_j, when the c_j are random) and the scale-recursive nature of (4), the prior expectation for the intensities $\mu_{j,k}$ is just

$$E[\mu_{j,k}] = E[\mu_{0,1}] \prod_{j'=0}^{j-1} E[\omega_{j',a_{j'}(k)}] = T\, e_{j,k} \ , \tag{6}$$

where $a_{j'}(k)$ is used here and throughout the paper generically to denote the location index of the ancestor of (j,k) at scale j'. That is, for a given (j,k), the subregion $B_{j',a_{j'}(k)}$ is the unique subregion at the coarser scale $j' < j$ that contains $B_{j,k}$. As a result of (6), we have $E[\theta_{j,k}] = E[\mu_{j,k}]/e_{j,k} \equiv T$ uniformly for all (j,k). In other words, the model we have defined in the space of multiscale parameters $\boldsymbol{\omega}_{j,k}$ corresponds to an induced model for the intensity process $\mu(\cdot)$, through scales $j \leq J$, such that upon normalization by the $e_{j,k}$ the corresponding distribution on the relative risk parameters is homogeneous in expectation at each scale.

It is important to note that under our model, although the multiscale parameters $\boldsymbol{\omega}_{j,k}$ are modeled as independent, the corresponding relative risk parameters $\theta_{j,k}$ are spatially correlated within each scale and also across scales. Similarly, the original measurements $Y_{J,1}, \ldots, Y_{J,N_J}$ are correlated (i.e., marginally). Generally speaking, the closer the nearest common ancestor (as a function of scale) of two subregions, say B_{J,k_1} and B_{J,k_2} the more highly correlated their measurements Y_{J,k_1}, Y_{J,k_2}. Scale thus serves as a proxy for distance in determining correlation in our model. The reader is referred to [5] for a detailed analytic examination of the correlation properties of $\boldsymbol{\mu}_J$ under models analogous to that defined here, for the case of fixed c_j, from which properties of our own model follow immediately. In particular, the role of the hyperparameters c_j is made clear and found to be not unlike that played by the canonical parameters in a standard Dirichlet process model.

3. Multiscale Estimation of Relative Risk

We propose to estimate the collection of relative risk parameters $\theta_{j,k} = \mu_{j,k}/e_{j,k}$ through the use of similarly normalized posterior expectations of the intensities $\mu_{j,k}$. We present two such multiscale SMR estimation strategies, distinguished by whether the hyperparameters c_j are taken as fixed or random.

3.1. Empirical Bayes Estimation: Fixed c_j

Consider first the case where the c_j are taken as fixed and known. Recall that our model then consists of (i) independent multiscale Dirichlet priors on the parameters $\boldsymbol{\omega}_{j,k}$, and (ii) a Poisson sampling model, given the finest scale intensity parameters $\boldsymbol{\mu}_J$ induced by the multiscale parameters $\boldsymbol{\omega}_{j,k}$. A key asset of multiscale factorizations like that in (3) is that, when combined with independent prior distributions on the multiscale parameters, the posterior is subject to an analogous factorization. In the context of the present paper, we have for our posterior of the intensity parameters at a given scale j the expression

$$p\left(\boldsymbol{\mu}_j \mid \boldsymbol{Y}_j, \mathbf{c}\right) = p\left(\mu_{0,1} \mid Y_{0,1}\right) \prod_{j'=0}^{j-1} \prod_{k'=1}^{N_{j'}} p\left(\boldsymbol{\omega}_{j',k'} \mid \boldsymbol{Y}_{j'+1,ch(k')}, c_{j'}\right) \quad , \tag{7}$$

where

$$\boldsymbol{\omega}_{j',k'} \mid \boldsymbol{Y}_{j'+1,ch(k')}, c_{j'} \sim Dirichlet(c_{j'} T \boldsymbol{e}_{j'+1,ch(k')} + \boldsymbol{Y}_{j'+1,ch(k')}) \tag{8}$$

and recall that $ch(k')$ denotes the location indices of the children of subregion $B_{j',k'}$. Therefore, we find that our proposed estimator takes the form

$$\hat{\theta}^{EB}_{j,k} \equiv \frac{E[\mu_{j,k} \mid \boldsymbol{Y}_j, \mathbf{c}]}{e_{j,k}} = \frac{E[\mu_{0,1} \mid Y_{0,1}]}{e_{j,k}} \times \left[\prod_{j'=0}^{j-1} \frac{c_{j'}\ T\ e_{j'+1,a_{j'+1}(k)} + Y_{j'+1,a_{j'+1}(k)}}{c_{j'}\ T\ e_{j',a_{j'}(k)} + Y_{j',a_{j'}(k)}} \right] \quad . \tag{9}$$

The entire collection $\{\hat{\theta}^{EB}_{j,k}\}$ may be calculated using a recursive algorithm of just $O(N_J)$ complexity, given values for T and the c_j. We have found that useful values for these parameters often may be elicited from the data. For example, the global SMR i.e., $SMR_{0,1} = Y_{0,1}/e_{0,1}$, is a natural estimate for T. And in the case of the c_j, we have used the method of marginal maximum likelihood in this paper. The relevant likelihood formulas for the latter are straightforward to derive, and their optimization may be accomplished through the use of either a modified Newton-Raphson algorithm (as we have done for the implementations in Sections 5 and 6) or an EM algorithm (as is detailed in [12]). The resulting estimators $\hat{\theta}^{EB}_{j,k}$ are therefore empirical Bayes in nature. We note that in a purely spatial (single scale) Bayesian analysis of disease maps, empirical Bayes methods have been shown to at times underestimate the variability in the relative risks, since the uncertainty associated with the hyperparameters is unaccounted for [13]. We therefore provide a fully Bayesian implementation of our multiscale model (as found in the subsequent Section 3.2), and we conduct a simulation study aimed at evaluating the performance of our estimator under both empirical Bayes and fully Bayesian analysis. Details and results from the simulation study can be found in Section 5.

Note that due to the recursive nature of our estimators, a sequence of disease maps generated by the collection $\{\hat{\theta}^{EB}_{j,k}\}$ will be coupled in a natural manner, with each map being a refinement of those maps at coarser scales before it. In order to better interpret the quantities $\hat{\theta}^{EB}_{j,k}$, we can re-express them as follows. Define $SMR_{j,k} \equiv Y_{j,k}/e_{j,k}$ to be the canonical SMR statistic for the data at scale j in location k i.e., the standard estimate of the relative risk $\theta_{j,k}$ on the subregion $B_{j,k}$. Normalizing numerator and denomenator in (9) by $e_{j'+1,a_{j'+1}(k)}$ and $e_{j',a_{j'}(k)}$, respectively, and noting the cancelation of the accompanying ratios $e_{j',a_{j'}(k)}/e_{j'+1,a_{j'+1}(k)}$

successively across scales, the expression in (9) becomes

$$\hat{\theta}^{EB}_{j,k} = \frac{E[\mu_{0,1} \mid Y_{0,1}]}{e_{0,1}} \times \left[\prod_{j'=0}^{j-1} \frac{c_{j'}\ T + SMR_{j'+1,a_{j'+1}(k)}}{c_{j'}\ T + SMR_{j',a_{j'}(k)}} \right] \quad . \tag{10}$$

Additionally, recalling our specification of a diffuse prior for $\mu_{0,1}$, it follows that $E[\mu_{0,1}|Y_{0,1}] \approx Y_{0,1}$, and therefore the first term in (10) is approximately equal to $SMR_{0,1} = Y_{0,1}/e_{0,1}$. So each estimate $\hat{\theta}^{EB}_{j,k}$ essentially begins with an estimate of the global relative risk and modifies that in a multiplicative fashion across scales $j' < j$, from coarse to fine. Ignoring the terms $c_{j'}T$, the multiplicative factors in (10) can be read as ratios of the SMR's for the ancestral subregions of each $B_{j,k}$. If consecutive SMR's are equal, then the corresponding factor contributes no change in moving between scales j' and $j'+1$. The terms $c_{j'}T$ in the numerator and denominator relate directly to the degree of uncertainty associated with our specification of prior information at scale j', and play the role of shrinkage factors.

Three special cases are of interest. First, if $c_j \to 0$ for all j, then $\hat{\theta}^{EB}_{j,k} \to SMR_{j,k}$ for each (j,k). That is, our estimators tend to the canonical SMR statistics at each scale. Second, if $c_j \to \infty$ for all j, then $\hat{\theta}^{EB}_{j,k} \to SMR_{0,1}$ for all (j,k). In other words, our estimators return the same measure of global relative risk for all scales and locations. Lastly, if $c_j \equiv 1$ for all j, and if $SMR_{0,1} \approx T$, then $\hat{\theta}^{EB}_{j,k} \approx (SMR_{0,1} + SMR_{j,k})/2$ for each (j,k), which is just an average of the two previous cases. More generally, if the c_j are quite large for some scales j and more modest for others, then at those scales for which they are large, the corresponding factor will be essentially equal to one, and as a result the refinement to estimates at finer scales than those will be negligible.

3.2. Fully Bayesian Inference: Random c_j

Now consider the case where the hyperparameters in $\mathbf{c}$ are random. Recall that our full model then consists of (i) independent hyperparameters c_j, drawn with respect to a Uniform$(0, c_{max})$ distribution, (ii) independent multiscale Dirichlet priors on the parameters $\boldsymbol{\omega}_{j,k}$, given the c_j, and (iii) a Poisson sampling model, given the finest scale intensity parameters $\boldsymbol{\mu}_J$ induced by the multiscale parameters $\boldsymbol{\omega}_{j,k}$.

The relevant posterior under this model has the form

$$p\left(\boldsymbol{\mu}_j \,|\, \boldsymbol{Y}_j\right) = \int p\left(\boldsymbol{\mu}_j \,|\, \boldsymbol{Y}_j, \mathbf{c}\right) p\left(\mathbf{c} \,|\, \boldsymbol{Y}_j\right) d\mathbf{c} \ , \tag{11}$$

which is just a combination of the conditional posteriors $p\left(\boldsymbol{\mu}_j \,|\, \boldsymbol{Y}_j, \mathbf{c}\right)$, weighted by the posterior evidence for each possible value of $\mathbf{c}$. However, note that while the conditional posteriors inside the integral in (11) still possess the factorization (7), the densities $p\left(\mathbf{c} \,|\, \boldsymbol{Y}_j\right)$ have no such property, and therefore the overall posterior $p\left(\boldsymbol{\mu}_j \,|\, \boldsymbol{Y}_j\right)$ does not either.

We again propose to estimate the relative risk parameters $\theta_{j,k} = \mu_{j,k}/e_{j,k}$ by the analogous normalization of the posterior expectation, yielding the estimators $\hat{\theta}^{FB}_{j,k} = E[\mu_{j,k}|\boldsymbol{Y}_j]/e_{j,k}$. In order to calculate these values, we can sample from the posteriors $p\left(\boldsymbol{\mu}_j \,|\, \boldsymbol{Y}_J\right)$, at each scale j, using an MCMC algorithm. We note that the factorization of the conditional posterior $p(\boldsymbol{\mu}_j \mid \boldsymbol{Y}_j, \mathbf{c})$, and the corresponding Dirichlet components in (7), result in easy composition sampling. Given our MCMC samples, the values $\hat{\theta}^{FB}_{j,k}$ then follow upon dividing the empirical marginal averages by the values $e_{j,k}$.

While the estimators $\hat{\theta}^{FB}_{j,k}$ of this section are not accompanied by nice analytic expressions, like (9) and (10) for $\hat{\theta}^{EB}_{j,k} \equiv \hat{\theta}^{EB}_{j,k}(\mathbf{c})$, intuitively it can be seen that there is a similar sharing of information across scales, since $\hat{\theta}^{FB}_{j,k} = \int \hat{\theta}^{EB}_{j,k}(\mathbf{c})\, p(\mathbf{c}|\boldsymbol{Y}_J)\, d\mathbf{c}$ is an average of the conditional posterior expectations.

4. Quantification of Uncertainty

Having developed posterior-based, multiscale estimates of relative risk, we now turn our attention to quantifying the uncertainty associated with these estimates. We take as our goal the calculation of central marginal posterior intervals for the $\theta_{j,k}$. Such intervals can then be used to create maps accompanying a basic disease map that indicate upper and lower levels of uncertainty.

The logical manner in which to obtain posterior intervals is through numerical sampling of the posterior distribution. In the case of random hyperparameters c_j, the intervals may be constructed from the same samples of $\Pr(\mu_{j,k} \,|\, \boldsymbol{Y}_j)$ that were used to obtain the estimates $\hat{\theta}^{FB}_{j,k}$. In the case of fixed c_j, the recursive form of equation (7) may be exploited to sample directly from the posteriors $\Pr(\boldsymbol{\mu}_j \,|\, \boldsymbol{Y}_j, \mathbf{c})$. Specifically, drawing samples from the posterior of each of the multiscale parameters $\boldsymbol{\omega}_{j,k}$, using the appropriate Dirichlet distributions dictated by (8), samples from the marginal posterior for each intensity vector $\boldsymbol{\mu}_j$ can be constructed through recursive application of (4). Intervals for $\theta_{j,k}$ then follow from normalizing intervals for $\mu_{j,k}$ by $e_{j,k}$.

On the other hand, there are contexts in which an approach based on posterior simulation may be computationally prohibitive. For example, for a user of a geographical information system wishing to generate a fluid succession of disease maps in an interactive fashion, such an approach can present a bottleneck to the overall process. Therefore, since the calculation of the empirical Bayes estimates $\hat{\theta}^{EB}_{j,k}$ is itself computationally efficient, we present here a computationally efficient method for calculating accurate approximations to the corresponding posterior intervals, based on the use of Johnson curves [14].

The system of Johnson curves is similar in principle to that of the frequency curves developed by Pearson. (See [15].) Consisting of tranformations of the standard normal distribution, the Johnson curves fall into three categories – termed lognormal, unbounded, and bounded – and can be expressed as densities

$$g(x) = \delta\, f'((x-\xi)/\lambda)\; p(z)\big|_{z=\gamma+\delta f(\frac{x-\xi}{\lambda})} \quad . \tag{12}$$

Here x is the Johnson variable, z is the standard normal variable, and $f((x-\xi)/\lambda)$ is $log\{(x-\xi)/\lambda\}$, $sinh^{-1}\{(x-\xi)/\lambda\}$, and $log\{(x-\xi)/(\xi+\lambda-x)\}$ for the lognormal, unbounded and bounded Johnson curves respectively. The parameter ξ is a location factor and λ a scale factor, while γ and δ help determine the shape of the distribution of x. The solution for these four parameters with respect to a defined relationship with the cumulants of a given distribution, results in an analytic approximation to the latter.

For our purposes, the distribution of interest is a given marginal posterior distribution $\theta_{j,k}|\boldsymbol{Y}_j, \mathbf{c}$, for which we have found the lognormal Johnson curve may be usefully employed. To sketch the method, define, as is standard, $\beta_1^{1/2} = \kappa_3/\kappa_2^{3/2}$ and $\beta_2 = \kappa_4/\kappa_2^2$, where $\kappa_r \equiv \kappa_r(\theta_{j,k}|\boldsymbol{Y}_j)$ denotes the r-th cumulant of $\theta_{j,k}$, conditional on $\boldsymbol{Y}_j$ and $\mathbf{c}$. Then the

lognormal curves lie on the line in the (β_1, β_2) plane where

$$\beta_1 = (\tau - 1)(\tau + 2)^2 \quad \text{and} \quad \beta_2 = \tau^4 + 2\tau^3 + 3\tau^2 - 3 \ , \tag{13}$$

for $\tau = e^{\delta^{-2}}$. For these curves, knowing β_1 determines β_2. The parameters of the lognormal Johnson curves are then related to τ and the cumulants through the equations

$$\begin{array}{rclcrcl} \delta & = & (ln\ \tau)^{-\frac{1}{2}}, & & \gamma & = & \frac{1}{2} ln\{\tau(\tau-1)/\kappa_2\} \\ \xi & = & \pm\kappa_1 - exp\{(1/2\delta - \gamma)/\delta\}, & & \lambda & = & \pm 1, \end{array} \tag{14}$$

where the $\pm$ is the sign of κ_3 [16].

Values for the cumulants κ_r of $\theta_{j,k}|\boldsymbol{Y}_j, \mathbf{c}$ are necessary to solve for the Johnson curve parameters. Under our model, these may be obtained in closed form from the cumulant generating function, the details for which are provided in the appendix. Having solved for these parameters, in order to obtain $100(1-\alpha)\%$ approximate central marginal posterior intervals for $\theta_{j,k}$, we need only solve the equation

$$x = \xi + \lambda f^{-1}\{(\pm z_{\alpha/2} - \gamma)/\delta\}, \tag{15}$$

where $z_{\alpha/2}$ is the $\alpha/2$ percentile of the standard normal distribution and f is as defined above for the lognormal curve. Simulation studies have found the approximation to be quite accurate [12].

Details on the use of Johnson curves, including efficient iterative algorithms for performing the necessary calculations, may be found in Hill, Hill and Holder [16], Elderton and Johnson [15], and Draper [17]. We note too that Barber, Nason and Silverman [18] have used this same approach to approximating posterior confidence intervals in the context of wavelet shrinkage estimation, and would appear to be the first to have done so in the literature on multiscale statistical methods. They cite reasons similar to those given here for pursuing this approach. However, the details necessarily differ and in fact they find, for technical reasons, that the cumulants in their setting must be estimated. These authors too find the method to yield quite accurate approximations.

5. Simulation Study

We conducted a simulation study, aimed at evaluating the performance of our multiscale SMR framework under a handful of scenarios. For simplicity, the initial dataspace D was taken to be a square region, and the nested hierarchy $\mathcal{B}^{(J)}$ was defined through the generic quad-tree structure common in image processing and computer vision. Specifically, D was partitioned into $2^j \times 2^j = 2^{2j}$ identical subregions, for $j = 0, 1, \ldots, 4$, so that each square subregion at scale j consisted of four sub-subregions at scale $j+1$. At the finest scale $J = 4$, there were a total of $16 \times 16 = 256$ subregions, which we refer to as 'pixels', as is standard.

Simulations were conducted under a total of eleven different scenarios of landscape design. The first landscape has uniform relative risk; that is, each pixel has a relative risk of one. The next nine landscapes each contain a single area of elevated relative risk. This area, which we will refer to as A4, is comprised of $4 \times 4 = 16$ pixels, each pixel with a relative risk of 2.5. The nine landscapes differ in the location of A4 within each. The different locations were chosen so as to allow for the examination of the effect on our multiscale framework of different

degrees of nesting of the elevated area within the elements of the nested hierarchy. The final landscape design combines five different patterns of elevated risk found in the aforementioned 9 landscapes.

Figure 1 illustrates all 11 landscape designs. If we call the pixels at scale $J = 4$ 'children', and subregions at scales $j = 3, 2$, and 1, 'parents', 'grandparents', and 'great-grandparents', respectively, then landscapes 2 through 10 may be characterized as follows. Landscapes 2 and 3 show A4 on the border and in the interior respectively, each of whose 16 children form four parents and a single grandparent. The A4 in landscapes 4 and 5 each form four parents who originate from two distinct grandparents. Landscapes 6, 7 and 8 each form four parents from four different grandparents who in turn are from one, two or four great-grandparents respectively. The A4 in landscapes 9 and 10 have the least amount of nesting (none, in fact, in the case of landscape 10). In these two landscapes, the children originate from nine distinct parents, four distinct grandparents, and one and four distinct great-grandparents, respectively.

We chose the pixel-level expected counts $e_{J,k}$ to be uniformly equal to 5, which resulted in pixel-level mean parameters $\mu_{J,k} = \theta_{J,k}\, e_{J,k}$ of either 5 or 12.5. We then generated 100 independent Poisson samples with intensity $\mu_{J,k}$ for every location k, resulting in 100 datasets of 256 observations for each of the eleven landscapes on D. For each dataset, for each landscape, the empirical Bayes and fully Bayes multiscale SMR estimators $\hat{\theta}^{EB}_{j,k}$ and $\hat{\theta}^{FB}_{j,k}$ were calculated at the finest scale $j = J = 4$, according to the formulas and methods described in Sections 3.1 and 3.2, respectively.

By way of comparison, we also computed estimates of the relative risks under a hierarchical Bayes Poisson-lognormal model. Specifically, in this model $Y_k \sim Poisson(\mu_k)$, as in our multiscale model, but the log relative risk is expressed as $log\ \mu_k = log\ e_k + \alpha + u_k + v_k$, where α captures the overall level of the log relative risk, and u_k and v_k are respectively the unstructured heterogeneity and clustering components of extra-Poisson variation. Under this framework, u_k is modeled as a zero mean Gaussian, and v_k as a conditionally autoregressive (CAR) Gaussian. Specifying a CAR prior for v_k allows one to model the assumption that pixels near each other will tend to be more similar than pixels that are farther apart. That is, the spatial correlation structure is based on the adjacency of pixels as opposed to the nested hierarchy. Further details on this type of model can be found in [19]. The posterior estimator under this model takes the form

$$\hat{\theta}^{PL}_k \quad = \quad E[\mu_k \mid \boldsymbol{Y}]/e_k. \tag{16}$$

Note that this model is location-sensitive only, being defined entirely at the finest scale J.

The performance of each of our three estimators was evaluated using an (approximate) integrated mean-squared-error (IMSE) criterion i.e.,

$$IMSE \quad = \quad \frac{1}{16^2} \times \frac{1}{100} \times \sum_{k=1}^{16^2} \sum_{i=1}^{100} (\hat{\theta}_k(i) - \theta_{J,k})^2, \tag{17}$$

where $\hat{\theta}_k(i)$ corresponds to the estimator $\hat{\theta}^{EB}_{J,k}$, $\hat{\theta}^{FB}_{J,k}$, or $\hat{\theta}^{PL}_k$ at the i-th trial. The resulting IMSE numbers, as well as their breakdown into bias and variance components, may be found in Table I. Similarly, a plot of the IMSE numbers alone is shown in Figure 2.

Consider first just the two multiscale estimators. While there is a slight improvement in performance from the fully Bayesian over the empirical Bayes analysis in the absence of any elevated risk (i.e., landscape 1), the empirical Bayes estimator consistently out-performs the

fully Bayesian estimator in the presence of elevated risk (i.e., landscapes 2 through 11). For both estimators, in the presence of elevated risk, the performance generally improves as the nesting of A4 within the hierarchy moves closer to the pixel level (i.e., roughly moving from right to left in Figure 2).

In comparison, since the spatial correlation for the CAR prior depends on the adjacency of pixels rather than on a nested hierarchy, we would expect little to no effect on the Poisson-lognormal estimator when the pattern of adjacencies remains the same. This is confirmed by the closeness of the results for landscapes 3 through 10. Note that the A4 in landscape 2 is located on the border, which changes the number of adjacencies, since the pixels have less neighbors than those of an A4 located in the interior. Similar to the multiscale estimators, landscape 11 exhibits the highest IMSE.

In comparing the multiscale estimators and the Poisson-lognormal estimator, with the exception of landscape 7, for $\hat{\theta}^{FB}_{j,k}$ in the fully Bayesian analysis, the former were found to out-perform the latter when the the spatial patterns of elevation were nested to some degree (i.e., landscapes 2 through 8), but the reverse was true when the spatial patterns were primarily or entirely non-nested (i.e., landscapes 9 and 10). In landscape 11, which combines both nested and non-nested areas of elevation, the multiscale estimators outperform the Poisson-lognormal estimator. Overall, these results suggest that our multiscale SMR framework offers estimators at the finest scale of observation that are competitive with standard monoscale estimators, while at the same time being accompanied by an entire hierarchy of inter-related estimators at the other coarser scales.

6. Analysis of Tuscany Gastric Cancer Mortality Data

We applied our multiscale method to male gastric cancer mortality data obtained from the Tuscan region of Italy during the period 1980-1989 [2, 20]. For our hierarchy of nested partitions we took the set of nested subregions displayed in Figure 3, corresponding to three of the five geo-political statistical units defined under the NUTS system (Nomenclature des Units Territoriales Statistiques) used by the European community. In particular, shown in Figure 3 are the single Tuscan region itself (NUTS Level II), the nine provinces within that region (NUTS Level III), and the 287 municipalities within those provinces (NUTS Level V). Italy itself corresponds to a single subregion at NUTS Level I, while NUTS Level IV is not used in that country. In the notation of this paper, the regional, provincial, and municipality levels correspond to $j = 0, 1$, and 2, respectively. Data were obtained for males over 35 years of age at the municipality level. Italian age specific rates for the same calendar period were used to obtain expected numbers of deaths [20].

Disease maps summarizing the empirical Bayes estimates $\hat{\theta}^{EB}_{j,k}$ of the relative risks $\theta_{j,k}$, at the municipal ($j = 1$) and provincial ($j = 2$) levels, are shown in Figures 4a and 4b, respectively. From Figure 4a it can be seen that there is evidence for some rough sense of a trend in relative risk across provinces, increasing from southwest to northeast and achieving a maximum in what is the province of Arrezo. Figure 4b provides a refinement of this information (i.e., recall the scale-recursive nature of (10)), through which it is found that the increased relative risk is particularly concentrated in spots along the northeast border, which is contained primarily within Arrezo but also partially within the neighboring province of Firenze to the northwest. This phenomenon was noted in [20] and judged not to be simply a distortion in rates due to the

proximity of those municipalities to the edge of the study boundary. We refer the reader to [2, Ch. 5.5] for background on the topic of edge effects and discussion of this particular dataset. Exact values for the empirical Bayes estimates of relative risk, as well as estimates from a full-Bayesian analysis and a Poisson-lognormal model, for the twelve municipalities along the northeast border, are presented in Table II. Although empirical Bayes methods have been shown to overshrink in some instances [13], the phenomenon is not reflected in the analysis of this dataset, as estimates from the fully Bayesian implementation are extremely close to the empirical Bayes estimates.

In Figures 5a and 5b are shown maps of the lower and upper bounds, respectively, of 95% central marginal posterior intervals for the relative risk parameters $\theta_{j,k}$ at the provincial level, computed using the method of approximation described in Section 4. The corresponding maps of the lower and upper bounds at the municipal level are shown in Figures 6a and 6b. The map in Figure 6b confirms the potential severity of the disease along the northeast border, as the municipalities throughout that entire area are assigned upper bounds on relative risk in excess of 2.3. The map of lower bounds in Figure 6a paints a somewhat less grim picture for this area, wherein most of the municipalities in Arrezo and part of neighboring Firenze are indicated as having at least slight elevations in risk (e.g., 1.15 - 1.72). However, there is also still persistent evidence for a more severe level of elevation in the two border municipalities of Poppi and Bibbiena (i.e., 1.92 - 2.10). At the provincial level, while Figures 5a and 5b do not show a pronounced elevated risk in Firenze and Arrezo, the map of the upper bounds still show Arrezo as having the highest elevated risk. Lower and upper bounds for all three methods of analysis, for the twelve municipalities along the northeast border, are presented in Table II.

7. Discussion

We have introduced in this paper a framework within which it is possible to derive an interrelated sequence of informative disease and confidence maps across a hierarchy of multiple spatial scales. We have concentrated our development of this framework on the case of tract count data, which is arguably the most commonly encountered in spatial epidemiology. The tools we provide here may be viewed as multiscale extensions of the canonical SMR statistic. We note, however, that our framework can be extended easily for estimating other measures of relative risk in Poisson-based measurement models. For example, the relevant extension to data arising in case-to-control studies may be found in [12].

There are various other manners in which it might be of interest to enhance the framework we offer. First, as currently posed, our methods do not explicitly adjust for the potential problem of edge-effects, caused by the use of a domain D that typically is artificially finite with respect to actual underlying disease mechanisms. Simple methods of re-weighting the original count data, based on some measure of distance of each region at scale J to the external boundary, can likely be applied here successfully. More interesting is the question of whether and how best to re-weight the data at more than just the original scale of measurement. It also seems possible that methods based on external guard areas could be similarly extended, although this would require that one consider how best to expand the original hierarchy of partitions.

Second, there is the issue of how to implement a multiscale analysis like that provided here when there are multiple and/or non-nested spatial hierarchies (e.g., [21]). A step in this direction likely would require that one begin with the use of multiple likelihood factorizations

and/or an extension of the basic multiscale likelihood factorization itself.

Third, it is natural to conceive of certain standard augmentations to the model, such as the incorporation of parameters for modeling over-dispersion. One might simply include a single, global over-dispersion parameter as a multiplicative factor of the variance of the Poisson sampling model or, alternatively, similarly modify the variances of each of the multinomial components in the multiscale likelihood factorization by a set of common scale-indexed over-dispersion parameters. The latter approach would be equivalent to specifying a single over-dispersion parameter for the Poisson sampling model that was in fact a particular multiplicative function of a sequence of scale-indexed parameters.

Fourth, on a related point, we note that our proposed methodology, with its emphasis on extensions of SMR, is by no means immune to issues of standardization like those raised in [23], and that the indirect standardization (as those authors call it) inherent in SMR-like statistics requires validity of an assumption of multiplicative structure that may or may not be valid. However, as the SMR is nevertheless a common, standard estimator, it is a natural place to start in introducing multiscale estimation strategies into disease mapping. The impact of standardization methods, and the possibility of extensions of our methodology to other standardizations, merits further study.

Finally, we point out that while our focus here has been on the particular task of disease mapping in spatial epidemiology, another task of equal importance is that of disease clustering i.e., the analysis of unusual aggregations of disease. This task generally is pursued using a hypothesis testing framework, as opposed to an estimation framework [22]. However, if a likelihood-based perspective is adopted, one may exploit the same multiscale likelihood factorization that underlies the work in this paper to develop a multiscale approach to disease clustering. The multiscale re-parameterization i.e., the mapping from means $\mu_{J,k}$ to parameters $\omega_{j,k}$, allows for an analogous recasting in the multiscale parameterization of hypotheses in the original spatial parameterization. It is shown in [12] how a Bayesian hypothesis testing approach may be used to derive a framework for disease clustering that should be sensitive to anomalies in both location and scale.

APPENDIX: CUMULANTS FOR LOG-$\boldsymbol{\mu}_J$

Without loss of generality (and for simplicity), we will evaluate the CGF of $log\,\boldsymbol{\mu}_j$, instead of $\boldsymbol{\mu}_j$, conditional on $\mathbf{Y}_j$ and $\mathbf{c}$. Let $\mu_{0,1} \mid Y_{0,1} \sim Gamma(\alpha + Y_{0,1}; \beta + 1)$, so that

$$M_{log\,\mu_{0,1}}(t) = (\beta+1)^{-t} \times \frac{\Gamma(\alpha+t)}{\Gamma(\alpha)}. \tag{18}$$

Similarly, the MGF of $log\,\boldsymbol{\omega}_{j,k}$ conditional on $\mathbf{Y}_j$ and $\mathbf{c}$, when $\boldsymbol{\omega}_{j,k} \mid \mathbf{Y}_j, \mathbf{c} \sim Dirichlet(c_j\ T\ \mathbf{e}_{j+1,ch(k)} + \mathbf{Y}_{j+1,ch(k)})$ is

$$\begin{aligned} M_{log\,\boldsymbol{\omega}_{j,k}}(\mathbf{t}) \;=\; & \frac{\Gamma(c_j\ T\ e_{j,k} + Y_{j,k})}{\prod_{i=1}^{n_{j,k}} \Gamma(c_j\ T\ e_{j+1,ch(k,i)} + Y_{j+1,ch(k,i)})} \times \\ & \frac{\prod_{i=1}^{n_{j,k}} \Gamma(c_j\ T\ e_{j+1,ch(k,i)} + Y_{j+1,ch(k,i)} + t_i)}{\Gamma(c_j\ T\ e_{j,k} + Y_{j,k} + \mathbf{1}_{n_{j,k}}^T \mathbf{t})}, \end{aligned} \tag{19}$$

where $(j+1, ch(k,i))$ denotes the ith child of parent (j,k) at scale $j+1$, and $n_{j,k}$ is the number of subregions at scale $j+1$ contained in $B_{j,k}$. It is not difficult to show that the MGF for

$log\,\boldsymbol{\mu}_j$ decouples across all scales and locations, that is,

$$M_{log\,\boldsymbol{\mu}_j}(\mathbf{t}) = c \times \frac{\Gamma(\alpha + Y_{0,1} + \mathbf{1}^T\mathbf{t})}{(\beta+1)^{\mathbf{1}^T\mathbf{t}}} \times \prod_{j'=0}^{j-1}\prod_{k'=1}^{N_{j'}} \frac{\prod_{i=1}^{n_{j',k'}} \Gamma(c_{j'}\ T\ e_{j'+1,ch(k',i)} + Y_{j'+1,ch(k',i)} + \mathbf{1}^T\mathbf{t}_{j',k',i})}{\Gamma(c_{j'}\ T\ e_{j',k'} + Y_{j',k'} + \sum_{i=1}^{n_{j',k'}} \mathbf{1}^T\mathbf{t}_{j',k',i})}, \quad (20)$$

where c is a constant with respect to $\mathbf{t}$, and the number of elements in $\mathbf{t}_{j',k',i}$ reflects the number of subregions of $B_{j',k'}$ at scale j. The CGF is then given by

$$K_{log\,\boldsymbol{\mu}_j}(\mathbf{t}) = log\,c + log\,\Gamma(\alpha + Y_{0,1} + \mathbf{1}^T\mathbf{t}) - \mathbf{1}^T\mathbf{t}\ log\,(\beta+1) + \sum_{j'=0}^{j-1}\sum_{k'=1}^{N_j}\left[\sum_{i=1}^{n_{j',k'}} log\,\Gamma(c_{j'}\ T\ e_{j'+1,ch(k',i)} + Y_{j'+1,ch(k',i)} + \mathbf{1}^T\mathbf{t}_{j',k',i}) - log\,\Gamma\left(c_{j'}\ T\ e_{j',k'} + Y_{j',k'} + \sum_{i=1}^{n_{j',k'}} \mathbf{1}^T\mathbf{t}_{j',k',i}\right)\right], \quad (21)$$

with the r-th marginal cumulant of $log\,\mu_{j,k}$ as

$$\left.\frac{d^r K_{log\,\mu_{j,k}}(\mathbf{t})}{dt_k^r}\right|_{\mathbf{t}=\mathbf{0}} = \psi^{(r-1)}(\alpha + Y_{0,1}) - \mathcal{I}_{(r=1)}[log\,(\beta+1)] + \sum_{j'=0}^{j-1}\left[\psi^{(r-1)}\left(c_{j'}\ T\ e_{j'+1,a_{j'}(k)} + Y_{j'+1,a_{j'}(k)}\right) - \psi^{(r-1)}\left(c_{j'}\ T\ e_{j',a_{j'}(k)} + Y_{j',a_{j'}(k)}\right)\right] \quad (22)$$

where $a_{j'}(k)$ denotes the ancestor of (j,k) at scale j'. Therefore, the cumulants for each $\mu_{j,k}$ are scale sensitive, containing information only from its ancestors.

ACKNOWLEDGEMENTS

This research was supported by NSF Grant BCS0079077 and ONR Award N00014-99-1-0219.

REFERENCES

1. Quattrochi DA, Goodchild MF, Eds. *Scale in Remote Sensing and GIS*. Boca Raton, FL: CRC Lewis, 1997.
2. Lawson AB. *Statistical Methods in Spatial Epidemiology*. Chichester: John Wiley, 2001.
3. De Cola L. Spatial forecasting of disease risk and uncertainty, *Cartography and Geographic Information Science* 2002; **29**(4):363-380.
4. Kolaczyk ED, Huang H. Multiscale statistical models for hierarchical spatial aggregation, *Geographical Analysis* 2001; **33**:95-118.
5. Louie MM, Kolaczyk ED. On the covariance properties of certain multiscale spatial processes. *Statistics, Probability Letters* 2004; **66**(4):407-416.
6. Hennekens CH, Buring JE. *Epidemiology and Medicine*. Lippincott Williams & Wilkins, 1987.
7. Donoho DL, Johnstone IM, Kerkyacharian G, Picard D. Wavelet shrinkage: Asymptopia? (with Discussion). *Journal of the Royal Statistical Society, Series B* 1995; **57**: 301-69.
8. Kolaczyk ED, Nowak RD. Multiscale likelihood analysis, complexity penalized estimation. *Annals of Statistics* 2004; **32**:500-527.

9. Antoniadis A, Bigot J, Sapatinas T. Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software* 2001; **6:6**: 1-83.
10. Zhang C-H. General empirical Bayes wavelet methods and exactly adaptive minimax estimation. *Annals of Statistics* 2005; **33**, 000-000.
11. Johnstone, IM, Silverman, BW. Empirical Bayes selection of wavelet thresholds. *Annals of Statistics* 2005; **33**, 000-000.
12. Louie, MM. *A Multiscale Approach to Disease Mapping.* Doctoral dissertation, Department of Mathematics and Statistics, Boston University. 2003.
13. Mollie A. Bayesian mapping of disease. In *Markov Chain Monte Carlo in Practice*, Gilks WR, Richardson S, Spiegelhalter DJ (eds). Chapman & Hall: London 1996; 359-379.
14. Johnson NL. Systems of frequency curves generated by methods of translation. *Biometrika* 1949; **36**:149-176.
15. Elderton WP, Johnson NL. *Systems of Frequency Curves*. Cambridge University Press, 1969.
16. Hill ID, Hill R, Holder RL. Algorithm AS99: Fitting Johnson curves by moments. *Applied Statistics* 1976; **25**(2):180-189.
17. Draper J. Properties of distributions resulting from certain simple transformations of the normal distribution. *Biometrika* 1952; **39**:290-301.
18. Barber S, Nason GP, Silverman BW. Posterior probability intervals for wavelet thresholding. *Journal of the Royal Statistical Society, Series B* 2002; **64**:189-205.
19. Banerjee S, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC: Boca Raton 2004; 158-167.
20. Dreassi E, Biggeri A. Edge Effect in disease mapping. *Journal of the Italian Statistical Society* 1998; **7**(3):267-283.
21. Mugglin AS, Carlin BP, Gelfand AE. Fully model based approaches for spatially misaligned data. *Journal of the American Statistical Association* 2000; **95**:877-887.
22. Diggle PJ. Overview of statistical methods for disease mapping and its relationship to cluster detection. In *Spatial Epidemiology: Methods and Applications*, Elliott P, Wakefield J, Best N, Briggs D (eds). Oxford Medical Publications: Oxford 2001; 87-103.
23. Pickle LW, White AA. Effects of the choice of age-adjustment method on maps of death rates. *Statistics in Medicine* 1995; **14**:615-627.

Table I. Integrated squared bias, variance and mean-squared error for the multiscale estimator under empirical Bayes analysis $\hat{\theta}^{EB}_{J,k}$, and fully Bayesian analysis $\hat{\theta}^{FB}_{J,k}$, and for the Poisson-lognormal estimator $\hat{\theta}^{PL}_{k}$, based on 100 trials per landscape design. Values are times 10^{-2}.

	Bias^2			Variance			MSE		
Landscape	$\hat{\theta}^{EB}_{J,k}$	$\hat{\theta}^{FB}_{J,k}$	$\hat{\theta}^{PL}_{k}$	$\hat{\theta}^{EB}_{J,k}$	$\hat{\theta}^{FB}_{J,k}$	$\hat{\theta}^{PL}_{k}$	$\hat{\theta}^{EB}_{J,k}$	$\hat{\theta}^{FB}_{J,k}$	$\hat{\theta}^{PL}_{k}$
1	1.8×10^{-3}	0.0	0.0	0.3	0.1	0.2	0.3	0.1	0.2
2	0.3	0.6	1.6	1.4	1.3	1.9	1.7	1.9	3.5
3	0.3	0.5	2.8	1.3	1.2	2.4	1.5	1.8	5.1
4	1.6	2.8	2.9	2.2	2.0	2.4	3.8	4.8	5.3
5	1.6	2.8	2.7	2.3	2.0	2.4	3.9	4.8	5.2
6	1.2	1.7	2.8	2.2	2.0	2.4	3.4	3.7	5.2
7	1.9	3.5	2.7	2.4	2.0	2.3	4.3	5.5	5.1
8	1.8	2.7	2.8	2.5	2.2	2.4	4.3	4.8	5.2
9	4.4	6.5	2.7	2.6	1.8	2.4	7.0	8.3	5.2
10	4.6	6.7	2.7	2.5	1.6	2.4	7.1	8.3	5.1
11	3.9	5.2	4.3	5.4	4.7	7.7	9.3	9.9	12.0

Table II. Comparison of different estimators and lower and upper bounds of 95% central marginal posterior for the twelve municipalities on the northeastern border of Tuscany, ordered from north to south. The estimates are based on the multiscale model using empirical Bayes ($\hat{\theta}^{EB}_{J,k}$) and fully Bayesian ($\hat{\theta}^{FB}_{J,k}$) analysis, and a Poisson-lognormal model ($\hat{\theta}^{PL}_{k}$).

Municipality	SMR	$\hat{\theta}^{EB}_{J,k}$ (lower,upper)	$\hat{\theta}^{FB}_{J,k}$ (lower,upper)	$\hat{\theta}^{PL}_{k}$ (lower,upper)
Firenzuola	2.74	2.08 (1.605, 2.654)	2.07 (1.579, 2.625)	2.21 (1.687, 2.829)
Palazzuolo	1.71	1.54 (0.770, 2.740)	1.54 (0.744, 2.607)	1.91 (1.270, 2.716)
Marradi	2.41	1.91 (1.388, 2.560)	1.89 (1.367, 2.519)	2.03 (1.525, 2.636)
S. Godenzo	2.39	1.90 (1.253, 2.756)	1.89 (1.229, 2.690)	2.13 (1.522, 2.870)
Stia	2.50	2.21 (1.531, 3.075)	2.19 (1.508, 3.015)	2.24 (1.602, 3.030)
Pratovecchio	2.00	1.91 (1.240, 2.800)	1.90 (1.219, 2.733)	2.07 (1.539, 2.693)
Poppi	3.08	2.56 (1.971, 3.256)	2.54 (1.941, 3.220)	2.67 (2.096, 3.331)
Bibbiena	3.09	2.56 (2.094, 3.092)	2.54 (2.066, 3.066)	2.85 (2.301, 3.468)
Chiusi Verna	1.61	1.67 (1.003, 2.621)	1.68 (0.983, 2.542)	2.01 (1.497, 2.610)
Pieve S. Stefano	1.72	1.74 (1.162, 2.499)	1.74 (1.149, 2.450)	1.80 (1.321, 2.359)
Badia Tedalda	1.70	1.73 (0.960, 2.859)	1.73 (0.941, 2.746)	1.83 (1.216, 2.624)
Sestino	2.16	2.00 (1.192, 3.142)	1.99 (1.174, 3.038)	2.01 (1.181, 3.115)

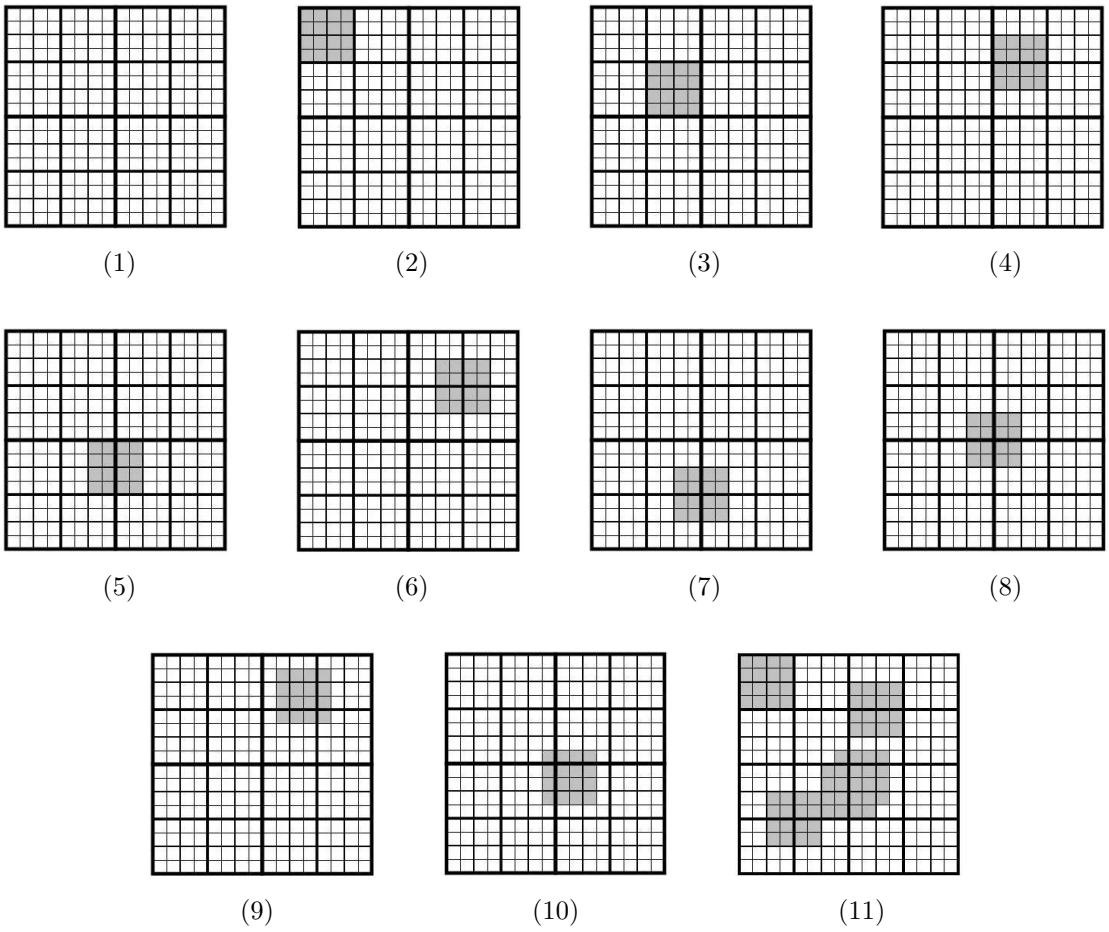


Figure 1. Simulated landscape designs for assessing the performance of the multiscale estimators to changes in the hereditary structure of the spatial pattern, and for comparison with a spatial (uni-scale) estimator. Design (1) corresponds to uniform relative risk of one. Designs (2) through (10) each contain a (shaded) area of elevated relative risk, comprised of 4x4 pixels. The locations of these areas were chosen to reflect the different hereditary structures. Design (11) combines patterns from (2), (4), (5), (6), and (10).

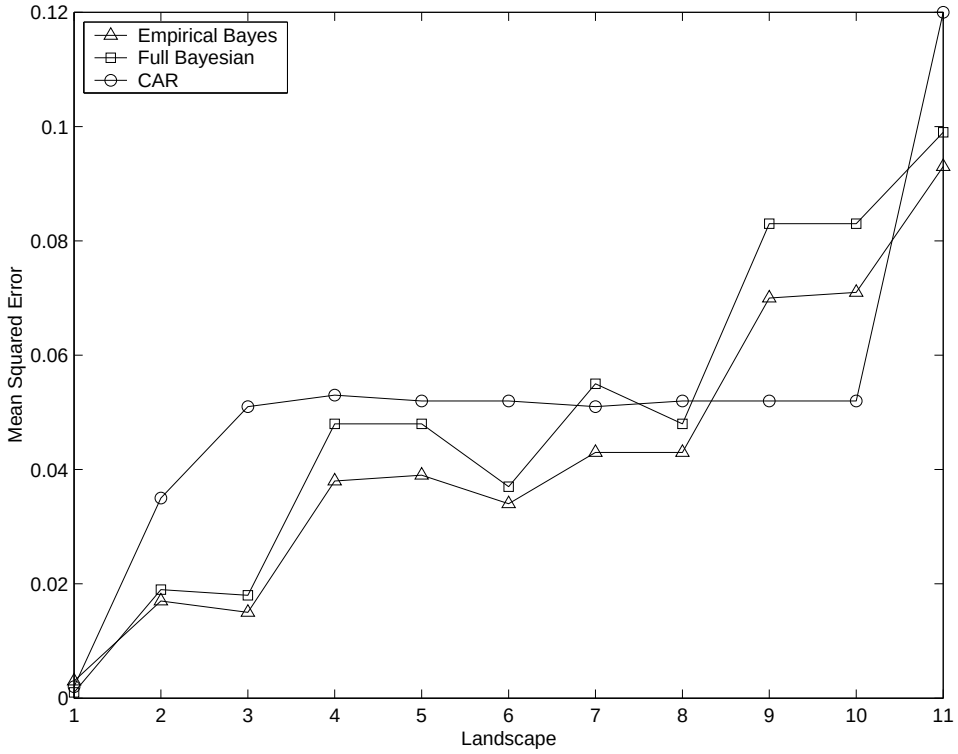
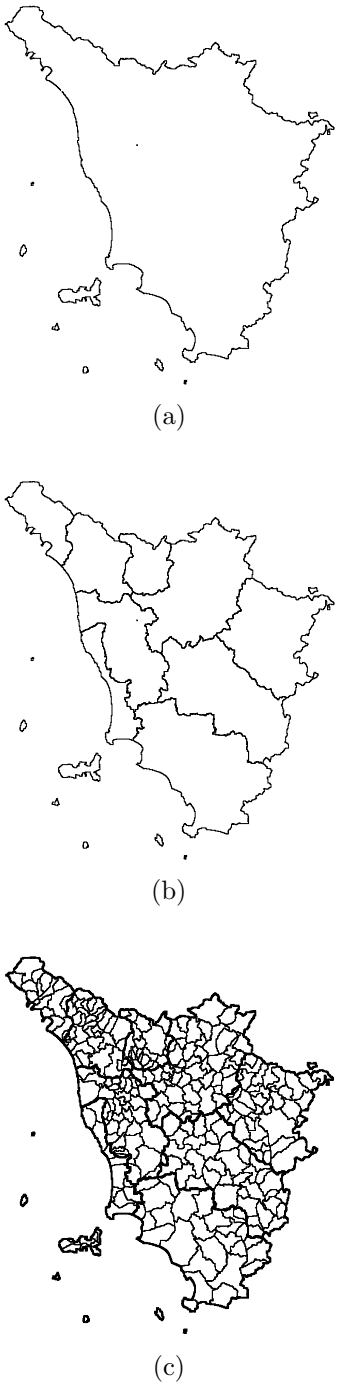


Figure 2. Plots of the integrated mean-squared errors in Table I



(a)

(b)

(c)

Figure 3. Hierarchy of three nested partitionings of the Tuscan region of Italy: (a) region level, (b) provincial level, and (c) municipality level.

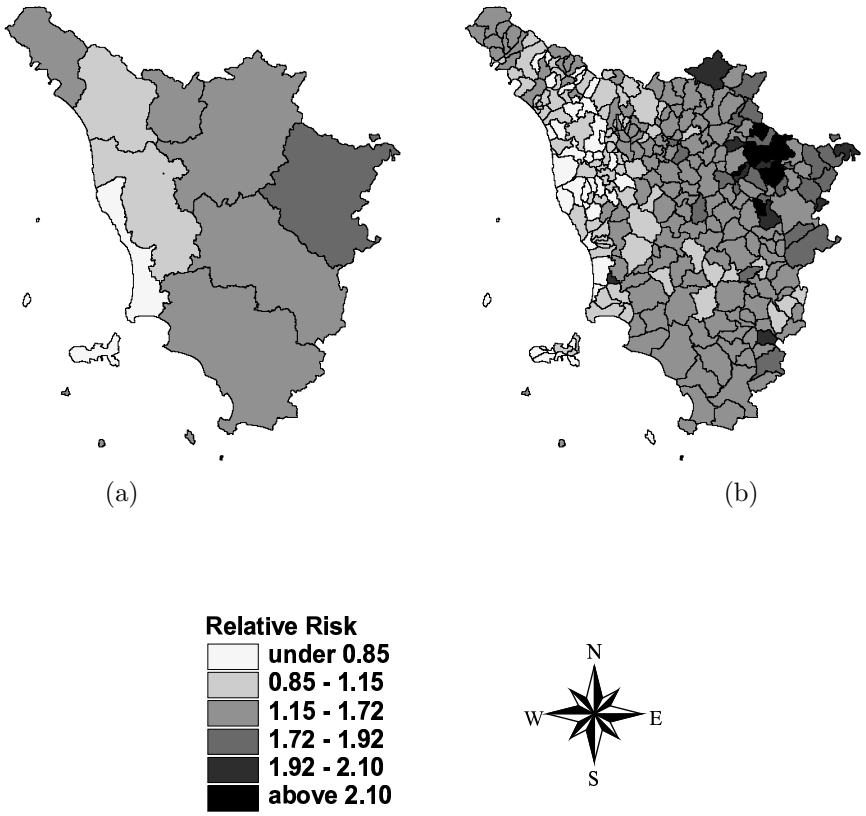


Figure 4. Posterior-based multiscale SMR estimates of relative risk for male gastric cancer in Tuscany at the (a) provincial and (b) municipality levels.

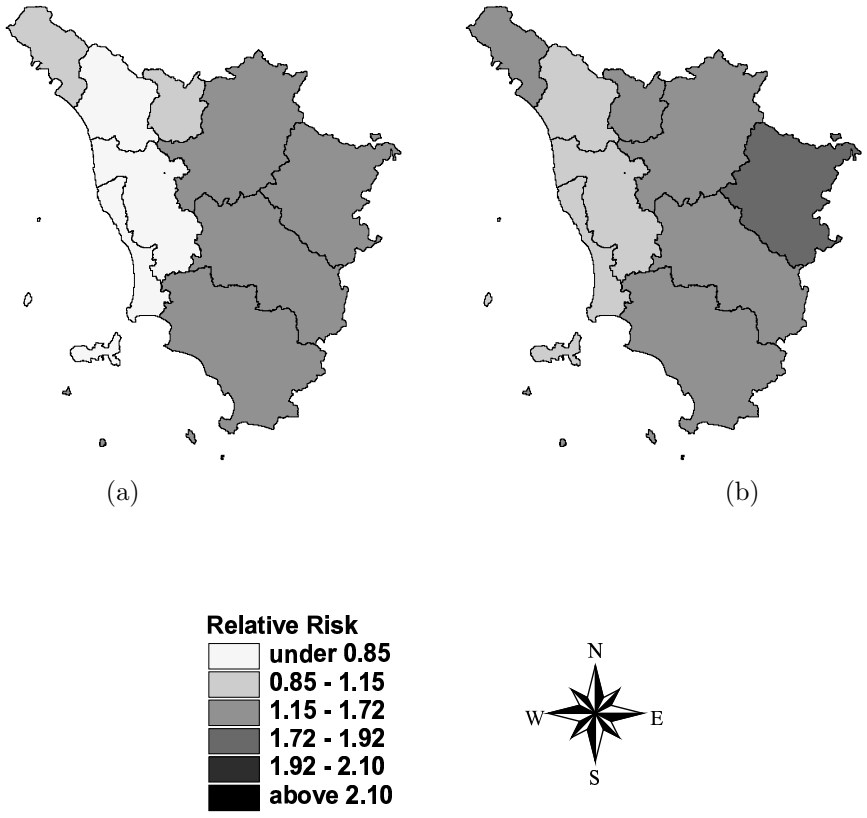


Figure 5. (a) Lower and (b) upper bounds of 95% central marginal posterior intervals for provincial-level relative risk, corresponding to the provincial-level posterior mean map of Figure 4a, obtained by analytic approximation to posterior distributions.

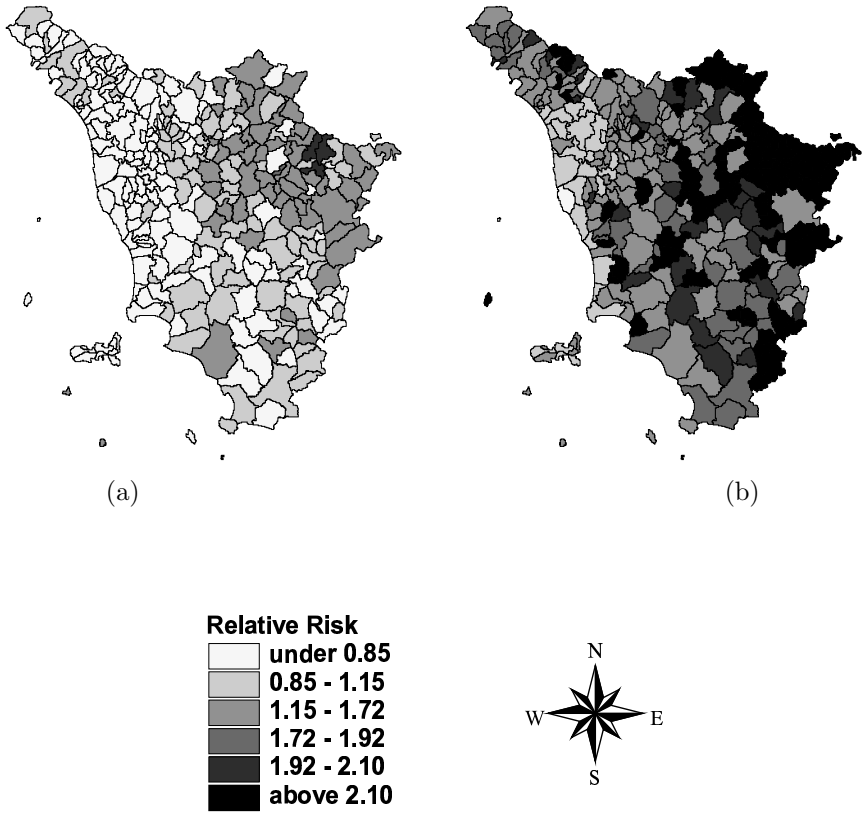


Figure 6. (a) Lower and (b) upper bounds of 95% central marginal posterior intervals for municipal-level relative risk, corresponding to the municipal-level posterior mean map of Figure 4b, obtained by analytic approximation to posterior distributions.