Published in final edited form as:

Stat Med. 2019 March 15; 38(6): 1013–1035. doi:10.1002/sim.8039.

Modeling a Bivariate Residential-Workplace Neighborhood Effect when Estimating the Effect of Proximity to Fast-Food Establishments on Body Mass Index

A. James O'Malley^{*,1,2}, Peter James^{3,4}, Todd A. MacKenzie^{1,2}, Jinyoung Byun¹, S. V. Subramanian^{3,4}, and Jason P. Block^{5,6}

¹Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, New Hampshire, USA

²The Dartmouth Institute of Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, New Hampshire, USA

³Department of Epidemiology and Environmental Health, Harvard TH Chan School of Public Health, Massachusetts, USA

⁴Department of Environmental Health, Harvard TH Chan School of Public Health, Massachusetts, USA

⁵Department of Population Medicine, Harvard Pilgrim Health Care Institute, Massachusetts, USA

⁶Department of Population Medicine, Harvard Medical School, Massachusetts, USA

Abstract

Hierarchical modeling is the preferred approach of modeling neighborhood effects. When both residential and workplace neighborhood are known, a bivariate (residential-workplace) neighborhood random effect that quantifies the extent that a neighborhood's residential and workplace effects are correlated may be modeled. However, standard statistical software for hierarchical models does not easily allow correlations between the random effects of distinct clustering variables to be incorporated. To overcome this challenge, we develop a Bayesian model and accompanying estimation procedure that allows for correlated bivariate neighborhood effects and also allows individuals to reside or work in multiple neighborhoods, cross-sectional and longitudinal heterogeneity between individuals, and serial correlation between repeated observations over time. Simulation studies that vary key model parameters evaluate how well each aspect of the model is identified by the data. We apply the model to the motivating Framingham Heart Study (FHS) linked food establishment data to examine whether (i) Proximity to fast-food establishments is associated with Body Mass Index (BMI); (ii) Workplace neighborhood exposure associations are larger than those for residential neighborhood exposure; (iii) Residential neighborhood exposure associations correlate with workplace neighborhood exposure.

*Correspondence: A. James O'Malley, The Dartmouth Institute of Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, 1 Medical Center Drive, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA. James.OMalley@Dartmouth.edu.

Conflict of interest

The authors declare no potential conflict of interests.

Comparisons of the full model to models with restricted versions of the covariance structure illustrate the impact of including each feature of the covariance structure.

Keywords

Bivariate neighborhood random effect; Body Mass Index; Framingham Heart Study; Hierarchical model; Random-intercept random-slope; Serial correlation

1 | INTRODUCTION

This paper is motivated by a longitudinal cohort study of the association of proximity to food establishments on an individual's BMI. However, in contrast to analyses that focus on estimation of direct effects of exposure variables, we focus on the unexplained variation in outcomes across neighborhoods. That is, do certain neighborhoods expose individuals to greater risk of an outcome (e.g., increased BMI) than others for no observable reason – a question that is of particular interest to geographers, sociologists, and other social scientists or policy-makers.¹ The primary hypothesized mechanism through which neighborhoods might influence BMI is the enhanced access that some neighborhoods provide to food. Some neighborhoods have copious availability of fast food restaurants and others supermarkets. To the extent that access leads to greater consumption of food, proximity may lead to higher or lower BMI.² Further, close proximity to certain types of food establishments might cue consumption, even if residents do not frequent the establishments in their neighborhood. Studies from neuroscience show that environmental food cues elicit a surge in brain dopamine which in turn stimulates the urge to consume food.³ Frequent visualization of unhealthy food establishments could elicit similar responses.⁴ Typically, neighborhood effects are modeled as univariate latent variables, also known as random effects, whose differences quantify unexplained heterogeneity between geographic or neighborhood units.^{5,6,7} But with knowledge of individuals' residential and workplace neighborhoods, a bivariate neighborhood random effect that distinguishes the effect of where an individual works (workplace exposure) from where they live (residential exposure) can be modeled.

In prior work, for multiple food establishment types (fast-food, supermarkets, grocery stores, convenience stores, full service restaurants, bakeries and coffee shops), we estimated the association between residential exposure to the closest food establishment of each type and BMI. We found that distance from home to the closest fast-food establishment had the only statistically significant association with BMI and that there was surprisingly small variation in BMI between neighborhoods.^{8,9} Based on these historical findings, we now perform a more intricate examination of the association of exposure to fast-food establishments using an expanded dataset consisting of our original data and recently obtained data on the distance from work to the closest fast-food establishment, the number of fast-food establishments passed on ones commute to and from work, and the neighborhood of working locations. The latter allows bivariate clustering due to residential and workplace neighborhood exposure to be modeled.

Although there already exist models with bivariate (or more generally multivariate) random effects, these differ from the situation considered here as they account for bivariate

outcomes,^{10,11} multiple random effects (e.g., random-intercepts, random-slopes) for the same clustering variable,¹² or stratified random effects such as allowing each neighborhood to have different effects at every observation time.¹³ To our knowledge, the case of a bivariate random effect for a clustering variable that classifies a single observation multiple ways (e.g., by residential and employment neighborhood) has not been considered previously. In this paper we develop and evaluate a statistical model to account for a bivariate random effect for neighborhood, a contribution that is both conceptual and methodological. The resulting model addresses the following important questions in our motivating application. (i) is proximity to fast-food establishments from home and work associated with BMI? (ii) does workplace neighborhood matter more than residential neighborhood? (iii) are neighborhoods that are high risk to residents gaining weight also high risk to employees of those neighborhoods gaining weight?

In addition to a bivariate random effect for neighborhood, the motivating food establishment data comprise several other complexities. Individuals may reside or work in different neighborhoods at different exams, individuals are clustered within neighborhoods, observations on individuals may be serially correlated, and multiple important predictors of BMI are measured. We develop a Bayesian statistical model that accommodates all of these features and use it to evaluate the impact of accounting for the correlation of the residential and workplace neighborhood effects, the movement of individuals between neighborhoods, heterogeneity of individuals' BMI trajectories across time, and the serial correlation between observations. A simulation study evaluates properties of the Bayesian estimators and provides insights into the extent to which each parameter is identified by the data.

The remainder of the paper is organized as follows. The BMI – food establishment proximity study motivating this paper is described in detail in Section 2. We introduce the notation and develop a Bayesian model for the analysis of (i) - (iii) above in Section 3. The estimation procedure and its implementation is described in Section 4; the conditional posterior distributions required to develop a self-programmed Markov chain Monte-Carlo estimation procedure are presented in Section A of the Appendix. To confirm that the model-fitting procedure works and explore properties of the posterior mean estimator of each model parameter, a simulation study is performed in Section 5. The empirical findings of the food proximity analysis including their sensitivity to the prior distribution of the covariance matrices, and thus the correlation parameter integral to question (iii) above, are in Section 6.1 while the impact of various alterations or simplifications of the data and the model are reported in Section 6.2. The paper concludes with a discussion in Section 7.

2 | PROXIMITY TO FAST-FOOD STUDY

We analyze Body Mass Index (BMI) – food establishment proximity data for the subjects in the Offspring and Omni Cohorts of the Framingham Heart Study (FHS). BMI was chosen to be the outcome because it was measured at all exams whereas measures such as waist circumference were intermittently collected. The Offspring Cohort comprises the 5,124 children of the Original Cohort members and their spouses. The offspring cohort began in 1971 with the data used in our analysis collected until the end of the 8th exam in 2008. The mid-points (and ranges) of the eight waves of exams for the FHS offspring cohort are 1973

(1971–1975), 1981 (1979–1983), 1985 (1983–1987), 1989 (1987–1991), 1993 (1991–1995), 1997 (1995–1998), 2000 (1998–2001), and 2006 (2005–2008). The Omni Cohort enrolled only racial minorities living in and around Framingham, Massachusetts, starting in 1994. We examine Omni Cohort data through 3 exams. Because they correspond to that of exams 6 through 8 of the Offspring Cohort, the Omni Cohort exams were coded as 6 through 8. Observations were excluded from our analysis if a cancer diagnosis was present or an individual resided in a nursing home at that exam. Individuals who were less than 18 at the time when they would have been enrolled in the Offspring or Omni Cohort were only available for inclusion after age 21 years. A feature of the FHS is the extremely high retention rate of 99% of participants regularly returning for scheduled exams.¹⁴

The majority of Offspring Cohort members live in the four principal FHS towns (Framingham, Holliston, Ashland, and Natick) or the 10 surrounding Massachusetts towns. We obtained food proximity data by visiting Boards of Health and libraries in each town, searching Framingham-area Yellow Pages & White Pages from 1971, and consulting the Commercial retail dataset, Dun and Bradstreet.⁸ The collection of proximity data began with the 4-town area and the portions of the 10 surrounding towns closest to the 4-town area. Although the multi-pronged approach of using Boards of Health data, the Yellow & White Pages, and the commercial retail dataset of Dun and Bradstreet was replicated, the level of scrutiny applied to the initial wave of data collection in the 4-town area was not able to be applied to the subsequent 10 towns. Because of the concern that the proximity data is less scrutinized for FHS members living and working in the surrounding 10 towns than for the 4-town area, for our “primary analysis” of the FHS data we analyze only observations in which the FHS member lived and worked (or alternative to working) in the 4-town area. However, in a supplemental analysis, the model is also fitted on the data from the 14-town area and the results compared.

All distance calculations were made by geocoding residences, workplaces, and food establishments and computing the minimum driving distance using ArcGIS. We reduced the commuting exposure to a scalar by counting the number of fast-food establishments passed (defined as being within a 60 meter radius) during the estimated shortest time commute to and from work. Thus, our exposure measure is threefold: residential proximity (distance to nearest fast-food establishment from home), workplace proximity (distance to nearest fast-food establishment from work), the number of fast-food establishments passed on shortest commutes to and from work. Due to prior studies showing that associations between residential neighborhood characteristics and health outcomes are stronger for women, the exposure measure was expanded to a 6-dimensional measure given by the three proximity to fast-food establishment measures stratified by sex (male and female).

Potential confounding variables of the BMI – food establishment proximity relationship are exam number (i.e., study wave), gender, year of birth, marital status, smoking status, education (high school or less, greater than high school, not reported), employed (versus not employed), residential poverty status, and workplace neighborhood poverty status. To account for a possible nonlinear population trend in BMI over time and the study member's age, we treat exam number as a categorical predictor and control for year of birth; we favored year of birth over the exam-specific age as a predictor because the latter is highly

collinear with exam number. Because the heterogeneity across individuals in the amount of time between the same pair of exams is minimal, it suffices to treat the data as comprising cross-sectional panels as opposed to representing time continuously.

Residential and workplace neighborhoods are indicated by census-tracts that are fixed to the 2000 census-tract borders. The restriction to the individual-exams in which individuals both resided and worked in the 4-town area reduced the total sample to 2,889 unique subjects in 23 neighborhoods contributing 11,056 (person-exam) observations. Of these, 2,586 subjects (10,490 observations) were enrolled in the OffSpring Cohort and 303 subjects (566 observations) were enrolled in the Omni Cohort. Employed observations accounted for 58.9% of the total. The surrounding 10 towns add an additional 39 neighborhoods, increasing the total sample size to 62 neighborhoods, 3,609 (3,234 in the Offspring Cohort) distinct subjects and 14,804 (14,010) observations.

Individual's had truncated records if they died, dropped out of the FHS, or permanently moved out of the area. A few had intermittent observations due to returning to the area after previously having left. Because it is unlikely that an individual moved due to their own or a family member's BMI, it is unlikely that any unmeasured factors associated with weight-change were also associated with that individual missing an exam. Therefore, the missing-at-random assumption^{15,16,17} likely holds implying that growth curve models (e.g., random-intercept, random-slope models) suffice to account for both dropout and intermittent unit non-response. Only 10 people dropped-out of the FHS, a single observation had missing covariates in the 4-town area data, and only 68 (< 0.5%) of observations had missing covariates in the 14-town area data (67 of which had missing workplace neighborhood poverty status). Because multiple imputation would have a trivial impact, we excluded observations with missing values yielding final sample sizes of 11,055 and 14,736 for the 4-town and 14-town areas, respectively.

3 | NOTATION AND MODEL

Let Y_{it} , X_{it} , R_{it} , and W_{it} denote the outcome (BMI), a vector of observed predictors, neighborhood of residence, and neighborhood of workplace for the i th ($i = 1, \dots, N$) individual at their t 'th FHS medical exam (measurement occasion referred to as "exam", $t = 1, \dots, T = 8$). The elements of X_{it} include the constant 1 for the intercept, control predictors, and the proximity to fast-food establishment variables for an individual's residence, workplace and commute to and from work (defined in Section 2). The availability of up to 8 exams per patient allows individual-specific affine departures from the population mean trend across time, which we model as random effects (random intercept, random slopes) to gain stability from shrinkage and to account for the fact that a few individuals had truncated or intermittent exam attendance. The number of observations per individual is not sufficient to allow individual specific random effects for nonlinear functions of t .

Because individual exams in which the individual is not employed yield undefined values of all terms involving workplace fast-food exposure or workplace neighborhood, we multiply the predictors involving workplace exposure and the workplace neighborhood random effects by Emp_{it} (=1 if employed, 0 otherwise) so that the corresponding effects are only

informed by observations of employed study members. This approach allows the predictors other than those involving workplace to be informed by all the data thereby ensuring greater statistical efficiency than an employed-only subgroup analysis. Therefore, the general form of the model for the food proximity analysis accounting for the various predictors, individual-specific random intercepts and slopes, and bivariate neighborhood clustering is given by:

$$Y_{it} = \beta^T X_{it} + \mu_{i0} + \mu_{i1}t + \theta_{R_{it}} + \gamma_{W_{it}} Emp_{it} + \epsilon_{it}. \quad (1)$$

where $\mu_i = (\mu_{i0}, \mu_{i1})^T \sim \text{Normal}_2(0, \text{ind})$, $\mathbf{n}_k = (\theta_k; \gamma_k)^T \sim \text{Normal}_2(0; \text{area})$ for $k = 1; \dots; K$ distinct areas, and $\epsilon_{it} | \epsilon_{i(t-1)} \sim \text{Normal}(\rho_e \epsilon_{i(t-1)}, \sigma_e^2(1 - \rho_e^2))$ for all i, t . The elements of X_{it} include an intercept, the binary indicator variables $I(t=2) \dots I(t=8)$ ($I(\text{event}) = 1$ if event is true and 0 otherwise), the other 9 control predictors described in Section 2, and the 6 proximity to fast-food establishment variables. There are 16 control variables: 15 single variables and the product of workplace neighborhood poverty status and Emp_{it} . Therefore, the linear predictor in (1) has the form.

$$\beta^T X_{it} = \beta_0 + \beta_{1:15}^T X_{0:15,it} + \beta_{16} WkPov_{it} Emp_{it} + (\beta_{17} Fem_{it} + \beta_{18} Male_{it}) DistHm_{it} + \{(\beta_{19} Fem_{it} + \beta_{20} Male_{it}) DistWk_{it} + (\beta_{21} Fem_{it} + \beta_{22} Male_{it}) NumDrv_{it}\} Emp_{it}, \quad (2)$$

where $WkPov$, Fem , $Male$, $DistHm$, $DistWk$, and $NumDrv$ denote workplace neighborhood poverty level, female, male, distance to the closest fast-food establishment from home, distance to the closest fast-food establishment from work, and the number of fast-food establishments passed on their likely commute to or from work.

The novel statistical feature of (1) is that R_{it} and W_{it} index random effects for census tract so that each individual-time observation depends on two neighborhood effects – one from a participant's residence and another from his/her workplace – that may be correlated if the same census tract is indexed. Although allowing for correlation between random effects for different classifications of the same units has been considered previously, this was only in the context of allowing time-varying random effects for the same neighborhood over time.¹³ Therefore, in this prior work each individual-time observation was linked to a single random effect for neighborhood, whereas under the model in (1) each observation is linked to residential and workplace neighborhoods (that may be the same). Hence, our scenario is structurally different. Furthermore, due to the involvement of Emp_{it} in (1), $\text{cov}(\theta_{R_{it}}, \gamma_{W_{it}})$ and $\text{var}(\gamma_{W_{it}})$ depend only on observations in which $Emp_{it} = 1$.

The variance and correlation parameters quantify the unexplained variation in BMI between neighborhoods (area), individuals (ind), and observations (σ_e^2 and ρ_e). For easy reference to terms specific to residential and workplace exposure, we denote the diagonal elements of

area as σ_{home}^2 and σ_{work}^2 and its correlation coefficient as ρ_{area} . Similarly, for easy reference to the variances of the individual main (intercept) and exam (slope) effects, the diagonal elements of Σ_{ind} are denoted σ_{int}^2 and σ_{slope}^2 slope and the associated correlation coefficient is denoted ρ_{ind} .

Individual heterogeneity is captured by a random intercept with variance σ_{int}^2 to account for differences in individuals' BMI at exam 1 and a random slope with variance σ_{slope}^2 to account for differences in individuals' average change in BMI over adjoining exams. The residual or error (serial) correlation ρ_e allows for correlation from one exam to the next beyond that captured in the individual-specific intercepts and slopes or the neighborhood effects. Under the AR(1) structure assumed here, ϵ_{it} is sufficient for predicting $\epsilon_{i(t+1)}$ and $\text{cor}(\epsilon_{it}, \epsilon_{i(t-k)}) = \rho_e^{|k|}$ is a decreasing function of $|k|$.

By the law of total variance, $\text{var}(\epsilon_{it}) = \sigma_e^2(1 - \rho_e^2) + \rho_e^2 \text{var}(\epsilon_{i(t-1)})$. Therefore, substituting the stationarity condition $\text{var}(\epsilon_{it}) = \text{var}(\epsilon_{i(t-1)})$ and solving for $\text{var}(\epsilon_{it})$, it follows that $\text{var}(\epsilon_{it}) = \sigma_e^2$ for $t = 2, \dots, T$. It is natural to further suppose that $\epsilon_{i1} \sim N(0, \sigma_e^2)$. Using analogous derivations, it follows that $\text{cov}(\epsilon_{it}, \epsilon_{i(t-1)}) = \sigma_e^2 \rho_e$, $\text{cor}(\epsilon_{it}, \epsilon_{i(t-1)}) = \rho_e$ and $\text{cor}(\epsilon_{it}, \epsilon_{i(t-k)}) = \rho_e^k$ for $0 < k < t$. See Chapter 7 of Hedeker and Gibbons¹⁸ for a general discussion of hierarchical models with autocorrelated errors.

3.1 | Limitation of current software

Although commonly-used statistical software packages may allow for multiple forms of clustering, they typically do not allow for correlations between the random effects of different clustering variables to be modeled as random variables. Allowing such a correlation is enabled through the bivariate (residential, workplace) neighborhood random effect in (1). To illustrate, suppose the identifying variables for individual and neighborhood (home, work) are personid and (homeid, workid), respectively. In addition, let observation time be the order of observations within individual, in this case from 1 to a maximum of 8 exams. Table 1 shows the SAS Proc Mixed, R lmer, and Stata code that most closely emulates the bivariate random area effect specification in (1). These are deficient because homeid and workid are distinct variables and each procedure assumes the random effects of the levels of distinct clustering variables are independent. Even a semi-programmable routine such as SAS Proc NLMixed fails to provide a means to overcome this problem; while general likelihood functions are permitted by NLMixed, its random statements do not allow cross-classified random effects related to two different clustering variables (e.g., residential and workplace neighborhood) to be correlated. Even if there was a means to implement Proc NLMixed, it relies on an analytical approximation (e.g., Gaussian quadrature) to account for the random-effects and requires excellent starting values to converge. As an alternative and more generalizable approach, we specify a Bayesian model and model estimation procedure.

3.2 | Prior Distribution

To complete a Bayesian model for (1), prior distributions must be specified for the remaining model parameters. Our prior specification assumes $p(\beta) \propto 1, \Sigma_{\text{ind}} \sim IW(a, A), \Sigma_{\text{area}} \sim IW(b, B), \sigma_e^2 \sim IG(\nu_1, \nu_2), \rho_e \sim U(-1, 1)$, where IW denotes inverse-Wishart and IG denotes inverse-gamma. The choice of at or improper priors for the location parameters is justified by the substantial size of the data set while the boundedness of the correlation coefficient makes a uniform distribution over $(-1, 1)$ natural to specify for the serial correlation parameter ρ_e . For all parameters other than ρ_e , the prior is conjugate in that the conditional posterior distributions have the same form as the prior, enabling the informativeness of the prior to be quantified in terms of numbers of observations (i.e., “a prior sample size”) (see Section A of Appendix). There has been extensive debate on appropriate prior distributions for variance parameters at higher levels than the observations.¹⁹ The debate emanates from concerns that highly diffuse distributions for a parameter with a restricted scale such as a variance parameter may end up being highly informative and lead to a posterior distribution with almost all of its mass concentrated in a spike near 0.²⁰ Under (1), the inverse-Wishart priors for Σ_{ind} and Σ_{area} are vulnerable to this concern. Notably, the standard non-hierarchical Jeffrey’s prior yields an improper posterior²⁰ and so should not be used. To specify the prior distribution we set $a = b = 4, A = 10^{-3}I$, and $B = 10^{-1}I$, where I denotes an identity matrix. The degrees-of-freedom $a = b = 4$ are the minimal value under which the mean of the bivariate Wishart prior distributions, $(a - 3)^{-1}A$ and $(b - 3)^{-1}B$ for $a, b > 3$, are defined. The matrix parameter B is 100-fold larger than A because the data contain substantially less information at the neighborhood-level such that a diffuse long-tailed prior might lead to model estimation problems. Because σ_e^2 is in the first-level of the model, an inverse-gamma prior with a large variance is unproblematic and so to minimize the impact of the prior we set $\nu_1 = \nu_2 = 10^{-3}$.

3.3 | Alternative Covariance Matrix Prior for Sensitivity Analysis

To assess the sensitivity of results to the priors for Σ_{ind} and Σ_{area} we also estimate (1) under the product-normal prior specification.^{21,22} The product-normal prior for Σ_{ind} is derived by specifying $\mu_{i0} \sim N(0, \tau_0^2), \mu_{i1} | \mu_{i0} \sim N(\phi \mu_{i0}, \tau_1^2), \phi \sim N(0, v_\phi^2)$ and $\tau_j^2 \sim IG(\omega_1, \omega_2), j = 0, 1$. We specify diffuse hyperpriors by setting $v_\phi = 10^6$ and $\omega_1 = \omega_2 = 10^{-3}$ (mean 1 and variance 1000). The implied prior for Σ_{ind} is derived from the expressions $\sigma_0^2 = \tau_0^2, \sigma_1^2 = \phi_{\text{ind}}^2 \tau_0^2 + \tau_1^2$ and

$\rho_{\text{ind}} = \phi_{\text{ind}} | \tau_0 | / (\phi_{\text{ind}}^2 \tau_0^2 + \tau_1^2)^{1/2}$. An advantage of the product-normal over the inverse-Wishart prior distribution for Σ_{ind} is that varying levels of prior precision can be imposed across ϕ_{ind} and $\tau_j^2, j = 0, 1$. An analogous specification applies to Σ_{area} .

4 | ESTIMATION

To estimate statistical models such as (1) we wrote JAGS code and called it from R (RJAGS”). To determine the number of iterations required to achieve a given level of precision and to bound the simulation-error below a given tolerance, we used the suite of diagnostic measures and tests available in the CODA (Convergence Diagnostics and Output

Delivery) R package²³ and plotted trace plots and history plots of the Markov-chain Monte Carlo (MCMC) draws. We found that running the code for a burn-in phase of 10,000 iterations followed by a further 50,000 draws allowed inferences to have the desired numerical accuracy. For our final calculations in both the applied analysis and in the simulations we used a burn-in of 20,000 draws and a further 50,000 draws from three different starting points and random number seeds. (See Figure B1 in Section B of the Appendix for a trace plot of the post burn-in draws of ρ_{area} from the three chains as well as results from the calculation of the Gelman-Rubin potential scale reduction factor, which confirm that the chain-lengths for both the burn-in and inferential phases of the chain were sufficient.) Instructions for accessing the code and synthetic data that allows our analyses to be replicated are in Section C.

For the simulations used to determine the operating characteristics of the Bayesian posterior mean estimator and the associated credible interval, the same MCMC model fitting procedure used for the actual data may be iterated for every randomly generated simulated dataset (Section 5). However, a more efficient computation may be performed by implementing a MCMC procedure manually in a lower level programming language such as C (although for the simulations herein we found that performing parallel computations in R sufficed). In Section A of the Appendix we derive the conditional posterior distribution for all parameters other than ρ_e , as its conditional posterior is not available in closed-form. To update ρ_e we instead derive a novel Metropolis-Hastings^{24,25,26} accept-reject independence sampler²⁷ updating step involving a candidate generating distribution for ρ_e that closely approximates the true conditional posterior of ρ_e . The approximation is obtained by substituting a conditional method-of-moments estimator of ρ_e for ρ_e in the variance of the distribution of $\epsilon_{it} | \epsilon_{i,t-1}$ (see Section A.7 of the Appendix). The resulting independence sampler draws candidate values of ρ_e with high Metropolis-Hastings (M-H) acceptance probabilities – in excess of 90% for the food proximity data – implying that new parameter values are drawn almost as frequently and thus that the updating step is almost as efficient as for a Gibbs (sampler) step. Whereas M-H steps with random walk proposal densities have an optimal acceptance rate for a univariate parameter of around 44%,²⁸ draws under an independence sampler are conditionally independent of prior values and so the greater the acceptance rate the faster the convergence of the MCMC chain. In this regard, independence-sampler M-H steps resemble an importance sampling procedure.²⁰

5 | OPERATING CHARACTERISTICS OF BAYESIAN ESTIMATORS

We performed a simulation study to evaluate the operating characteristics of the Bayesian estimation procedure under various conditions. Because Bayesian estimators are not designed to adhere to traditional frequentist operating characteristics such as 0 bias of point estimators and nominal coverage of interval estimators, we use these as points of comparison as opposed to optimality criteria. In each set of simulations we vary one model parameter at a time, the “focal parameter,” while fixing all other features of the model so that an operating characteristic profile can be determined for the focal parameter. The values of the non-focal parameters used in generating simulated data were based on the posterior means obtained from the analysis of the FHS food proximity 4-town area data (see Section 6, “primary analysis”) under the model specified in Section 3. In all simulations, X_t provides

the covariate values and the prior distributions were as specified in Section 3.2, including the use of inverse-Wishart priors for the covariance matrices. Under these specifications the simulated data emulates the actual data and so the operating characteristics of the Bayesian estimators are directly relevant to the analysis of the food proximity data. Because equivalent results obtain at other values of the model parameters and predictors, there is minimal loss of generality under these specifications. For each set of parameter values, the random effects followed by the outcome variables were each randomly generated multiple times to form 1,000 simulated data sets. The closeness of the fitted posterior distributions to the true value of the model parameters are summarized in terms of the bias and mean squared error (MSE) of the parameters' posterior means and the coverage of their equal-tailed 95% credible intervals whose end-points are the 2.5% and 97.5% quantiles of the empirical distribution of each parameter's posterior draws.

Because the operating characteristics were similar for each element of β , the operating characteristics for β were summarized in terms of the average absolute bias,

$$\text{AABias} = p^{-1} \sum_{k=1}^p |\text{bias}(\beta_k)|$$

average MSE and average coverage across the elements of β .

Across the values of ρ_{area} and ρ_e used to generate data, AABias was small (0.001 to 0.003), average MSE was just below 0.06 and average coverage ranged from 0.943 to 0.951 (first rows of results in Table 2). Thus, as expected, the posterior mean estimator of β has excellent frequentist properties. Furthermore, when the covariance structure (bivariate clustering by neighborhood; random-intercepts and random-slopes for individuals; serially correlated errors) was incorrectly specified, the average MSE of the posterior mean of β was slightly larger (results not reported) than when it was correctly specified – a result consistent with the property of weighted linear regression that the most precise estimates of the regression parameters coincide with correct specification of the covariance structure.

The remainder of the discussion in this section focuses on the variance and correlation parameters with particular focus on the ability of the model to reproduce the true value of ρ_{area} . In the first series of simulations we consider $\rho_{\text{area}} \in \{0, \pm 0.2, \pm 0.4\}$ (five left-hand columns of results in Table 2). Across these values, the posterior mean of ρ_{area} exhibited a pattern of bias consistent with shrinkage towards the prior mean of 0 (Table 2). For example, bias was only -0.025 when $\rho_{\text{area}} = 0$ but 0.211 when $\rho_{\text{area}} = -0.4$ and -0.269 when $\rho_{\text{area}} = 0.4$; MSE was higher at the extremes than when $\rho_{\text{area}} = 0$; and coverage was close to 1 when $\rho_{\text{area}} = 0$, as the prior for ρ_{area} reinforces the data, but attenuated the further ρ_{area} was from 0, as the data is incongruent with the prior (Figure 1). These results suggest that the data contain a modest amount of information about ρ_{area} , allowing the prior to pull the posterior distribution of ρ_{area} towards 0. The operating characteristics of all other parameters were largely invariant to ρ_{area} and were close to the nominal levels (small bias, coverage close to 0.95).

Because ρ_e is the lone parameter for which a closed-form conditional posterior is not available, a second series of simulations with $\rho_e \in \{0, \pm 0.15, \pm 0.30, \pm 0.45\}$ was performed and to confirm the adequacy of the computational procedure to investigate the relationship of the operating characteristics to ρ_e . The posterior means of σ_e^2 and ρ_e have sufficiently small bias and coverage near enough to 0.95 (Figure 1 and the seven right-side columns of

results in Table 2) to confirm the adequacy of the computational procedure. However, there are some interesting observations. Firstly, as ρ_e increases in value beyond 0, the bias and MSE of both its own value and, to a greater extent, that of σ_e^2 increase. This phenomena is a consequence of the improved ability to distinguish the various model parameters when $\rho_e = 0$, resulting in more symmetric and tighter posterior distributions that lead to close to 0 bias and minimal MSE, than when $\rho_e > 0$, and the posterior correlation between σ_e^2 and ρ_e exceeding 0.90 when $\rho_e > 0.30$, resulting in the combined finding of small positive bias of ρ_e and much greater positive bias (and MSE) of σ_e^2 . Second, the coverage of ρ_e appears slightly greater at extreme values of ρ_e (i.e., $\rho_e > 0.45$) than when $\rho_e = 0$. We attribute this latter observation to the Bayesian credible intervals typically being asymmetric when ρ_e is far from 0, which results in elevated coverage, compared to when the intervals are symmetric, as typically occurs when

Crucially, the posterior mean estimates of σ_{home}^2 , σ_{work}^2 and ρ_{area} are largely invariant to ρ_e . However, because the true value of ρ_{area} for these simulations was -0.051 (close to the prior distribution mean) and the prior for ρ_{area} supports shrinkage towards its mean of 0, it is not surprising that the posterior 95% credible interval for ρ_{area} had an estimated coverage close to 1. The coverage levels of the posterior means of σ_{home}^2 and σ_{work}^2 were, however, close to 0.95.

It is clear from the results in this section that all variance and correlation parameters are well identified and yield intuitive results across the parameter space. The fact that the limited information in the data about ρ_{area} results in shrinkage towards 0 is appealing due to the desirable properties of shrinkage estimators and the use of priors for ρ_{area} with means of 0. In general, the simulations confirm that the posterior mean point estimators and associated credible intervals of all model parameters in the BMI – fast-food proximity study will be statistically valid as long as the assumed model is correctly specified.

6 | RESULTS. PROXIMITY TO FAST-FOOD STUDY

The results presented in this section first report the results for the scenario referred to as the “primary analysis” in which the 4-town area data is analyzed using the full model with inverse-Wishart prior distributions for the covariance matrix parameters of the individual and neighborhood random effects. At the end of this section, we evaluate the sensitivity of the results to the use of a different prior distribution for the covariance matrix parameters. In Section 6.2 the sensitivity of the results to four different simplifications of the data or the model are considered to explore the impact of ignoring features of either. The regression coefficients of all models represent the change in BMI associated with a one-unit change in the corresponding predictor; 95% equal-tailed credible intervals accompany the results and the posterior probability that the coefficient exceeds 0 is also reported.

6.1 | Primary Analysis

The results of the primary analysis appear in the left-hand-side of Table 3. These suggest that the further women live from their closest fast-food establishment, the lower their

expected BMI (estimate (95% equal-tailed credible interval) of $\beta_{17} = -0.105(-0.228, 0.019)$) with the posterior probability $p_{\beta_{17},0} = \text{pr}(\beta_{17} > 0 \mid \text{data}) = 0.048$ while the association of workplace proximity is much lower ($\beta_{19} = -0.058(-0.201, 0.086)$), $p_{\beta_{19},0} = 0.215$. In contrast, the workplace proximity association for men is stronger, $\beta_{20} = -0.142(-0.313, 0.028)$, $p_{\beta_{20},0} = 0.050$ than their residential proximity association ($\beta_{18} = 0.040(-0.103, 0.184)$), $p_{\beta_{18},0} = 0.711$. The number of fast-food establishments that women pass on their commutes has a modest association with greater BMI ($\beta_{21} = 0.025(-0.000, 0.051)$), $p_{\beta_{21},0} = 0.975$.

Therefore, there are three modestly strong findings. closer female residential proximity and closer male workplace proximity are associated with greater BMI while greater female commuting exposure (more fast-food establishments passed) is associated with greater BMI.

The regression coefficients for the remaining predictors cohere with established findings for BMI. For example, average BMI has increased over time, is greater for males and individuals born earlier but lower for smokers. BMI is greater for married individuals and there is an indication that neighborhood poverty is associated with the BMI of residents but not of workers. Most of the posterior distributions are symmetric in shape (Figure 2). There is no discernible pattern or any hint of model lack-of-fit in residual analysis performed for the fitted values and each predictors, suggesting that the model fits the data well. In addition, the posterior mean of the R-square (R^2) statistic was 0.872, illustrating the predictability within-individuals of BMI across time.

The results for the random part of the model (see the bottom three segments of Table 3) suggest that the amount of unexplained variation between neighborhoods (area segment of Table 3) is of comparable numerical magnitude for home ($\sigma_{\text{home}}^2 : 0.122(0.058, 0.246)$) and workplace ($\sigma_{\text{work}}^2 : 0.100(0.049, 0.195)$). The estimated correlation between the neighborhood latent variables is negative but small ($\rho_{\text{area}} : -0.051$), $p_{\rho_{\text{area}},0} = \text{pr}(\rho_{\text{area}} > 0 \mid \text{data}) = 0.128$. An initial explanation of the near 0 correlation was that it was an artifact of neighborhoods with high workplace exposure having few residents while those with high residential exposure have few workers in them. To explore this theory we tabulated the occurrence of each (home, work) neighborhood combination. The sum of the diagonal elements (i.e., where residential and workplace are in the same neighborhood) of the resulting two-way table is 5,534 (50.1% of the total number of observations) implying that individuals often reside and are employed at a location within the same neighborhood. However, 81.2% of the person-exam observations with concordant residential and workplace neighborhoods are for individuals not employed (e.g., unemployed, student, retired, housewife, househusband); recall that 58.9% of observations were for employed individuals. After subtraction of the number of individuals in the sample who live and work in the same neighborhood or who are unemployed, the row and column margins are strictly positive confirming that each neighborhood includes individuals who reside and work in them, reside only, and work only (Figure 3). The number of different workplace census tracts observed for each residential census tract ranged from 3 to 20 and had a mean of 13.8. Conversely, the number of different residential census tracts observed for each workplace census tract ranged from 1 to 20 (same

mean of 13.8 by construction). These descriptive analyses provide reassurance that there exists sufficient mixing of individuals' residential and workplace neighborhoods to robustly identify ρ_{area} from the data.

The variances of the individual random intercepts and slopes (σ_{area} segment of Table 3) are substantial ($\sigma_{\text{int}}^2 : 14.29(13.10, 15.52)$, $\sigma_{\text{slope}}^2 : 0.393(0.341, 0.448)$), revealing substantial heterogeneity in individuals' BMI cross-sectionally and over time, and minimally correlated ($\rho_{\text{ind}} : 0.001(-0.078, 0.084)$) $p_{\rho_{\text{ind}}, 0} = 0.502$. Individuals with initial high BMI experience more modest increases over time, reflecting some regression to the mean.

The posterior mean residual variance ($\sigma_e^2 : 3.132(2.864, 3.452)$) is small compared to the variation between individuals implying that there is substantially more unexplained variation between than within individuals (bottom segment of Table 3). The serial (AR(1)) correlation parameter is moderately large ($\rho_e : 0.393(0.341, 0.448)$), $p_{\rho_e, 0} = 1.000$ revealing that the impact of an idiosyncratic increase in BMI at one exam lingers quite substantially at the following exam.

6.1.1 | Sensitivity of results to prior for σ_{area} and σ_{ind} —To study the impact of the prior, the model was re-estimated using the product-normal prior specified in Section 3.3. The differences in the results were minimal except for the posterior inferences about σ_{area} . The posterior means of both σ_{home}^2 and σ_{work}^2 are smaller. Although the change is small in absolute terms, the posterior means of σ_{home}^2 and σ_{work}^2 are 50% to 40% smaller, respectively, under the product-normal prior than under the inverse-Wishart prior (right-side of Table 3). In contrast, the posterior mean of σ_{int}^2 is slightly higher under the product-normal prior illustrating that the different priors impact the relative variation at each level of the model more than they impact the sum of the variances at the area, individual and observation levels. A consequence of the lower posterior variances is that $E[\rho_{\text{area}} | \text{data}]$ is further from 0 (-0.152 versus -0.051) but it still has a wide 95% credible region.

It is not surprising that estimates of σ_{area} were noticeably dependent on the prior as there are only 23 neighborhoods. Because the inverse-Wishart is the conjugate prior, the prior parameter b functions like a prior sample-size. When $b = 4$ the information in the prior is about 17.3% that of the information in the data. Confirmation of this point was demonstrated by setting $b = 13$, which reduced the prior means of σ_{home}^2 and σ_{work}^2 from 0.1 to 0.01 and resulted in substantially smaller posterior means as the prior had about 56.5% as much influence on the posterior as the data. In contrast, setting $a = 13$ had virtually no impact on the posterior for σ_{ind} . We favor the results for σ_{home}^2 and σ_{work}^2 obtained with the inverse-Wishart ($b = 4$) prior for the primary analysis as this specification allowed the data to most overwhelm the prior.

6.2 | Impact on results of key elements of data and covariance structure

Five analyses designed to allow the impact of using two alternative specifications of the data and of removing three key elements of the statistical model are reported in Table 4 and Table 5, respectively. In the former, the 14-town area and the restricted form of the 4-town area data in which the first observed residential and workplace neighborhood is carried-forward as if there was no movement of individuals between neighborhoods are analyzed. In the latter, three constraints are imposed on the model: statistically independent (residential, workplace) neighborhood effects ($\text{cov.}(\theta_k, \gamma_k) = 0$ for $k = 1, \dots, K$); parallel linear individual growth curves ($\mu_{i1} = 0$ for $i = 1, \dots, N$); no serial correlation between the residuals of an individual ($\rho_e = 0$). The results of these analyses are compared to those of the primary analysis in Section 6.1.

6.2.1 | Impact of different data specifications

Analysis of 14-town area data: In general, the results for the 14-town area are similar to those for the 4-town area (middle columns versus left-most columns of Table 4). Although the negative residential proximity association for females strengthens ($\beta_{17} = -0.094(-0.171, -0.015)$), $p_{\beta_{17}, 0} = 0.009$ in statistical significance, the estimated workplace proximity association for males and the commuting exposure association for females are smaller than for the 4-town area analysis. The posterior variances of σ_{home}^2 and σ_{work}^2 are smaller in the 14-town analysis, reflecting that the tripling of the number of neighborhoods leads to tighter posterior distributions. The posterior means of the other random effect parameters change minimally between the 4- and 14-town analyses.

Ignoring movement of individuals between neighborhoods: When individuals' neighborhoods were fixed at those for their first exam and carried-forward, the posterior mean of σ_{home}^2 increases substantially to 0.220(0.084, 0.497) while that of σ_{work}^2 increases minimally (right-most columns of Table 4). This result for σ_{home}^2 was surprising as the misclassification of neighborhood in later exams that occurs from carrying forward the neighborhood of the first exam was expected to reduce the unobserved heterogeneity between neighborhoods. An explanation is that the posterior mean of ρ_{area} of -0.191 is further from 0 under this new specification, which inflates the posterior means of σ_{home}^2 and σ_{work}^2 while $\text{var.}(\theta_k + \gamma_k)$ remains close to constant. Therefore, not accounting for time-varying neighborhood membership induces noisy, possibly biased estimates of $\sigma_{\text{home}}^2, \sigma_{\text{work}}^2$ and ρ_{area} although there was almost no change in the posterior mean of R^2 from that of the primary analysis reflecting the change in these parameters is an artifact of a different data set being analyzed as opposed to indicating that a different model is needed.

6.2.2 | Impact of removing features from the model

No correlation between residential and workplace neighborhoods: Under the restriction $\text{cov}(\theta_k, \gamma_k) = 0$, $k = 1, \dots, K$, the estimates of σ_{home}^2 and σ_{work}^2 decrease in magnitude (first versus second set of three columns of Table 5). This result reflects that $\text{var.}(\theta_k + \gamma_k)$

decreases the more $\rho_{\text{area}} < 0$. Hence, setting $\rho_{\text{area}} = 0$ leads to smaller estimates of σ_{home}^2 and σ_{work}^2 and narrower credible regions especially for area-level parameters although there was no change in the posterior mean of R^2 from that of the primary analysis.

No random slope: Setting $\mu_{ji} = 0$, $i = 1, \dots, N$ leads to increases in σ_{int}^2 , σ_e^2 and in particular ρ_e (third set of three columns of Table 5). In addition, the posterior means of β_{17} and β_{21} , the coefficients of the residential and commuting proximity terms for females increased. Thus, individual-specific trends appear to capture some of the association that would otherwise be attributed to the observed proximity measures, suggesting that the individual by time interaction represented by the individual random slopes confounds the food proximity effects if excluded from the model. The absence of the random-slope reduced the posterior mean of R^2 from 0.872 to 0.580, a massive reduction, implying that there is substantial heterogeneity in the slope of the linear trend of individuals' BMI trajectories across time. Ignoring such heterogeneity inflates ρ_e to very high values (posterior mean of 0.803 with 95% credible region .0.754, 0.904)) implying that other terms cannot overcome the exclusion of the random slopes.

No residual serial correlation: Setting $\rho_e = 0$ leads to an increase in the posterior variances of σ_{slope}^2 and ρ_{ind} (right-most three columns of Table 5). However, setting $\rho_e = 0$ results in the posterior mean of σ_e^2 reducing in magnitude due to the positive posterior correlation of these parameters. These results again illustrates that the exclusion of one of the key features of the random part of the model leads to a compensatory increase in the prominence of the remaining terms, although the reduced model does not fit the data nearly as well.

7 | DISCUSSION

Identifying relationships between contextual factors (e.g., the food environment) and individual health characteristics (e.g., BMI) requires accounting for the complex dynamic interplay of geography with individual characteristics. In this paper we developed a novel Bayesian model in which each neighborhood had a dual-purpose (residential, workplace) bivariate latent effect and conducted a thorough analysis of the motivating FHS-food proximity data. Developing the model and estimation procedure to estimate ρ_{area} is an important contribution as it summarizes the extent to which neighborhoods present overlapping exposure risks versus context-specific (residential versus workplace) risks. To our knowledge, a bivariate neighborhood random effect has not previously been used to model neighborhood effects. The methodology developed herein has broad potential appeal as it may be applied to any situation in which an observation is grouped or clustered multiple ways by the same unit or clustering variable. It is possible that this data structure has been previously encountered but investigators in the past did not recognize the subtle conceptual point that multivariate clustering is present and so did not capture the complexity of the data in their analyses.

Although likelihood-based procedures for hierarchical linear models generally yield consistent estimators, the complex array of correlation parameters at multiple levels of the

model motivated simulation studies to assess the operating characteristics of the Bayesian estimators, including whether there was sufficient information in the data to identify all of the model parameters. The ensemble of simulation results confirm that the Bayesian model and the estimation procedure yield accurate results for the FHS food proximity data. Although the posterior mean estimator of ρ_{area} was biased towards the prior mean of 0 when the true value differed from 0, this is understandable and desirable given the limited amount of information across the 23 areas and the fact that shrinkage estimators are more robust and have lower mean squared errors.

In the proximity to fast-food establishment analysis, the impact of accounting for the correlation between residential and workplace effects was minor as the posterior distribution of ρ_{area} substantially overlapped 0. However, allowing the uncertainty in $\rho_{\text{area}} = \text{cor}(\theta_k, \gamma_k)$ to propagate through the analysis is a more desirable statistical practice than fixing $\rho_{\text{area}} = 0$. Accounting for the uncertainty in ρ_{area} across the model components is an important contribution in-and-of-itself as it allows the statistical significance of all results based on the model to be correctly calibrated, even if the true correlation is close to 0.

We found that proximity to fast-food establishments from home had a positive association with the BMI of females; conversely, workplace proximity had a positive association with BMI of males but not females. Furthermore, the workplace neighborhood appears to at least have a similar overall effect as the residential neighborhood with minimal correlation between the risk of a neighborhood to its residents and its risk employees. These findings are novel for workplace proximity. Very few prior studies have examined this association. In a cross-sectional study of over 5,000 adults living in Cambridgeshire, United Kingdom, participants with the highest quartiles of exposure to fast food or take-away restaurants at work, home and on commutes had higher BMI of 0.99, 0.92, and 0.65 compared to those in the least exposed quartile.²⁹ An important distinction of this study to ours is that it is cross-sectional whereas ours is longitudinal and so allows more rigorous control for spurious effects.

Many more studies have examined residential exposure to food establishments with BMI; however, most of these studies have been rated as low quality, in large part because of cross-sectional designs. In a systematic review of 71 studies, most found no relationship between proximity to food establishments and body weight.³⁰ The best rated studies among these had mixed results. For example, a longitudinal study of 13,000 adults from 1999 to 2005 found that BMI was slightly higher for women exposed to more fast food restaurants and convenience stores; low-income women had slightly lower BMIs when they lived closer to supermarkets.³¹ More recent studies examining the introduction of supermarkets in low access, low income areas (food deserts) as natural experiments have been disappointing. These studies in Pittsburgh, New York City, and Philadelphia have not found any association with BMI and there has been limited to no increases in healthy diets after the opening of supermarkets.^{32,33,34}

Overall, these prior studies and our current study show effects that are quite small, even when statistically significant, and perhaps are not consequential. When effects are present, results support the historical theory that residential proximity to fast food matters more for

the BMI of females than males.^{8,35,31} In contrast, based on our findings in this study, workplace proximity to fast-food establishments might matter more for the BMI of males. More research needs to be done to confirm these conclusions, including using better measures of food establishment exposure, perhaps through use of mobile devices.³⁶

It may appear surprising that the correlation between the neighborhood random effects on residents and workers of the same neighborhood should be 0 (or negative). We conjecture that this reflects a reliance on distinct fast-food outlets or the same outlets at different times by residents and employees of the same neighborhood. Although there is some concern of confounding between the distance measures and neighborhood, the fact that the distance measures vary within neighborhood alleviates this concern. The exposure results were robust to different prior specifications, different study population specifications, and to the impact of excluding terms from the random-effect component of the model.

Because the statistical model derived and applied to the food proximity analyses has multiple components, there are a multitude of alternative specifications all of which are worthy of further study. Except in the aspects of the model that are central to this paper, we have used standard assumptions such as the random effects being normally distributed. In the context of BMI, the assumption of normal distributions appear to fit the data well. We examined using log-normally distributed random effect distributions but found that model fit was compromised. However, in other applications it may be beneficial to allow the distributions of the random effects to be skewed or to have semi-parametric forms, such as under Dirichlet process priors.

In conclusion, neighborhood effects have important implications for policy-makers. The development of a model with a bivariate neighborhood effect has the potential to advance the detection, measurement, and impact of neighborhood effects. Such advances may ultimately lead to improved neighborhood-level interventions to reduce obesity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Michael Seward of Department of Population Medicine at the Harvard Pilgrim Health Care Institute for administrative work and proof-reading. The research in this paper was supported by a grant from the NIH National Heart, Lung, and Blood Institute (R01 HL109263–01A1, PI. O'Malley, Subramanian), a mentored career development award from the same division of NIH (K23 HL111211, PI. Block), and the Harvard NHLBI Cardiovascular Epidemiology Training Grant (T32 HL 098048, NIH National Cancer Institute Grant K99 CA201542).

Financial disclosure

None reported.

APPENDIX

A SELF-DEVELOPED MCMC PROCEDURE. CONDITIONAL POSTERIOR DISTRIBUTIONS

Although we were able to estimate the model in JAGS (and WinBUGS), deriving the conditional posterior distributions provides important insights into the manner by which the data informs the model parameters. The availability of the conditional posterior distributions also allow the calculations performed in JAGS to be checked and the computational time for the simulation to be dramatically reduced. The following describes our hybrid Gibbs sampler^{37,38} – Metropolis-Hastings^{24,25,26} Markov-chain Monte-Carlo (MCMC) algorithm.

A.1 Overview and novel feature of MCMC procedure

The prior distributions for ϵ_{ind} and ϵ_{area} have conjugate forms in all analyses, allowing the conditional posterior distribution of every parameter other than ρ_e to be evaluated in closed form. The marginal distribution for ϵ_{it} combined with the AR(1) structure on ϵ_{it} , $t = 2, \dots, T = 8$, inhibits closed-form determination of the conditional posterior for ρ_e . However, we derive a close approximation to the conditional posterior distribution for ρ_e by substituting ρ_e with its conditional method-of-moments estimator in the expression for the variance and then use the resulting closed-form approximation to accept or reject the randomly generated candidate values. The resulting independence sampler²⁷ draws candidate values of ρ_e with high Metropolis-Hastings (M-H) acceptance probabilities. For the analysis of the food proximity data, the M-H acceptance rate for ρ_e under (1) exceeds 90% implying that new parameter values are drawn almost as frequently as for a Gibbs (sampler) step. Unlike Gibbs steps, which are random walks and thus yield serially correlated draws, the accept-reject MCMC step yields new values of ρ_e that are independent of prior values, improving the efficiency of the computation.

A.2 Notation and set-up

Recall from Section 3 that Y_{it} , X_{it} , R_{it} , and W_{it} denote the outcome (BMI), a vector of fixed-effect predictors, neighborhood of residence, and neighborhood of workplace for the i th ($i = 1, \dots, N$) individual at their t th medical exam (measurement occasion referred to as “exam”, $t = 1, \dots, T = 8$). The general form of the model for the proximity to food analysis is given in (1) and unless otherwise stated, the following prior distributions are assumed.

$$p(\beta) \propto 1, \Sigma_{ind} \sim IW(a, A), \Sigma_{area} \sim IW(b, B), \sigma_e^2 \sim IG(\nu_1, \nu_2), \text{ and } \rho_e \sim U(-1, 1).$$

Let $S_i \subset \{1, T = 8\}$ for $i = 1, \dots, N$ denote the set of exams individual i attended, $S_r = \{i: \sum_{t=1}^T I(R_{it} = r) > 0\}$ for $r = 1, \dots, K$ denote the set of individuals who reside in neighborhood r at some point, and $S_w = \{i: \sum_{t=1}^T I(W_{it} = w) > 0\}$ for $w = 1, \dots, K$ denote the set of individuals who work in neighborhood w at some point. The cardinality of S_i , S_r , and S_w are denoted n_i , n_r , and n_w , respectively. Then let X_i denote the $n_i \times p$, $p = \dim(X_{it})$, matrix of predictors for the observations on individual i (the k 'th row of X_i contains the observations from individual i 's k 'th exam, which differs from study exam k if they missed a prior exam), $Z_{0i} = (j_i, e_i)$ be the $n_i \times 2$ matrix in which j_i is a vector of ones of length n_i and e_i

is the vector counterpart of S_i . Also let $\mathbf{Z}_{1i}$ and $\mathbf{Z}_{2i}$ be the n_i by K matrices containing 1 in element (t, k) if $R_{it} = k$ and $W_{it} = k$, respectively. The vectorized form of the model is then

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_{0i} \boldsymbol{\mu}_i + \mathbf{Z}_{1i} \boldsymbol{\theta} + \mathbf{Z}_{2i} \boldsymbol{\gamma} + \boldsymbol{\epsilon}_i, \quad (\text{A1})$$

where $\boldsymbol{\epsilon}_i \sim \text{Normal}(0, \sigma_e^2 \mathbf{V}_i(\rho_e))$, $V(\rho_e)_{hj} = \rho_e^{lh-jl}$ and $\mathbf{V}_i = \mathbf{V}_i(\rho_e)$ is the sub-matrix of $\mathbf{V}(\rho_e)$ corresponding to S_i .

For separate updates of $\boldsymbol{\theta}_k$ and $\boldsymbol{\gamma}_k$, the conditional priors $p(\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k)$ and $p(\boldsymbol{\gamma}_k | \boldsymbol{\theta}_k)$ are needed. It follows from results for normal distributions that $\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k \sim N(\lambda_{\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k}, \tau_{\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k}^2)$ and

$\boldsymbol{\gamma}_k | \boldsymbol{\theta}_k \sim N(\lambda_{\boldsymbol{\gamma}_k | \boldsymbol{\theta}_k}, \tau_{\boldsymbol{\gamma}_k | \boldsymbol{\theta}_k}^2)$, where $\lambda_{\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k} = (\rho_{\text{area}} \sigma_{\text{home}} / \sigma_{\text{work}}) \boldsymbol{\gamma}_k$, $\tau_{\boldsymbol{\theta}_k | \boldsymbol{\gamma}_k}^2 = \sigma_{\text{home}}^2 - \rho_{\text{area}} \sigma_{\text{home}} / \sigma_{\text{work}}$, $\lambda_{\boldsymbol{\gamma}_k | \boldsymbol{\theta}_k} = (\rho_{\text{area}} \sigma_{\text{work}} / \sigma_{\text{home}}) \boldsymbol{\theta}_k$, and $\tau_{\boldsymbol{\gamma}_k | \boldsymbol{\theta}_k}^2 = \sigma_{\text{work}}^2 - \rho_{\text{area}} \sigma_{\text{work}} / \sigma_{\text{home}}$.

A.3 Conditional Posterior of $\boldsymbol{\beta}$

Let $\tilde{\mathbf{Y}}_i^{(1)} = \mathbf{Y}_i - \mathbf{Z}_{0i} \boldsymbol{\mu}_i - \mathbf{Z}_{1i} \boldsymbol{\theta} - \mathbf{Z}_{2i} \boldsymbol{\gamma}$. Under the at non-informative prior specified above, the conditional posterior of $\boldsymbol{\beta}$ given the data $\mathbf{D} = (\mathbf{Y}, \mathbf{X}, \mathbf{R}, \mathbf{W})_{i=1..N, t=1..8}$ and the other parameters is

$$(\boldsymbol{\beta} | \cdot) \sim \text{Normal}_p \left(\hat{\boldsymbol{\beta}}, \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \right),$$

where $\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \tilde{\mathbf{Y}}_i^{(1)}$ is the weighted least squares estimator.

A.4 Conditional Posterior of $\boldsymbol{\mu}_i$

Let $\tilde{\mathbf{Y}}_i^{(2)} = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_{1i} \boldsymbol{\theta} - \mathbf{Z}_{2i} \boldsymbol{\gamma}$. Then the conditional posterior of $\boldsymbol{\mu}_i$ is

$$(\boldsymbol{\mu}_i | \cdot) \sim \text{Normal}_2 \left(\left(\mathbf{Z}_{0i}^T \mathbf{V}_i^{-1} \mathbf{Z}_{0i} + \boldsymbol{\Sigma}_{\text{ind}}^{-1} \right)^{-1} \mathbf{Z}_{0i}^T \mathbf{V}_i^{-1} \tilde{\mathbf{Y}}_i^{(2)}, \left(\mathbf{Z}_{0i}^T \mathbf{V}_i^{-1} \mathbf{Z}_{0i} + \boldsymbol{\Sigma}_{\text{ind}}^{-1} \right)^{-1} \right).$$

A.5 Conditional Posterior of $\boldsymbol{\theta}_r$

Let $\tilde{\mathbf{Y}}_i^{(3)} = \mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\beta} - \mathbf{Z}_{0i} \boldsymbol{\mu}_i - \mathbf{Z}_{2i} \boldsymbol{\gamma}$. In addition, let $\mathbf{j}_{ir}$ and $\mathbf{V}_{ir}$ be the vector of ones and the sub-matrix of $\mathbf{V}(\rho)$ corresponding to the elements of $S_{ir} = \{t : R_{it} = r\}$, respectively. Then the conditional posterior of $\boldsymbol{\theta}_r$ given all observed data and all other parameters (including $\boldsymbol{\gamma}_r$) is.

$$(\boldsymbol{\theta}_r | \cdot) = \text{Normal}_2 \left(\hat{\boldsymbol{\theta}}_r, \left(\sum_{i \in S_r} \mathbf{j}_{ir}^T \mathbf{V}_{ir}^{-1} \mathbf{j}_{ir} + \tau_{\boldsymbol{\theta}_r | \boldsymbol{\gamma}_r}^{-2} \right)^{-1} \right),$$

where $S_r = \{i: \sum_{t=1}^T I(R_{it} = r) > 0\}$ and

$\hat{\theta}_r = \left(\sum_{i \in S_r} j_{ir}^T V_{ir}^{-1} j_{ir} + \tau_{\theta_r}^{-2} \right)^{-1} \left(\sum_{i \in S_r} j_{ir}^T V_{ir}^{-1} \tilde{Y}_i^{(3)} + \tau_{\theta_r}^{-2} \lambda_{\theta_r} \right)$. Thus, $\hat{\theta}_r$ is a weighted average of the adjusted outcomes attributed to neighborhood r based on workplace.

A.6 Conditional Posterior of γ_w

The conditional posterior of γ_w is a mirror image of that for $\tilde{Y}_i^{(4)} = Y_i - X_i \beta - Z_{0i} \mu_i - Z_{1i} \theta$. In addition, let j_{iw} and V_{iw} be the vector of ones and the sub-matrix of $V(\rho)$ corresponding to the elements of $S_{iw} = \{t: w_{it} = w\}$. Then the conditional posterior γ_w of given all observed data and other parameters (including θ_w) is:

$$(\gamma_w | \cdot) = \text{Normal}_2 \left(\hat{\gamma}_w, \left(\sum_{i \in S_w} j_{iw}^T V_{iw}^{-1} j_{iw} + \tau_{\gamma_w}^{-2} \right)^{-1} \right).$$

Where $S_w = \{i: \sum_{t=1}^T I(W_{it} = w) > 0\}$ and

$\hat{\gamma}_w = \left(\sum_{i \in S_w} \left(j_{iw}^T V_{iw}^{-1} j_{iw} + \tau_{\gamma_w}^{-2} \right)^{-1} \right) \left(\sum_{i \in S_w} j_{iw}^T V_{iw}^{-1} \tilde{Y}_i^{(4)} + \tau_{\gamma_w}^{-2} \lambda_{\gamma_w} \right)$. Thus, $\hat{\gamma}_w$ is a weighted average of the adjusted outcomes attributed to neighborhood w based on workplace.

A.7 Novel step. Conditional Posterior of ρ_e

The serial correlation ρ_e is a measure of dependence between the residuals, expressed in vectorized form as $\epsilon_t = X_t \beta - Z_{0t} \mu_t - Z_{1t} \theta - Z_{2t} \gamma$. In the case of a uniform $(-1, 1)$ prior for ρ_e , the conditional posterior distribution of ρ_e is proportional to the density of the joint distribution of ϵ_t truncated to the $(-1, 1)$ interval. Under the AR(1) model, the joint distribution of ϵ_t reduces to the product of the densities for $\epsilon_{i1}, (\epsilon_{i2} | \epsilon_{i1}), \dots, (\epsilon_{iT} | \epsilon_{i(T-1)})$ illustrating that ρ_e is informed by ϵ_{i1} in addition to the sample correlation between ϵ_{it} and $\epsilon_{i(t-1)}$ for $t > 1$. Because the marginal variance of the residual at $t = 1$ is given by $\text{var}(\epsilon_{i1}) = \sigma_e^2$ the conditional posterior of ρ_e is proportional to.

$$p(\rho_e | \cdot) \propto \text{Normal} \left(\rho_e; \hat{\rho}_e \left(\frac{\sum_{i,t=2:T} \epsilon_{it}^2}{\sum_{i,t=2:T} \epsilon_{i(t-1)}^2} \right)^{1/2}, \frac{\sigma_e^2 (1 - \rho_e^2)}{\sum_{i,t=2:T} \epsilon_{i(t-1)}^2} \right) I(-1, 1),$$

where $\hat{\rho}_e = \sum_{i,t=2:T} \epsilon_{it} \epsilon_{i(t-1)} / \left(\sum_{i,t=2:T} \epsilon_{i(t-1)}^2 \sum_{i,t=2:T} \epsilon_{it}^2 \right)^{1/2}$ is the Pearson correlation coefficient evaluated on $\{\epsilon_{i(t-1)}, \epsilon_{it}\}_{i=1, \dots, N; t=2, \dots, T}$.

Because $p(\rho_e | \cdot)$ is not available in closed-form, to draw values of ρ_e we use the Metropolis-Hastings algorithm with an independence candidate generating distribution based on the approximation of the true conditional posterior for $p(\rho_e | \cdot)$ given in step 3 below. Let $\rho_{e, \text{old}}$ denote the current value of ρ_e . The M-H step is then.

1. Generate the candidate value $\rho_{e,\text{new}} \sim \text{Normal}$

$$\rho_{e,\text{new}} \sim \text{Normal} \left(\left(\frac{\sum_{i,t=2:T} \epsilon_{it}^2}{\sum_{i,t=2:T} \epsilon_{i(t-1)}^2} \right)^{1/2} \hat{\rho}_e, \frac{\sigma_e^2 (1 - \hat{\rho}_e^2)}{\sum_{i,t=2:T} \epsilon_{i(t-1)}^2} \right)$$

2. Repeat step 1 until $\rho_{e,\text{new}} \in (-1, 1)$ is obtained.
3. Evaluate the acceptance probability

$$q = \left(\frac{1 - \rho_{e,\text{old}}^2}{1 - \rho_{e,\text{new}}^2} \right)^{1/2} \exp \left[- \left(\frac{1}{1 - \rho_{e,\text{new}}} - \frac{1}{1 - \hat{\rho}_e} \right) \left(\rho_{e,\text{new}} - \left(\frac{\sum_{i,t=2:T} \epsilon_{it}^2}{\sum_{i,t=2:T} \epsilon_{i(t-1)}^2} \right)^{1/2} \hat{\rho}_e \right)^2 + \left(\frac{1}{1 - \rho_{e,\text{old}}} - \frac{1}{1 - \hat{\rho}_e} \right) \left(\rho_{e,\text{old}} - \left(\frac{\sum_{i,t=2:T} \epsilon_{it}^2}{\sum_{i,t=2:T} \epsilon_{i(t-1)}^2} \right)^{1/2} \hat{\rho}_e \right)^2 \right]$$

4. If $q \geq 1$ accept $\rho_{e,\text{new}}$ as the new value of ρ_e . If $q < 1$ generate $u \sim U(0, 1)$ and accept $\rho_{e,\text{new}}$ as the new value of ρ_e if $u < q$; otherwise, retain $\rho_{e,\text{old}}$ as the value of ρ_e .

We obtain values of ϵ_{it} by substituting the current values of the model parameters to evaluate the vector $\epsilon_i = Y_i - X_i \beta - Z_0 \mu_i - Z_1 \theta - Z_2 \gamma$.

The rationale for the above procedure is that the true posterior distribution is closely approximated by the posterior generated from the Pearson correlation coefficient evaluated on $\{\epsilon_{i(t-1)}, \epsilon_{it}\}_{i=1,\dots,N; t=2,\dots,T}$. The quantity $(\sum_{i,t=2:T} \epsilon_{i(t-1)}^2) / (\sum_{i,t=1:(T-1)} \epsilon_{it}^2)$ should be very close to 1 as the numerator and denominator are summing over many of the same terms. Therefore, the candidate generating density for ρ_e is a distribution whose mean is close to $\hat{\rho}_e$ the Pearson correlation coefficient. Therefore, we use the Metropolis-Hastings independence sampler with candidate values of ρ_e drawn from the approximate conditional posterior distribution as performed in steps 1 and 2 of the above accept-reject algorithm²⁰ to update ρ_e . We anticipate that the acceptance probability, q , will be close to 1 in practical applications such as those considered in this paper.

A.8 Conditional Posterior of σ_e^2

As for the conditional posterior of ρ_e , to determine the conditional posterior of σ_e^2 we use the fact that the joint distribution of ϵ_i expands into the product of the marginal distribution of ϵ_{i1} and the conditional densities of $(\epsilon_{it} | \epsilon_{i(t-1)})$ for $i = 2, \dots, T$. Let $\delta_{it} = \epsilon_{it} - \rho_e \epsilon_{i(t-1)}$ and $N_{\text{tot}} = \sum_{i=1}^n n_i$. Then

$$(\sigma_e^2 | \cdot) = IG \left(\nu_1 + \frac{N_{\text{tot}}}{2}, \left(\frac{1}{\nu_2} + \frac{1}{2} \sum_i \epsilon_{i1}^2 + \frac{1}{2(1 - \rho_e^2)} \sum_{i,t=2} \delta_{it}^2 \right)^{-1} \right).$$

The middle term in the evaluation of the second parameter of the above inverse-gamma distribution obtains from the first residual for each subject whereas the final term aggregates squared difference, δ_{it}^2 of the current residual and its conditional mean given the prior residual. These both contribute information about σ_e^2 .

A.9 Conditional Posterior of $\mathbf{\Sigma}_{\text{ind}}$

The latent variables (or level II data) that inform the conditional posterior of $\mathbf{\Sigma}_{\text{ind}}$ are the bivariate random effects $\boldsymbol{\mu}_i, i = 1, \dots, N$. Because the inverse-Wishart is the conjugate prior, the conditional posterior distribution is given by:

$$(\boldsymbol{\Sigma}_{\text{ind}} | \cdot) = IW \left(a + N, \mathbf{A} + \sum_{i=1}^N \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T \right).$$

A.10 Conditional Posterior of $\mathbf{\Sigma}_{\text{area}}$

The latent variables that inform the conditional posterior of $\mathbf{\Sigma}_{\text{area}}$ are the bivariate random effects $\boldsymbol{\eta}_k = (\theta_k, \gamma_k), k = 1, \dots, K$. Because the inverse-Wishart is the conjugate prior, the conditional posterior distribution is given by:

$$(\boldsymbol{\Sigma}_{\text{area}} | \cdot) = IW \left(b + K, \mathbf{B} + \sum_{k=1}^K \boldsymbol{\eta}_k \boldsymbol{\eta}_k^T \right).$$

A.11 Conditional Posterior under Product-Normal Priors—Under the product-normal prior $\mu_{i0} \sim N(0, \tau_0^2), \mu_{i1} | \mu_{i0} \sim N(\phi \mu_{i0}, \tau_1^2), \phi \sim N(0, v_\phi^2)$ and $\tau_j^2 \sim \text{IG}(\omega_1, \omega_2), j = 0, 1$ The conditional posterior for $\boldsymbol{\mu}_i = (\mu_{i0}, \mu_{i1})$ is structurally unchanged from the inverse-Wishart prior specification for $\mathbf{\Sigma}_{\text{ind}}$. It follows that

$$(\tau_0^2 | \cdot) = \text{IG} \left(\omega_1 + \frac{N}{2}, \left(\frac{1}{\omega_2} + \frac{1}{2} \sum_{i=1}^N \mu_{0i}^2 \right)^{-1} \right),$$

$$(\tau_1^2 | \cdot) = \text{IG} \left(\omega_1 + \frac{N}{2}, \left(\frac{1}{\omega_2} + \frac{1}{2} \sum_{i=1}^N (\mu_{1i} - \phi \mu_{i0})^2 \right)^{-1} \right),$$

and

$$(\phi | \cdot) = \text{Normal} \left(\frac{\sum_{i=1}^N \mu_{i0} \mu_{i1}}{\sum_{i=1}^N \mu_{i0}^2 + \tau_1^2 / v_\phi^2}, \frac{\tau_1^2}{\sum_{i=1}^N \mu_{i0}^2 + \tau_1^2 / v_\phi^2} \right).$$

Notably, as $v_{\phi}^2 \rightarrow \infty$ the conditional prior mean of ϕ converges to the ratio of the covariance between the elements of $\mu_1 = (\mu_{11}, \dots, \mu_{N1})$ and $\mu_0 = (\mu_{10}, \dots, \mu_{N0})$ and the variance of the elements of μ_0 . The elements of σ_{ind} are then determined via $\sigma_{\text{int}}^2 = \tau_0^2$, $\sigma_{\text{slope}}^2 = \psi^2 \tau_0^2 + \tau_1^2$ and $\text{cov}(\mu_{i0}, \mu_{i1}) = \rho_{\text{ind}} \sigma_{\text{int}} \sigma_{\text{slope}} = \psi \tau_0^2$.

Under a product-normal prior for $\mathbf{n}$, the conditional posterior distributions for the elements of ρ_{area} are analogous to those for σ_{ind} .

B CONVERGENCE DIAGNOSTICS FOR MCMC

The convergence of the MCMC procedure used for Bayesian estimation was examined using CODA (Convergence Diagnostics and Output Delivery) software.²³ The behavior of the chain was monitored using trace plots³⁹ of the sequence of draws for each parameter for different starting values and random number seeds. These suggested that 10,000 was a sufficient number of iterations for the burn-in phase and that a subsequent 50,000 iterations was sufficient for posterior inferences to converge with minimal Monte-Carlo error. To err on the side of caution, our final models were fitted using a burn-in phase of 20,000 iterations and inferences based on 150,000 draws obtained by extracting 50,000 draws across each of three independent chains. The Gelman-Rubin corrected scale reduction factor⁴⁰ obtained a maximum value of 1.015 for the 0.975 upper confidence interval across all of the model parameters confirming both that the burn-in and inferential phases were sufficiently long. Further evidence of the highly satisfactory convergence is evident in Figure B1, which shows the trace plot from three different starting points and random number seeds of the posterior mean for ρ_{area} (the parameter least informed by the data) in the primary analysis. The numerically very similar results confirm convergence of the chain and that the Monte-Carlo standard error had largely dissipated after 50,000 draws had been obtained.

FIGURE B1.

Trace plot of ρ_{area} for the primary analysis evaluated on the 50,000 post burn-in draws (i.e., after the 20,000 burn-in phase iterations)

C ONLINE SUPPLEMENTARY MATERIALS

The Supplementary Materials available at the GitHub site (<https://github.com/kiwijomalley/ProximityToFood>) contain details of the R and JAGS code for estimating the model, and instructions for running the code on the synthetic data set we provide.

References

1. Block JP, Subramanian SV. Reorienting U.S. policies to reduce disparities in diet quality. *PLoS Medicine*. 2015;12(12):e1001914. [PubMed: 26645285]

2. Block JP, Seward MW, James P. The food environment In. Duncan DT, Kawachi I, eds. *Neighborhood and Health*, Oxford University Press; 2017; New York, NY.
3. Volkow ND, Wise RA. How can drug addiction help us understand obesity?. *Nature Neuroscience*. 2005;8(5).555–560.
4. Cohen DA. Obesity and the built environment. changes in environmental cues cause energy imbalances. *International Journal of Obesity (London)*. 2008;32(Supp 7).S137–142.
5. Jon Rasbash, George Leckie, Rebecca Pillinger, Jennifer Jenkins. Children's educational progress. Partitioning family, school and area effects. *Journal of the Royal Statistical Society, Series A. Statistics in Society*. 2010;173(3).657–682.
6. McGarrell Edmund F, Corsaro Nicholas, Hipple Natalie Kroovand, Bynum Timothy S.. Project safe neighborhoods and violent crime trends in US cities. Assessing violent crime impact. *Journal of Quantitative Criminology*. 2010;26(2).165–190.
7. Sampson RJ, Raudenbush S, Earls F. Neighborhoods and violent crime. A multilevel study of collective efficacy. *Science*. 2003;277918–924.
8. Block JP, Christakis NA, O'Malley AJ, Subramanian SV. Proximity to Food Establishments and Body Mass Index in the Framingham Heart study Offspring Cohort over 30 Years. *American Journal of Epidemiology*. 2011;1741108–1114. [PubMed: 21965186]
9. Block JP, Subramanian SV, Christakis NA, O'Malley AJ. Population Trends and Variation in Body Mass Index From 1971 to 2008 in the Framingham Heart Study Offspring Cohort. *PLoS One*. 2013;8(5).e63217. [PubMed: 23675464]
10. Fan Li, Zaslavsky Alan M.. Using a short screening scale for small-area estimation of mental illness prevalence for schools. *Journal of the American Statistical Association*. 2010;105(492). 1323–1332. [PubMed: 25477700]
11. Bello Nora M, Steibel Juan P, Tempelman Robert J. Hierarchical Bayesian modeling of random and residual variance-covariance matrices in bivariate mixed effects models. *Biometrical Journal*. 2010;52(3).297–313. [PubMed: 20544726]
12. Harvey Goldstein, William Browne, Jon Rasbash. Multilevel modelling of medical data. *Statistics in Medicine*. 2002;21(21).3291–3315. [PubMed: 12375305]
13. Leyland AH, Naess O. The effect of area of residence over the life course on subsequent mortality. *Journal of the Royal Statistical Society, Series A. Statistics in Society*. 2009;172555–578.
14. Tsao CW, Vasan RS. Cohort Profile. The Framingham Heart Study (FHS). overview of milestones in cardiovascular epidemiology. *International Journal of Epidemiology*. 2015;44(6).1800–1813. [PubMed: 26705418]
15. Little Roderick JA, Rubin Donald B. *Statistical analysis with missing data*. New York; Chichester John Wiley & Sons; 2002.
16. Rubin DB. Inference and missing data. *Biometrika*. 1976;63581–590.
17. Duncan Terry E, Duncan Susan C. Modeling the processes of development via latent variable growth curve methodology. *Structural Equation Modeling. A multidisciplinary Journal*. 1995;2187–213.
18. Hedeker D, Gibbons R. *Longitudinal Data Analysis*. Wiley; 2006.
19. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006;1(3).515–533.
20. Berger JO, Strawderman W, Tang D. Posterior Propriety and Admissibility of Hyperpriors in Normal Hierarchical Models. *The Annals of Statistics*. 2005;33606–646.
21. Spiegelhalter David J. Bayesian Graphical Modelling. A Case-study in Monitoring Health Outcomes. *Journal of the Royal Statistical Society, Series C. Applied Statistics*. 1998;47115–133.
22. Cooper Nicola J, Lambert Paul C, Abrams Keith R, Sutton Alexander J. Predicting costs over time using Bayesian Markov chain Monte Carlo methods. an application to early in ammatory polyarthRitus. *Health Economics*. 2007;1637–56. [PubMed: 16981192]
23. Best N, Cowles MK, Vines K. *Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK; 1995.

24. Chib S, Greenberg E. Understanding the Metropolis-Hastings Algorithm. *The American Statistician*. 1995;49:327–335.
25. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*. 1953;21:1078–92.
26. Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and their Applications. *Biometrika*. 1970;57:97–109.
27. Tierney L. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*. 1994;22:1701–1762.
28. Roberts GO, Gelman A, Gilks WR. Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *Annals of Applied Probability*. 1997;7:110–120.
29. Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK. population based, cross sectional study. *British Medical Journal*. 2014;348:g1464. [PubMed: 24625460]
30. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CA. The relationship of the local food environment with obesity. A systematic review of methods, study quality, and results. *Obesity (Silver Spring)*. 2015;23(7):1331–1344. [PubMed: 26096983]
31. Powell LM, Han E. Adult obesity and the price and availability of food in the United States. *American Journal of Agricultural Economics*. 2011;93(2):378–384.
32. Dubowitz T, Ghosh-Dastidar M, Cohen DA, et al. Diet And Perceptions Change With Supermarket Introduction In A Food Desert, But Not Because Of Supermarket Use. *Health Affairs (Millwood)*. 2015;34(11):1858–1868.
33. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health Affairs (Millwood)*. 2014;33(2):283–291.
34. Elbel B, Moran A, Dixon LB, et al. Assessment of a government-subsidized supermarket in a high-need area on household food availability and children's dietary intakes. *Public Health Nutrition*. 2015;18(15):2881–2890. [PubMed: 25714993]
35. Morenoff J, Diez Roux AV, Oyaspuk T, Hansen B. Residential Environments and Obesity. What Can We Learn About Policy Interventions From Observational Studies? Reorienting U.S. policies to reduce disparities in diet quality. 2006;. Unpublished conference paper. National Poverty Center's Health Effects of Non-Health Policy, Bethesda, MD.
36. Cetateanu A, Jones A. How can GPS technology help us better understand exposure to the food environment? A systematic review. *SSM Population Health*. 2016;2:196–205. [PubMed: 28018957]
37. Gelfand Alan E, Smith Adrian FM. Sampling-based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*. 1990;85:398–409.
38. Stuart Geman, Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1984;6:721–741. [PubMed: 22499653]
39. Hellmich M, Abrams KR, Jones DR, Lambert PC. A Bayesian Approach to a General Regression Model for ROC Curves.. *Academic Radiology*. 1998;18:436–443.
40. Brooks Stephen P, Gelman Andrew. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*. 1998;7:434–455.

FIGURE 1.

Operating characteristics of the posterior mean estimators of ρ_{area} and ρ_{ind} across likely values of ρ_{area} and of the posterior mean estimators of σ_e^2 and ρ_e across likely values of ρ_e .

FIGURE 2.

Posterior distributions for selected model parameters under inverse-Wishart prior distributions for ind and area

FIGURE 3.

Counts of individuals by status (resident, employee, both) for each neighborhood in the four-town area for employed individuals. Approximately 19.6% of individual-exams had the same neighborhood for residence and work, a number inflated to 50.6% if non-employed people were assigned the same “workplace” location as their residential location.

TABLE 1

SAS, R, and Stata syntax for estimating the model closest to that in Equation (1)

SAS Proc Mixed	R lmer	Stata
random int exam/subject=personid type=un;	(1 exam personid)	personid: exam
random int/subject = homeid;	+ (1 homeid)	homeid:
random int/subject = workid;	+ (1 workid)	workid:

Note. MLWiN is another package for which estimation of the full model is elusive. It only allows correlated random effects across homeid and workid on data restricted to a few areas and only when using iterative generalized least squares for estimation. MLWiN also does not allow serially correlated errors.

TABLE 2

Simulation results for ρ_{area} , ρ_{ind} , σ_e^2 and ρ_e .

Term	True ρ_{area} values					True ρ_e values						
	-0.400	-0.200	0.000	0.200	0.400	-0.459	-0.306	-0.153	0.000	0.153	0.306	0.459
Bias												
	Regression parameters ($\text{AABias} = p^{-1} \sum_{k=1}^p \text{bias}(\beta_k) $)											
β	0.003	0.001	0.002	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001
	Neighborhood covariance parameters (Σ_{area})											
σ_{home}^2	0.025	0.025	0.025	0.026	0.027	0.024	0.025	0.026	0.026	0.026	0.025	0.024
σ_{work}^2	0.020	0.021	0.021	0.022	0.023	0.020	0.021	0.022	0.022	0.022	0.021	0.020
ρ_{area}	0.211	0.094	-0.025	-0.146	-0.269	-0.008	-0.010	-0.013	-0.015	-0.016	-0.017	-0.018
	Individual covariance parameters (Σ_{ind})											
σ_{int}^2	-0.189	-0.180	-0.177	-0.181	-0.190	-0.174	-0.161	-0.156	-0.156	-0.163	-0.177	-0.197
σ_{slope}^2	-0.004	-0.003	-0.003	-0.004	-0.004	-0.002	-0.002	-0.002	-0.002	-0.003	-0.003	-0.004
ρ_{ind}	0.000	0.000	0.000	0.000	0.000	-0.003	-0.002	-0.002	-0.001	-0.001	0.000	0.001
	Error-term parameters											
σ_e^2	0.016	0.018	0.019	0.018	0.017	0.001	0.002	0.002	0.004	0.010	0.019	0.030
ρ_e	0.003	0.003	0.003	0.003	0.003	0.001	0.001	0.001	0.001	0.002	0.003	0.004
MSE												
	Regression parameters ($\text{AMSE} = p^{-1} \sum_{k=1}^p \text{mse}(\beta_k)$)											
β	0.057	0.057	0.058	0.058	0.058	0.057	0.058	0.059	0.059	0.059	0.058	0.057
	Neighborhood covariance parameters (Σ_{area})											
σ_{home}^2	0.037	0.036	0.037	0.037	0.039	0.035	0.037	0.038	0.038	0.037	0.036	0.035
σ_{work}^2	0.029	0.029	0.030	0.030	0.031	0.029	0.030	0.031	0.031	0.031	0.030	0.028
ρ_{area}	0.277	0.203	0.187	0.228	0.326	0.184	0.188	0.190	0.190	0.189	0.186	0.182
	Individual covariance parameters (Σ_{ind})											
σ_{int}^2	0.547	0.548	0.549	0.548	0.547	0.527	0.527	0.530	0.534	0.540	0.549	0.559
σ_{slope}^2	0.014	0.014	0.014	0.014	0.014	0.011	0.012	0.012	0.013	0.014	0.014	0.015
ρ_{ind}	0.021	0.021	0.021	0.021	0.021	0.018	0.019	0.020	0.020	0.021	0.021	0.022
	Error-term parameters											
σ_e^2	0.075	0.080	0.082	0.080	0.075	0.034	0.036	0.041	0.044	0.060	0.082	0.111
ρ_e	0.017	0.017	0.017	0.017	0.017	0.009	0.010	0.012	0.013	0.015	0.017	0.019
Coverage												

Term	True ρ_{area} values					True ρ_{e} values						
	-0.400	-0.200	0.000	0.200	0.400	-0.459	-0.306	-0.153	0.000	0.153	0.306	0.459
Regression parameters ($\text{ACoverage} = p^{-1} \sum_{k=1}^p \text{coverage}(\beta_k)$)												
	0.947	0.951	0.944	0.948	0.950	0.949	0.948	0.943	0.951	0.948	0.949	0.951
Neighborhood covariance parameters (Σ_{area})												
σ_{home}^2	0.947	0.950	0.951	0.948	0.942	0.950	0.953	0.955	0.955	0.954	0.951	0.947
σ_{work}^2	0.901	0.900	0.897	0.893	0.888	0.866	0.868	0.871	0.878	0.886	0.898	0.912
ρ_{area}	0.876	0.977	1.000	0.947	0.816	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Individual covariance parameters (Σ_{ind})												
σ_{int}^2	0.923	0.933	0.935	0.929	0.914	0.934	0.940	0.943	0.943	0.941	0.935	0.926
σ_{slope}^2	0.907	0.912	0.915	0.917	0.916	0.950	0.926	0.911	0.904	0.906	0.915	0.933
ρ_{ind}	0.934	0.945	0.950	0.939	0.934	0.939	0.953	0.922	0.957	0.946	0.943	0.953
Error-term parameters												
σ_{e}^2	0.922	0.943	0.950	0.944	0.925	0.933	0.936	0.939	0.943	0.946	0.950	0.950
ρ_{e}	0.950	0.947	0.948	0.952	0.961	0.977	0.954	0.939	0.933	0.936	0.948	0.968

TABLE 3

Posterior estimates for the primary analysis (Inverse-Wishart prior) and based on the alternative (Product-Normal) prior for area and ind

Variable	Inverse-Wishart				Product-Normal			
	Mean	95% CR		Pr > 0	Mean	95% CR		Pr > 0
Intercept (β_0)	24.676	24.329	25.024	1.000	24.651	24.274	24.958	1.000
Exam 2 (β_1)	0.484	0.378	0.591	1.000	0.488	0.380	0.595	1.000
Exam 3 (β_2)	1.004	0.871	1.136	1.000	1.008	0.872	1.142	1.000
Exam 4 (β_3)	1.632	1.480	1.784	1.000	1.638	1.481	1.793	1.000
Exam 5 (β_4)	2.215	2.041	2.390	1.000	2.223	2.047	2.397	1.000
Exam 6 (β_5)	2.649	2.452	2.847	1.000	2.656	2.459	2.854	1.000
Exam 7 (β_6)	2.931	2.708	3.154	1.000	2.938	2.714	3.160	1.000
Exam 8 (β_7)	3.003	2.741	3.266	1.000	3.013	2.754	3.274	1.000
Male (β_8)	1.876	1.551	2.208	1.000	1.878	1.556	2.196	1.000
Year of Birth (β_9)	-0.054	-0.070	-0.039	0.000	-0.055	-0.071	-0.040	0.000
Smokes (β_{10})	-0.812	-0.955	-0.670	0.000	-0.809	-0.951	-0.665	0.000
Married (β_{11})	0.221	0.060	0.383	0.996	0.226	0.070	0.386	0.998
High school or less (β_{12})	-0.181	-0.448	0.089	0.095	-0.184	-0.451	0.092	0.100
Greater than high school (β_{13})	0.678	0.219	1.139	0.998	0.692	0.238	1.151	0.998
Employed (β_{14})	0.207	0.021	0.397	0.985	0.202	0.043	0.364	0.994
Tract poverty home (β_{15})	0.026	-0.021	0.073	0.870	0.026	-0.014	0.064	0.918
Tract poverty work (β_{16})	0.005	-0.037	0.047	0.597	0.005	-0.029	0.043	0.620
Fast-food proximity associations								
Home Distance & Female (β_{17})	-0.105	-0.228	0.019	0.048	-0.105	-0.231	0.017	0.048
Home Distance & Male (β_{18})	0.040	-0.103	0.184	0.711	0.045	-0.102	0.191	0.730
Work Distance & Female (β_{19})	-0.058	-0.201	0.086	0.215	-0.052	-0.194	0.090	0.238
Work Distance & Male (β_{20})	-0.142	-0.313	0.028	0.050	-0.136	-0.306	0.030	0.054
Drive Count & Female (β_{21})	0.025	0.000	0.051	0.975	0.025	-0.001	0.051	0.968
Drive Count & Male (β_{22})	-0.018	-0.049	0.013	0.128	-0.017	-0.048	0.014	0.141
Variation between neighborhoods (Σ_{area})								
Residence (σ_{home}^2)	0.123	0.058	0.246	1.000	0.069	0.023	0.164	1.000
Workplace (σ_{work}^2)	0.100	0.049	0.195	1.000	0.064	0.022	0.153	1.000
Correlation (ρ_{area})	-0.051	-0.510	0.431	0.415	-0.152	-0.774	0.618	0.338
Variation between individuals (Σ_{ind})								
Intercept (σ_{int}^2)	14.290	13.103	15.522	1.000	14.341	13.166	15.554	1.000
Exam (slope) (σ_{slope}^2)	0.303	0.266	0.343	1.000	0.305	0.270	0.344	1.000
Correlation (ρ_{ind})	0.001	-0.078	0.084	0.502	-0.002	-0.079	0.079	0.471

Variable	Inverse-Wishart				Product-Normal			
	Mean	95% CR		Pr > 0	Mean	95% CR		Pr > 0
Variation between observations								
Residual variance (σ_e^2)	3.132	2.864	3.452	1.000	3.127	2.872	3.420	1.000
Serial correlation (ρ_e)	0.393	0.341	0.448	1.000	0.393	0.343	0.444	1.000

Note. The excluded category of education is “not reported”, tract poverty is a percentage, and the proximity variables are scaled so that the units are kilometers (1000s of meters).

TABLE 4

Impact of use of 14-town area data and in the 4-town data to carrying forward the residential and workplace locations of the first exam to all exams compared to the primary analysis

Variable	Primary analysis			14-town			4-town, No movement					
	Mean	95% CR	Pr >0	Mean	95% CR	Pr >0	Mean	95% CR	Pr >0			
Fast-food exposure associations												
Home Distance & Female (β_{17})	-0.105	-0.228	0.019	0.048	-0.171	-0.015	0.009	-0.094	-0.215	0.026	0.062	
Home Distance & Male (β_{18})	0.040	-0.103	0.184	0.711	-0.018	-0.092	0.057	0.318	0.044	-0.095	0.188	0.725
Work Distance & Female (β_{19})	-0.058	-0.201	0.086	0.215	-0.015	-0.095	0.064	0.359	-0.035	-0.170	0.100	0.301
Work Distance & Male (β_{20})	-0.142	-0.313	0.028	0.050	-0.058	-0.139	0.021	0.075	-0.112	-0.273	0.050	0.089
Drive Count & Female (β_{21})	0.025	0.000	0.051	0.975	0.013	-0.005	0.031	0.925	0.022	-0.003	0.047	0.961
Drive Count & Male (β_{22})	-0.018	-0.049	0.013	0.128	-0.008	-0.027	0.012	0.213	-0.017	-0.048	0.014	0.141
Variation between neighborhoods (Σ_{neigh})												
Residence (σ_{home}^2)	0.123	0.058	0.246	1.000	0.086	0.046	0.154	1.000	0.220	0.084	0.497	1.000
Workplace (σ_{work}^2)	0.100	0.049	0.195	1.000	0.076	0.042	0.133	1.000	0.115	0.054	0.229	1.000
Correlation (ρ_{neigh})	-0.051	-0.510	0.431	0.415	0.012	-0.395	0.408	0.522	-0.191	-0.641	0.344	0.229
Variation between individuals (Σ_{ind})												
Intercept (σ_{int}^2)	14.290	13.103	15.522	1.000	15.005	13.991	16.050	1.000	14.243	13.063	15.458	1.000
Exam (slope) (σ_{slope}^2)	0.303	0.266	0.343	1.000	0.316	0.285	0.348	1.000	0.305	0.268	0.342	1.000
Correlation (ρ_{ind})	0.001	-0.078	0.084	0.502	-0.024	-0.088	0.040	0.229	0.004	-0.075	0.088	0.523
Variation between observations												
Residual variance (σ_e^2)	3.132	2.864	3.452	1.000	2.861	2.681	3.064	1.000	3.121	2.872	3.393	1.000
Serial Correlation (ρ_e)	0.393	0.341	0.448	1.000	0.349	0.308	0.391	1.000	0.391	0.342	0.439	1.000

Note: Distance variables are in units of kilometers (1000s of meters)

TABLE 5

Impact of simplifying the random component of the model for the primary analysis

Variable	Primary analysis			No area cor ($\rho_{\text{area}} = 0$)			No random slope ($\rho_{\text{nd}} = 0$)			No serial cor ($\rho_e = 0$)		
	Mean	95% CR		Mean	95% CR		Mean	95% CR		Mean	95% CR	
Fast-food exposure associations												
Home Distance & Female (β_{17})	-0.105	-0.228	0.019	-0.105	-0.226	0.017	-0.138	-0.265	-0.010	-0.098	-0.220	0.024
Home Distance & Male (β_{18})	0.040	-0.103	0.184	0.043	-0.101	0.188	0.078	-0.072	0.227	0.038	-0.109	0.177
Work Distance & Female (β_{19})	-0.058	-0.201	0.086	-0.058	-0.203	0.085	-0.002	-0.155	0.147	-0.095	-0.234	0.046
Work Distance & Male (β_{20})	-0.142	-0.313	0.028	-0.141	-0.314	0.035	-0.136	-0.315	0.036	-0.184	-0.357	-0.015
Drive Count & Female (β_{21})	0.025	0.000	0.051	0.025	-0.001	0.050	0.034	0.008	0.060	0.013	-0.011	0.038
Drive Count & Male (β_{22})	-0.018	-0.049	0.013	-0.018	-0.049	0.013	-0.011	-0.043	0.021	-0.025	-0.055	0.005
Variation between neighborhoods (Σ_{area})												
Residence (σ_{home}^2)	0.123	0.058	0.246	0.115	0.055	0.229	0.121	0.057	0.242	0.128	0.059	0.256
Workplace (σ_{work}^2)	0.100	0.049	0.195	0.096	0.048	0.182	0.105	0.051	0.204	0.104	0.050	0.200
Correlation (ρ_{area})	-0.051	-0.510	0.431				-0.049	-0.509	0.425	-0.056	-0.524	0.436
Variation between individuals (Σ_{ind})												
Intercept (σ_{int}^2)	14.290	13.103	15.522	14.285	13.098	15.534	11.224	0.415	14.273	16.473	15.360	17.643
Exam (slope) (σ_{slope}^2)	0.303	0.266	0.343	0.302	0.265	0.342				0.383	0.348	0.419
Correlation (ρ_{ind})	0.001	-0.078	0.084	0.002	-0.077	0.088				-0.127	-0.185	-0.067
Variation between observations												
Residual variance (σ_e^2)	3.132	2.864	3.452	3.142	2.868	3.455	11.589	8.837	22.505	2.112	2.041	2.187
Serial Correlation (ρ_e)	0.393	0.341	0.448	0.395	0.343	0.449	0.803	0.754	0.904			

Note: Distance variables are in units of kilometers (1000s of meters). The posterior probabilities that the above parameters exceed 0 are not presented for ease of presentation.