

Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing

YUFEN ZHANG, JAMES S. HODGES, AND SUDIPTO BANERJEE¹

*Division of Biostatistics, School of Public Health, University of Minnesota,
Mayo Mail Code 303, Minneapolis, Minnesota 55455-0392, U.S.A.* Correspondence

sudiptob@biostat.umn.edu December 11, 2008
author: Sudipto Banerjee telephone: (612) 624-0624 fax: (612) 626-0660 email:

¹Yufen Zhang is a Research Assistant, James S. Hodges and Sudipto Banerjee are Associate Professors in the Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, 55455. The work of the third author was supported in part by NIH grant 1-R01-CA95995.

Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing

SUMMARY

Rapid developments in geographical information systems (GIS) and advanced spatial statistics continue to generate interest in analyzing complex spatial datasets. One area of activity is in creating smoothed disease maps to describe the geographic variation of disease and generate hypotheses for apparent differences in risk. With multiple diseases, a multivariate conditionally autoregressive model (MCAR) is often used to smooth across space while accounting for associations between the diseases (see, e.g., Jin et al 2007). The MCAR, however, imposes complex covariance structures that are difficult to interpret and estimate. This article develops a much simpler alternative approach building upon the techniques of smoothed ANOVA (SANOVA) developed by Hodges et al (2007). Instead of simply shrinking effects without any structure as in Hodges et al (2007), here we use SANOVA to smooth spatial random effects by taking advantage of the spatial structure. This paper extends SANOVA to cases in which one factor is a spatial lattice, which is smoothed using a CAR model, and a second factor is, for example, type of cancer. Datasets routinely lack enough information to identify the additional structure of MCAR. SANOVA offers a simpler and more intelligible structure than the MCAR while performing as well. We demonstrate our approach with simulation studies designed to compare SANOVA with different design matrices versus MCAR with different priors. Subsequently a cancer-surveillance dataset, describing incidence of 3 cancers in Minnesota's 87 counties, is analyzed using both approaches, showing the competitiveness of the SANOVA approach.

KEY WORDS: Analysis of variance, Bayesian inference, conditionally autoregressive model, hierarchical model, smoothing.

1 Introduction

Statistical modelling and analysis of spatially referenced data receive considerable interest due to the increasing availability of geographical information systems (GIS) and spatial databases. For data aggregated over geographic regions such as counties, census tracts or ZIP codes (often called *areal* data), with individual identifiers and precise locations removed, inferential objectives focus on models for spatial clustering and variation. Such models are often used in epidemiology and public health to understand geographical patterns in disease incidence and morbidity. Recent reviews of methods for such data include Lawson et al. (1999), Elliott et al. (2000), Waller & Gotway (2004) and Rue & Held (2004). Traditionally such data have been modelled using conditionally specified probability models that shrink or smooth spatial effects by borrowing strength from neighboring regions. Perhaps the most pervasive model is the conditionally autoregressive (CAR) family pioneered by Besag (1974), which has been widely investigated and applied to spatial epidemiological data (Wakefield, 2007, gives an excellent review). Recently the CAR has been extended to multivariate responses, building on multivariate conditional autoregressive (MCAR) models described by Mardia (1988). Gelfand & Vonatsou (2003) and Carlin & Banerjee (2003) discussed their use in hierarchical models, while Kim, Sun & Tsutakawa (2001) presented a different “twofold CAR” model for counts of two diseases in each areal unit. Other extensions allowing flexible modelling of cross-correlations include Sain & Cressie (2002), Jin, Carlin & Banerjee (2005) and Jin, Banerjee & Carlin (2007). The MCAR can be viewed as a conditionally specified probability model for interactions between space and an attribute of interest. For instance, in disease mapping interest often lies in modelling geographical patterns in disease rates or counts of several diseases. The MCAR acknowledges dependence between the diseases as well as dependence across space. However, practical difficulties arise from MCAR’s elaborate dependence structure: most interaction effects will be weakly identified by the data, so the dependence structure is poorly identified. In hierarchical models (e.g., Gelfand & Vonatsou, 2003; Jin et al., 2005, 2007), strong prior distributions may improve identifiability, but this is not uncontroversial as inferences are sensitive to the prior and perhaps unreliable without genuine prior information. This article proposes a much simpler and more interpretable alternative to the MCAR, modelling multivariate

spatial effects using smoothed analysis of variance (SANOVA) as developed by Hodges, Cui, Sargent & Carlin (2007), henceforth HCSC. Unlike an ANOVA that is used to identify some interaction effects to retain and others to remove, SANOVA mostly retains effects that are large, mostly removes those that are small, and partially retains middling effects. (Loosely speaking, “large”, “middling”, and “small” describe the size of the unsmoothed effects compared to their standard errors.) To accommodate rich dependence structures, MCAR introduces weakly identifiable parameters that complicate estimation. SANOVA, on the other hand, focuses instead on smoothing interactions to yield more stable and reliable results. Our intended contribution is to show how SANOVA can solve the multiple disease mapping problem while avoiding the dauntingly complex covariance structures imposed by MCAR and its generalizations. We demonstrate that SANOVA produces inference that is largely indistinguishable from MCAR, yet SANOVA is simpler, more explicit, easier to put priors on and easier to estimate. The rest of the article is as follows. Section 2 reviews SANOVA and MCAR, identifying SANOVA as a special case of MCAR. Section 3 is a “tournament” of simulation experiments comparing SANOVA with MCAR for normal and Poisson data, while Section 4 analyzes data describing the number of deaths from lung, larynx and esophagus cancer in Minnesota between 1990 and 2000. A summary and discussion of future research in Section 5 concludes the paper. The Appendix gives computational and technical details.

2 The Competitors

2.1 Smoothing spatial effects using SANOVA

2.1.1 SANOVA for balanced, single-error-term models (HCSC 2007)

Consider a balanced, single-error-term analysis of variance, with M_1 degrees of freedom for main effects and M_2 degrees of freedom for interactions. Specify this ANOVA as a linear model: let A_1 denote columns in the design matrix for main effects, and A_2 denote columns in the design matrix for interactions. Assume the design has c cells and n observations per cell, giving cn observations in total. To simplify later calculations, normalize the columns of A_1 and A_2 so $A_1'A_1 = I_{M_1}$ and $A_2'A_2 = I_{M_2}$. (Note: HCSC normalized columns differently,

fixing $A_1'A_1 = cnI_{M_1}$ and $A_2'A_2 = cnI_{M_2}$.) Then write the ANOVA as:

$$\mathbf{y} = [A_1|A_2] \begin{bmatrix} \boldsymbol{\Theta}_1 \\ \boldsymbol{\Theta}_2 \end{bmatrix} + \boldsymbol{\epsilon} = A_1\boldsymbol{\Theta}_1 + A_2\boldsymbol{\Theta}_2 + \boldsymbol{\epsilon} \quad (1)$$

where $\boldsymbol{\epsilon} \sim N(0, \frac{1}{\eta_0}I)$ with η_0 being a precision, \mathbf{y} is $cn \times 1$, A_1 is $cn \times M_1$, A_2 is $cn \times M_2$, $\boldsymbol{\Theta}_1$ is $M_1 \times 1$, $\boldsymbol{\Theta}_2$ is $M_2 \times 1$, and $\boldsymbol{\epsilon}$ is $cn \times 1$. This ANOVA is smoothed by further modeling $\boldsymbol{\Theta}$. HCSC emphasized smoothing interactions, although main effects can be smoothed by exactly the same means. Following HCSC, we add constraints (or a prior) on $\boldsymbol{\Theta}_2$ as $\theta_{M_1+j} \sim N(0, 1/\eta_j)$ for $j = 1, \dots, M_2$, written as:

$$\mathbf{0}_{M_2} = [0_{M_2 \times M_1} | I_{M_2}] \begin{bmatrix} \boldsymbol{\Theta}_1 \\ \boldsymbol{\Theta}_2 \end{bmatrix} + \boldsymbol{\delta} \quad (2)$$

where $\boldsymbol{\delta} \sim N(0, \text{diag}(\frac{1}{\eta_j}))$, in the manner of Lee & Nelder (1996) and Hodges (1998). Combining (1) and (2), express this hierarchical model as a linear model:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{M_2} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ 0_{M_2 \times M_1} & I_{M_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Theta}_1 \\ \boldsymbol{\Theta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\delta} \end{bmatrix}. \quad (3)$$

More compactly, write:

$$\mathbf{Y} = X\boldsymbol{\Theta} + \mathbf{e} \quad (4)$$

where \mathbf{Y} has dimension $(cn + M_2) \times 1$ and \mathbf{e} 's covariance Γ is block diagonal with blocks $\Gamma_1 = \frac{1}{\eta_0}I_{cn}$ for the data cases (rows of X corresponding to the observation \mathbf{y}) and $\Gamma_2 = \text{diag}(1/\eta_1, \dots, 1/\eta_{M_2})$ for the constraint cases (rows of X with error term $\boldsymbol{\delta}$). For convenience, define the matrix $X_D = [A_1|A_2]$, the data-case part of X . This development can be done using the mixed linear model (MLM) formulation traditionally written as $\mathbf{y} = X\boldsymbol{\beta} + Z\mathbf{u} + \boldsymbol{\epsilon}$, where our (1) supplies this equation and $\mathbf{u} = \boldsymbol{\Theta}_2 \sim N(\mathbf{0}, \Gamma_2)$. The development to follow can also be done using the MLM formulation at the price of slightly greater complexity, so we omit it. HCSC developed SANOVA for exchangeable priors on groups formed from components of $\boldsymbol{\Theta}_2$. The next section develops the extension to spatial smoothing.

2.1.2 What is CAR?

Suppose a map has N regions, each with an unknown quantity of interest ϕ_i , $i = 1, \dots, N$. A conditionally autoregressive (CAR) model specifies the full conditional distribution of each ϕ_i as

$$\phi_i \mid \phi_j, j \neq i, \sim N \left(\frac{\alpha}{m_i} \sum_{i \sim j} \phi_j, \frac{1}{\tau m_i} \right), \quad i, j = 1, \dots, N, \quad (5)$$

where $i \sim j$ denotes that region j is a *neighbor* of region i (typically defined as spatially adjacent), and m_i is the number of region i 's neighbors. Equation (5) reduces to the well-known intrinsic conditionally autoregressive (ICAR) model (Besag et al., 1991) if $\alpha = 1$ or an independence model if $\alpha = 0$. The ICAR model induces “local” smoothing by borrowing strength from neighbors, while the independence model assumes spatial independence and induces “global” smoothing. The CAR prior’s smoothing parameter α also controls the strength of spatial dependence among regions, though it has long been appreciated that a fairly large α may be required to induce large spatial correlation; see Wall (2004) for recent discussion and examples. It is well-known (e.g., Besag, 1974) that the conditional specifications in (5) lead to a valid joint distribution for $\boldsymbol{\phi} = (\phi_1, \dots, \phi_N)'$ expressed in terms of the map’s neighborhood structure. If Q is an $N \times N$ matrix such that $Q_{ii} = m_i$, $Q_{ij} = -\alpha$ whenever $i \sim j$ and $Q_{ij} = 0$ otherwise, then the intrinsic CAR model (Besag et al 1991) has density:

$$p(\boldsymbol{\phi} \mid \tau) \sim \tau^{\frac{N^*}{2}} \exp \left(-\frac{\tau}{2} \boldsymbol{\phi}' Q \boldsymbol{\phi} \right), \quad \text{with } N^* = \begin{cases} N, & \text{if } \alpha \in (0, 1) \\ N - G, & \text{if } \alpha = 1 \end{cases}. \quad (6)$$

In (6), τ is the spatial precision parameter, τQ is the precision matrix in this multivariate normal distribution and G is the number of “islands” (disconnected parts) in the spatial map (Hodges et al 2003). When $\alpha \in (0, 1)$, (6) is a proper multivariate normal distribution. When $\alpha = 1$, Q is singular with $Q\mathbf{1} = \mathbf{0}$; Q has rank $N - G$ in a map with G islands, therefore the exponent on τ becomes $(N - G)/2$. In hierarchical models, the CAR model is usually used as a prior on spatial random effects. For instance, let Y_i be the observed number of cases of a disease in region i , $i = 1, \dots, N$, and E_i be the expected number of cases in region i . Here the Y_i are treated as random variables, while the E_i are treated as fixed and known, often simply proportional to the number of persons at risk in region i .

For rare diseases, a Poisson approximation to a binomial sampling distribution for disease counts is often used, so a commonly used likelihood for mapping a single disease is

$$Y_i \stackrel{ind}{\sim} \text{Poisson}(E_i e^{\mu_i}), \quad i = 1, \dots, N, \quad (7)$$

where $\mu_i = \mathbf{x}_i' \boldsymbol{\beta} + \phi_i$. The \mathbf{x}_i are explanatory, region-specific regressors with coefficients $\boldsymbol{\beta}$ and the parameter μ_i is the log-relative risk describing departures of observed from expected counts, that is, from E_i . The hierarchy's next level is specified by assigning the CAR distribution to $\boldsymbol{\phi}$ and a hyper-prior to the spatial precision parameter τ . In the hierarchical setup, the improper ICAR with $\alpha = 1$ gives proper posterior distributions for spatial effects. In practice, Markov Chain Monte Carlo (MCMC) algorithms are designed for estimating posteriors from such models and the appropriate number of linear constraints on the $\boldsymbol{\phi}$ suffices to ensure sampling from proper posterior distributions (Banerjee, Carlin & Gelfand, 2004, pp 163–164, gives details).

2.1.3 How does CAR fit into SANOVA?

To use CAR in SANOVA, the key is re-expressing the improper CAR, i.e., (6) with $\alpha = 1$. Let Q have spectral decomposition $Q = VDV'$, where V is an orthogonal matrix with columns containing Q 's eigenvectors and D is diagonal with non-negative diagonal entries. D has G zero diagonal entries, one of which corresponds to the eigenvector $\frac{1}{\sqrt{N}} \mathbf{1}_N$, by convention the N^{th} (right-most) column in V . Define a new parameter $\boldsymbol{\Theta} = V' \boldsymbol{\phi}$, so $\boldsymbol{\Theta}$ has dimension N and precision matrix τD . Giving an N -vector $\boldsymbol{\Theta}$ a normal prior with mean zero and precision τD is equivalent to giving $\boldsymbol{\phi} = V \boldsymbol{\Theta}$ a CAR prior with precision τQ . $\boldsymbol{\Theta}$ consists of $\Theta_N = \frac{1}{\sqrt{N}} \mathbf{1}_N' \boldsymbol{\phi} = \sqrt{N} \bar{\phi}$, the scaled average of the ϕ_i , along with $N - 1$ contrasts in $\boldsymbol{\phi}$, which are orthogonal to $\frac{1}{\sqrt{N}} \mathbf{1}_N$ by construction. Thus the CAR prior is informative (has positive precision) only for contrasts in $\boldsymbol{\phi}$, while putting zero precision on $\Theta_{GM} = \Theta_N = \frac{1}{\sqrt{N}} \mathbf{1}_N' \boldsymbol{\phi}$, the overall level, and on $G - 1$ orthogonal contrasts in the levels of the G islands. In other words, the CAR model can be thought of as a prior distribution on the contrasts rather than individual effects (hence the need for the sum-to-zero constraint). A related result, discussed in Besag et al. (1995), shows the CAR to be a member of a family of “pairwise difference” priors. This reparameterization allows the CAR model to

fit into the ANOVA framework, with Θ_{GM} corresponding to the ANOVA's grand mean and the rest of Θ , Θ_{Reg} , corresponding to $V^{(-)'}\phi$, where $V^{(-)}$ is V excluding the column $\frac{1}{\sqrt{N}}\mathbf{1}_N$, consisting of $N - 1$ orthogonal contrasts among the N regions and giving the $N - 1$ degrees of freedom in the usual ANOVA:

$$\begin{aligned}\phi &= \left[\phi_1, \phi_2, \dots, \phi_N \right]' \\ &= V\Theta \\ &= \left[V^{(-)} \quad \frac{1}{\sqrt{N}}\mathbf{1}_N \right] \begin{bmatrix} \Theta_{Reg} \\ \Theta_{GM} \end{bmatrix}.\end{aligned}$$

Giving ϕ a CAR prior is equivalent to giving Θ a $N(\mathbf{0}, \tau D)$ prior; the latter are the “constraint cases” in HCSC's SANOVA structure. The precision $D_{NN} = 0$ for the overall level is equivalent to a flat prior on Θ_{GM} , though Θ_{GM} could alternatively have a normal prior with mean zero and finite variance. If $G > 1$, the CAR prior also puts zero precision on $G - 1$ contrasts in ϕ , which are contrasts in the levels of the G islands (Hodges et al 2003).

2.2 SANOVA as a competitor to MCAR

2.2.1 Multivariate conditionally autoregressive (MCAR) models

With multiple diseases, we have unknown ϕ_{ij} corresponding to region i and disease j , where $i = 1, \dots, N$ and $j = 1, \dots, n$. Letting Ω be a common precision matrix (i.e., inverse of the covariance matrix) representing correlations between the diseases in a given region, MCAR distributions arise through conditional specifications for $\phi_i = (\phi_{i1}, \dots, \phi_{in})'$:

$$\phi_i | \{\phi_{i'}\}_{i' \neq i} \sim MVN \left(\frac{\alpha}{m_i} \sum_{i' \sim i} \phi_{i'}, \frac{1}{m_i} \Omega^{-1} \right). \quad (8)$$

These conditional distributions yield a joint distribution for $\phi = (\phi_1', \dots, \phi_N')$:

$$f(\phi | \Omega) \propto \|\Omega\|^{\frac{N-G}{2}} \exp \left(-\frac{1}{2} \phi' (Q \otimes \Omega) \phi \right), \quad (9)$$

where Q is defined as in Section 2.1.2 and again (9) is an improper density when $\alpha = 1$. However, as for the univariate CAR, this yields proper posteriors in conjunction with a proper likelihood. The specification above is a “separable” dispersion structure, i.e., the covariances between the diseases are invariant across regions. This may seem restrictive, but

relaxing this restriction gives even more complex dispersion structures (see Jin, Banerjee & Carlin, 2007 and references therein). As mentioned earlier, our focus is to retain the model’s simplicity without compromising the primary inferential goals. We propose to do this using SANOVA and will compare it with the separable MCAR only.

2.2.2 SANOVA with Minnesota counties as one factor

We now describe the SANOVA model using the Minnesota 3-cancer dataset. Consider the Minnesota map with $N = 87$ counties, and suppose each county has counts for $n = 3$ cancers. County i has an n -vector of parameters describing the n cancers, $\phi_i = (\phi_{i1}, \phi_{i2}, \dots, \phi_{in})'$; define the Nn vector ϕ as $\phi = (\phi_1', \phi_2', \dots, \phi_N')'$. For now, we are vague about the specific interpretation of ϕ_{ij} ; the following description applies to any kind of data. Assume the $N \times N$ matrix Q describes neighbor pairs among counties as before. The SANOVA model for this problem is a 2-way ANOVA with factors cancer (“CA”, n levels) and county (“CO”, N levels) and no replication. As in section 2.1.1, we model ϕ with a saturated linear model and put the grand mean and the main effects in their traditional positions as in ANOVA (matrix dimensions and definitions appear below the equation):

$$\phi = [\phi_1', \phi_2', \dots, \phi_N']' = [A_1 | A_2] \Theta =$$

$$\left[\begin{array}{c|c} \underbrace{\frac{1}{\sqrt{Nn}} \mathbf{1}_{Nn}} & \underbrace{\frac{1}{\sqrt{N}} \mathbf{1}_{N \otimes H_{CA}}} \\ \hline \underbrace{V^{(-)} \otimes \frac{1}{\sqrt{n}} \mathbf{1}_n} & \underbrace{V^{(-)} \otimes H_{CA}^{(1)} \dots V^{(-)} \otimes H_{CA}^{(n-1)}} \end{array} \right] \begin{bmatrix} \Theta_{GM} \\ \Theta_{CA} \\ \Theta_{CO} \\ \Theta_{CO \times CA} \end{bmatrix} \quad (10)$$

	Cancer	County	Cancer × County
Grand mean	main effect	main effect	interaction
$Nn \times 1$	$Nn \times (n-1)$	$Nn \times (N-1)$	$Nn \times (N-1)(n-1)$

where H_{CA} is an $n \times (n-1)$ matrix whose columns are contrasts among cancers, so $\mathbf{1}_n' H_{CA} = \mathbf{0}'_{n-1}$, and $H_{CA}' H_{CA} = I_{n-1}$; $H_{CA}^{(j)}$ is the j^{th} column of H_{CA} ; and $V^{(-)}$ is V without its N^{th} column $\frac{1}{\sqrt{N}} \mathbf{1}_N$, so it has $N-1$ columns, each a contrast among counties, i.e., $\mathbf{1}_N' V^{(-)} = \mathbf{0}'_{N-1}$, and $V^{(-)'} V^{(-)} = I_{N-1}$. The column labelled “Grand mean” corresponds to the ANOVA’s grand mean and has parameter Θ_{GM} ; the other blocks of columns labeled

as main effects and interactions correspond to the analogous ANOVA effects and to their respective parameters $\Theta_{CA}, \Theta_{CO}, \Theta_{CO \times CA}$. Defining prior distributions on Θ completes the SANOVA specification. We put independent flat priors (normal with large variance) on Θ_{GM} and Θ_{CA} , which are, therefore, not smoothed. This is equivalent to putting a flat prior on each of the n cancer-specific means. To specify the smoothing priors, define $H_{CA}^{(0)} = \frac{1}{\sqrt{n}}\mathbf{1}_n$. Let the county main effect parameter Θ_{CO} have prior $\Theta_{CO} \sim N_{N-1}(\mathbf{0}, \tau_0 D^{(-)})$, where $D^{(-)}$ corresponds to $V^{(-)}$ i.e., D without its N^{th} row and column, $\tau_0 > 0$ is unknown and $\tau_0 D^{(-)}$ is a precision matrix. Similarly, let the j^{th} group of columns in the cancer-by-county interaction, $V^{(-)} \otimes H_{CA}^{(j)}$, have prior $\Theta_{CO \times CA}^{(j)} \sim N_{N-1}(\mathbf{0}, \tau_j D^{(-)})$, for $\tau_j > 0$ unknown. Each of the priors on Θ_{CO} and the $\Theta_{CO \times CA}^{(j)}$ is a CAR prior; the overall level of each CAR, with prior precision zero, has been included in the grand mean and cancer main effects.

To compare this to the MCAR model, use SANOVA's priors on Θ to produce a marginal prior for ϕ comparable to the MCAR's prior on ϕ (section 2.2.1); in other words, integrate Θ_{CO} and $\Theta_{CO \times CA}$ out of the foregoing setup. *A priori*,

$$\left[\begin{array}{c} \Theta_{CO} \\ \Theta_{CO \times CA}^{(1)} \\ \vdots \\ \Theta_{CO \times CA}^{(n-1)} \end{array} \right] \quad (11)$$

has precision $Q \otimes (H_A^{(+)} \text{diag}(\tau_j) H_A^{(+)'})$, where K is the columns of the design matrix for the county main effects and cancer-by-county interactions—the right-most $n(N-1)$ columns in equation (10)'s design matrix—and $H_A^{(+)} = (\frac{1}{\sqrt{n}}\mathbf{1}_n | H_{CA})$ is an orthogonal matrix. Appendix A gives a proof.

2.2.3 Comparing SANOVA vs MCAR

Defining ϕ as in sections 2.2.1 and 2.2.2, consider the MCAR prior for ϕ , with within-county precision matrix Ω . Let Ω have spectral decomposition $V_\Omega D_\Omega V_\Omega'$, where D_Ω is $n \times n$ diagonal and V_Ω is $n \times n$ orthogonal. Then the prior precision of ϕ is $Q \otimes (V_\Omega D_\Omega V_\Omega')$, where Q is the known neighbor relations matrix and V_Ω and D_Ω are unknown. Comparing MCAR to SANOVA, the prior precision matrices for the vector ϕ are as in Figure 1. SANOVA is clearly a special case of MCAR in which $H_A^{(+)}$ is known. Also, as described so far, $H_A^{(+)}$ has

one column proportional to $\mathbf{1}_n$ with the other columns being contrasts, while MCAR avoids this restriction. MCAR is thus more flexible, while SANOVA is simpler, presumably making it better identified and easier to set priors for. MCAR should have its biggest advantage over SANOVA when the “true” V_Ω is not like $H_A^{(+)}$ for any specification of the smoothing precisions τ_j . However, because datasets often have modest information about higher-level variances, it may be that using the wrong $H_A^{(+)}$ usually has little effect on the analysis. In other words, SANOVA’s performance may be relatively stable despite having to specify $H_A^{(+)}$, while MCAR may be more sensitive to Ω ’s prior.

2.3 Setting priors in MCAR and SANOVA

2.3.1 Priors in SANOVA

For the case of normal errors, based on equations (1) and (10), setting priors for Θ, τ_j, η_0 completes a Bayesian specification. Since τ and η_0 are precision parameters, one possible prior is Gamma; this paper uses a Gamma with mean 1 and variance 10. As mentioned, the grand mean and cancer main effects $\theta_1, \theta_2, \theta_3$ have flat priors with $\pi(\theta) \propto 1$, though they could have proper informative priors. The priors for θ_4, \dots are set according to the SANOVA structure as in section 2.2.2. We ran chains drawing in the order θ, τ , and η_0 (Appendix B gives details). Hodges et al. (2007) also considered priors on the degrees of freedom in the fitted model, some conditioned so the degrees of freedom in the model’s fit were fixed at a certain degree of smoothness. The present paper emphasizes comparing MCAR and SANOVA, so we do not consider such priors. For the case of Poisson errors, we use a normal prior with mean 0 and variance 10^6 for the grand mean and cancer main effects $\theta_1, \theta_2, \theta_3$. The other θ_i s are given normal CAR priors as discussed in section 2.2.2. For the prior on the smoothing precisions τ_j , we use Gamma with mean 1 and variance 10. To reduce high posterior correlations among the θ s, we used a transformation during MCMC; Appendix C gives details.

2.3.2 Priors in MCAR

MCAR models were fitted in WinBUGS. For the normal-error case, we used this model and parameterization:

$$\begin{aligned} Y_{ij} &\sim N(\mu_{ij}, \frac{1}{\eta_0}) \\ \mu_{ij} &= \beta_j + S_{ij}, \end{aligned} \tag{12}$$

$i = 1, \dots, N; j = 1, \dots, n$, where η_0 has a gamma prior with mean 1 and variance 10 as for SANOVA. To satisfy WinBUGS's constraint that $\sum_i S_{ij} = 0$, we add cancer-specific intercepts β_j . We give β_j a flat prior and for S , the spatial random effects, we use an intrinsic multivariate CAR prior. Similarly, in the Poisson case

$$\begin{aligned} Y_{ij} &\sim \text{Poisson}(\mu_{ij}) \\ \log(\mu_{ij}) &= \log(E_{ij}) + \beta_j + S_{ij} \end{aligned} \tag{13}$$

where E_{ij} is an offset. Prior settings for β_j and S_{ij} are as in the normal case. For MCAR priors, the within-county precision matrix Ω needs a prior; a Wishart distribution is an obvious choice. If $\Omega \sim \text{Wishart}(R, \nu)$, then $E(\Omega) = \nu R^{-1}$. We want a “vague” Wishart prior; usually $\nu = n$ is used but little is known about how to specify R . Thus, we considered three different R s, each proportional to the identity matrix. One of these priors sets R 's diagonal entries to $R_{ii} = 0.002$, close to the setting used in an example in the GeoBUGS manual (oral cavity cancer and lung cancer in West Yorkshire). The other two R s are the identity matrix and 200 times the identity. For the special case $n = 1$, where the Wishart reduces to a Gamma, these Wisharts are $\Gamma(0.5, 0.001)$, $\Gamma(0.5, 0.5)$ and $\Gamma(0.5, 100)$, respectively.

3 Simulation Experiment

For this simulation experiment, artificial data were simulated from the model used in SANOVA with a spatial factor, as described in section 2.2.2. Three different types of Bayesian analysis were applied to the simulated data: SANOVA with the same $H_A^{(+)}$ used to generate the simulated data (called “SANOVA correct”); SANOVA with incorrect $H_A^{(+)}$;

and MCAR. SANNOVA correct is a theoretical best possible analysis in that it takes as known things that MCAR estimates, i.e., it uses additional correct information. SANNOVA correct cannot be used in practice, of course, because the true $H_A^{(+)}$ is not known. MCAR vs SANNOVA with incorrect $H_A^{(+)}$ is the comparison relevant to practice, and comparing them to SANNOVA correct shows how much each method pays for its “deficiency” relative to SANNOVA correct. Obviously it is not enough to test the SANNOVA model using only data generated from a similar SANNOVA model. To avoid needless computing and facilitate comparisons, instead of generating data from an MCAR model and fitting a SANNOVA model as specified above, we use a trick that is equivalent to this. Section 3.1.2 gives the details.

3.1 Design of the simulation experiment

We simulated both normally-distributed and Poisson-distributed data. For both types of data, we considered two different true sets of smoothing parameters $\boldsymbol{r} = \boldsymbol{\tau}/\eta_0$ or $\boldsymbol{\tau}$ (Table 1). For the normal data, we considered $\boldsymbol{\tau}/\eta_0$, since this ratio determines smoothing in normal models, and we also considered two error precisions η_0 (Table 1).

3.1.1 Generating the simulated datasets

To generate data from the SANNOVA model, we need to define the true $H_A^{(+)}$. Let

$$HA_1 = \begin{pmatrix} 1 & -2 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

We used HA_1 as the correct $H_A^{(+)}$; its columns are scaled to have length 1. Given $V^{(-)}$ and with $H_A^{(+)}$ known, one draw of $\boldsymbol{\Theta}$ and ϵ produces a draw of $X_D\boldsymbol{\Theta} + \epsilon$, therefore a draw of \boldsymbol{y} . In the simulation, we let the grand mean and main effects, which are not smoothed, have true value 5. Each observation is simulated from a 3×20 factorial design, where 3 is the number of cancers and 20 is the number of counties. We used the 20 counties in the right lower corner of Minnesota’s map, with their actual neighbor relations. Thus, the dimension of each artificial dataset is 60. The simulation experiment is a repeated-measures design, in which a “subject” s in the design is a draw of $(\boldsymbol{\delta}^{(s)}, \boldsymbol{\gamma}^{(s)})$, referring to equation (3), where $\boldsymbol{\delta}_{1-3}^{(s)} = 5$ and $\boldsymbol{\delta}_{4-60}^{(s)} \sim N_{57}(\mathbf{0}, I_3 \otimes D^{(-)})$ specify $\boldsymbol{\Theta}$ and $\boldsymbol{\gamma}^{(s)} \sim N_{60}(\mathbf{0}, I_{60})$ gives ϵ . For the normal-errors case, 100 such “subjects” were generated. Given a design cell in the simulation experiment

with $\boldsymbol{\tau} = (a, b, c)$ and $\eta_0 = d$, the artificial dataset for subject s is $\mathbf{y}^{(s)} = X_D \text{diag}(\mathbf{1}'_3, \frac{1}{\sqrt{a}}\mathbf{1}'_{19}, \frac{1}{\sqrt{b}}\mathbf{1}'_{19}, \frac{1}{\sqrt{c}}\mathbf{1}'_{19})\boldsymbol{\delta}^{(s)} + \frac{1}{\sqrt{d}}\boldsymbol{\gamma}^{(s)}$. All factors of the simulation experiment were applied to each of the 100 “subjects”. For the normally-distributed data, the simulation experiment had these factors: (a) the true $(\tau_0/\eta_0, \tau_1/\eta_0, \tau_2/\eta_0) : (100, 100, 0.1)$ or $(0.1, 100, 0.1)$; (b) the true error precision η_0 : 1 or 10; and (c) six statistical methods, described below in section 3.1.2. Each design cell described in Table 1 thus had 100 simulated data sets. Similarly, for the Poisson-data experiment, another 100 “subjects” were generated, but now there is no $\boldsymbol{\gamma}^{(s)}$. Thus, each “subject” s is a vector $\boldsymbol{\delta}^{(s)}$, where $\boldsymbol{\delta}^{(s)}$ is as described above. For the design cell with $\boldsymbol{\tau} = (100, 100, 0.1)$, the artificial data for subject s is $\mathbf{y}^{(s)} \sim \text{Poisson}(\boldsymbol{\mu}^{(s)})$, where $\log(\boldsymbol{\mu}^{(s)}) = \log(\mathbf{E}) + X_D \text{diag}(\mathbf{1}'_3, \frac{1}{10}\mathbf{1}'_{19}, \frac{1}{10}\mathbf{1}'_{19}, \frac{1}{\sqrt{0.1}}\mathbf{1}'_{19})\boldsymbol{\delta}^{(s)}$. In the simulation experiment, we use “internal standardization” of the Minnesota 3-cancer data to supply the expected numbers of cancers E_{ij} . Among the 20 extracted counties, Hennepin county has the largest average population over 11 years, about 1.1 million; its cancer counts are 5294, 119, 439 for lung, larynx, and esophagus respectively. Faribault county has the smallest average population, 16,501, with cancer counts 110, 7 and 13 respectively. The E_{ij} have ranges 80 to 5275, 2 to 113 and 7 to 449 for lung, larynx, and esophagus cancer respectively. For the Poisson data, the simulation experiment had these factors: (a) the true $\tau_0, \tau_1, \tau_2 : (100, 100, 0.1)$ or $(0.1, 100, 0.1)$; and (b) six statistical methods described below in section 3.1.2. Again, each of the two design cells in Table 1 had 100 simulated datasets.

3.1.2 The six methods (procedures)

For each simulated dataset, we did a Bayesian analysis for each of six different models described in Table 2. The six models are: SANOVA with the correct $H_A^{(+)}$, HA_1 ; SANOVA with a somewhat incorrect $H_A^{(+)}$, HA_2 given below; a variant SANOVA with a very incorrect $H_A^{(+)}$, H_{AM} given below; MCAR with $R_{ii} = 0.002$; MCAR with $R_{ii} = 1$; and MCAR with $R_{ii} = 200$ (see section 2.3.2). HA_2 and H_{AM} are

$$HA_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}, \quad H_{AM} = \begin{pmatrix} 0.56 & -0.64 & -0.52 \\ -0.53 & -0.77 & 0.36 \\ -0.63 & 0.07 & -0.77 \end{pmatrix}.$$

The incorrect HA_2 has the same first column (grand mean) as the correct HA_1 , so it

differs from the correct HA_1 , though less than it might. As noted above, we need to see how the SANOVA model performs for data generated from an MCAR model in which V_Ω from Figure 1 does not have a column proportional to $\mathbf{1}_n$. To do this without needless computing, we used a trick: we used the data generated from a SANOVA model with HA_1 and fit the variant SANOVA mentioned above, in which $H_A^{(+)}$ is replaced by the orthogonal matrix H_{AM} with no column proportional to $\mathbf{1}_n$, chosen to be very different from HA_1 . For normal errors (Data1 through Data4), this is precisely equivalent to fitting a SANOVA with $H_A^{(+)} = HA_1$ to data generated from an MCAR model with $V_\Omega = BHA_1$, for $B = HA_1H_{AM}^{-1}$, i.e.,

$$V_\Omega = \begin{bmatrix} 0.43 & -0.74 & -0.52 \\ -0.13 & -0.63 & 0.77 \\ -0.89 & -0.26 & -0.37 \end{bmatrix} \quad (14)$$

(to 2 decimal places). For Poisson errors (Data5, Data6), the equivalence is no longer precise but the divergence of fitted SANOVA (using $H_A^{(+)} = H_{AM}$) and generated data (using $H_A^{(+)} = HA_1$) is quite similar. Finally, we considered three priors for MCAR because little is known about how to set this prior and we did not want to hobble MCAR with an ill-chosen prior. For the SANOVA and variant SANOVA analyses, we gave τ_j a $\Gamma(0.1, 0.1)$ prior with mean 1 and variance 10 for both the normal data and the Poisson data.

3.2 Outcome measures

To compare the six different methods for normal and Poisson data, we consider three criteria. The first is average mean squared error (AMSE). For each of the 60 $(X_D\Theta)_{ij}$, the mean squared error is defined as the average squared error over the 100 simulated datasets. AMSE for each design cell in the simulation experiment is defined as the average of mean squared error over the 60 $(X_D\Theta)_{ij}$. Thus for the design cell labelled DataK in Table 1, define

$$\widehat{AMSE}_K = \frac{1}{L} \sum_{d=1}^L \sum_{i=1}^N \sum_{j=1}^n [(X_D\widehat{\Theta})_{ij}^d - (X_D\Theta)_{ij}^d]^2 / Nn, \quad (15)$$

where $L = 100, N = 20, n = 3, K = 1, \dots, 4$ for Normal, $K = 5, 6$ for Poisson, Θ is the true value and $\widehat{\Theta}$ is the posterior median of Θ . For each design cell (K), the Monte Carlo

standard error for AMSE is $(100)^{-0.5}$ times the standard deviation, across DataK’s 100 simulated datasets, of $\sum_{i=1}^N \sum_{j=1}^n [(X_D \widehat{\Theta})_{ij}^d - (X_D \Theta)_{ij}^d]^2 / Nn$. The second criterion is the bias of $X_D \Theta$. For each of DataK’s 100 simulated datasets, first compute posterior medians of $(X_D \Theta)_{1,1}, \dots, (X_D \Theta)_{20,3}$, then average each of those posterior medians across the 100 simulated datasets. From this average, subtract the true $(X_D \Theta)_{ij}$ s to give the estimated bias for each of the 60 $(X_D \Theta)_{ij}$ s. MBIAS is defined as the 2.5th, 50th and 97.5th percentiles of the 60 estimated biases. More explicitly, for design cell DataK, MBIAS is:

$$\widehat{MBIAS}_K = 2.5^{th}, 50^{th}, 97.5^{th} \text{ percentiles of } \left(\frac{1}{L} \sum_{d=1}^L (X_D \widehat{\Theta}^d - X_D \Theta^d) \right). \quad (16)$$

Finally, the coverage rate of Bayesian 95% equal-tailed posterior intervals, “PI rate”, is the average coverage rate for the 60 individual $(X_D \Theta)_{ij}$ s.

3.3 Markov chain Monte Carlo specifics

While the MCAR models were implemented in WinBUGS, our SANOVA implementations were coded in R and run on Unix. The different architectures do not permit a fair comparison between the run times of SANOVA and MCAR. However, the SANOVA models have lower computational complexity than the MCAR models: MCAR demands a spectral decomposition in every iteration, while SANOVA does not. For each of our models, we ran three parallel MCMC chains for 10000 iterations. The CODA package in R (www.r-project.org) was used to diagnose convergence by monitoring mixing using Gelman-Rubin diagnostics and autocorrelations (e.g., Gelman et al., 2003, Section 11.6). Sufficient mixing was seen within 500 iterations for the SANOVA models, while 200 iterations typically revealed the same for the MCAR models; we retained 8000×3 samples for the posterior analysis.

3.4 Results

Table 3 and Figures 2 and 3 show the simulation experiment’s results. Table 3 shows AMSE; for all methods and design cells, the standard Monte Carlo errors of AMSE are small, less than 0.07, 0.005, 0.025 for Data1/Data2, Data3/Data4 and Data5/Data6 respectively. Figure 2 shows MBIAS, where the middle symbols represent the median bias and the line

segments represent the 2.5th and 97.5th percentiles. Figure 3 shows coverage of the 95% posterior intervals. Denote SANNOVA with the correct $H_A^{(+)}$ (HA_1) as “SANNOVA correct”, SANNOVA with HA_2 as “SANNOVA incorrect”, the variant SANNOVA with H_{AM} as “SANNOVA variant”, MCAR with $R_{ii} = 0.002$ as “MCAR0.002” and so on.

3.4.1 As expected, SANNOVA with correct $H_A^{(+)}$ performs best.

For normal data, SANNOVA correct has the smallest AMSE for all true η_0 and τ (Table 3). The advantage is larger in Data1 and Data2 where the error precision η_0 is 1, than in Data3 and Data4 where η_0 is 10 (i.e., error variation is smaller). For Poisson data, SANNOVA correct also has the smallest AMSE. Considering MBIAS (Figure 2), SANNOVA correct has the narrowest MBIAS intervals for all cases. In Figure 3, the posterior coverage for SANNOVA correct is nearly nominal. As expected, then, SANNOVA correct performs best among the six methods.

3.4.2 SANNOVA with incorrect HA_2 and H_{AM} perform very well.

Table 3 shows that for normal data, both SANNOVA incorrect and SANNOVA variant have smaller AMSEs than MCAR200 and MCAR0.002, and AMSEs at worst close to MCAR1’s. For Poisson data, Table 3 shows that MCAR0.002 and MCAR1 do somewhat better than SANNOVA incorrect and variant SANNOVA. Considering MBIAS in normal data (Figure 2a), the width of the 95% MBIAS intervals for SANNOVA incorrect are the same as or smaller than for all three MCAR procedures. Similarly, SANNOVA variant has MBIAS intervals better than MCAR0.002 and MCAR200 and almost as good as MCAR1. Figure 2b for Poisson data shows SANNOVA correct, MCAR0.002 and MCAR1 have similar MBIAS intervals. SANNOVA variant in Data5 and SANNOVA incorrect in Data6 show the worst performance for MBIAS apart from MCAR200, whose MBIAS interval is much the widest. Figure 3a shows that for normal data, interval coverage for SANNOVA incorrect and SANNOVA variant is very close to nominal. It appears that the specific value of $H_A^{(+)}$ has little effect on PI coverage rate for the cases considered here. Apart from MCAR200 for Data1/Data2 and MCAR0.002 for Data1 through Data4, which show low coverage, all the other methods have coverage rates greater than 90% for normal data, most close to 95%. For Data3

and Data4, PI rates for MCAR200 reach above 99%. For Poisson data, the PI rates for SANNOVA incorrect and SANNOVA variant are close to nominal and better than MCAR0.002 and MCAR200. In particular, all SANNOVAs have the closest to nominal coverage rates for both normal and Poisson data, which again shows the stability of SANNOVA under different $H_A^{(+)}$ settings.

3.4.3 MCAR is sensitive to the prior on Ω .

To fairly compare SANNOVA and MCAR, we considered MCAR under three different prior settings. For normal data, MCAR1 has the smallest AMSEs and narrowest MBIAS intervals among the MCARs considered, while MCAR0.002 has the largest and widest, respectively. For Poisson data, however, MCAR0.002 has the best AMSE and MBIAS among the MCARs. MCAR200 performs poorly for both Normal and Poisson. The coverage rates in Figure 3 show similar comparisons. These results imply that the prior matters for MCAR: no single prior was always best. By comparison, SANNOVA seems more robust, at least for the cases considered.

3.5 Summary

As expected, SANNOVA correct had the best performance because it uses more correct information. For normal data, SANNOVA incorrect and SANNOVA variant had similar AMSEs, better than two of the three MCARs for the datasets considered. For Poisson data, SANNOVA incorrect and SANNOVA variant had AMSEs as good as those of MCAR0.002 and MCAR1 for Data5 and somewhat worse for Data 6, while showing nearly nominal coverage rates in all cases and less tendency to bias than MCAR in most cases. Replacing the $\Gamma(0.1, 0.1)$ prior for τ with $\Gamma(0.001, 0.001)$ left AMSE and MBIAS almost unchanged and coverage rates a bit worse (data not shown). MCAR, on the other hand, seems more sensitive to the prior on Ω . MCAR0.002 tends to smooth more than MCAR1, more so in normal models where the prior is more influential than in Poisson models. (The latter is true because data give more information about means than variances, and the Poisson model's error variance is the same as its mean, while the normal model's is not.) For the normal data, MCAR0.002's tendency to extra shrinkage appears to make it oversmooth

and perform poorly for Data2 and Data4, where the truth is least smooth. For the Poisson data, MCAR0.002 and MCAR1 give results similar to each other and somewhat better than the SANOVAs except for interval coverage. Therefore SANOVA, with stable results under different $H_A^{(+)}$ and with parameters that are easier to understand and interpret, may be a good competitor to MCAR in multivariate spatial smoothing.

4 Example: Minnesota 3-cancer data

Researchers in different fields have illustrated that accounting for spatial correlation could provide insights that would have been overlooked otherwise, while failure to account for spatial association could potentially lead to spurious and sometimes misleading results (see, e.g. Turechek and Madden, 2002; Ramsay et al., 2003; Lichstein et al., 2002). Among the widely investigated diseases are the different types of cancers. We applied SANOVA and MCAR to a cancer-surveillance dataset describing total incidence counts of 3 cancers (lung, larynx, esophagus) in Minnesota's 87 counties for the years 1990 to 2000 inclusive.

Minnesota's geography and history make it plausible that disease incidence would show spatial association. Three major North American land forms meet in Minnesota: the Canadian Shield to the north, the Great Plains to the west, and the eastern mixed forest to the southeast. Each of these regions is distinctive in both its terrain and its predominant economic activity: mining and outdoors tourism in the mountainous north, highly mechanized crop cultivation in the west, and dairy farming in the southeast. The different regions were also settled by somewhat different groups of in-migrants, e.g., disproportionately many Scandinavians in the north. These factors imply spatial association in occupational hazards as well as culture, weather, and access to health care especially in the thinly-populated north, which might be expected to produce spatial association in diseases.

With multiple cancers one obvious option is to fit a separate univariate model for each cancer. But diseases may share the same spatially distributed risk factors, or the presence of one disease might encourage or inhibit the presence of another in a region, for example, larynx and esophagus cancer have been shown to be closely related spatially (Baron et al 1993). Thus, we may need to account for dependence among the different cancers while

maintaining spatial dependence between sites.

Although the dataset has counts broken out by age groups, for the present purpose we ignore age standardization and just consider total counts for each cancer. Age standardization would affect only the expected cancer counts E_{ij} , while other covariates could be added to either SANOVA or MCAR as unsmoothed fixed effects (i.e., in the A_1 design matrix).

Given the population and disease count of each county, we estimated the expected disease count for each cancer in each county using the Poisson model. Denote the 87×3 counts as $y_{1,1}, \dots, y_{87,3}$; then the model is

$$\begin{aligned} y_{ij} | \mu_{ij} &\sim \text{Poisson}(\mu_{ij}) \\ \log(\mu_{ij}) &= \log(E_{ij}) + (X_D \Theta)_{ij}, \end{aligned} \tag{17}$$

where $X_D \Theta$ is the SANOVA structure and Θ has priors as in section 2.2.2. For disease j in county i , $E_{ij} = P_i \frac{\sum_i O_{ij}}{\sum_i P_i}$, where O_{ij} is the disease count for county i and disease j and P_i is county i 's population. For the SANOVA design matrix, we consider HA_1 and HA_2 from the simulation experiment, though now neither is known to be correct. We also consider a variant SANOVA analysis using $H_A^{(+)}$ estimated from the MCAR1 model, to test the stability of the SANOVA results. Appendix D describes the latter analysis.

Figures 4 to 6 show the data and results for MCAR1 and SANOVA with HA_1 . In each Figure, the upper left plot shows the observed y_{ij}/E_{ij} ; the two lower plots show the posterior median of μ_{ij}/E_{ij} for MCAR1 and SANOVA with HA_1 . Lung cancer counts tended to be high and thus were not smoothed much by any method, while counts of the other cancers were much lower and thus smoothed considerably more (see also Figure 7). Since SANOVA with HA_1 , HA_2 and estimated $H_A^{(+)}$ gave very similar results, only those for HA_1 are shown. Results for MCAR0.002 are similar to those for MCAR1, so they are omitted. As expected, MCAR200 shows the least shrinkage among the three MCARs and gives some odd μ_{ij}/E_{ij} .

To compare models, we calculated the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002). To define DIC, define the deviance $D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y} | \boldsymbol{\theta}) + 2 \log h(\mathbf{y})$, where $\boldsymbol{\theta}$ is the parameter vector in the likelihood and $h(\mathbf{y})$ is a function of the data. Since h does not affect model comparison, we set $\log h(\mathbf{y})$ to 0. Let $\bar{\boldsymbol{\theta}}$ be the posterior mean of $\boldsymbol{\theta}$ and \bar{D} the posterior expectation of $D(\boldsymbol{\theta})$. Then define $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$ to be a measure

of model complexity and define $DIC = \overline{D} + p_D$.

Table 4 shows \overline{D} , p_D and DIC for nine analyses, SANOVA with 3 different $H_A^{(+)}$, MCAR with 3 different priors for Ω , and 3 fits of univariate CAR models to the individual diseases, discussed below. Considering \overline{D} , the three SANOVAs and MCAR1 are similar; Figures 4 to 6 show the fits are indeed similar. Figure 7 reinforces this point, showing that MCAR1 and SANOVA with HA_1 induce similar smoothing for the three cancers. SANOVA with $H_A^{(+)}$ estimated from MCAR has the smallest \overline{D} (1458), though its model complexity penalty ($p_D = 103$) is higher than MCAR0.002's ($p_D = 79$). Despite having the second worst fit (\overline{D}), MCAR0.002 has the best DIC, and the three SANOVAs have DICs much closer to MCAR0.002's than to the other MCARs'. Generally, all SANOVA models have similar \overline{D} (≈ 1460) and DIC (≈ 1562), while MCAR results are sensitive to Ω 's prior, consistent with the simulation experiment.

For comparison, we fit separate univariate CAR models to the three diseases considering three different priors for the smoothing precision, $\tau \sim \text{Gamma}(a, a)$ for $a = 0.001, 1$, and 1000. For each prior, we added up \overline{D} , p_D and DIC for three diseases (see Table 4). With $a = 0.001$ and 1, we obtained \overline{D} 's (1461 and 1453 respectively) competitive with SANOVA, MCAR0.002 and MCAR1 but with considerably greater complexity penalties (141 and 149 respectively) and thus DICs slightly larger than 1600. For $a = 1000$, we obtained an even lower \overline{D} (1432) but an increased penalty (180) resulted in a poorer DIC score. Figure 7 shows fitted values for CAR1, which were smoothed like MCAR1 and SANOVA for lung and esophagus cancers but smoothed rather more for larynx cancer. Overall, these results reflect some gain in performance from accounting for the space-cancer interactions/associations.

To further examine the smoothing under SANOVA, Figure 8 shows separate maps for the county main effect and interactions from the SANOVA fit with HA_1 . The upper left plot is the cancer main effect, the mean of the three cancers; the lower left plot is the comparison of lung versus average of larynx and esophagus; the lower right plot is the comparison of larynx versus esophagus. All values are on the same scale as y_{ij}/E_{ij} in Figures 4 to 6 and use the same legend. The two interaction contrasts are smoothed much more than the county main effect, agreeing with previous research that larynx and esophagus cancer are closely related spatially (Baron et al 1993). To see whether the interactions are necessary, we fit

a SANOVA model (using HA_1) without the interactions. As expected, model complexity decreased ($p_D = 77$), while \bar{D} increased slightly, so DIC became 1558, a bit better than SANOVA with interactions.

Now consider the posterior of the MCAR’s precision matrix Ω . The posterior mean of Ω is much larger for MCAR0.002 than MCAR200; the diagonal elements are larger by 4 to 5 orders of magnitude. This may explain the poor coverage for MCAR0.002 in the simulation. Further, consider the correlation matrix arising from the inverse of Ω ’s posterior mean. As the diagonals of R change from 0.002 to 200, the correlation between any two cancers decreases and the complexity penalty p_D increases. By comparison, the three SANOVAs have similar model fits and complexity penalties, leading to similar DICs. So again, in this sense SANOVA shows greater stability.

5 Discussion and future work

We used SANOVA to do spatial smoothing and compared it with the much more complex MCAR model. For the cases considered here, we found SANOVA with spatial smoothing to be an excellent competitor to MCAR. It yielded essentially indistinguishable inference, while being easier to fit and interpret. In the SANOVA model, $H_A^{(+)}$ is assumed known. For most of the SANOVA models considered, $H_A^{(+)}$ ’s first column was fixed to represent the average over diseases, while other columns were orthogonal to the first column. Alternatively, $H_A^{(+)}$ could be treated as unknown and estimated as part of the analysis. With this extension, SANOVA with spatial effects is a re-parameterization of the MCAR model and gains the MCAR model’s flexibility at the price of increased complexity. This extension would be non-trivial, involving sampling from the space of orthogonal matrices while avoiding identification problems arising from, e.g., permuting columns of $H_A^{(+)}$.

Other covariates can be added to a spatial SANOVA. Although (10) is a saturated model, spatial smoothing “leaves room” for other covariates. Such models would suffer from collinearity of the CAR random effects and the fixed effects, as discussed by Reich et al. (2006), who gave a variant analysis that avoids the collinearity.

For datasets with spatial and temporal aspects, e.g., the 11 years in the Minnesota 3-

cancer data, interest may lie in the counts' spatial pattern and in their changes over time. By adding a time effect, SANOVA can be extended to a spatiotemporal model. Besides spatial and temporal main effects, their interactions can also be included and smoothed. There are many modeling choices; the simplest model is an additive model without space-time interactions, where the spatial effect has a CAR model and the time effect a random walk, which is a simple CAR. But many other choices are possible.

We have examined intrinsic CAR models, where $Q_{ij} = -1$ if region i and region j are connected. SANOVA with spatial smoothing could be extended to more general CAR models. Banerjee et al. (2004) replaced Q with the matrix $D_w - \rho W$, where D_w is diagonal with the same diagonal as Q and $W_{ij} = 1$ if region i is connected with region j , otherwise $W_{ij} = 0$. Setting $\rho = 1$ gives the intrinsic CAR model considered in this paper. For known ρ , the SANOVA model described here is easily extended by replacing Q in Section 2 with $D_w - \rho W$. However, for unknown ρ , our method cannot be adjusted so easily, because updating ρ in the MCMC would force V and the design matrix to be updated as well, but this would change the definition of the parameter Θ . Therefore, a different approach is needed for unknown ρ .

A different extension of SANOVA would be to survival models for areal spatial data (e.g., Li & Ryan, 2002; Banerjee, Wall & Carlin, 2003; Diva, Dey and Banerjee, 2008). If the regions are considered strata, then random effects corresponding to nearby regions might be similar. In other words, we can embed the SANOVA structure in a spatial frailty model. For example, the Cox model with SANOVA structure for subject j in stratum i is

$$h(t_{ij}, X_{ij}) = h_0(t_{ij}) \exp(X_{ij}\beta), \quad (18)$$

where X is the design matrix, which may include a spatial effect, a temporal effect, their interactions and other covariates. Banerjee et al. (2004) noted that in the CAR model, considering both spatial and non-spatial frailties, the frailties are identified only because of the prior, so the choice of priors for precisions is very important.

Besides the above extensions, HCSC introduced tools for normal SANOVA models that can be extended to non-normal SANOVA models. For example, HCSC defined the degrees of freedom in a fitted model as a function of the smoothing precisions. This can be used as a measure of the fit's complexity, or a prior can be placed on the degrees of freedom as

a way of inducing a prior on the unknowns in the variance structure. The latter is under development and will be presented soon.

APPENDICES

A Derivation of the precision matrix for (11)

Proof. Let $\mathbf{T} = (\boldsymbol{\Theta}'_{CO}, \boldsymbol{\Theta}'_{CO \times CA_1}, \dots, \boldsymbol{\Theta}'_{CO \times CA_{n-1}})'$; $f(\mathbf{T}|W) \propto \exp(-\frac{1}{2}\mathbf{T}'W\mathbf{T})$ with precision matrix W . Define $\mathbf{F} = K\mathbf{T}$, where $K = (V^{(-)} \otimes \frac{1}{\sqrt{n}}\mathbf{1}_n | V^{(-)} \otimes H_{CA}^{(1)} | \dots | V^{(-)} \otimes H_{CA}^{(n-1)})$ as in section 2.2.2. Because $K'K = I_{n(N-1)}$, therefore $K'\mathbf{F} = K'K\mathbf{T} = \mathbf{T}$. Changing variables in f from \mathbf{T} to \mathbf{F} , $f(\mathbf{F}|W) \propto \exp(-\frac{1}{2}\mathbf{F}'KWK'\mathbf{F})$, so \mathbf{F} has precision matrix KWK' . With

$$\mathbf{W} = \begin{bmatrix} D^{(-)} \otimes \tau_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D^{(-)} \otimes \tau_{n-1} \end{bmatrix},$$

$K\mathbf{T}$ has precision KWK'

$$\begin{aligned} &= (V^{(-)} \otimes \frac{1}{\sqrt{n}}\mathbf{1}_n | \dots | V^{(-)} \otimes H_{CA}^{(n-1)}) \begin{bmatrix} D^{(-)} \otimes \tau_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D^{(-)} \otimes \tau_{n-1} \end{bmatrix} \begin{pmatrix} V^{(-)'} \otimes \frac{1}{\sqrt{n}}\mathbf{1}'_n \\ \vdots \\ V^{(-)'} \otimes H_{CA}^{(n-1)'} \end{pmatrix} \\ &= V^{(-)}D^{(-)}V^{(-)'} \otimes \frac{\tau_0}{n}\mathbf{1}_n\mathbf{1}'_n + V^{(-)}D^{(-)}V^{(-)'} \otimes \tau_1 H_{CA}^{(1)} H_{CA}^{(1)'} + \dots \\ &= Q \otimes H_A^{(+)} \text{diag}(\tau_j) H_A^{(+)'} \end{aligned}$$

where $H_A^{(+)} = (\frac{1}{\sqrt{n}}\mathbf{1}_n | H_{CA})$ is an orthogonal matrix; τ_j is unknown and $H_A^{(+)} \text{diag}(\tau_j) H_A^{(+)}'$ is a precision matrix; $Q = VDV' = V^{(-)}D^{(-)}V^{(-)'}$ is a function of the neighborhood structure as defined in section 2.1.2. \square

B MCMC algorithms

For the normal error case, we used the conditional distribution of η_0 as in HCSC. If $r_j = \frac{\eta_j}{\eta_0}$ and η_0 has a $\Gamma(\alpha, \beta)$ prior, it follows that

$$f(\eta_0 | \mathbf{r}, \mathbf{Y}) = \frac{\beta + \frac{1}{2}W(\mathbf{r})^\xi}{\Gamma(\xi)} \eta_0^{\xi-1} \exp(-\eta_0(\beta + W(\mathbf{r})/2)) \quad (19)$$

where $\xi = \frac{cn-M_1}{2} + \alpha$; $\mathbf{r} = \{r_j\}$; and $W(\mathbf{r}) = \mathbf{Y}'S^{-1}\mathbf{Y} - \mathbf{Y}'S^{-1}X(X'S^{-1}X)^{-1}X'S^{-1}\mathbf{Y}$, with $S^{-1} = \Gamma^{-1}/\eta_0$, which is a function of \mathbf{r} . The only change here compared to HCSC is that we normalized the columns of X_D (data cases) so each column of X_D has length 1. To do MCMC for both the normal and Poisson models, we first draw Θ 's components univariately using an M-H algorithm. For a gamma prior on τ_j , τ_j 's full conditional posterior is a gamma, so we draw τ_j using a Gibbs sampler. For normal data, an additional step is to draw η_0 by Gibbs sampler using (19). For Poisson data, a transformation of Θ was used in the chain, so the new parameter Ξ was drawn first, then Θ was calculated from Ξ ; Appendix C gives details. Since simulation steps for the normal and Poisson data are similar, we only list them for normal data. For current draws θ_i , τ_j and η_0 :

- for $i = 1, \dots, 60$ (Nn):
 - Generate $\theta_i^* \sim N(\theta_i, 4)$
 - Calculate the MH ratio as $\frac{f(\theta_i^*|\mathbf{Y}, \boldsymbol{\tau})}{f(\theta_i|\mathbf{Y}, \boldsymbol{\tau})}$; accept θ_i^* with probability $\min(1, \text{MH ratio})$.
- Generate $\tau_j^* \sim \Gamma(\alpha + \frac{19}{2}, \beta + \frac{1}{2} \sum_{i=1}^{19} D_{ii}^{(-)} \theta_{(i+3+19 \times j)}^2)$, where $j = 0, 1, 2$
- Sample η_0^* from the gamma distribution in (19).

All computations were carried out on a dual-core 2.20GHz AMD Athlon processor with 1.93GB of RAM physical address extension. MCMC code for SANOVA models was written in R; MCAR models were analyzed in WinBUGS. For each model we ran three parallel chains for 10,000 iterations and discarded the first 2,000 iterations as “burn-in”. The combined draws of the three chains are used for posterior summaries. Trace plots showed good convergence for all the models.

C Mean-structure transformation for the Poisson case

The Poisson model in the Θ parameterization has high posterior correlations among the θ s, which leads to slow MCMC mixing. To see this, let $\tilde{y}_{ij} = \log y_{ij} - \log E_{ij}$; then from the normal approximation to the Poisson,

$$\tilde{\mathbf{Y}} = \begin{bmatrix} \tilde{\mathbf{y}}_{Nn} \\ \mathbf{0}_{Nn} \end{bmatrix} \approx X\Theta + \mathbf{e} = \begin{bmatrix} X_D \\ I_{Nn} \end{bmatrix} \Theta + \mathbf{e} \quad (20)$$

where $\mathbf{e} \sim N(\mathbf{0}, \Gamma)$ with

$$\Gamma = \begin{bmatrix} \text{diag}(1/y_i) & 0 & 0 \\ 0 & 10^6 I_n & 0 \\ 0 & 0 & \text{diag} \frac{1}{\tau D^{(-)}} \end{bmatrix} \quad (\text{see section 2.3.1}).$$

As a result,

$$\begin{aligned} \text{cov}(\Theta | \tilde{\mathbf{Y}}, \Gamma) &\approx (X'(\text{cov}(\tilde{\mathbf{Y}}))^{-1}X)^{-1} \\ &= \left[(X'_D | I_{Nn}) \Gamma^{-1} \begin{pmatrix} X_D \\ I_{Nn} \end{pmatrix} \right]^{-1} \\ &= [X'_D \text{diag}(y_i) X_D + A^{-1}]^{-1} \\ &\equiv \Upsilon, \end{aligned} \quad (21)$$

where

$$A = \begin{bmatrix} 10^6 I_n & 0 \\ 0 & \text{diag} \frac{1}{\tau D^{(-)}} \end{bmatrix}.$$

Although A is diagonal, $X'_D \text{diag}(y_i) X_D$ is not. The approximate data-case precision matrix $\text{diag}(y_i)$ amplifies the off-diagonal entries of Υ substantially for the Minnesota cancer data because E_i and y_i range over a few orders of magnitude, so the posterior correlations among the θ s are large. To minimize this effect, we re-parameterized as follows. Let $\Delta = (\mathbf{e}_1 | \mathbf{e}_2 | \dots | \mathbf{e}_p)'$, where \mathbf{e}_i is the i^{th} eigenvector of Υ for a particular choice of τ ; we discuss this below. Therefore, Δ is a $p \times p$ matrix, where $p = \text{dim}(\Theta) = Nn$ in the Poisson simulations. By the spectral decomposition, if $\lambda_1, \dots, \lambda_p$ are the eigenvalues corresponding to $\mathbf{e}_1, \dots, \mathbf{e}_p$,

$$\begin{aligned} \text{Cov}(\Theta | \tilde{\mathbf{Y}}, \Gamma) &\approx \Upsilon \\ &= \Delta' \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \Delta. \end{aligned}$$

Let $\Xi = \Delta \Theta$; then

$$\tilde{\mathbf{Y}} \approx X \Theta + \mathbf{e} = X \Delta' \Delta \Theta + \mathbf{e} = X^* \Xi + \mathbf{e}. \quad (22)$$

From (21),

$$\begin{aligned}
Cov(\Xi|Y, \Theta, \Gamma) &= \Delta cov(\Theta|Y, \Gamma) \Delta' \\
&\approx \Delta \Upsilon \Delta' \\
&= \Delta \Delta' \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix} \Delta \Delta' \\
&= \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_p \end{bmatrix}.
\end{aligned}$$

We can then sample Ξ instead of Θ . Ξ and Θ are one-to-one, so a draw of Ξ gives a draw of Θ . Since A is a function of τ , Δ is also a function of τ . Any MCMC routine will update τ frequently, so it is not obvious how to specify a general Δ to use in this transformation. Fortunately, τ had only a small effect on off-diagonal entries of Υ in our datasets, so we used $\tau = (100, 100, 100)$ to calculate a value of Δ which was then fixed and used in the MCMC. With this τ , the conditional posterior correlations among the Ξ s were mostly below 0.1, which is good enough to solve the mixing problem.

D Estimating $H_A^{(+)}$ from MCAR1

We used the posterior median of Ω , $\hat{\Omega}$, which had spectral decomposition

$$\begin{aligned}
\hat{\Omega} &= \begin{bmatrix} 11.85 & -2.21 & -5.05 \\ -2.21 & 5.24 & -2.09 \\ -5.05 & -2.09 & 6.88 \end{bmatrix} \\
&= \begin{pmatrix} 0.86 & 0.26 & -0.45 \\ -0.08 & -0.78 & -0.62 \\ -0.51 & 0.57 & -0.65 \end{pmatrix} \begin{bmatrix} 15.06 & 0 & 0 \\ 0 & 7.47 & 0 \\ 0 & 0 & 1.43 \end{bmatrix} \begin{pmatrix} 0.86 & -0.08 & -0.51 \\ 0.26 & -0.78 & 0.57 \\ -0.45 & -0.62 & -0.65 \end{pmatrix}. \tag{23}
\end{aligned}$$

Let Λ be the eigenvector matrix of $\hat{\Omega}$, the left-most matrix in the second row of (23). Then in the design matrix in (10), for the data cases (X_D) replace $H_A^{(+)} \otimes V^{(-)}$ with $\Lambda \otimes V^{(-)}$. Unlike HA_1 and HA_2 , this specification of $H_A^{(+)}$ has no county main effect.

REFERENCES

- Banerjee S, Carlin BP, Gelfand AE (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman & Hall/CRC Press.
- Banerjee S, Wall MM, Carlin BP (2003). Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*, 4:123-142.
- Baron AE, Franceschi S, Barra S, Talamini R, La Vecchia C (1993). Comparison of the joint effect of alcohol and smoking on the risk of cancer across sites in the upper aerodigestive tract. *Cancer Epidemiology Biomarkers and Prevention*, 2:519-523.
- Besag J (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Stat. Soc., Ser. B*, 36:192-236.
- Besag, J, Green, P, Higdon, D, Mengersen, K (1995). Bayesian computation and stochastic systems (with discussion). *Statist. Science* 10:3–66.
- Besag, J, York, JC, and Mollié, A (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, **43**:1–59.
- Carlin, BP and Banerjee, S (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. In *Bayesian Statistics 7*, eds. J.M. Bernardo et al., Oxford: Oxford University press, pp.45–64.
- Elliott P, Wakefield JC, Best NG, Briggs DJ, eds. (2000). *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press.
- Gelfand AE, Vounatsou P (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4:11-25.
- Haran M, Hodges JS, Carlin BP (2002). Accelerating computation in Markov random field models for spatial data via structured MCMC. *J. Comp. Graph. Stat.*, 12:249-264.
- He Y, Hodges JS, Carlin BP (2007). Re-considering the variance parameterization in multiple precision models. *Bayes. Anal.*, 2:1-28.

- Hodges JS (1998). Some algebra and geometry for hierarchical models. *J. Roy. Stat. Soc., Ser. B*, 60:497-536.
- Hodges JS, Sargent DJ (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika*, 88:367-379.
- Hodges JS, Carlin BP, Fan Q (2003). On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, 59:317-322.
- Hodges JS, Cui Y, Sargent DJ, Carlin BP (2007). Smoothing balanced single-error-term analysis of variance. *Technometrics*, 49:12-25.
- Jin X, Carlin BP, Banerjee S (2005) Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, 61:950-961.
- Jin X, Banerjee S, Carlin BP (2007). Order-free coregionalized lattice models with application to multiple disease mapping. *J. Roy. Stat. Soc., Ser. B*, 69:817-838.
- Kim H, Sun D, Tsutakawa RK (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *J. Amer. Stat. Assn.*, 96:1506-1521.
- Lawson AB, Biggeri AB, Bohning D, Lesaffre E, Viel JF, Bertollini R (1999). *Disease Mapping and Risk Assessment for Public Health*, New York: John Wiley.
- Lee Y, Nelder JA (1996). Hierarchical generalized linear models (with discussion). *J. Roy. Stat. Soc., Ser. B*, 58:619-673.
- Li Y, Ryan L (2002). Modeling spatial survival data using semi-parametric frailty models. *Biometrics*, 58:287-297.
- Lichstein JW, Simons TR, Shriner SA, Franzreb KE (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72:445-463.
- Lu H, Hodges JS, Carlin BP (2004). Measuring the complexity of generalized linear hierarchical models. *Can. J. Stat.*, 19:69-87.

- Mardia KV (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J. Multivar. Anal.*, 24:265-284.
- Ramsay, T, Burnett, R, and Krewski, D (2003). Exploring bias in a generalized additive model for spatial air pollution data. *Environmental Health Perspectives*, 111:1283-1288.
- Reich BJ, Hodges JS, Zadnik V (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, 62:1197-1206.
- Rue H, Held L (2005). *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman and Hall/CRC.
- Sain SR, Cressie N (2002). Multivariate lattice models for spatial environmental data. In *Proc. A.S.A. Section on Statistics and the Environment*. Alexandria, VA: American Statistical Association, pp. 2820-2825.
- Sargent DJ, Hodges JS, Carlin BP (2000). Structured Markov chain Monte Carlo. *J. Comp. Graph. Stat.*, 9:217-234.
- Turechek, WW and Madden, LV (2002). A generalized linear modeling approach for characterizing disease incidence in spatial hierarchy. *Phytopathology*, 93:458-466.
- Wakefield J (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8:158-183.
- Wall MM (2004). A close look at the spatial structure implied by the CAR and SAR models. *J. Stat. Plan. Inf.* 121:311-324.
- Waller LA, Gotway CA (2004). *Applied Spatial Statistics for Public Health Data*. New York: Wiley.

Error distribution	η_0	$(\tau_0/\eta_0, \tau_1/\eta_0, \tau_2/\eta_0) (\tau_0, \tau_1, \tau_2)$	Data Name
Normal	1	(100,100,0.1)	Data1
	1	(0.1,100,0.1)	Data2
	10	(100,100,0.1)	Data3
	10	(0.1,100,0.1)	Data4
Poisson	NA	(100,100,0.1)	Data5
	NA	(0.1,100,0.1)	Data6

Table 1: Experimental conditions in the simulation experiments

procedure	prior
SANOVA with correct $H_A^{(+)}$	$\eta_0, \tau_j \sim \Gamma(0.1, 0.1)$ for $j = 0, 1, 2$
SANOVA with incorrect $H_A^{(+)}$	$\eta_0, \tau_j \sim \Gamma(0.1, 0.1)$ for $j = 0, 1, 2$
variant SANOVA with H_{AM}	$\eta_0, \tau_j \sim \Gamma(0.1, 0.1)$ for $j = 0, 1, 2$
MCAR	$\Omega \sim Wishart(R, 3), R = 0.002I_3$
MCAR	$\Omega \sim Wishart(R, 3), R = I_3$
MCAR	$\Omega \sim Wishart(R, 3), R = 200I_3$

Table 2: The six statistical methods used in the simulation experiment

model	normal-error model				Poisson-error model	
	Data1	Data2	Data3	Data4	Data5	Data6
SANOVA with HA_1	0.34	0.60	0.04	0.06	0.02	0.04
SANOVA with HA_2	0.47	0.84	0.05	0.07	0.02	0.14
SANOVA with HA_{AM}	0.48	0.74	0.05	0.06	0.03	0.11
MCAR with $R_{ii} = 0.002$	0.66	1.88	0.04	0.13	0.02	0.04
MCAR with $R_{ii} = 1$	0.36	0.84	0.04	0.06	0.02	0.06
MCAR with $R_{ii} = 200$	0.93	0.92	0.09	0.09	0.24	0.36

Table 3: AMSE for simulated normal and Poisson data

Model	\bar{D}	p_D	DIC
SANOVA with HA_1	1461	103	1564
SANOVA with HA_2	1463	102	1565
SANOVA with H_A estimated from MCAR1	1458	103	1561
MCAR0.002	1476	79	1555
MCAR1	1459	132	1591
MCAR200	1559	356	1915
CAR0.001	1461	141	1602
CAR1	1453	149	1602
CAR1000	1432	180	1612

Table 4: Model comparison using DIC

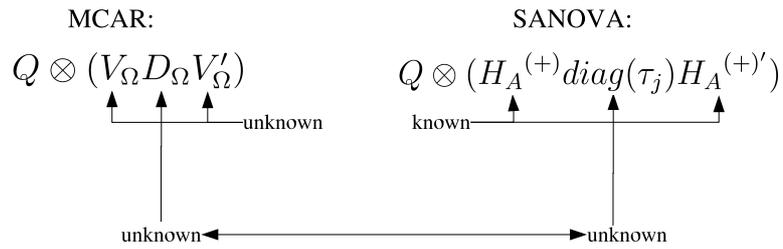
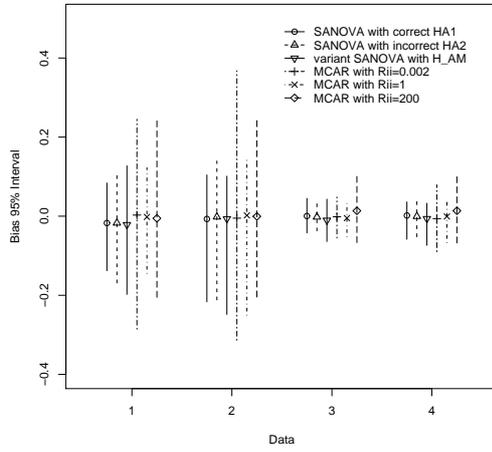
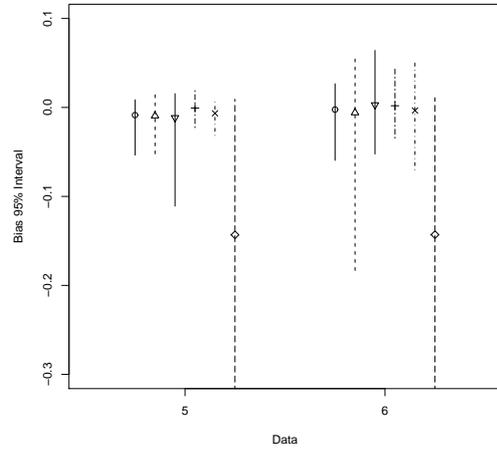


Figure 1: Comparing prior precision matrices for ϕ in MCAR and SANOVA.

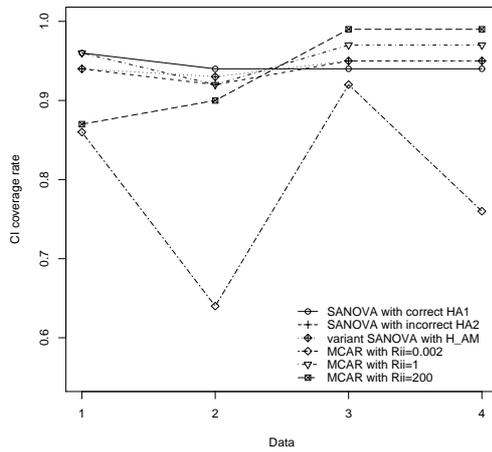


(a) Normal

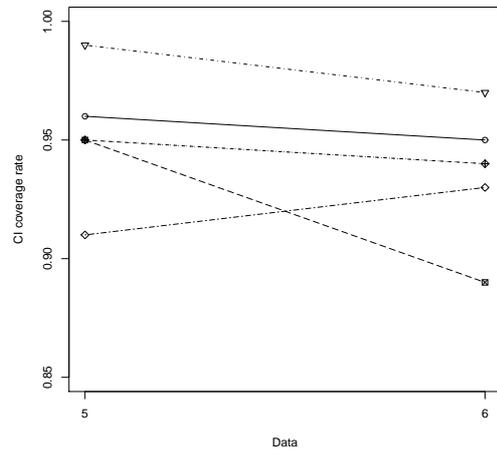


(b) Poisson

Figure 2: MBIAS for simulated normal and Poisson data

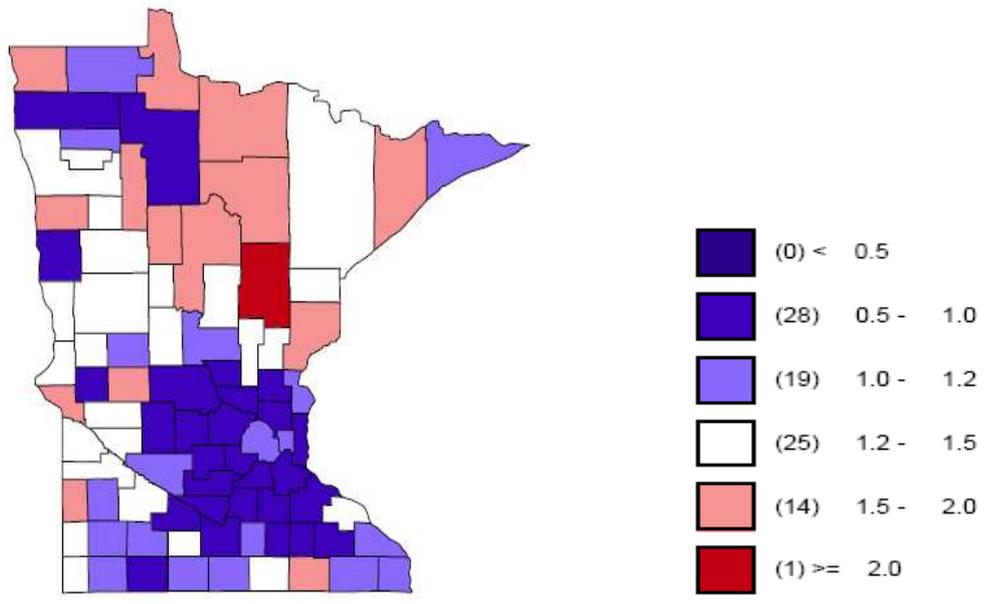


(a) Normal

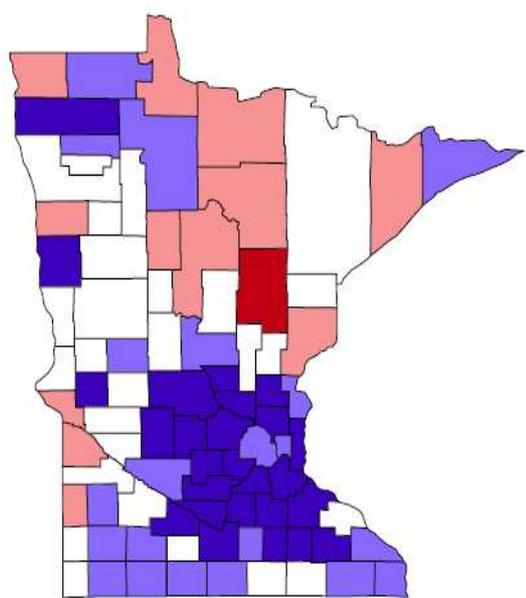


(b) Poisson

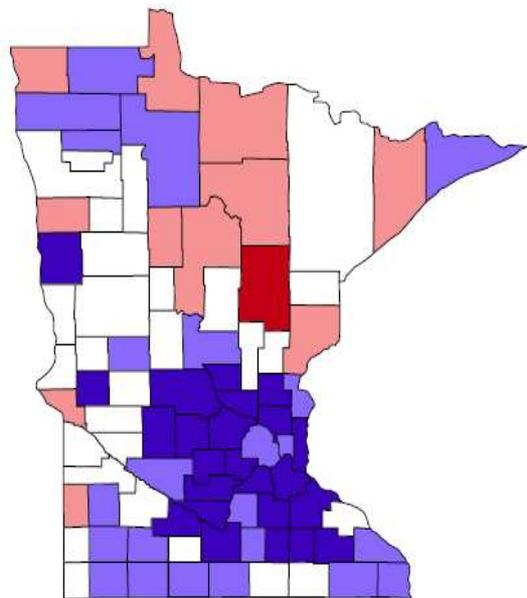
Figure 3: PI rate for simulated normal and Poisson data



(a) data: y_{ij}/E_{ij}

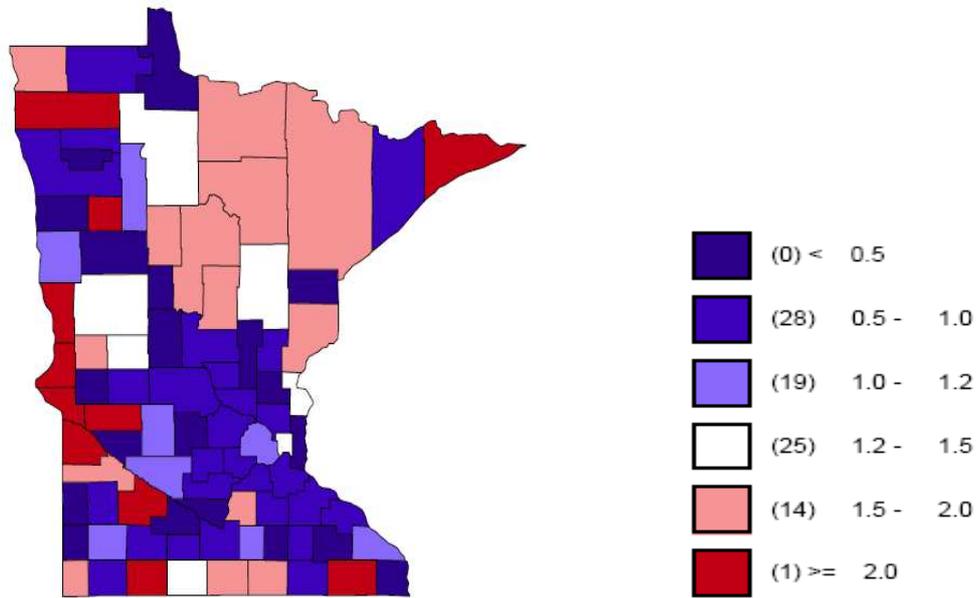


(b) fitted: SANOVA with HA_1

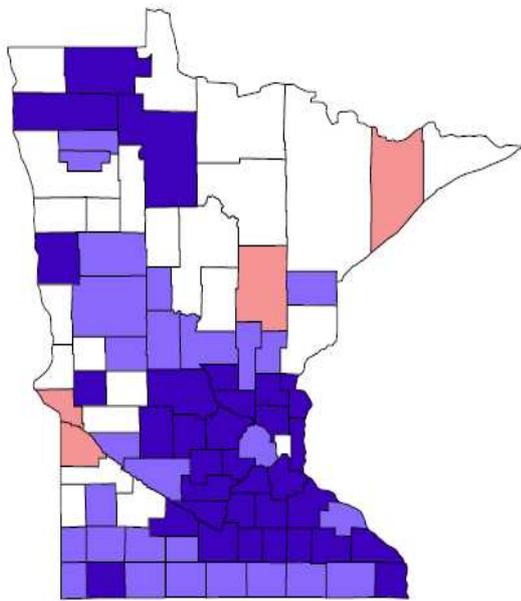


(c) fitted: MCAR with $R_{ii} = 1$

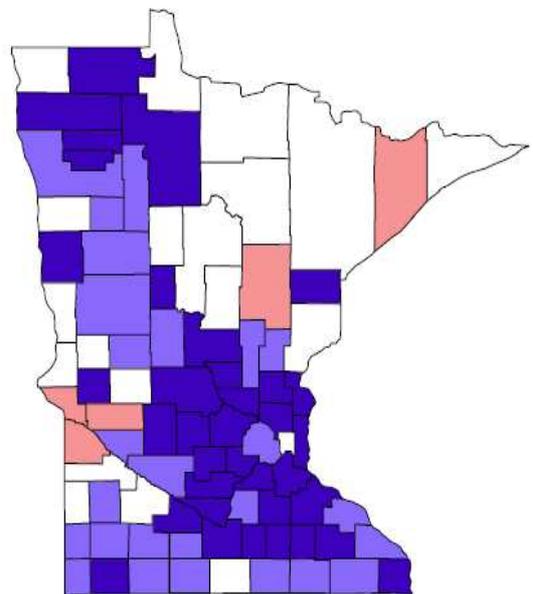
Figure 4: Lung cancer data and fitted values



(a) data: y_{ij}/E_{ij}

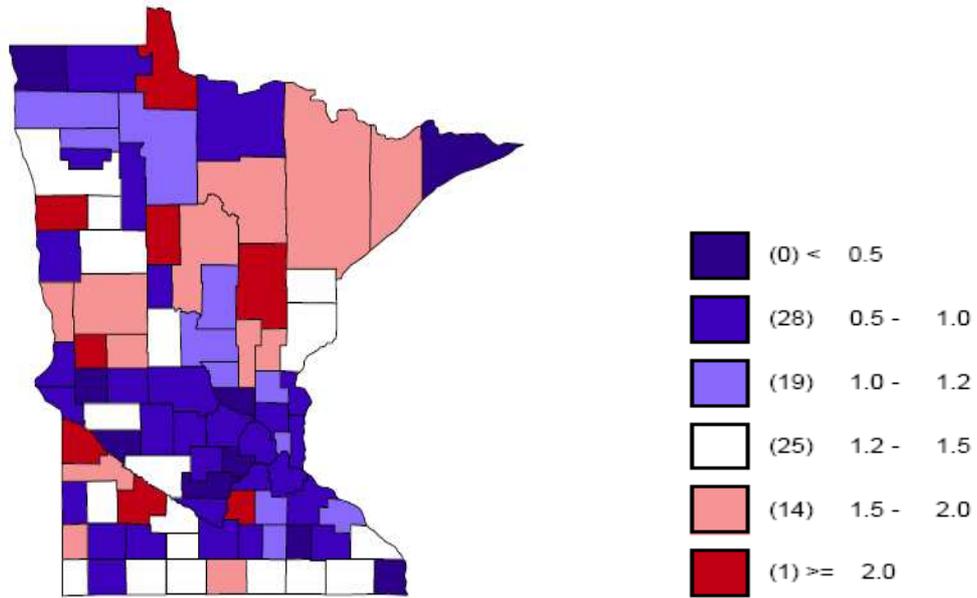


(b) fitted: SANOVA with HA_1

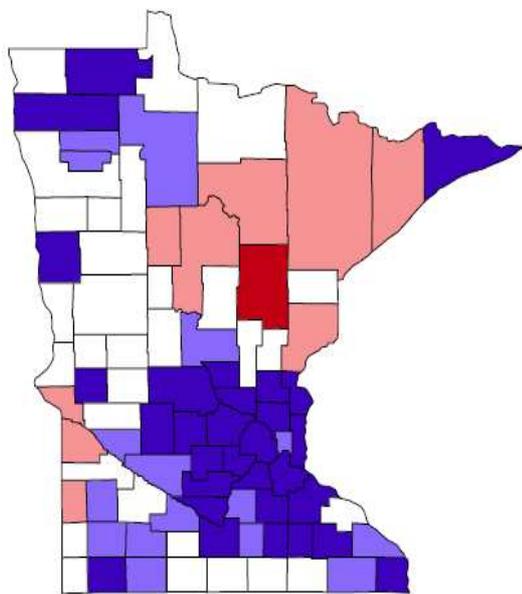


(c) fitted: MCAR with $R_{ii} = 1$

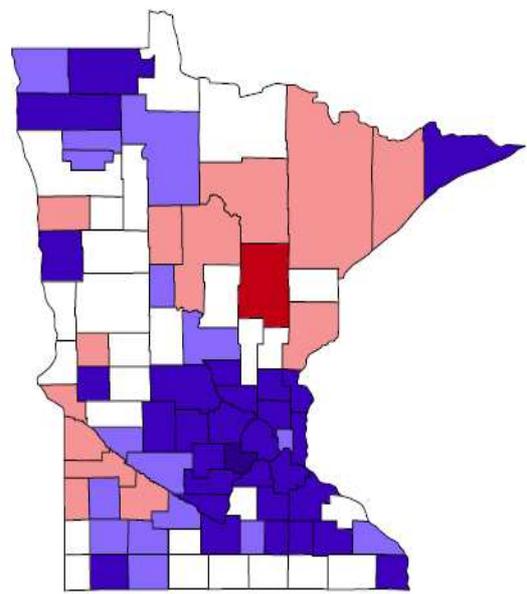
Figure 5: Larynx cancer data and fitted values



(a) data: y_{ij}/E_{ij}



(b) fitted: SANOVA with HA_1



(c) fitted: MCAR with $R_{ii} = 1$

Figure 6: Esophagus cancer data and fitted values

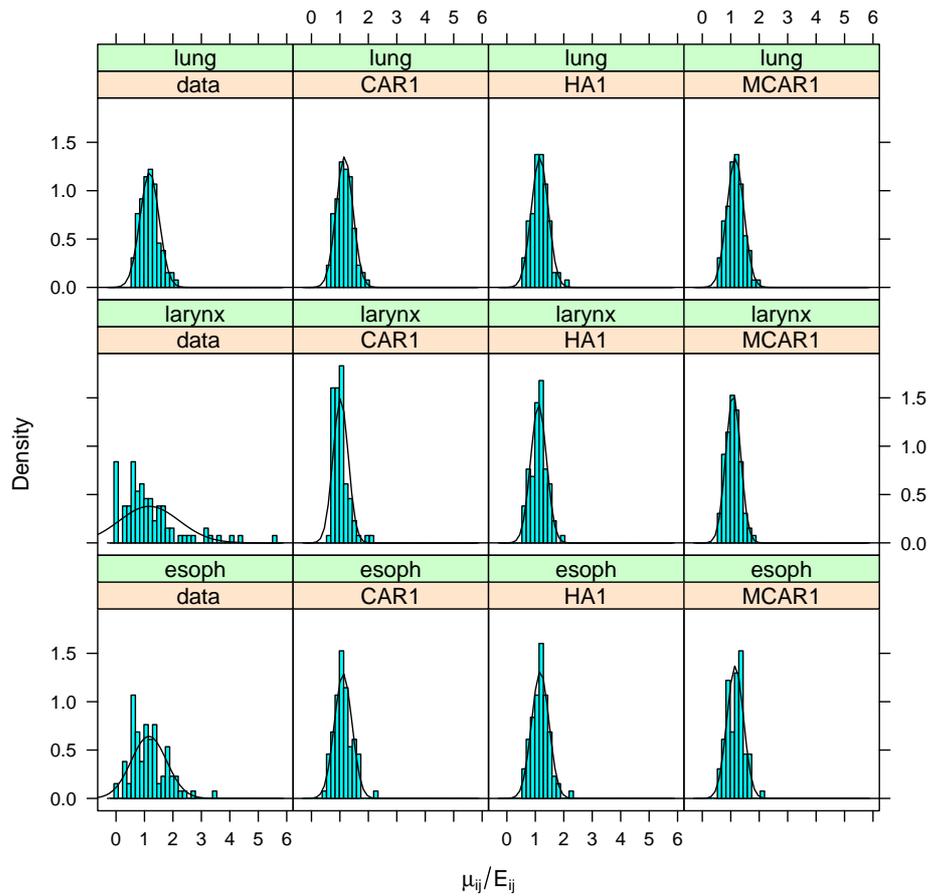


Figure 7: Comparing data and fitted values for each cancer. The “Data” panel shows the density for y_{ij}/E_{ij} , while the other three panels show the posterior median of μ_{ij}/E_{ij} for univariate CAR, SANOVA and MCAR;

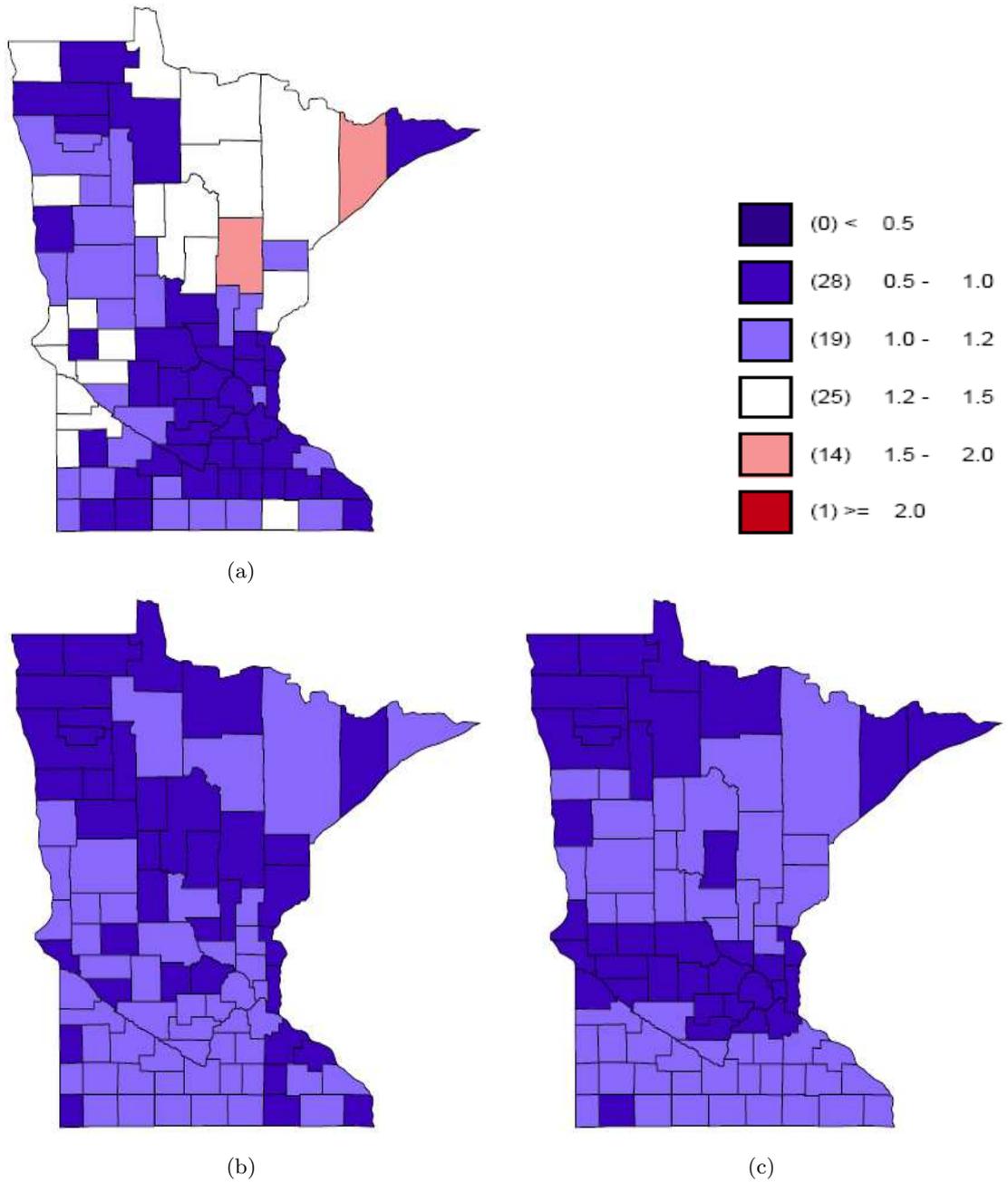


Figure 8: SANOVA with HA_1 : (a) county main effect; (b) cancer \times county interaction 1 for larynx; (c) cancer \times county interaction 2 for esophagus.