Statistics
in Medicine

Research Article

Received 1 October 2012, Accepted 20 August 2013 Published online 13 September 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5967

Hierarchical multiple informants models: examining food environment contributions to the childhood obesity epidemic‡

Jonggyu Baek,[a*†] Brisa N. Sánchez[a] and Emma V. Sanchez-Vaznaugh[b]

Methods for multiple informants help to estimate the marginal effect of each multiple source predictor and formally compare the strength of their association with an outcome. We extend multiple informant methods to the case of hierarchical data structures to account for within cluster correlation. We apply the proposed method to examine the relationship between features of the food environment near schools and children's body mass index z-scores (BMIz). Specifically, we compare the associations between two different features of the food environment (fast food restaurants and convenience stores) with BMIz and investigate how the association between the number of fast food restaurants or convenience stores and child's BMIz varies across distance from a school. The newly developed methodology enhances the types of research questions that can be asked by investigators studying effects of environment on childhood obesity and can be applied to other fields. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: generalized estimating equations; multiple informants; hierarchical data structure

1. Introduction

The childhood obesity epidemic has led several researchers to examine factors beyond the individual as possible causes of obesity. For instance, because children spend large amounts of time in schools, there is increased interest in environmental factors in or around schools. The presence of food stores such as fast food restaurants (FFR) or convenience stores (CS) has received attention as children may purchase, or be exposed to advertising of energy-dense, nutrient-poor foods on their way to or from school. These features of the environment near schools are typically operationalized as the number of food stores within a specific distance from a school (e.g., number of stores falling within a circle of $1/2$ mile radius around a school, also known as $1/2$ mile 'buffer'). The associations between each feature and children's body weight are examined in separate models because the marginal association between each feature of the environment and children's body weight is of substantive interest or because features are strongly correlated making it difficult to include them simultaneously in one model. Comparing the strength of associations between one food store type and another is of particular interest (e.g., CS vs. FFR) because limiting certain types of food stores versus others may need to be considered from a policy perspective. An overarching limitation of the methods presently employed in these prior studies is that they do not rigorously examine or test differences among associations between environment features and outcomes. Furthermore, the specific distance from a school at which the circle is drawn (also known as buffer size) is typically chosen in an ad hoc manner, because the distance from a school at which the presence of food stores may influence children's body weight is unknown. Some researchers have

[a]*Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.*
[b]*Department of Health Education, San Francisco State University, San Francisco, CA, U.S.A.*
**Correspondence to: Jonggyu Baek, Department of Biostatistics, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, U.S.A.*
†*E-mail: jongguri@umich.edu*
‡*Supporting information may be found in the online version of this article.*

compared associations of the outcome with a feature at different buffer sizes by examining the extent of overlap of CIs of measures of association obtained from several buffer sizes or examined the distance at which the association is, or ceases to be, significant [1]. However, comparing the extent of overlap of CIs is problematic because the estimates are correlated. The purpose of the present research is to develop a hierarchical multiple informant model (HMIM) that will facilitate comparing differences in the associations of the same type of food store at several buffer sizes from a school and/or differences in the associations of two or more types of food stores.

Methods for multiple informant data were independently proposed by Horton *et al.* [2] and Pepe *et al.* [3] and have been comprehensively reviewed by Horton and Fitzmaurice [4]. The term 'multiple informants' refers to information from multiple sources used to measure the same construct. Horton *et al.* [2] give an example of multiple informants, such as information collected from a child's teacher and parent to assess the child's psychopathology. In our setting, the multiple informant predictors are features of the environment (e.g., multiple store types or number of a given store type at several buffer sizes) that may affect children's weight.

Models for multiple informants can be constructed using non-standard generalized estimating equation (GEE) methods to estimate the marginal association between each multiple source predictor and an outcome, and provide a formal comparison of the strength of the associations between each predictor with the outcome [2, 3]. Alternatively, Litman *et al.* [5] developed a maximum likelihood estimation (MLE) approach that, under a joint normality of predictors and an outcome, can accommodate more general models than can be estimated with a GEE method. The MLE approach can incorporate multiple informants measured in different scales and enable estimation of a common 'standardized' association (e.g., adjusted correlation coefficient), and incorporate data missing at random. However, existing multiple informant methods are limited to non-hierarchical data where univariate outcomes are measured on independent subjects. Although Horton and Fitzmaurice [4] stressed the importance of complex survey designs, the estimating equations they employed assume independent subjects.

In Section 2, we briefly review multiple informant methods for univariate outcomes, and extend multiple informant approaches to a hierarchical data setting in Section 3. In Section 4, we present a small-scale simulation study to highlight properties of the proposed methods. In Section 5, we apply the methods to examine the association between the presence of food stores near schools and child's body mass index z-score (BMIz) using a surveillance dataset from all 5th, 7th, and 9th grade children enrolled in public schools in the state of California. We use two different features of the food environment: FFR and CS. Section 6 concludes with a discussion.

2. Review of univariate multiple informant models and generalized estimating equations

2.1. Non-standard generalized estimating equation approach for multiple informant models with independent subjects

Based on a non-standard application of GEE methods, Pepe *et al.* [3] and Horton *et al.* [2] developed a multiple informant model (MIM) to estimate the association between univariate outcomes and multiple informant predictors. For the i^{th} $(i = 1, \ldots, n)$ subject, let Y_i be an outcome and X_{ki} be multiple informants, $k = 1, \ldots, K$. The marginal associations between the outcome and each predictor, X_{ki}, are defined by separate regressions

$$E\left[Y_i | X_{ki}\right] = \beta_{0k} + \beta_{1k} X_{ki}, \tag{1}$$

where β_{0k} and β_{1k} are the intercept and the slope parameter in the k^{th} regression, $k = 1, \ldots, K$. Joint estimation of model parameters can be accomplished by re-structuring the data as

$$\tilde{\boldsymbol{Y}}_i = \begin{bmatrix} Y_i \\ Y_i \\ \vdots \\ Y_i \end{bmatrix}_{(K\times 1)}, \tilde{\boldsymbol{X}}_i = \begin{bmatrix} 1 & X_{1i} & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & X_{2i} & \ldots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & X_{Ki} \end{bmatrix}_{(K\times 2K)}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \\ \vdots \\ \beta_{0K} \\ \beta_{1K} \end{bmatrix}_{(2K\times 1)}. \tag{2}$$

Note that $\tilde{\boldsymbol{Y}}_i$ has K copies of the same outcome Y_i, and covariate vectors, $[1 X_{ki}], k = 1, \ldots, K$, are diagonally stacked in $\tilde{\boldsymbol{X}}_i$; correspondingly, $\boldsymbol{\beta}$ is a vector with all coefficients β_{0k} and β_{1k} stacked. Essentially, each subject is treated as an independent cluster with K repeated measures (which are in fact K copies of the same outcome).

Under the assumption of the identity link, constant variance, and the working independence correlation matrix, the GEE for $\boldsymbol{\beta}$ is

$$\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i^T (\tilde{\boldsymbol{Y}}_i - \tilde{\boldsymbol{X}}_i \boldsymbol{\beta}) = 0. \tag{3}$$

By solving (3), the regression parameters $\boldsymbol{\beta}$ can be estimated, and the variance–covariance matrix for the $2K$ parameter estimates, $\hat{\boldsymbol{\beta}}$, can be derived by either the empirical variance estimator or the model-based variance via the GEE approach [5]. Because the multiple informant model basically employs GEE with re-structured data, binary or count data can be also fitted by changing the link function (e.g., logit, log) [6].

Litman *et al.* [5] demonstrated that assuming the working independence correlation is optimal for certain models because the non-standard GEE approach and MLE approach yield the same estimator. Further, the working independence structure within cluster is necessary to ensure consistency in the non-standard GEE approach [7, 8]. Indeed, without a zero constraint to off-diagonal terms, joint modeling of the same outcome on multiple informants is invalid. For instance, suppose that we have an outcome y_{it} and a predictor x_{it} for the i^{th} subject at two occasions $t = 1, 2$. Under a normal assumption of y_{it} conditional on x_{it}, the joint distribution of y_{it} given x_{it} for $t = 1, 2$ can be expressed as

$$\begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix} \Big| x_{i1}, x_{i2} \sim N\left(\begin{bmatrix} \beta_{01} + \beta_{11} x_{i1} \\ \beta_{02} + \beta_{12} x_{i2} \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \right). \tag{4}$$

An implicit assumption of GEE is that covariates at a given occasion are not related to the outcome given the same covariate measured at another occasion, i.e., $E[y_{i1}|x_{i1}, x_{i2}] = E[y_{i1}|x_{i1}]$ and $E[y_{i2}|x_{i1}, x_{i2}] = E[y_{i2}|x_{i2}]$ [7, 8]. Another implicit assumption here is $y_{i1} \neq y_{i2}$. When $y_{i1} = y_{i2}$, the non-standard GEE approach needs to impose a zero constraint to σ_{12} (all off-diagonal terms in the covariance matrix).

In our motivating study, we are interested in the associations between multiple correlated predictors and weight status of children nested in schools (i.e., hierarchical data). We next review hierarchical modeling using well-developed GEE methods and subsequently extend the multiple informant model to hierarchical data.

2.2. *The generalized estimating equation model with exchangeable correlation structure*

Generalized estimating equation methods have been well established and are extensively used to model hierarchical data. We briefly review the specific case of GEE with an exchangeable correlation structure as a building block for our proposed models in Section 3. Consider a simple case where data consist of J clusters, each with n_j units with measures on an outcome and a covariate: $\{y_{ij}, x_{ij}\}, i = 1, \ldots, n_j$ for each of $j = 1, 2, \ldots, J$ clusters. Units are assumed to be correlated within clusters, but independent across clusters. A common correlation structure used for this data is an exchangeable correlation—i.e., $corr(y_{ij}, y_{i'j}) = \rho, i \neq i'$ in the j^{th} cluster. A generalized linear model is commonly used to relate the mean of $y_{ij}, \mu_{ij} = E[y_{ij}]$, to a covariate, x_{ij}, via a link function $g(\cdot)$

$$g(\mu_{ij}) = \beta_0 + \beta_1 x_{ij}, \tag{5}$$

and the variance of y_{ij} is $Var(y_{ij}) = \phi v(\mu_{ij})$, where $v(\cdot)$ is a known variance function, and ϕ is a dispersion parameter. Similar to (3), GEE estimates, $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1\right)^T$ are given by solving

$$\sum_{j=1}^{J} \boldsymbol{D}_j \boldsymbol{V}_j^{-1} (\boldsymbol{Y}_j - \boldsymbol{\mu}_j) = 0, \tag{6}$$

where $\boldsymbol{Y}_j = \left(y_{1j}, \ldots, y_{n_j j}\right)^T, \boldsymbol{\mu}_j = E[\boldsymbol{Y}_j], \boldsymbol{D}_j = \partial \boldsymbol{\mu}_j / \partial (\beta_0, \beta_1)^T, \boldsymbol{V}_j = \boldsymbol{A}_j^{1/2} \boldsymbol{R}_j \boldsymbol{A}_j^{1/2}, \boldsymbol{A}_j = \phi diag\left\{v\left(\mu_{1j}\right), \ldots, v\left(\mu_{n_j,j}\right)\right\}$, and $\boldsymbol{R}_j$ is a working correlation matrix [6]. For a continuous outcome

y_{ij} with the identity link function and an exchangeable correlation assumption, the solution of (6) for $\boldsymbol{\beta}$ with known ρ is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J} \boldsymbol{X}_j^T \boldsymbol{V}_j^{-1} \boldsymbol{X}_j\right)^{-1} \left(\sum_{j=1}^{J} \boldsymbol{X}_j^T \boldsymbol{V}_j^{-1} \boldsymbol{Y}_j\right). \tag{7}$$

The empirical or 'sandwich' variance of $\hat{\boldsymbol{\beta}}$ is

$$\widehat{Var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{BFB}, \tag{8}$$

where $\boldsymbol{B} = \left(\sum_{j=1}^{J} \boldsymbol{X}_j^T \boldsymbol{V}_j^{-1} \boldsymbol{X}_j\right)^{-1}$, $\boldsymbol{F} = \sum_{j=1}^{J} \boldsymbol{X}_j^T \boldsymbol{V}_j^{-1} \left(\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_j\right) \left(\boldsymbol{Y}_j - \hat{\boldsymbol{\mu}}_j\right)^T \boldsymbol{V}_j^{-1} \boldsymbol{X}_j$.

Because ϕ and ρ are generally unknown, $\hat{\boldsymbol{\beta}}$ needs to be iteratively re-estimated to update the estimated variance–covariance matrix, $\hat{\boldsymbol{V}}_j$. To estimate the dispersion parameter ϕ and correlation ρ, refer to Liang and Zeger [6]. Because the empirical variance estimator (8) protects against a misspecified working correlation and variance structure, inference based on GEE estimators is robust to departures from the true covariance structure [5, 6].

In Section 2.1, it is necessary to assume the working independence structure for a MIM for consistency of the estimators, but for hierarchical models reviewed here, the working independence assumption may be inefficient in some situations. Mancl and Leroux [9] demonstrated that loss of efficiency for the working independence correlation assumption can be substantial even for small correlation when the coefficient of variation in the cluster sizes (CV) is greater than 0.5. In our motivating data, the number of children varies largely across schools (CV $\approx$ 1.1). We extend the MIM to hierarchical data structures by incorporating a block diagonal working correlation to make the model valid but with diagonal blocks of exchangeable correlation structures to model correlations within clusters to enhance efficiency.

3. Hierarchical multiple informants model

3.1. Data structure and model

Let Y_{ij} be an outcome of the i^{th} unit (e.g., child's BMIz) within the j^{th} cluster (e.g., school) and denote the mean of Y_{ij} as $\mu_{ij} = E[Y_{ij}], i = 1, \ldots, n_j$ for each of $j = 1, 2, \ldots, J$. For simplicity, assume there are two multiple informant predictors measured at the cluster level, X_{1j} and X_{2j} (e.g., X_{1j} is the number of FFR, and X_{2j} is the number of CS within d miles from the j^{th} school). Given a link function $g(\cdot)$, μ_{ij} can be modeled as

$$\begin{cases} g(\mu_{ij}) = \beta_{01} + \beta_{11} X_{1j} \\ g(\mu_{ij}) = \beta_{02} + \beta_{12} X_{2j} \end{cases}, \tag{9}$$

where β_{0k} and β_{1k} for $k = 1, 2$ are the population-level intercept and the slope parameter for the k^{th} regression.

Similar to (2), the data are re-structured as

$$\tilde{\boldsymbol{Y}}_j = \begin{bmatrix} Y_{1j} \\ \vdots \\ Y_{n_j,j} \\ Y_{1j} \\ \vdots \\ Y_{n_j,j} \end{bmatrix}, \tilde{\boldsymbol{X}}_i = \begin{bmatrix} 1 & X_{1j} & 0 & 0 \\ \vdots & \vdots & 0 & 0 \\ 1 & X_{1j} & 0 & 0 \\ 0 & 0 & 1 & X_{2j} \\ 0 & 0 & \vdots & \vdots \\ 0 & 0 & 1 & X_{2j} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}^1 \\ \boldsymbol{\beta}^2 \end{bmatrix}. \tag{10}$$

Note that two copies of the outcome vector $Y_j = \left(Y_{1j}, \ldots, Y_{n_j,j}\right)^T$ for all subjects $i = 1, 2, \ldots, n_j$ within the j^{th} cluster are stacked. The covariate matrices $\boldsymbol{X}_{kj}, k = 1, 2$, consist of n_j copies of the vector $[1 X_{kj}]$, and are diagonally stacked in $\tilde{\boldsymbol{X}}_j$ Accordingly, $\boldsymbol{\beta}$ contains all β_{0k} and β_{1k}, the population-level intercept and slope parameter.

Including individual-level predictors and other cluster-level variables are straightforward. For instance, let $\boldsymbol{W}_{ij} = \begin{bmatrix} \boldsymbol{Z}_{ij} & \boldsymbol{Z}_j \end{bmatrix}$ include individual-level predictors $\boldsymbol{Z}_{ij}$ and other cluster-level variables $\boldsymbol{Z}_j$. Then, the covariate matrices, $\begin{bmatrix} \boldsymbol{X}_{1j} & \boldsymbol{W}_{ij} \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{X}_{2j} & \boldsymbol{W}_{ij} \end{bmatrix}$, within the j^{th} cluster can be re-structured as in (10).

Careful modeling of the correlation within clusters can improve inference of the population-level parameters $\boldsymbol{\beta}$. However, because of the implicit assumptions of the GEE as discussed in Section 2.1, we restrict the working covariance structure for the HMIM to a block diagonal matrix where the diagonal blocks are the correlation structures given each correlated predictor. Let $\boldsymbol{V}_{kj} = \phi_k \boldsymbol{R}_{kj}$, where $\boldsymbol{V}_{kj}$ consists of constant variance and a correlation matrix $\boldsymbol{R}_{kj}$ for $k = 1, 2$. In the motivating example, we use an exchangeable correlation structure with correlation ρ_k to model $\boldsymbol{R}_{kj}$, because children within schools can be assumed exchangeable. Hence, the working covariance matrix of the HMIM can be

$$\tilde{\boldsymbol{V}}_j = \begin{bmatrix} \boldsymbol{V}_{1j} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{V}_{2j} \end{bmatrix}_{(2n_j \times 2n_j)}. \tag{11}$$

Note that $\tilde{\boldsymbol{V}}_j$ consists of a block diagonal of distinct exchangeable covariance matrices, $\boldsymbol{V}_{1j}$ and $\boldsymbol{V}_{2j}$, given X_{1j} and X_{2j}, respectively.

With $\tilde{\boldsymbol{V}}_j$ and the identity link function, and by virtue of the block diagonal covariates and covariance matrices, the estimator for $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \left(\sum_{j=1}^{J} \tilde{\boldsymbol{X}}_j^T \tilde{\boldsymbol{V}}_j^{-1} \tilde{\boldsymbol{X}}_j\right)^{-1} \left(\sum_{j=1}^{J} \tilde{\boldsymbol{X}}_j^T \tilde{\boldsymbol{V}}_j^{-1} \tilde{\boldsymbol{Y}}_j\right),$$

yields equivalent estimates to fitting a separate model for each predictor, $\hat{\boldsymbol{\beta}}^1$, $\hat{\boldsymbol{\beta}}^2$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^1 \\ \hat{\boldsymbol{\beta}}^2 \end{bmatrix} = \begin{bmatrix} \left(\sum_{j=1}^{J} \boldsymbol{X}_{1j}^T \boldsymbol{V}_{1j}^{-1} \boldsymbol{X}_{1j}\right)^{-1} \left(\sum_{j=1}^{J} \boldsymbol{X}_{1j}^T \boldsymbol{V}_{1j}^{-1} \boldsymbol{Y}_j\right) \\ \left(\sum_{j=1}^{J} \boldsymbol{X}_{2j}^T \boldsymbol{V}_{2j}^{-1} \boldsymbol{X}_{2j}\right)^{-1} \left(\sum_{j=1}^{J} \boldsymbol{X}_{2j}^T \boldsymbol{V}_{2j}^{-1} \boldsymbol{Y}_j\right) \end{bmatrix}. \tag{12}$$

The empirical or 'sandwich' variance–covariance for the estimated parameters is

$$\widehat{Var}\left(\hat{\boldsymbol{\beta}}\right) = \tilde{\boldsymbol{B}} \tilde{\boldsymbol{F}} \tilde{\boldsymbol{B}},$$

where $\tilde{\boldsymbol{B}} = \left(\sum_{j=1}^{J} \tilde{\boldsymbol{X}}_j^T \tilde{\boldsymbol{V}}_j^{-1} \tilde{\boldsymbol{X}}_j\right)^{-1}$, and $\tilde{\boldsymbol{F}} = \sum_{j=1}^{J} \tilde{\boldsymbol{X}}_j^T \tilde{\boldsymbol{V}}_j^{-1} \left(\tilde{\boldsymbol{Y}}_j - \tilde{\boldsymbol{X}}_j \hat{\boldsymbol{\beta}}\right) \left(\tilde{\boldsymbol{Y}}_j - \tilde{\boldsymbol{X}}_j \hat{\boldsymbol{\beta}}\right)^T \tilde{\boldsymbol{V}}_j^{-1} \tilde{\boldsymbol{X}}_j$. Equivalently, if the models for each multiple informant are fitted separately,

$$\widehat{Var}\left(\hat{\boldsymbol{\beta}}\right) = \tilde{\boldsymbol{B}}^* \tilde{\boldsymbol{F}}^* \tilde{\boldsymbol{B}}^*, \tag{13}$$

where $\tilde{\boldsymbol{B}}^* = \begin{bmatrix} \left(\sum_{j=1}^{J} \boldsymbol{X}_{1j}^T \boldsymbol{V}_{1j}^{-1} \boldsymbol{X}_{1j}\right)^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\sum_{j=1}^{J} \boldsymbol{X}_{2j}^T \boldsymbol{V}_{2j}^{-1} \boldsymbol{X}_{2j}\right)^{-1} \end{bmatrix}$

and $\tilde{\boldsymbol{F}}^* = \sum_{j=1}^{J} \left(\begin{bmatrix} \boldsymbol{X}_{1j} \boldsymbol{V}_{1j}^{-1} \boldsymbol{r}_j^1 \\ \boldsymbol{X}_{2j} \boldsymbol{V}_{2j}^{-1} \boldsymbol{r}_j^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_{1j} \boldsymbol{V}_{1j}^{-1} \boldsymbol{r}_j^1 \\ \boldsymbol{X}_{2j} \boldsymbol{V}_{2j}^{-1} \boldsymbol{r}_j^2 \end{bmatrix}^T \right)$, where $\begin{bmatrix} \boldsymbol{r}_j^1 \\ \boldsymbol{r}_j^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_j - \boldsymbol{X}_{1j} \hat{\boldsymbol{\beta}}^1 \\ \boldsymbol{Y}_j - \boldsymbol{X}_{2j} \hat{\boldsymbol{\beta}}^2 \end{bmatrix}$.

That is, the empirical variance/covariance for $\hat{\boldsymbol{\beta}}$ can be calculated using results from each fitted marginal GEE model. From a practical point of view, fitting each marginal GEE model has computational efficiencies: (i) the dimension of the data will be smaller for any one model and (ii) available GEE software can be implemented to obtain the empirical covariance matrix. Example R code for calculating the empirical variance/covariance matrix (13) is provided in Appendix A.

3.2. *Hypothesis testing*

One advantage of the HMIM is that it gives a formal test to compare the association among multiple predictors on a univariate outcome while taking into account the correlation within clusters. In our motivating example, we seek to compare the association of two different features of the food environment (FFR vs. CS) with BMIz and to compare the associations between the number of FFR (or CS) and child's BMIz across several buffers. These tests can be conducted using general linear

hypotheses expressed in the form $H_0 : \boldsymbol{L\beta} = \boldsymbol{L}_0$, where $\boldsymbol{L}$ consists of l linearly independent constraints on $\boldsymbol{\beta}$, and $\boldsymbol{L}_0$ is a vector of constant terms (usually a zero vector). The Wald test statistic is $T = \left(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{L}_0\right)^T \left(\boldsymbol{L}\widehat{Var}\left(\hat{\boldsymbol{\beta}}\right)\boldsymbol{L}^T\right)^{-1}\left(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{L}_0\right)$ which, given the asymptotic normality of $\hat{\boldsymbol{\beta}}$, asymptotically follows a chi-squared distribution with l DOF. We next describe a strategy to conduct hypothesis tests.

Two approaches can be followed to compare the associations between two (or more) predictors on an outcome (e.g., different features of food environment, $X_1 =$ FFR and $X_2 =$ CS on child's BMIz within a given buffer size). The first is to use the predictors in their original scales and test for equality of coefficients, $H_0 : \beta_{11} = \beta_{12}$. Alternatively, if the scales are different (e.g., there is an overall preponderance of one feature compared with the other), the predictors can be standardized so that the coefficients are in standard deviation units (i.e., one standard deviation increase, or interquartile range increase). If we fail to reject the null hypothesis that the effects of multiple informants are the same, then, as suggested by Litman *et al.* [5], a constrained model (i.e., a model that assumes $\beta_{11} = \beta_{12}$) could be used to increase power.

In our motivating example, we are also interested in comparing the effects of a given environmental feature (e.g., FFR) across several buffers on child's BMIz. Suppose that there are a priori specified distances of interest, $d_1 < d_2 < \cdots < d_K$ from a school, and let $X_1, X_2, \ldots, X_K$ be number of FFR within the corresponding buffers. For exposition suppose $K = 3$. Then, let $\beta_{1k}, k = 1, 2, 3$, be the corresponding marginal regression coefficients. We are interested in testing whether the effects differ, i.e., the overall test $H_0 : \beta_{11} = \beta_{12} = \beta_{13}\ vs.\ H_1$: at least one differs. Failure to reject the null hypothesis suggests that the most appropriate buffer size is at least up to d_3 miles from a school. However, if the overall null hypothesis is rejected, we suggest the following subsequent tests. First, test the one-sided null hypothesis, $H_0 : \beta_{11} \leqslant \beta_{12}$. If the null hypothesis is rejected, then we decide that the buffer size d_1 miles from a school has the strongest association, and testing stops. Otherwise, conduct a second one-sided test $H_0 : \beta_{12} \leqslant \beta_{13}$. If the null hypothesis is rejected, stop and conclude the buffer with size d_2 has strongest effects. Otherwise, buffer with size d_3 is most relevant.

3.3. *AR(1) and three-level nested structures*

Given our motivating example, we consider hierarchical structures where individuals are nested in larger units in which an exchangeable correlation structure is natural. However, hierarchical data can also arise in longitudinal repeated measures for which other correlation structures may be better suited. The HMIM can be applicable to this setting as well. For instance, let Y_{it} be the i^{th} child's BMIz at time $t, i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, and the two covariates be X_{1it} and X_{2it}. Data can be re-structured in a similar manner as (10)

$$\tilde{\boldsymbol{Y}}_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \\ Y_{i1} \\ \vdots \\ Y_{iT} \end{bmatrix}, \tilde{\boldsymbol{X}}_i = \begin{bmatrix} 1 & X_{1i1} & 0 & 0 \\ \vdots & \vdots & 0 & 0 \\ 1 & X_{1iT} & 0 & 0 \\ 0 & 0 & 1 & X_{2i1} \\ 0 & 0 & \vdots & \vdots \\ 0 & 0 & 1 & X_{2iT} \end{bmatrix}.$$

Note that $\tilde{\boldsymbol{Y}}_i$ has two copies of the vector of repeated measures $\boldsymbol{Y}_i^T = [Y_{i1} \cdots Y_{iT}]$ and $\tilde{\boldsymbol{X}}_i$ has a block diagonal structure of the two covariates including the intercepts. To reflect within-cluster correlation over time in a longitudinal study, an AR(1) correlation structure can be used

$$\tilde{\boldsymbol{V}}_i = \begin{bmatrix} \sigma_1^2 AR(1, \rho_1)_{T\times T} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_2^2 AR(1, \rho_2)_{T\times T} \end{bmatrix},$$

where $AR(1, \rho_1)$ and $AR(1, \rho_2)$ are AR(1) correlation structures with autocorrelation ρ_1 and ρ_2, respectively, and the variances are σ_1^2 and σ_2^2. Other variances can also be incorporated (e.g., non-constant variance if an outcome is binary).

Another extension is for a three-level nested structure. For instance, suppose Y_{ijl} is the i^{th} child's BMI in the j^{th} school in the l^{th} county, $i = 1, \ldots, n_{jl}, j = 1, \ldots, n_l, l = 1, \ldots, L$. Assume that there are two county-level covariates $X_{kl}, k = 1, 2$, of interest for comparison. With the independence assumption across counties and given each predictor X_{kl}, let the correlation within schools be $Corr\left(Y_{ijl}|X_{kl}, , Y_{i'jl}|X_{kl},\right) = \rho_k^w, i \neq i'$, and let the correlation between schools within counties be $Corr\left(Y_{ijl}|X_{kl}, , Y_{i'j'l}|X_{kl},\right) = \rho_k^b, j \neq j'$. Data can be re-structured in a similar way as (10) where a vector of an outcome is replicated twice at the county level, and the re-arranged covariate matrix has a block diagonal structure. To account for the correlations within schools and the correlation between schools in a county, a three-level exchangeable working covariance matrix can be expressed as

$$\tilde{\boldsymbol{V}}_l = \begin{bmatrix} \sigma_1^2 \boldsymbol{R}_1 & 0 \\ 0 & \sigma_2^2 \boldsymbol{R}_2 \end{bmatrix}, \boldsymbol{R}_k = \begin{bmatrix} \boldsymbol{ex}\left(\rho_k^w\right) & \rho_k^b \mathbf{1}_{n_{jl} \times n_{jl}} & \cdots & \rho_k^b \mathbf{1}_{n_{jl} \times n_{jl}} \\ \rho_k^b \mathbf{1}_{n_{jl} \times n_{jl}} & \boldsymbol{ex}\left(\rho_k^w\right) & \cdots & \rho_k^b \mathbf{1}_{n_{jl} \times n_{jl}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k^b \mathbf{1}_{n_{jl} \times n_{jl}} & \rho_k^b \mathbf{1}_{n_{jl} \times n_{jl}} & \cdots & \boldsymbol{ex}\left(\rho_k^w\right) \end{bmatrix}_{n_{jl} n_l \times n_{jl} n_l}, k = 1, 2,$$

where $\boldsymbol{ex}\left(\rho_k^w\right)$ is a $n_{jl} \times n_{jl}$ exchangeable correlation structure with the correlation ρ_k^w. The matrix $\mathbf{1}_{n_{jl} \times n_{jl}}$ is a $n_{jl} \times n_{jl}$ one matrix, and the dispersion parameter or variance parameter are σ_k^2.

4. Simulation study

We conducted a small-scale simulation study to provide guidance on practical approaches to estimate model parameters and to examine properties of estimators and hypothesis tests. In available software, the most straightforward way to implement the proposed method is to use the working independence assumption within clusters. In standard GEEs, this approach is fully efficient when cluster sizes are equal and covariates are invariant or mean-balanced within clusters but can suffer severe efficiency loss otherwise [9].

Because the environmental effects in our motivating study will typically be small, loss or gain in efficiency may have important implication for derived inferences. Hence, we examine statistical power of detecting a small degree of differences of environmental effects assuming unequal large cluster sizes and invariant covariates within clusters.

4.1. Simulation setup

We set up the simulations to reflect two possible scenarios of the comparison of the marginal effects of FFR across several buffers on child's BMIz: (i) marginal effects of FFR are diminishing with distance and (ii) marginal effects of FFR have threshold at some distance. For both simulation scenarios, sample size, nesting structure (i.e., number of clusters and subjects per cluster), and distribution of the multiple informants (the number of restaurants within 1/4, 1/2, and 3/4 miles from each school) were the same as observed in the data example. For instance, in our motivating data, the average number of children per school and its standard deviation are 145.6 and 159.5, respectively, yielding the CV of 1.1. This means that unbalance of cluster size is large. For each simulation scenario we simulated 1000 datasets where each data contain 926,018 observations nested in 6323 clusters. Multiple informants were fixed to the observed number of FFR in the motivating example (Table I), thus we only generated outcome data conditional on the predictors $\text{FFR}_k, k = 1, 2, 3$.

To simulate data for diminishing effects of FFR with distance, the marginal effect of FFR on child's BMIz within 1/4 miles was fixed at the observed value in our motivating example ($\beta_{11} = 0.0234$). True parameters of FFR within distance 1/2 and 3/4 miles, β_{12} and β_{13}, were set to $\beta_{12} = a\beta_{11}$ and $\beta_{13} = 0.8a\beta_{11}$, where $0 \leqslant a \leqslant 1$, i.e., the effects of FFR consistently decrease over distance. Similarly, for threshold effects with distance, $\beta_{12} = a\beta_{11}$ and $\beta_{13} = a\beta_{11}$, where $0 \leqslant a \leqslant 1$, i.e., the effects of FFR decrease at some distance and continue to be constant. Here, the constant a controls the differences across regression parameters. Figure 1 shows regression parameter values used in the simulations for a range of values of a, for both diminishing effects and threshold effects.

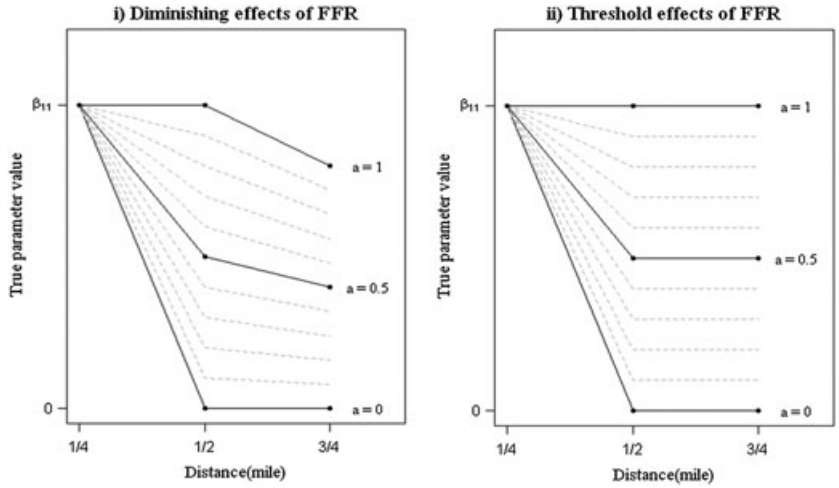

Figure 1. Regression parameter values used in simulations for i) diminishing and ii) threshold effects of fast food restaurants at distance 1/4, 1/2, and 3/4 miles. True parameter settings are $\beta_{12} = a\beta_{11}$ at 1/2 mile, $\beta_{13} = 0.8a\beta_{11}$ at 3/4 mile for diminishing effects and $\beta_{12} = a\beta_{11}$ at 1/2 mile, $\beta_{13} = a\beta_{11}$ at 3/4 mile for threshold effects, such that $a\,(0 \leqslant a \leqslant 1)$ controls the differences across parameters.

Note that, given the observed variances of the predictors (Table I), these effects constrain the marginal covariances between predictors and outcome to $\sigma_{X_1Y} = \beta_{11}\sigma^2_{X_1} = 0.011, \sigma_{X_2Y} = a\beta_{11}\sigma^2_{X_2}$, and $\sigma_{X_3Y} = a\beta_{11}\sigma^2_{X_3}$ for threshold effects, and $\sigma_{X_3Y} = 0.8a\beta_{11}\sigma^2_{X_3}$ for diminishing effects. Further, we assumed the marginal outcome mean was $\mu_Y = 0$ (centered) and had marginal variance $\sigma^2_Y = 0.14$ (as observed in our motivating data).

To simulate outcomes, we first generated cluster-level values from a normal distribution with mean $E\left[\bar{Y}_{\cdot j}|FFR_{1j;}, FFR_{2j}, FFR_{3j}\right] = \gamma_0 + \gamma_1 FFR_{1j} + \gamma_2 FFR_{2j} + \gamma_3 FFR_{3j}$ and variance $Var\left[\bar{Y}_{\cdot j}|FFR_{1j}, FFR_{2j}, FFR_{3j}\right] = \sigma^2_\delta\left(\sigma^2_\delta = \sigma^2_Y - \gamma^2_1 Var\left(FFR_{1j}\right) - \gamma^2_2 Var\left(FFR_{2j}\right) - \gamma^2_3 Var\left(FFR_{3j}\right) - 2\gamma_1\gamma_2 Cov\left(FFR_{1j}, FFR_{2j}\right) - 2\gamma_2\gamma_3 Cov\left(FFR_{2j}, FFR_{3j}\right) - 2\gamma_3\gamma_1 Cov\left(FFR_{3j}, FFR_{1j}\right)\right)$. The conditional mean of the cluster level given all three predictors was used because the outcome needs to be simulated only once given all three predictors. We used the Sweep operator [10, 11] to derive conditional associations, $\gamma_0, \gamma_1, \gamma_2, \gamma_3$ given the specified marginal associations $\beta_{11}, \beta_{12}, \beta_{13}$ (see Appendix B). Given the cluster mean, we generated subject-level observations as $Y_{ij} = \bar{Y}_{\cdot j} + \epsilon_{ij}$, where $\epsilon_{ij} \sim N\left(0, \sigma^2_\epsilon\right), \sigma^2_\epsilon = \frac{1-\rho_y}{\rho_y}\sigma^2_\delta$, and $\rho_y = Corr\left(Y_{ij}, Y_{i'j}\right)$ set to 0.05 for $i \neq i'$ in the j^{th} cluster. With this parameter setting, the true marginal covariance matrix for the HMIM (11) has non-equal blocks (see Appendix C).

4.2. Simulation results

Let $\hat{\boldsymbol{\beta}}_{Ex}$ denote the estimator for $\boldsymbol{\beta}$ in (12) when using $\boldsymbol{V}_{kj} = \phi_k \boldsymbol{R}_{kj}$ as the diagonal blocks of $\boldsymbol{V}_j$ (11), with $\boldsymbol{R}_{kj}$ being an exchangeable correlation structure with parameter $\rho_k, k = 1, 2, 3$. Similarly, let $\hat{\boldsymbol{\beta}}_I$ denote the estimator for $\boldsymbol{\beta}$ when $\boldsymbol{V}_j$ consists of blocks of $\boldsymbol{V}_{kj} = \phi_k \boldsymbol{R}_{kj}$ with $\boldsymbol{R}_{kj}$ being the independence correlation structure. From the 1000 datasets, the empirical power was calculated as the rate of rejecting the overall test for comparing marginal effects β_{1k} for $k = 1, 2, 3$, i.e., $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ versus H_1 : at least one differs, for both estimators. For a given data set, the null hypothesis was rejected when the Wald test statistic T (Section 3.2) exceeded the critical value for a chi-squared distribution with 2 DOF. As shown in Figure 2, the empirical power of $\hat{\boldsymbol{\beta}}_{Ex}$ was uniformly higher than $\hat{\boldsymbol{\beta}}_I$ for diminishing effects of FFR. Power was always greater than the significance level (0.05) because in the range of a $(0 \leqslant a \leqslant 1)$, true parameters of FFR were always distinguishable. The U-shape of the power function within the range of a is due to the non-centrality parameter of the test statistic, T, being a quadratic function of a under the alternative hypothesis.

For the threshold effects of FFR, the empirical powers of both estimators $\hat{\boldsymbol{\beta}}_{Ex}$ and $\hat{\boldsymbol{\beta}}_I$ go to nominal value (0.05) of Type I error rate when $a = 1$, or $\beta_{11} = \beta_{12} = \beta_{13}$. When β_{1k}'s are distinguishable or a goes to 0, the power for $\hat{\boldsymbol{\beta}}_{Ex}$ increases faster than for $\hat{\boldsymbol{\beta}}_I$. The crossing of the power curves of

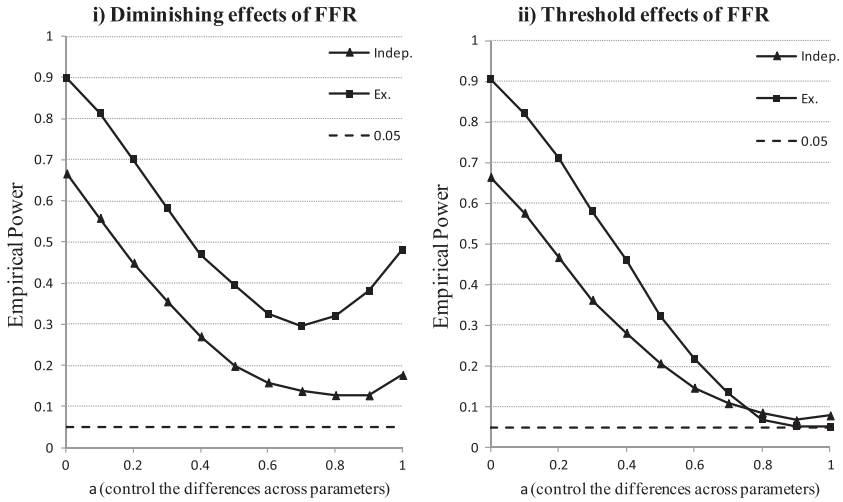

Figure 2. Simulation results assessing power for the hypothesis test $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ for i) diminishing and ii) threshold effects of fast food restaurants using hierarchical multiple informants model with exchangeable (Ex.) and independence (Indep.) correlation structures. True parameter settings are $\beta_{12} = a\beta_{11}, \beta_{13} = 0.8a\beta_{11}$ for diminishing effects and $\beta_{12} = a\beta_{11}, \beta_{13} = a\beta_{11}$ for threshold effects.

the estimators may be due to Monte Carlo errors from the simulation. The Monte Carlo errors will be negligible with an increased number of simulations.

Because of the large number of clusters, the empirical power from the overall test when $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$ is true preserved 5% Type I error rate. When the number of clusters is small, a bias corrected sandwich estimator could be used [12].

4.3. Simulation conclusions

Accounting for correlation within clusters is important to better detect small differences between marginal effects in an environmental study of clustered or hierarchical data. For instance, for threshold and diminishing effects, this simulation shows that, if using $\hat{\boldsymbol{\beta}}_{Ex}$ 80% power was achieved when $a < 0.2$. That is, to be statistically distinguishable, the association between the outcome and number of FFR at the outer buffers needs to be at most 20% of the association with the number of FFR in the inner buffer. However, note that if using $\hat{\boldsymbol{\beta}}_I$, 80% power could not be reached for any value of a.

The fact that power using $\hat{\boldsymbol{\beta}}_{Ex}$ is higher than $\hat{\boldsymbol{\beta}}_I$ can be explained by applying previous work on asymptotic relative efficiency (ARE) of Mancl and Leroux [9]. According to their formula for ARE and given that we have the conditions: (i) invariant covariates within clusters, (ii) unequal cluster sizes (CV ≈ 1.1), (ii) large cluster size ($\bar{J} = 145.6$), and (iv) intra-cluster correlation $\rho_k \approx 0.05, k = 1, 2, 3$, the ARE of $\hat{\boldsymbol{\beta}}_{Ex}$ to $\hat{\boldsymbol{\beta}}_I$ in the current data is about 0.55, meaning approximately 45% loss of efficiency by employing the independence correlation structure even for small intra-cluster correlation and invariant covariates within clusters.

This simulation study shows that an HMIM should be employed for formal testing of associations of an outcome among correlated predictors in clustered or hierarchical data to increase power.

5. Data example

We used data for children who participated in the 2007 California physical fitness test (also known as FitnessGram), which contains direct measures of children's weight and height, among all children attending 5th, 7th, and 9th grade, as well as other covariates such as age, sex, and race. Following prior exclusion criteria, we used data on 926,018 children nested in 6,323 schools [13]. The location of FFR and CS in California was purchased from InfoUSA, a commercial source. Geocodes for schools and food stores were cross-referenced to obtain the counts of stores within 1/4, 1/2, and 3/4 mile of a school, denoted by FFR_1, FFR_2, FFR_3 and CS_1, CS_2, CS_3. We obtained data from the California Department of Education's databases, and the 2000 US Census to characterize the size and composition of the schools, as well as the socio-economic conditions of the neighborhoods in which schools were located.

Table I. Descriptive statistics for body mass index z-scores (BMIz)*, number of fast food restaurants and convenience stores at three distances and their pairwise correlations.

Distance	Variable	Mean	SD	Corr.	BMIz*	FFR_1	FFR_2	FFR_3	CS_1	CS_2	CS_3
	BMIz*	0.744	0.374	BMIz*	1	0.04	0.09	0.13	0.13	0.20	0.25
1/4 mile	FFR_1	0.233	0.679	FFR_1		1	0.55	0.36	0.25	0.23	0.21
1/2 mile	FFR_2	1.150	1.758	FFR_2			1	0.73	0.19	0.38	0.37
3/4 mile	FFR_3	2.709	2.915	FFR_3				1	0.15	0.37	0.50
1/4 mile	CS_1	0.178	0.473	CS_1					1	0.53	0.39
1/2 mile	CS_2	0.774	1.083	CS_2						1	0.76
3/4 mile	CS_3	1.697	1.845	CS_3							1

BMIz* is the mean of child's BMIz within schools.

Body mass index z-score was used as a continuous outcome. BMIz was derived by calculating body mass index (weight in kg/ height in meters squared) and standardizing it according to an age and gender-specific BMI distribution. In other words, BMIz indicates how much a child's BMI differs from a reference group of the same age and gender [14]. In contrast to BMI among adults, BMI among children needs to be standardized to a reference population because they are still growing and their body composition is changing as they grow [15] such that the meaning of BMI is not the same across age and sex. Following prior analyses [13], we included individual-level and school-level covariates as adjustment factors in models. The individual-level covariates are grade, age, gender, and race/ethnicity. The school-level covariates are school's racial composition, school's neighborhood-level education, school's total enrollment, and percent of children enrolled in the free or reduced price meal program.

Descriptive statistics of child's BMIz, the number of FFR and CS within distance 1/4, 1/2, and 3/4 miles are summarized in Table I. The average number of children per school and its standard deviation are 145.6 and 159.5, respectively, yielding the CV of 1.1.

We conducted two sets of analyses: (i) the comparison of two different features of food stores within the same buffer, and (ii) the comparison of a food environment feature across several buffer sizes. In both sets of analyses, we fitted an HMIM with both exchangeable and independence structures, and, for comparison, also MIM without accounting for cluster correlation. Further, the individual-level and school-level covariates described previously were included.

First, for the comparison of two different features of food stores within the same buffer, the counts of FFR and CS were standardized to a mean of zero and a standard deviation of one because of potentially different scales (e.g., an overall preponderance of one feature may be different compared with the other) so that the coefficients are in standard deviation units. We use $F^s_{1j}, F^s_{2j}, F^s_{3j}$ and $C^s_{1j}, C^s_{2j}, C^s_{3j}$ to denote the standardized number of FFR and CS within 1/4, 1/2, and 3/4 miles from the j^{th} school, respectively, and $\boldsymbol{Z}^T_{ij}$ to denote the vector of the individual-level and the school-level covariates or confounders. For each of three buffers, $k = 1, 2, 3$, the fitted models are

$$E\left[BMIz_{ij}|F^s_{kj}, \boldsymbol{Z}_{ij}\right] = \beta^{F^s}_{0k} + \beta^{F^s}_{1k} F^s_{kj} + \boldsymbol{Z}^T_{ij}\boldsymbol{\beta}^{F^s}_{(k)},$$

$$E\left[BMIz_{ij}|C^s_{kj}, \boldsymbol{Z}_{ij}\right] = \beta^{C^s}_{0k} + \beta^{C^s}_{1k} C^s_{kj} + \boldsymbol{Z}^T_{ij}\boldsymbol{\beta}^{c^s}_{(k)}.$$

The null hypotheses of interest are whether for each buffer, the association between the number of CS and BMIz is the same as the association between the number of FFR and BMIz, i.e., $H_0 : \beta^{F^s}_{1k} = \beta^{C^s}_{1k}$ for $k = 1, 2, 3$.

Table II provides the results for the comparison of FFR and CS within the same buffer size. For all buffer sizes, the adjusted associations of CS with BMIz are significantly greater than those of FFR with BMIz. For example, given the 1/4 mile buffer size, child's BMIz increases 0.77×10^{-3} and 10.82×10^{-3} per one standard deviation increase of FFR (= 0.679) and CS (= 0.473), respectively, after adjusting for individual and school factors. Using either an exchangeable correlation structure or the independence structure, we reached the same substantive conclusion, although the point estimates are slightly different.

Second, to investigate how the association between the number of FFR (or CS) and BMIz varies across several buffers, we fitted models $E[BMIz_{ij}|FFR_{kj}, \boldsymbol{Z}_{ij}] = \beta_{0k} + \beta_{1k} FFR_{kj} + \boldsymbol{Z}^T_{ij}\boldsymbol{\beta}_{(k)}$ for $k = 1, 2, 3$ (similar for CS).

Table II. Estimated associations* of two different features of the food environment (fast food restaurants vs. convenience stores) within the same buffer on body mass index z-scores and hypotheses tests of equality of the associations.

		HMIM					
		Exchangeable		Independence		MIM	
Distance	Association* or test	Est.	(SE)	Est.	(SE)	Est.	(SE)
1/4 mile	Std. fast food $\left(\beta_{11}^{Fs}\right)$	0.77	(2.54)	−0.11	(2.33)	−0.11	(0.99)
	Std. convenience stores $\left(\beta_{11}^{Cs}\right)$	10.82	(2.44)	7.78	(2.40)	7.78	(1.12)
	$H_0 : \beta_{11}^{Fs} = \beta_{11}^{Cs}$	P = 0.001		P = 0.009		P < 0.0001	
1/2 mile	Std. fast food $\left(\beta_{12}^{Fs}\right)$	1.96	(3.46)	2.75	(2.66)	2.75	(1.10)
	Std. convenience stores $\left(\beta_{12}^{Cs}\right)$	11.56	(2.88)	10.35	(2.72)	10.35	(1.14)
	$H_0 : \beta_{12}^{Fs} = \beta_{12}^{Cs}$	P = 0.003		P = 0.009		P < 0.0001	
3/4 mile	Std. fast food $\left(\beta_{13}^{Fs}\right)$	5.02	(3.36)	5.59	(2.73)	5.59	(1.15)
	Std. convenience stores $\left(\beta_{13}^{Cs}\right)$	13.24	(3.00)	13.80	(2.82)	13.80	(1.18)
	$H_0 : \beta_{13}^{Fs} = \beta_{13}^{Cs}$	P = 0.005		P = 0.004		P < 0.0001	

Associations are estimated on the basis of standardized number of fast food restaurants and convenience stores, adjusting for individual-level and school-level covariates, using the proposed hierarchical multiple informants model and the multiple informants model without accounting for within cluster correlation.
* Estimate and SE were multiplied by 10^3 to enhance readability.

The question of interest is whether the associations between the number of a given foods store type (FFR or CS) varies across buffer sizes, i.e., the overall null hypothesis $H_0 : \beta_{11} = \beta_{12} = \beta_{13}$. Note that the coefficients are expressed in units of BMIz per one unit increase in the number of stores, because the same feature is being compared across buffers. The parameter estimates and the p-values for the overall hypotheses tests are given in Table III.

The overall test for the associations of FFR across several buffers is not rejected ($p = 0.895$), meaning the number of FFR within 1/4, 1/2, and 3/4 mile from a school do not have significantly different associations with child's BMIz. This implies that the most relevant buffer size is 3/4 mile from a school, or potentially further. Child's BMIz increases 1.72×10^{-3} per one FFR increment within 3/4 miles from a school ($p = 0.862$, not reported in Table III). By employing either an exchangeable correlation matrix or the independence structure, the same conclusion is derived.

Unlike the associations of FFR, the associations of CS across the buffers are significantly different based on HMIM with the exchangeable correlation matrix ($p = 0.004$). On the basis of the result of HMIM with the exchangeable correlation matrix, we performed subsequent hypothesis test as described in Section 3.2. i.e., test the one-sided null hypothesis, $H_0 : \beta_{11} \leqslant \beta_{12}$. We rejected the one-sided null hypothesis ($p = 0.017$) and concluded that the most relevant buffer size for the association between CS and child's BMIz is 1/4 mile from a school. Child's BMIz increases 0.022 per one CS increment within 1/4 miles from a school after adjusting other covariates.

Note that in Tables II and III, the estimates and standard errors using an exchangeable correlation matrix differ from those using the independence assumption. These changes result in test statistics (e.g., a ratio of the difference between two regression parameters to its standard error) that are larger when using the exchangeable correlation matrix and thus smaller p-values. For instance, in Table III the overall null hypothesis yields a p-value = 0.118 when using the independence assumption within cluster while an exchangeable assumption yielded a p−value = 0.004, highlighting the gain in efficiency when using the exchangeable versus independence assumption.

Lastly, as shown in Tables II and Table III, the MIM without accounting for within-cluster correlation provides the same point estimates of HMIM with the independence structure, but the failure to account for hierarchical structures yields the underestimated standard errors, resulting in invalid inference.

Table III. Estimated associations* between number of fast food restaurants (or convenience stores) and body mass index z-scores across three distances and test of equality of association across distances.

		HMIM					
		Exchangeable		Independence		MIM	
Distance	Association* or Test	Est	(SE)	Est	(SE)	Est	(SE)
1/4 mile	Fast Food (β_{11})	1.14	(3.73)	−0.17	(3.21)	−0.17	(1.36)
1/2 mile	Fast Food (β_{12})	1.12	(1.97)	1.56	(1.42)	1.56	(0.59)
3/4 mile	Fast Food (β_{13})	1.72	(1.15)	1.92	(0.88)	1.92	(0.37)
	$H_0 : \beta_{11} = \beta_{12} = \beta_{13}$	P = 0.895		P = 0.782		P = 0.265	
1/4 mile	Convenience store (β_{11})	22.89	(5.16)	16.47	(4.77)	16.47	(2.23)
1/2 mile	Convenience store (β_{12})	10.67	(2.66)	9.55	(2.36)	9.55	(0.99)
3/4 mile	Convenience store (β_{13})	7.17	(1.62)	7.48	(1.44)	7.48	(0.60)
	$H_0 : \beta_{11} = \beta_{12} = \beta_{13}$	P = 0.004		P = 0.118		P =<0.001	

Associations are estimated from three different models, adjusting for individual-level and school-level covariates.
* Estimate and SE was multiplied by 10^3 to enhance readability

6. Discussion

We extended multiple informants methods to a hierarchical data setting to enable comparison of the associations between multiple correlated predictors on a univariate outcome measured in clustered sets of individuals. The method is based on a non-standard application of generalized estimating equations and can be applied to settings where the outcome is continuous, count, or binary. In the simulation study, we showed the improved power and efficiency of estimators based on using a block diagonal of exchangeable correlation matrices instead of the working independence correlation structure. A practical advantage of an HMIM is that it can be fitted using available GEE software. A marginal GEE model for each multiple informant can be separately fitted, and then a joint empirical variance estimator can be calculated to conduct hypothesis tests involving the marginal effects of predictors. We applied HMIMs to examine how the association between the number of FRR (or CS) and child's BMIz varies across several buffers from a school and to compare the association of two different features of the food environment (FFR vs. CS) with child's BMIz. The overall hypothesis that the association between number of FFR and child's BMIz across several buffers is the same was not rejected, suggesting that the association of the count of FFR up to 3/4 mile from a school does not differ significantly from the association at small buffer sizes with accounting for individual-level and school-level covariates. In contrast, the association of BMIz with the count of CS differs depending on distance from a school, with 1/4 mile being most relevant buffer size. We also showed the association between the count of CS and BMIz is much stronger compared with the association between FFR and BMIz.

We proposed a testing strategy that may be helpful in selecting an appropriate buffer size at which to estimate the association between an environmental feature and an outcome. However, there are some extreme cases where the hypothesis testing strategy may fail. For instance, if there are few or no additional of food stores between distances d_{k-1} and d_k, then the marginal association between the number of food stores and child's BMIz across buffers d_{k-1} and d_k would likely to be the same due to $X_{k-1,j} \approx X_{kj}$. An alternative might be to define multiple informant covariates as the new information not contained in the previous buffer, e.g., $X_j(\delta_k) = X_{kj} - X_{k-1,j}$, and fit a model $g(\mu_{ij}) = \beta_0^* + \beta_1^* X_j(\delta_1) + \beta_2^* X_j(\delta_2) + \beta_3^* X_j(\delta_3)$. If $\beta_3^* \neq 0$, then the buffer of size is d_3 from a school still provides information on child's BMIz, and similarly for the other buffer sizes. However, the interpretation of these coefficients is that of conditional associations, not marginal associations.

We also confirmed the underestimated variances of the estimators from the MIM because of the failure to incorporate hierarchical structures, which provide us invalid inference. The bootstrap method, as pointed out by a referee, may be employed for valid inference. For instance, suppose that we have 5000 bootstrap estimates of regression parameters from the MIM. Then, the empirical variance/covariance of the estimates can be used for hypothesis testing because the failure to account for hierarchical structures has little impact on the population point parameter estimates. The bootstrap method, however, may require extensive computational time.

The main idea of MIM is very similar to seemingly unrelated regression methods (SUR) [16]. The main difference between MIM and SUR is that in MIM the same outcome is replicated to form an

outcome vector for the cluster with predictors changing from one replicate to another, whereas different outcomes form an outcome vector in SUR. Similarly, an HMIM is related to the model structure described by Rochon [17]. The author employed SUR in a repeated measures setting for discrete and continuous outcome variables; nevertheless, here we extended MIM to hierarchical data to enable estimation and testing of marginal effects of several correlated factors or multiple informants or predictors.

Possible extensions of this work may include estimating non-linear or semi-parametric associations between multiple predictors and outcomes (e.g., penalized splines). Furthermore, the size of the buffer may vary by whether schools are located in urban or rural locations, or other spatially varying features of the environment. Hence, a future area of research can be to investigate the spatial variation of the coefficients and/or tests presented in the present article.

The methods presented are motivated by school environment contributions to childhood obesity, yet such methods apply more broadly to other environmental features and other types of nested outcome data. For example, the proposed approach can be applied to study questions concerning the relative influence of home versus work environment on adult health outcomes such as cardiovascular disease or quality of life.

Acknowledgements

The authors acknowledge salary support by grants from the National Heart, Lung, and Blood Institute of the National Institutes of Health (K01HL115471, Sanchez-Vaznaugh) and the Robert Wood Johnson Foundation (69599, Sanchez). The content is solely the responsibility of the authors and does not necessarily represent the official views of those institutions.

References

1. Davis B, Carpenter C. Proximity of fast-food restaurants to schools and adolescent obesity. *American Journal of Public Health* 2009; **99**(3):505–510.
2. Horton NJ, Laird NM, Zahner GEP. Use of multiple informant data as a predictor in psychiatric epidemiology. *International Journal of Methods in Psychiatric Research* 1999; **8**(1):6–18.
3. Pepe MS, Whitaker RC, Seidel K. Estimating and comparing univariate associations with application to the prediction of adult obesity. *Statistics in Medicine* 1999; **18**(2):163–173.
4. Horton NJ, Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. *Statistics in Medicine* 2004; **23**(18):2911–2933.
5. Litman HJ, Horton NJ, Hernández B, Laird NM. Estimation of marginal regression models with multiple source predictors. *Handbook of Statistics* 2008; **27**(7):730–746.
6. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1):13–22.
7. Pan W, Louis TA, Connett JE, Connett E. A note on marginal linear regression with correlated response data. *The American Statistician* 2000; **54**(3):191–195.
8. Sullivan PM, Anderson GL. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics* 1994; **23**(4):939–951.
9. Mancl LA, Leroux BG. Efficiency of regression estimates for clustered data. *Biometrics* 1996; **52**(2):500–511.
10. Beaton AE. The use of special matrix operators in statistical calculus. *Research Bulletin RB–64–51*, Princeton, NJ, 1964.
11. Dempster AP. *Elements of Continuous Multivariate Analysis*. Addison-Wesley: Reading, MA, 1969.
12. Mancl LA, Derouen TA. A covariance estimator for GEE with improved small-sample properties. *Biometrics* 2001; **57**:126–134.
13. Sánchez BN, Sanchez-Vaznaugh EV, Uscilka A, Baek J, Zhang L. Differential associations between the food environment near schools and childhood overweight across race/ethnicity, gender, and grade. *American Journal of Epidemiology* 2012; **175**(12):1284–1293.
14. Centers for Disease Control and Prevention. CDC growth charts, United States. *Atlanta (GA): CDC* 2005; Retrieved January 28, 2010, from http://www.cdc.gov/GROWTHcharts/.
15. Must A, Anderson SE. Body mass index in children and adolescents: considerations for population-based applications. *International Journal of Obesity* 2006; **30**(4):590–594.
16. Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 1962; **57**(298):348–368.
17. Rochon J. Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics* 1996; **52**(2):740–750.