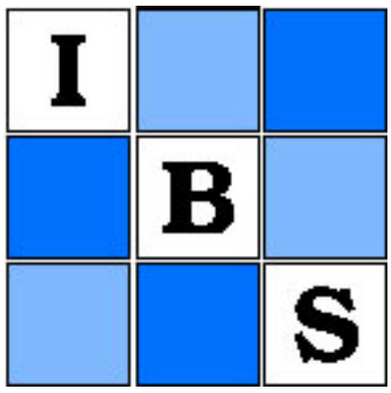

A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data
Author(s): Paul S. Albert and Lisa M. McShane
Source: *Biometrics*, Vol. 51, No. 2 (Jun., 1995), pp. 627-638
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2532950
Accessed: 01/10/2014 12:49

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*.

http://www.jstor.org

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

BIOMETRICS 51, 627–638
June 1995

A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data

Paul S. Albert* and Lisa M. McShane‡

Biometry and Field Studies Branch,
National Institute of Neurological Disorders and Stroke,
National Institutes of Health,
Bethesda, Maryland 20892, U.S.A.

SUMMARY

This paper proposes a generalized estimating equations approach for the analysis of spatially correlated binary data when there are large numbers of spatially correlated observations on a moderate number of subjects. This approach is useful when the scientific focus is on modeling the marginal mean structure. Proper modeling of the spatial correlation structure is shown to provide large efficiency gains along with precise standard error estimates for inference on mean structure parameters. Generalized estimating equations for estimating the parameters of both the mean and spatial correlation structure are proposed. The use of semivariogram models for parameterizing the correlation structure is discussed, and estimation of the sample semivariogram is proposed as a technique for choosing parametric models and starting values for generalized estimating equations estimation. The methodology is illustrated with neuroimaging data collected as part of the National Institute of Neurological Disorders and Stroke (NINDS) Stroke Data Bank. A simulation study demonstrates the importance of accurate modeling of the spatial correlation structure in data with large numbers of spatially correlated observations such as those found in neuroimaging studies.

1. Introduction

Neuroimaging data are often collected as a spatial array of numbers with each array element corresponding to an observed data value at a particular brain location. In the example illustrating the methodology developed in this paper, CT (computer tomography) scans were collected on a large group of stroke patients for the purpose of examining the characteristics of stroke-induced lesions. The specific scientific interest here is to describe lesion frequency as a function of spatial location and subject-specific covariates. A grid was superimposed on every slice of the brain image, and the data were collected by recording binary variates indicating the presence or absence of a lesion in each box of the grid. Because lesions may extend beyond one box, the binary lesion indicators are highly correlated across brain regions. This correlation must be taken into account when making inferences about lesion distributions. This article proposes a generalized estimating equations (GEE) approach for modeling the effect of spatial location and subject-specific covariates on spatially correlated binary data. The interest here is on making inferences on the marginal mean structure in the presence of this spatial correlation.

A generalized estimating equations approach allows us to concentrate on the marginal mean structure, treating the spatial correlation as a nuisance. There is an extensive literature on using GEEs to model discrete data in a longitudinal data setting (Liang and Zeger, 1986; Zeger and Liang, 1986; Zeger, Liang, and Albert, 1988; Prentice, 1988); Generalized estimating equations have also

* *Current address:* Office of Biostatistics Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892, U.S.A.

‡ *Current address:* Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Maryland 20892, U.S.A.

Key words: Binary data; Generalized estimating equations; Neuroimaging data; Semivariogram estimates; Spatial correlation.

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

been applied to the analysis of discrete time series (Zeger, 1988), to the analysis toxicological data (Lefkopoulou, Moore, and Ryan, 1989), and to the analysis of periodontal disease data (Pack, Coxhead, and McDonald, 1990). The current paper considers the use of GEEs for analyzing spatially correlated binary data, where there are a large number of spatially correlated observations on a moderate number of subjects.

A particular challenge in adapting GEE methodology to the analysis of spatial data is the selection of sensible correlation structures. Some of the standardly assumed correlation structures in GEE are not appropriate for spatial data. Semivariogram estimation has proven useful for characterizing the correlation structure in spatial data sets (Cressie, 1991). Diggle (1988) discusses the importance of accurately modeling the correlation structure for efficient estimation of mean structure parameters in a repeated measures setting; we expect this to be important in our case as well, since we have many spatially correlated observations on each subject.

This paper demonstrates how semivariogram estimates can be used to choose sensible parametric forms for the variance structure that satisfy positive-definiteness conditions. These parametric forms can then be incorporated into a GEE approach for estimation. In Section 2 we discuss a model for the mean structure and discuss the use of semivariogram models for parameterizing the correlation structure. A GEE approach is then proposed for estimation in Section 3. Section 4 discusses the use of exploratory analyses for choosing a semivariogram model parameterization for GEE estimation, for choosing parameter starting values, and for assessing the fit of the GEE estimated spatial correlation structure. The methodology is illustrated with CT scan data on patients from the National Institute of Neurological Disorders and Stroke (NINDS) Stroke Data Bank (Foulkes et al., 1988) in Section 5. Section 6 presents a simulation study examining the small-sample properties and demonstrates the efficiency gains obtained by modeling the correlation structure. A discussion follows in Section 7 which includes a summary of our GEE approach and mentions alternative approaches.

2. Model

Let $Y_i(\mathbf{s})$ denote the binary response for subject i at spatial location $\mathbf{s}$, and let $\mathbf{X}_i(\mathbf{s})$ denote the corresponding vector containing spatial location information and subject-specific covariates. For conceptual simplicity, we will consider binary data recorded in two-dimensional space where each subject is observed at the same spatial locations; however, the model extends to data with higher dimensionality and to situations where spatial locations vary across subjects. The scientific interest is to relate both spatial location and subject-specific covariates to marginal response frequency, denoted by $P_i(\mathbf{s}) = P(Y_i(\mathbf{s}) = 1)$, where $i = 1, 2, 3, \ldots, K$ is an index of the K subjects, n is the number of spatial observations on each subject, and $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ are two-dimensional vectors containing the spatial locations at which observations are made. The vector of spatially correlated binary random variables for the ith subject is represented by $\mathbf{Y}_i = (Y_i(\mathbf{s}_1), Y_i(\mathbf{s}_2), Y_i(\mathbf{s}_3), \ldots, Y_i(\mathbf{s}_n))'$ with the mean vector $\mathbf{P}_i = (P_i(\mathbf{s}_1), P_i(\mathbf{s}_2), \ldots, P_i(\mathbf{s}_n))'$.

A generalized estimating equations approach for modeling the marginal mean structure is proposed. This approach is a multivariate extension of the quasi-likelihood model (Wedderburn, 1974), in the spirit of Zeger and Liang (1986), where we parameterize the marginal mean structure as well as the variance structure. For the mean model, we assume

$$h(P_i(\mathbf{s}_j)) = \mathbf{X}_i(\mathbf{s}_j)\boldsymbol{\beta} \quad \text{or} \quad P_i(\mathbf{s}_j) = h^{-1}(\mathbf{X}_i(\mathbf{s}_j)\boldsymbol{\beta}),$$

where h is a link function, and $\boldsymbol{\beta}$ is a $p \times 1$ dimensional vector of unknown parameters describing the effect of spatial location and subject-specific covariates on the mean. In this paper, we consider the commonly used logit link function ($h(P) = \text{logit}(P)$), although other link functions are possible. The covariance structure of $\mathbf{Y}_i$ denoted by $\mathbf{V}_i = \text{Var}(\mathbf{Y}_i)$ can be expressed as

$$\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}(\boldsymbol{\alpha})\mathbf{A}_i^{1/2},$$

where $\mathbf{A}_i = \text{diag}[P_i(\mathbf{s}_1)(1 - P_i(\mathbf{s}_1)), P_i(\mathbf{s}_2)(1 - P_i(\mathbf{s}_2)), \ldots, P_i(\mathbf{s}_n)(1 - P_i(\mathbf{s}_n))]$, and the spatial correlation is specified via the $n \times n$ correlation matrix $\mathbf{R}(\boldsymbol{\alpha}) = \text{Corr}(\mathbf{Y}_i)$, where $\boldsymbol{\alpha}$ is a vector of parameters characterizing the correlation structure.

The semivariogram (Matheron, 1962) is a useful construct for parameterizing the spatial correlation. Define the standardized residual for subject i at spatial location $\mathbf{s}_j$ as

$$r_i(\mathbf{s}_j) = \frac{Y_i(\mathbf{s}_j) - P_i(\mathbf{s}_j)}{\sqrt{P_i(\mathbf{s}_j)(1 - P_i(\mathbf{s}_j))}} \quad i = 1, 2, \ldots, K; j = 1, 2, \ldots, n. \tag{2.1}$$

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

A semivariogram of these residuals, which is assumed constant across subjects, is defined as

$$\gamma(\mathbf{h}_{lm}) = \gamma(\mathbf{s}_l - \mathbf{s}_m) = \text{var}(r_i(\mathbf{s}_l) - r_i(\mathbf{s}_m))/2 \quad i = 1, 2, \ldots, K;\ l, m = 1, 2, \ldots, n,$$

where $\mathbf{h}_{lm} = \mathbf{s}_l - \mathbf{s}_m$. When a semivariogram depends only on Euclidean distance (i.e., $\gamma(\mathbf{h}_{lm}) = \gamma(\|\mathbf{h}_{lm}\|)$) it is called isotropic. Semivariogram models must satisfy negative-definiteness conditions in order to guarantee valid correlation structures. Several valid parametric families of isotropic semivariogram models are given by Cressie (1991, pp. 61–63), and general techniques for assessing the validity of semivariogram models are given by Christakos (1984). If $\gamma(h, \boldsymbol{\alpha})$ is a parametric isotropic model for $\gamma(h)$, where $h = \|\mathbf{h}\|$, then for a process with unit variance, the correlation matrix $\mathbf{R}(\boldsymbol{\alpha})$ has parametric form with (l, m)-th element given by $R_{lm}(\boldsymbol{\alpha}) = 1 - \gamma(h_{lm}, \boldsymbol{\alpha})$, $l, m = 1, 2, \ldots, n$, where $h_{lm} = \|\mathbf{s}_l - \mathbf{s}_m\|$.

A typical example of an isotropic semivariogram model that exhibits positive spatial correlation is the exponential model,

$$\gamma(h; \boldsymbol{\alpha}) = \begin{cases} 0 & h = 0 \\ c + b[1 - \exp(-h/a)] & h > 0, \end{cases} \tag{2.2}$$

where $\boldsymbol{\alpha} = (c, b, a)$, $c \geq 0$, $b \geq 0$, $a \geq 0$, and $c + b \leq 2$. This semivariogram provides for a flexible correlation structure that decreases exponentially as a function of h. The parameters b and c allow for additional flexibility in the spatial correlation structure, although a commonly used one-parameter exponential semivariogram is established by setting $c = 0$ and $b = 1$. In this case, the correlation matrix $\mathbf{R}(\boldsymbol{\alpha})$ has (l, m)-th element equal to $\exp(-h_{lm}/a)$, where $h_{lm} = \|\mathbf{s}_l - \mathbf{s}_m\|$ and a is a parameter to be estimated.

3. Estimation

We employ a generalized estimating equations approach for estimating the parameters of the mean as well as those of the correlation structure. The generalized estimating equation for the mean parameters $\boldsymbol{\beta}$ are given by

$$\sum_{i=1}^{K} \mathbf{D}_i' \mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{P}_i) = 0, \tag{3.1}$$

where $\mathbf{D}_i = \partial \mathbf{P}_i / \partial \boldsymbol{\beta}$ (Zeger and Liang, 1986).

Let $\mathbf{Z}_i$ be the sample covariance matrix on the ith subject with (j, k)-th element given by

$$Z_i(j, k) = (Y_i(\mathbf{s}_j) - P_i(\mathbf{s}_j))(Y_i(\mathbf{s}_k) - P_i(\mathbf{s}_k)),$$

and let $\mathbf{z}_i$ and $\mathbf{v}_i$ be the vectors of $n(n-1)/2$ elements consisting of all entries below the diagonal of the symmetric matrices $\mathbf{Z}_i$ and $\mathbf{V}_i$, respectively. The generalized estimating equations for the correlation model parameters, $\boldsymbol{\alpha}$, are given by

$$\sum_{i=1}^{K} \mathbf{E}_i' \mathbf{W}_i^{-1}(\mathbf{z}_i - \mathbf{v}_i) = 0, \tag{3.2}$$

where $\mathbf{E}_i = \partial \mathbf{v}_i / \partial \boldsymbol{\alpha}$ and $\mathbf{W}_i$ is the working variance matrix for z_i (Prentice, 1988). As suggested by Prentice (1988), we take $\mathbf{W}_i = \text{diag}(w_{i21}, w_{i31}, w_{i41}, \ldots)$, where

$$w_{ijk} = \text{var}(Z_i(j, k)) = (1 - 2P_i(\mathbf{s}_j))(1 - 2P_i(\mathbf{s}_k))v_{ijk} + v_{ijj}v_{ikk} - v_{ijk}^2,$$

and where v_{ijk} is the (j, k)-th element in $\mathbf{V}_i$.

Starting with initial estimates of $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ ($\hat{\boldsymbol{\beta}}_0$, and $\hat{\boldsymbol{\alpha}}_0$), equations (3.1) and (3.2) can be solved by iterating between modified scoring algorithms for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

First, given $\hat{\boldsymbol{\alpha}}_m$, an updated $\hat{\boldsymbol{\beta}}_{m+1}$ is given by

$$\hat{\boldsymbol{\beta}}_{m+1} = \hat{\boldsymbol{\beta}}_m + \left(\sum_{i=1}^{K} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right)^{-1} \sum_{i=1}^{K} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1}(\mathbf{Y}_i - \hat{\mathbf{P}}_i), \tag{3.3}$$

where $\hat{\mathbf{D}}_i = \mathbf{D}_i(\hat{\boldsymbol{\beta}}_m)$, $\hat{\mathbf{V}}_i = \mathbf{V}_i(\hat{\boldsymbol{\beta}}_m, \hat{\boldsymbol{\alpha}}_m)$, and $\hat{\mathbf{P}}_i = \mathbf{P}_i(\hat{\boldsymbol{\beta}}_m)$. Second, given $\hat{\boldsymbol{\beta}}_{m+1}$, an updated $\hat{\boldsymbol{\alpha}}_{m+1}$ is obtained by

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

$$\hat{\boldsymbol{\alpha}}_{m+1} = \hat{\boldsymbol{\alpha}}_m + \left(\sum_{i=1}^{K} \hat{\mathbf{E}}_i' \hat{\mathbf{W}}_i^{-1} \hat{\mathbf{E}}_i \right)^{-1} \sum_{i=1}^{K} \hat{\mathbf{E}}_i' \hat{\mathbf{W}}_i^{-1} (\hat{\mathbf{z}}_i - \hat{\mathbf{v}}_i), \tag{3.4}$$

where $\hat{\mathbf{E}}_i = \mathbf{E}_i(\hat{\boldsymbol{\beta}}_{m+1}, \hat{\boldsymbol{\alpha}}_m)$, $\hat{\mathbf{W}}_i = \mathbf{W}_i(\hat{\boldsymbol{\beta}}_{m+1}, \hat{\boldsymbol{\alpha}}_m)$, $\hat{\mathbf{z}}_i = \mathbf{z}_i(\hat{\boldsymbol{\beta}}_{m+1})$, and $\hat{\mathbf{v}}_i = \mathbf{v}_i(\hat{\boldsymbol{\beta}}_{m+1}, \hat{\boldsymbol{\alpha}}_m)$. Estimates of $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\alpha}}$ are obtained by iterating between equations (3.3) and (3.4) until convergence ($m = 1, 2, 3, \ldots$).

The asymptotic properties of the parameter estimators for the mean and correlation structure follow the work of Liang and Zeger (1986), Zeger and Liang (1986), and Prentice (1988). The estimator $\hat{\boldsymbol{\beta}}$ is consistent (fixed n, $K \to \infty$) even for misspecified spatial correlation structure (semivariogram model). The estimator $\hat{\boldsymbol{\alpha}}$ is consistent under a correctly specified mean and semivariogram model. If the semivariogram model is correctly specified, a model-based consistent estimator of the variance of $\hat{\boldsymbol{\beta}}$ is

$$\widehat{\text{var}}_{\text{mod}}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{K} \widehat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \widehat{\mathbf{D}}_i \right)^{-1}. \tag{3.5}$$

As an alternative, the robust estimator of variance,

$$\widehat{\text{var}}_{\text{rob}}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{K} \widehat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \widehat{\mathbf{D}}_i \right)^{-1} \left(\sum_{i=1}^{K} \widehat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{Z}_i \hat{\mathbf{V}}_i^{-1} \widehat{\mathbf{D}}_i \right) \left(\sum_{i=1}^{K} \widehat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \widehat{\mathbf{D}}_i \right)^{-1}, \tag{3.6}$$

is consistent even when the semivariogram model is misspecified. Similarly, estimators of var($\hat{\boldsymbol{\alpha}}$) can be obtained as described in Prentice (1988); these may be useful in testing for anisotropy in the spatial correlation structure when scientifically reasonable classes of anisotropic semivariogram models (Journel and Huijbregts, 1978, p. 179) can be proposed.

Starting values for the mean structure parameters ($\hat{\boldsymbol{\beta}}_0$) can be obtained by assuming spatial independence and fitting a generalized linear model with a logit link function (logistic regression). The logistic regression model can be fit using GLIM (McCullagh and Nelder, 1989), and it is identical to GEE estimation under spatial independence. These estimates are consistent and have asymptotic variance

$$\text{var}_{\text{indep}}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{K} \mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \right)^{-1} \left(\sum_{i=1}^{K} \mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{V}_i \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \right) \left(\sum_{i=1}^{K} \mathbf{D}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{D}_i \right)^{-1}, \tag{3.7}$$

where $\boldsymbol{\Sigma}_i = \mathbf{A}_i$, and $\mathbf{V}_i$ is the true spatial covariance structure.

The above GEE methodology involves only first and second order moment assumptions about the marginal distribution of $Y_i(\mathbf{s}_j)$, and the specification of a working spatial correlation structure $\mathbf{R}(\boldsymbol{\alpha})$. Extensions of GEE that involve a joint estimating equation for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, require assumptions about higher order moments, and result in efficiency gains (small for $\boldsymbol{\beta}$ and more substantial for $\boldsymbol{\alpha}$) have been proposed (Zhao and Prentice, 1990; Liang, Zeger, and Qaqish, 1992). The price paid for the small efficiency gains is that $\hat{\boldsymbol{\beta}}$ is only consistent if the correlation structure has been correctly specified. When, as in our case, the correlation structure is considered a nuisance, it is preferable to forego small efficiency gains for guaranteed consistency.

In the next section, we discuss the use of semivariogram estimation techniques for choosing an appropriate class of parametric semivariogram models for inclusion in the correlation structure of the GEE model. These techniques allow for checking the assumption of isotropy and provide starting values for the correlation parameter estimates.

4. Exploratory Analyses for Modeling the Spatial Correlation Structure

Define $\widehat{r_i(\mathbf{s}_j)}$ as the estimated standardized residual for the ith subject at the jth spatial location from a fitted logistic regression model assuming no spatial correlation; these are computed using (2.1) but substituting $\hat{\boldsymbol{\beta}}_0$ for $\boldsymbol{\beta}$ and rescaling so they have unit variance. Since $\hat{\boldsymbol{\beta}}_0$ computed under the independence assumption results in consistent estimation of the mean structure, $\widehat{r_i(\mathbf{s}_j)}$ asymptotically has zero mean. Dividing by the square root of the factor $(nK - 1)^{-1} \sum_{i=1}^{K} \sum_{j=1}^{n} (r_i(\mathbf{s}_j)^* - \bar{r})^2$, where $r_i(\mathbf{s}_j)^*$ is defined by substituting $\hat{\boldsymbol{\beta}}_0$ for $\boldsymbol{\beta}$ in (2.1) and $\bar{r}$ is the mean of the values of $r_i(\mathbf{s}_j)^*$, rescales the residuals by their sample standard deviation, and a sample semivariogram on these residuals is a useful technique for choosing models to parameterize the correlation structure. It has an advantage over the sample autocorrelation function in that it can be estimated more stably when the spatial

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

locations of observations vary across subjects and are unequally spaced (as discussed by Diggle (1988) in a repeated measures setting), and is less sensitive to misspecification of the mean structure model. We used method of moments techniques to estimate the common semivariogram across subjects by

$$\widehat{\gamma(\mathbf{h})} = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (\widehat{r_i(\mathbf{s}_l)} - \widehat{r_i(\mathbf{s}_m)})^2, \tag{4.1}$$

where,

$$N(\mathbf{h}) = \{(i, \mathbf{s}_l, \mathbf{s}_m): \mathbf{s}_l - \mathbf{s}_m = \mathbf{h};\ l, m = 1, 2, \ldots, n;\ i = 1, 2, \ldots, K\},$$

and $|N(\mathbf{h})|$ is the cardinality of $N(\mathbf{h})$. For only one subject this expression is identical to the method of moments estimator proposed by Matheron (1962). As recommended by Journel and Huijbregts (1978, p. 194), estimation of the semivariogram is usually only attempted for distances not exceeding half the maximum possible distance. Typically $\widehat{\gamma(\mathbf{h})}$ is computed in several different directions to examine for anisotropy. Similarity of the directional semivariograms provides evidence for isotropy. If the assumption of isotropy appears reasonable, we estimate the common isotropic semivariogram across subjects by

$$\widehat{\gamma(h)} = \frac{1}{2|N(h)|} \sum_{N(h)} (\widehat{r_i(\mathbf{s}_l)} - \widehat{r_i(\mathbf{s}_m)})^2, \tag{4.2}$$

where,

$$N(h) = \{(i, \mathbf{s}_l, \mathbf{s}_m): \|\mathbf{s}_l - \mathbf{s}_m\| = h;\ l, m = 1, 2, \ldots, n;\ i = 1, 2, \ldots, K\},$$

and $h = \|\mathbf{h}\|$. Within a class of isotropic semivariogram models, starting values for GEE estimation ($\hat{\boldsymbol{\alpha}}_0$) can be obtained by approximate weighted least squares (Cressie, 1991, pp. 95–97) fitted to (4.2).

After fitting a GEE model, the semivariogram model can be reassessed by computing sample semivariogram estimates (expressions (4.1) and (4.2)) using standardized residuals estimated (re-scaled so they have unit variance) from the GEE fit rather than under the independence model. These sample semivariograms can then be compared to the GEE-fitted parametric semivariogram model.

5. Example

Data from a subset of patients collected as part of the NINDS Stroke Data Bank (Foulkes et al., 1988) are used to illustrate the methodology presented in this paper. The subset consisted of 193 patients with complete CT scan data, and a single lesion corresponding to a first ever cerebral infarction. The details of the selection of this subset are provided in Morris et al. (unpublished manuscript). For every subject, the CT data consisted of nine two-dimensional binary arrays, each corresponding to one of nine transcranial brain CT slices. Each array was constructed by superimposing a grid over the corresponding CT slice, and recording a 1 when more than 50% of the grid box contained lesion, and recording a 0 otherwise. Our goal was to model lesion frequency (the probability on the ith subject of a lesion occupying a particular grid box) as a function of spatial location and the subject-specific covariates age and sex. Due to technical difficulties in aligning slices, imaging data is frequently analyzed on a slice-by-slice basis. Therefore, to illustrate the methods in this paper, we focused on a single transcranial slice consisting of a mix of subcortical and cortical regions, whose corresponding grid was represented by an 88-element binary array.

Since positive correlation is induced from lesions intersecting the CT slice at multiple grid boxes, the GEE approach will be useful in making inference about the marginal mean structure in the presence of this spatial correlation. Figure 1 shows the sum (over the 193 patients) of the binary arrays, where each sum is associated with a set of spatial coordinates, $\mathbf{s}_j = (s_{1j}, s_{2j})$, and where s_{1j} and s_{2j} index location on the anterior-posterior axis and on the lateral (left-right) axis, respectively; row and column means are also presented. These sums are a measure of lesion frequency and aid in the formulation of a mean structure model. Lesion frequency appears to increase as you move outwards from the midline with the rate of increase possibly differing in the two hemispheres. The frequency increases and then decreases moving from the most anterior to the most posterior regions. Our interest is in assessing the statistical significance of these effects and the effects of subject-specific covariates. A model for the mean structure which incorporates the above features is

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

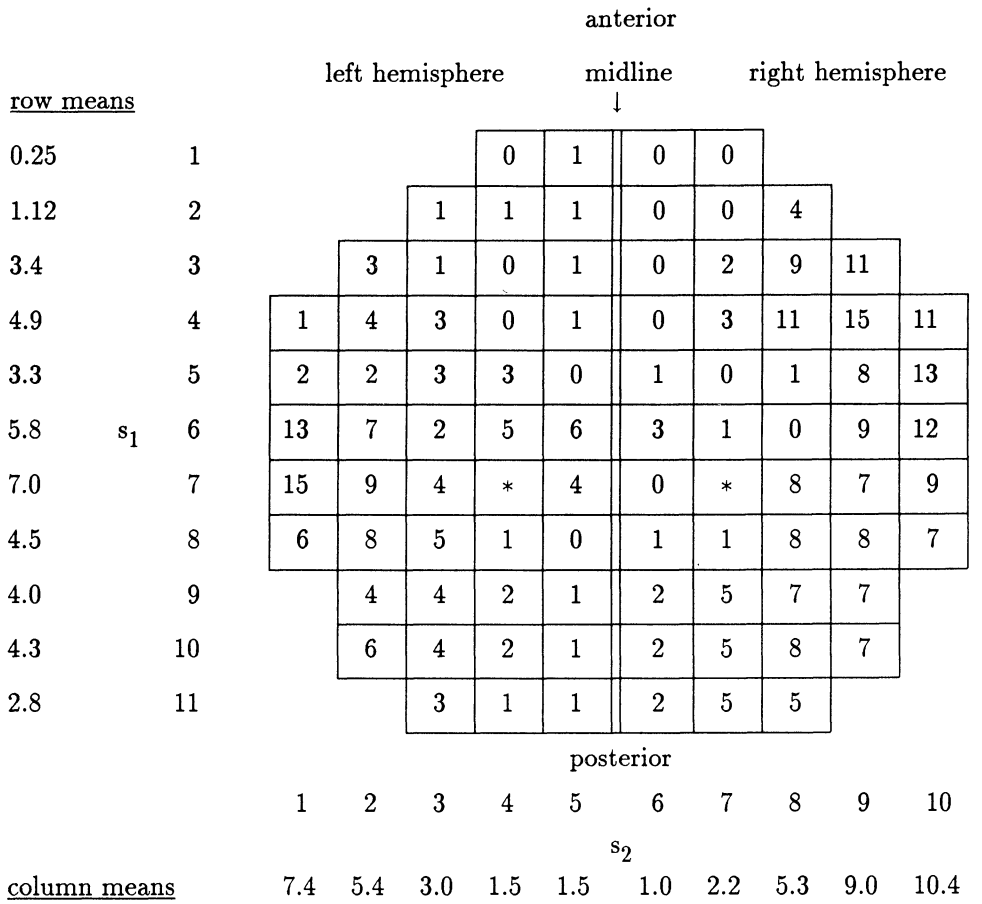

Figure 1. Lesion Frequency by Spatial Location. Frequency in each box is the number of subjects (out of 193) whose CT scan showed a lesion overlapping at least half of the box. s_1, position along the anterior-posterior axis; s_2, position along the lateral (left-right) axis; *, boxes that are at least half occupied by ventricles and are not included in the analysis.

$$\text{logit}P_i(\mathbf{s}_j) = \beta_0 + \beta_1 s_{1j} + \beta_2 s_{1j}^2 + \beta_3 I_{(s_{2j} \leq 5)}(5 - s_{2j}) + \beta_4 I_{(s_{2j} > 5)} + \beta_5 I_{(s_{2j} > 5)}(s_{2j} - 6) + \beta_6 \text{Age}_i + \beta_7 \text{Sex}_i,$$

where $I_{(s_{2j} > 5)} = 1$ if $s_{2j} > 5$ (right hemisphere) or 0 (left hemisphere) otherwise, where $\text{Sex}_i = 1$ if the ith subject is male or 0 if female, and where Age_i is recorded in years. This model allows us to examine for lateral changes in the two hemispheres (β_3, β_4, and β_5), changes in the anterior-posterior direction (β_1 and β_2), and subject-specific covariates (β_6 and β_7). The model is parameterized to reflect the anatomical symmetry of the brain about the midline. The parameters β_3 and β_5 represent the slopes moving from the midline outward in either the left or right directions, respectively. Symmetry of lesion frequency would be indicated if $\beta_3 = \beta_5$ and $\beta_4 = 0$.

Initial estimates of $\boldsymbol{\beta}$ for the above mean structure were obtained by logistic regression and were used to compute sample semivariograms on the estimated standardized residuals. These semivariograms were computed using equations (4.1) and (4.2) on all pairs of spatial coordinates up to a distance of seven units (approximately one-half the maximum possible distance). Directional sample semivariograms (Figure 2A) provide strong evidence for isotropy, since they are nearly identical. An isotropic sample semivariogram (shown in Figure 2B) suggests the class of exponential semivariograms for GEE estimation, corresponding to an exponentially decaying spatial correlation. Using an approximate weighted least-squares criterion, we fit a one-parameter exponential model (setting $c = 0$ and $b = 1$ in expression (2.2)) to the sample isotropic semivariogram, yielding a parameter estimate of $\hat{a} = 1.11$. Fitted values obtained from this model are close to the sample estimates (Figure 2B) suggesting that the class of one parameter exponential semivariogram models provides a good representation of the spatial correlation structure for this data. In cases where there is only one lesion and the marginal lesion frequencies are neither very low nor very high, correlations may

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

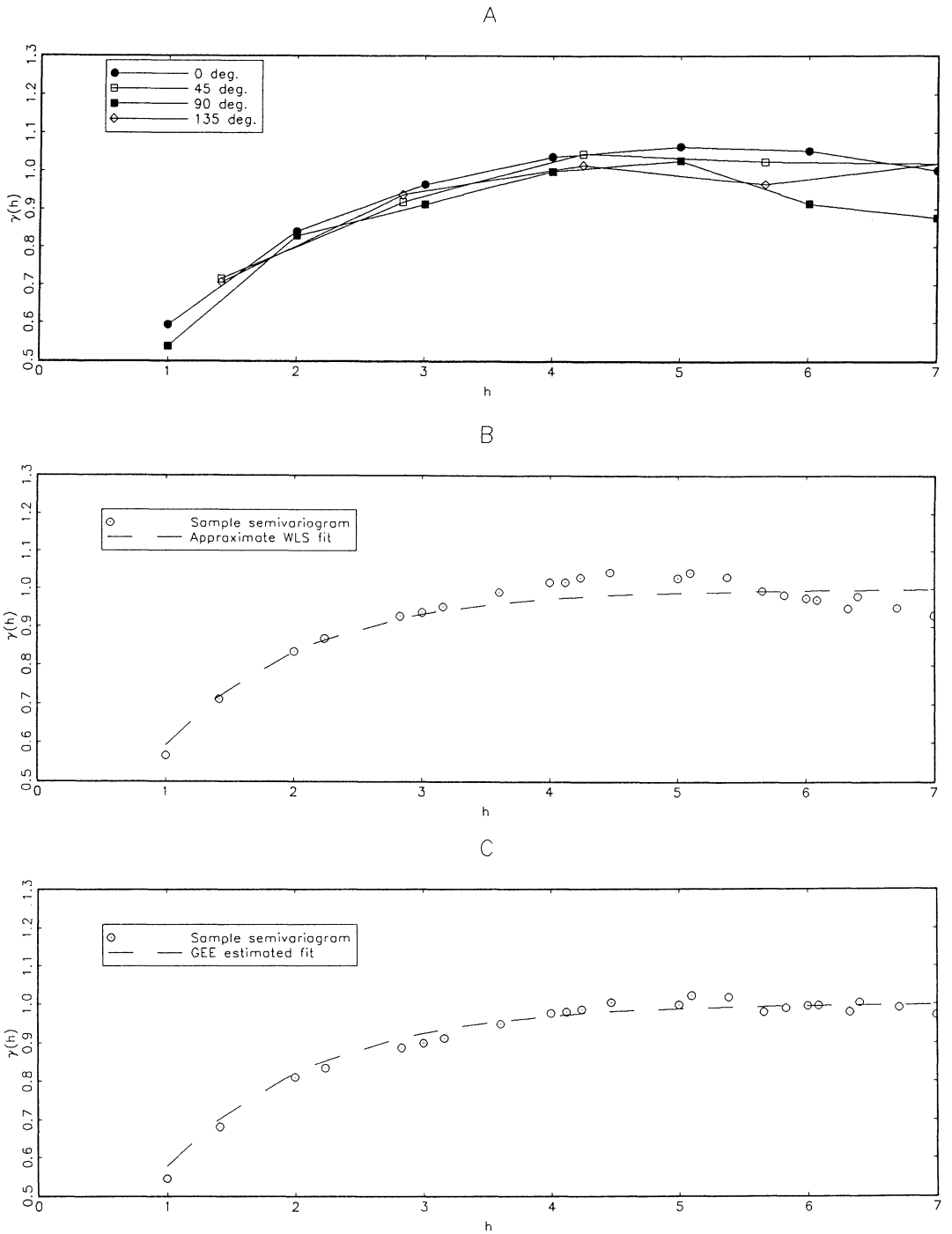

Figure 2. A, sample directional semivariograms using standardized residuals estimated from a logistic regression model assuming spatial independence. B, sample isotropic semivariogram and approximate weighted least-squares fit using standardized residuals estimated from a logistic regression model assuming spatial independence. C, sample isotropic semivariogram and GEE estimated fit using standardized residuals estimated from the GEE-exponential model.

become substantially negative at moderate distances. In these situations, semivariogram models that exhibit negative correlations can be used. For our case, the marginal lesion frequencies were very low; hence, spatial correlations were bounded below very close to 0 due to inherent constraints for binary random variables (Prentice, 1988).

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

Table 1 shows parameter estimates from GEE estimation assuming (a) independence and (b) an exponential semivariogram model; we will refer to these models as GEE-independence and GEE-exponential, respectively. For both models, model-based and robust estimators of the standard errors (denoted by SE_{mod} and SE_{rob}, respectively) are computed as the square roots of expressions (3.5) and (3.6). In addition, estimators of the standard errors for the GEE-independence parameter estimates are computed assuming that the true mean structure and spatial correlations are described by the GEE-exponential model; these are denoted by SE_{indep} and computed as the square roots of estimated expression (3.7). The GEE-independence model falsely assumes spatial independence, but results in consistent estimation of the mean structure parameters. Note that the parameter estimates, $\hat{\boldsymbol{\beta}}$, for the GEE-independence model are similar to those computed for the GEE-exponential model. As a diagnostic tool for assessing the fit of the exponential semivariogram model, we recomputed the sample semivariogram using the estimated standardized residuals (rescaled so they have unit variance) computed with GEE-exponential estimated parameters instead of with GEE-independence mean parameters. It can be seen from Figure 2C that the GEE estimated parametric semivariogram provides an excellent fit to the updated sample semivariogram estimate, indicating that the one-parameter exponential semivariogram is a good model for the spatial correlation structure.

Table 1
GEE estimation of stroke lesion frequency

Model: $\text{logit}P_i(\mathbf{s}_j) = \beta_0 + \beta_1 s_{1j} + \beta_2 s_{1j}^2 + \beta_3 I_{(s_{2j}\leq 5)}(5 - s_{2j}) + \beta_4 I_{(s_{2j}>5)}$
$+ \beta_5 I_{(s_{2j}>5)}(s_{2j} - 6) + \beta_6 \text{Age}_i + \beta_7 \text{Sex}_i$

$\text{Var}(Y_i(\mathbf{s}_j)) = P_i(\mathbf{s}_j)(1 - P_i(\mathbf{s}_j))$

Coefficient	Estimate	SE_{mod}[a]	SE_{rob}[b]	SE_{indep}[c]
a) Spatial independence				
β_0	−5.76	.469	.823	.937
β_1	.340	.125	.191	.217
β_2	−.0210	.00945	.0147	.0162
β_3	.403	.0697	.104	.106
β_4	−.223	.245	.411	.347
β_5	.507	.0585	.0736	.0993
β_6	−.00872	.00398	.00864	.00966
β_7	.298	.111	.281	.269
b) Exponential semivariogram model[d]				
β_0	−5.47	.857	.773	
β_1	.416	.199	.170	
β_2	−.0282	.0149	.0134	
β_3	.266	.0982	.0776	
β_4	−.126	.297	.320	
β_5	.451	.0925	.0653	
β_6	.00906	.00887	.00902	
β_7	.267	.247	.292	

[a] Computed using estimator (3.5).
[b] Computed using estimator (3.6).
[c] Computed using estimator (3.7) and assuming the GEE-exponential is the true model.
[d] Exponential semivariogram parameter estimated to be $\hat{a} = 1.161$.

Using the GEE-exponential model for inference (with either the robust or model-based standard error estimates), we find an anterior-posterior effect, a lateral effect in both the left and right direction with a suggestion of asymmetry ($\hat{\beta}_5 - \hat{\beta}_3 = .185$, $SE_{mod} = .100$), and no significant effects of age and sex on lesion frequency. Note that if we ignore the spatial correlation and incorrectly use the model-based standard errors computed under GEE-independence, we would conclude that both age and sex significantly affect lesion frequency.

Additional terms were added to our model in order to examine for an interaction between s_{1j} and s_{2j}. Addition of the terms $s_{1j}I_{(s_{2j}>5)}$ and $s_{1j}^2 I_{(s_{2j}>5)}$ allowed for different quadratic functions of s_{1j} in the two hemispheres, and the terms $s_{1j}I_{(s_{2j}\leq 5)}(5 - s_{2j})$ and $s_{1j}I_{(s_{2j}>5)}(s_{2j} - 6)$ allowed for lateral changes in the two hemispheres to depend on anterior-posterior location (s_{1j}). Of these terms, only

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

the coefficient of $s_{1j}I_{(s_{2j}>5)}(s_{2j} - 6)$ was marginally significant ($Z = -2.0$) using model-based estimates of standard error and not adjusting for multiple comparisons. This interaction suggests that lateral changes in right hemisphere lesion frequency were greatest in the more anterior regions and decreased as one moves to more posterior regions. Because this effect was only marginally significant, and for the sake of a more parsimonious model, our efficiency calculations and simulation study are based on the original model.

The robust standard errors for the mean parameters involving spatial effects are generally larger for the GEE-independence model than for GEE-exponential; in fact, anterior-posterior effects are nonsignificant when using GEE-independence parameter estimates and corresponding robust estimates of the standard errors for inference. Furthermore, GEE-exponential estimation provides an approximate 21% gain in asymptotic relative efficiency over the GEE-independence estimators (obtained by comparing SE_{indep} with SE_{mod} under the GEE-exponential model in Table 1), assuming that the true model for the mean and correlation structure is the estimated GEE-exponential model. Although robust estimation of the standard errors protects against the misspecification of the spatial correlation structure, the robust estimators rely on the sample covariance matrix, and therefore, may have poor finite sample properties. The finite sample properties of the spatial GEE approach are investigated in the next section.

6. Finite Sample Properties

We investigated the finite sample properties of GEE by a small simulation study. Using an algorithm proposed by Emrich and Piedmonte (1991), we generated 100 samples ($n = 88$, $K = 193$) of correlated binary data with a marginal mean structure and exponential semivariogram model as estimated in our example. From each simulated realization, we computed GEE-independence and GEE-exponential estimates of $\boldsymbol{\beta}$ as well as model-based and robust estimates of the standard error of $\hat{\boldsymbol{\beta}}$. For both models, nominal 90% confidence intervals of the form $\hat{\beta}_l \pm 1.645\ SE(\hat{\beta}_l)$ were computed for each mean structure parameter ($l = 0, 1, 2, \ldots, 7$), using both model-based and robust estimates of the standard errors. The simulation results are summarized for GEE-independence and GEE-exponential estimation in Table 2. For each model, Table 2 lists the mean estimated parameter values, the mean model-based and robust estimates of standard errors, Monte Carlo estimates of standard errors, and the estimated coverage probabilities of the nominal 90% confidence intervals (i.e., the proportion of the 100 intervals containing the true parameter). Estimated standard deviations of the model-based and robust estimates of the standard errors are also presented for both models.

The simulations suggest there is little small-sample bias in the mean structure parameter estimates under either GEE-independence or GEE-exponential estimation. However, model-based estimates of the standard errors computed under GEE-independence estimation severely underestimate the true standard errors (as estimated by Monte Carlo), and result in confidence interval coverage probabilities that are much too small. That is, tests of the mean parameters using the independence model-based standard errors would too frequently show significant results. A finite sample efficiency comparison (comparing Monte Carlo estimates of standard error between GEE-independence and GEE-exponential) show an average efficiency gain of 30% for exploiting our knowledge about the spatial correlation structure.

GEE estimation under the correctly specified exponential correlation structure (Table 2b) resulted in model-based estimates of the standard error with little bias, and estimated coverage probabilities close to the intended 90% (within the uncertainty of Monte Carlo), suggesting that asymptotic inferences on the mean structure parameters using these estimates are very reasonable. There was little bias in the robust estimators of the standard errors under either GEE-independence or GEE-exponential estimation. However, the precision of the various standard error estimates differed. Comparing the estimated standard deviations of the standard error estimates given in the parentheses in Table 2, we find that GEE-exponential model-based estimators had an average efficiency gain of 106% over GEE-exponential robust estimators, which in turn, had an average efficiency gain of 41% over GEE-independence robust estimation. Estimated coverage probabilities obtained with robust standard errors tended to be lower than the intended 90%, suggesting that inferences on the mean structure parameters with robust standard errors tend to be anticonservative. These simulation results highlight the importance of correctly specifying the spatial correlation structure in data with large numbers of spatially correlated observations.

7. Discussion

This paper proposes a generalized estimating equations approach for the analysis of spatially correlated binary data. This approach is useful when the scientific interest is on modeling the

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

Table 2
Simulation results using GEE-independence and GEE-exponential estimation on correlated binary data generated from the exponential semivariogram model

True coefficient		Average estimated value	Average estimated SE_{mod}[a]	Average estimated SE_{rob}[b]	Monte Carlo SE	Coverage probabilities (in %) for nominal 90% intervals,[c] using SE_{mod};	using SE_{rob}
a) GEE-independence							
β_0	−5.47	−5.69	.47 (.048)[d]	.91 (.17)	.98	54	87
β_1	.416	.464	.13 (.013)	.21 (.031)	.25	61	84
β_2	−.0282	−.0320	.0096 (.00093)	.016 (.0019)	.018	62	83
β_3	.266	.255	.066 (.0070)	.11 (.014)	.12	60	87
β_4	−.126	−.161	.23 (.024)	.34 (.052)	.35	77	90
β_5	.451	.456	.060 (.0053)	.097 (.014)	.10	64	89
β_6	−.00906	−.00834	.0041 (.00028)	.0091 (.0017)	.0097	43	91
β_7	.267	.311	.11 (.0075)	.27 (.027)	.27	50	88
b) GEE-exponential							
β_0	−5.47	−5.70	.87 (.086)[d]	.85 (.13)	.86	90	88
β_1	.416	.457	.20 (.019)	.20 (.025)	.21	89	85
β_2	−.0282	−.0315	.015 (.0013)	.015 (.0017)	.016	88	83
β_3	.266	.257	.099 (.0095)	.099 (.013)	.10	91	89
β_4	−.126	−.148	.30 (.037)	.30 (.055)	.29	90	89
β_5	.451	.453	.093 (.0070)	.091 (.012)	.097	89	87
β_6	−.00906	−.00789	.0088 (.00079)	.0086 (.0013)	.0086	94	91
β_7	.267	.319	.25 (.022)	.25 (.021)	.24	88	87

[a] Computed using estimator (3.5).
[b] Computed using estimator (3.6).
[c] Computed as $\hat{\beta}_l \pm 1.645\ \mathrm{SE}(\hat{\beta}_l)$.
[d] Numbers in parentheses are estimated standard deviations of estimators.

marginal mean structure and the variance is considered a nuisance, but the results presented suggest that accurate modeling of the spatial correlation structure may be important for efficient estimation of mean structure parameters when there are large numbers of correlated observations per subject.

There are alternative approaches for the analysis of spatially correlated binary data. Without subject-specific covariates and when binary observations are recorded at the same spatial locations across individuals, the data can be viewed as a spatial data set of counts, rates, or proportions. In this case, one could apply techniques such as empirical Bayes methods using spatial autoregressive priors (Clayton and Kaldor, 1987; Mollie and Richardson, 1991) or more general spatial priors (Cressie, 1992). Alternatively, one may model the data (or transformed data) using simultaneous or conditional Gaussian spatial autoregressive models, or Markov random fields (Besag, 1974; Cliff and Ord, 1981, pp. 145–149; Cressie and Chan, 1989). None of these approaches allowed for the incorporation of subject-specific covariates, and would require common observation locations

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

across subjects. Approaches for modeling the gridded binary data directly based on the autologistic model (Besag, 1974; Ising, 1925) yield mean structures with conditional rather than marginal interpretations.

The results of our simulation suggest that for the low marginal probability of a lesion, large numbers of spatial locations, and moderate number of patients in our example, choosing a good model for the spatial correlation was particularly important. First, we found large efficiency gains in estimating $\boldsymbol{\beta}$ using GEE with a correct spatial correlation structure as compared with GEE-independence. Second, by correctly specifying the spatial correlation structure we gain efficiency in estimating the standard errors of $\hat{\boldsymbol{\beta}}$ and reduce bias in the model-based standard error estimators. This results in more accurate inferences on $\boldsymbol{\beta}$.

To aid in more accurately characterizing the correlation structure, sample semivariogram estimation is proposed as a tool for choosing an appropriate parametric form for the correlation function and checking for isotropy. Following GEE estimation, the sample semivariogram can be recomputed from the updated estimated residuals as a diagnostic for the fit of the final correlation model. In our application, we had clear indication that the spatial correlation structure was very nicely described by a one-parameter, isotropic, exponential model; for other applications more complex anisotropic models may be required.

The issue of whether to use the model-based or robust estimates of the standard errors of $\hat{\boldsymbol{\beta}}$ for inference comes down to a trade-off between the lack of fit of the spatial correlation model (bias) and the precision of the standard error estimators (variance). Since in our example, the isotropic exponential model fit well, we would recommend using model-based estimates for inference in our case. In cases where the semivariogram is borderline isotropic, and no anisotropic model can be found that describes the data well, we suggest using the isotropic exponential model with the robust estimator of variance for inference. In this case, the best fitting isotropic model is closer to the true correlation structure than an independence model, and its use will buy a sizable efficiency gain as compared with making inference using the independence model with the robust estimator of variance.

The particular example of neuroimaging data described in this paper consisted of binary data collected on a moderate number of boxes. Advanced neuroimaging techniques are becoming less expensive and more popular as outcome measures in medical studies. The spatial GEE methods presented here can be extended to situations in which these outcome measures may be continuous rather than binary, may be collected over a finer grid consisting of a very large number of pixels, and the images may be three-dimensional rather than two-dimensional.

Résumé

Cet article propose l'approche des équations d'estimation généralisées pour analyser des données binaires présentant une corrélation spatiale quand il existe un grand nombre d'observations corrélées spatialement et un nombre modéré d'individus. Cette approche est utile quand l'intérêt scientifique porte sur la modélisation de la structure de l'espérance marginale. Il est montré qu'une modélisation juste de la structure de la corrélation spatiale produit des gains d'efficience importants en particulier sur la précision des estimations des écarts-type utilisés pour l'inférence des paramètres de la structure de l'espérance. Des équations d'estimation généralisées pour estimer les paramètres à la fois de l'espérance et de la structure de la corrélation spatiale sont proposées. L'utilisation de modèles de semivariogramme pour paramétrer la structure de corrélation est discutée, et l'estimation du semivariogramme empirique est proposé comme une technique pour choisir les modèles paramétriques et les valeurs de départ pour l'estimation des GEE. La méthodologie est illustrée avec des données de neuroimagerie collectées dans la banque de données "NINDS Stroke Data Bank". Une étude de simulation démontre l'importance d'une modélisation précise de la structure de corrélation spatiale quand les données sont composées d'un grand nombre d'observations corrélées spatialement comme on les trouve dans les études de neuroimagerie.

References

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36,** 192–225.

Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research* **20,** 251–265.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43,** 671–681.

Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*. London: Pion.

Cressie, N. (1991). *Statistics for Spatial Data*. New York: Wiley.

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions

Cressie, N. (1992). Smoothing regional maps using empirical Bayes predictors. *Geographical Analysis* **24,** 75–95.
Cressie, N. and Chan, N. H. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association* **84,** 393–401.
Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44,** 959–971.
Emrich, L. J. and Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *The American Statistician* **45,** 302–304.
Foulkes, M. A., Wolf, P. A., Price, T. R., Mohr, J. P., and Hier, D. B. (1988). The stroke data bank: Design, methods and baseline characteristics. *Stroke* **19,** 547–554.
Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift fur Physik* **31,** 253–258.
Journel, A. G. and Huijbregts, C. T. (1978). *Mining Geostatistics*. London: Academic Press.
Lefkopoulou, M., Moore, D., and Ryan, L. (1989). The analysis of multiple correlated binary outcomes: Application to rodent teratology experiments. *Journal of the American Statistical Association* **84,** 810–815.
Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.
Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analysis for categorical data (with discussion). *Journal of the Royal Statistical Society, Series B* **54,** 3–40.
Matheron, G. (1962). *Traite de Geostatistique Appliquee, Tome I. Memoires du Bareau de Recherches Geologiques et Minieres 14*. Paris: Editions Technip.
McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
Mollie, A. and Richardson, S. (1991). Empirical Bayes estimates of cancer mortality rates using spatial models. *Statistics in Medicine* **10,** 95–112.
Pack, A. R. C., Coxhead, L. J., and McDonald, B. W. (1990). The prevalence of overhanging margins in posterior amalgam restorations and periodontal consequences. *Journal of Clinical Periodontology* **17,** 145–152.
Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44,** 1033–1048.
Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61,** 439–447.
Zeger, S. L. (1988). A regression model for time series count data. *Biometrika* **75,** 621–629.
Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.
Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equations approach. *Biometrics* **44,** 1049–1060.
Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77,** 642–648.

Received August 1993; *revised January and June* 1994; *accepted June* 1994.

This content downloaded from 72.33.153.245 on Wed, 1 Oct 2014 12:49:22 PM
All use subject to JSTOR Terms and Conditions