Published in final edited form as:

J R Stat Soc Ser A Stat Soc. 2018 June ; 181(3): 783–802. doi:10.1111/rssa.12330.

Factors associated with supermarket and convenience store closure: a discrete time spatial survival modelling approach

Joshua L. Warren and

Yale University, New Haven, USA

Penny Gordon-Larsen

University of North Carolina at Chapel Hill, Chapel Hill, USA

Summary

While there is a literature on the distribution of food stores across geographic and social space, much of this research uses cross-sectional data. Analyses attempting to understand whether the availability of stores across neighborhoods is associated with diet and/or health outcomes are limited by a lack of understanding of factors that shape the emergence of new stores and the closure of others. We used quarterly data on supermarket and convenience store locations spanning seven years (2006–2012) and tract-level census data in four US cities: Birmingham, Alabama; Chicago, Illinois; Minneapolis, Minnesota; San Francisco, California. A spatial discrete-time survival model was used to identify factors associated with an earlier and/or later closure time of a store. Sales volume was typically the strongest indicator of store survival. We identified heterogeneity in the association between tract-level poverty and racial composition with respect to store survival. Stores in high poverty, non-White tracts were often at a disadvantage in terms of survival length. The observed patterns of store survival varied by some of the same neighborhood sociodemographic factors associated with lifestyle and health outcomes, which could lead to confusion in interpretation in studies of the estimated effects of introduction of food stores into neighborhoods on health.

Keywords

Bayesian modeling; integrated nested Laplace approximations (INLA); neighborhood sociodemographics; sales volume; spatially correlated random effects

1. INTRODUCTION

Research on the association between access to neighborhood supermarkets and convenience stores in relation to diet behaviors and obesity is mixed, despite discussion of policies supporting the introduction of supermarkets in underserved areas (Morland et al., 2002; Cummins et al., 2005; Pearce et al., 2008; Duffey et al., 2009; Larson et al., 2009; Pearce et al., 2009; Powell et al., 2010; Duran et al., 2013; Powell and Nguyen, 2013; Hollands et al.,

[†]Address for correspondence: Joshua L. Warren, Department of Biostatistics, Yale University, 60 College St., 213 LEPH, New Haven, Connecticut 06510, USA. joshua.warren@yale.edu.

The authors have no financial or other conflicts of interest to disclose.

2014; Richardson et al., 2014). Yet there is little known or understood about factors that shape the emergence of certain new stores and the closure of others. Furthermore, the patterning of opening/closure of food stores in relation to geographic and social space is not well understood. Thus, it is unclear whether low income and minority populations are truly underserved by the emergence and closure of more versus less healthy food stores in their neighborhoods.

There are also significant methodological limitations of the current literature, which is largely cross-sectional (see recent reviews Fleischhacker et al. (2011), Gordon-Larsen (2014), and Martin et al. (2014)), thus limiting understanding of temporality in the association between neighborhood sociodemographics and availability of neighborhood food stores. In addition, the bulk of the literature includes no direct accounting for spatial correlation (see Caspi et al. (2012) for a review). Furthermore, there is little understanding of the role of internal store characteristics (e.g., square footage, sales volume) or external characteristics (e.g., density of other stores, distance to core business district) in the emergence and closure of specific types of food stores. Thus, the application of advanced statistical methods to examine spatially referenced events through time is much needed in this literature.

To address these gaps, we used quarterly data on supermarket and convenience store locations from Nielsen TDLinx (Nielsen Holdings NV, New York, NY) spanning seven years (2006–2012) and neighborhood tract-level census data from the American Community Survey (2006–2010) in metropolitan statistical areas (MSAs) of four US cities including data on internal store characteristics (e.g. square footage, sales volume) and external data (e.g. neighborhood sociodemographics, density of other stores, distance to the core business district). We used a spatial discrete-time survival model within a Bayesian hierarchical framework to determine which internal and external store factors were associated with an earlier and/or later than expected closure time of a store, while accounting for potential spatial correlation in the data.

2. DATA

We include data from four US MSAs including Birmingham, Alabama (AL); Chicago, Illinois (IL); Minneapolis, Minnesota (MN); and San Francisco, California (CA) (Department of Commerce, 2012). Supermarket and convenience store data were obtained from Nielsen TDLinx (Nielsen Holdings N.V., 2010) and have been previously described (Lamichhane et al., 2014). We had access to store information from 27 time points between March 2006 and August 2012. We included data on: latitude/longitude of the store location, square footage, sales volume (thousands of dollars), distance in kilometers (km) to the central business district of the MSA containing the store, number of stores of the same type (supermarket or convenience store) within one km of the store location, number of stores of the other type (supermarket or convenience store) within one km of the store location, and census tract variables including racial composition (predominantly White, predominantly Black, racially mixed, predominantly Hispanic), population density (km^2), and poverty level (proportion of people living below the poverty line and calculated tertiles of this variable).

Additional details on the selection and transformation of the covariates in the models are given in Section S1 of the supplementary on-line material.

Tables 1–4 display summary information for our analyzed datasets averaged over all time periods and subset by the survival status (open/closed) for each of the MSAs and store types. Generally, supermarkets have larger square footage, larger sales volume, are located in tracts with a higher population density, and have less nearby competition (number of supermarkets within one km of the store location) than convenience stores. Table 5 displays the average number (and standard deviation) of open stores and store closings across time for the four MSAs and two store types. Results suggest that in each of the four MSAs there are more convenience stores on average than supermarkets and the variability in these totals across time is also larger for convenience stores, indicating greater turnover across time.

Convenience stores also have more closures on average than supermarkets, largely due to a higher number of currently open stores. We display the spatial location of all stores (supermarkets (Figure 1) and convenience stores (Figure 2)) observed during the entire study period for each MSA. Stores that closed during any time interval are shown in red. Figures 1 and 2 do not suggest any obvious spatial patterning of store closings across the four MSAs.

3. METHODS

3.1. Statistical model

We model the data from supermarkets and convenience stores and from each of the four MSAs in separate analyses given the large geographic distances between the locations and differences in store types. We use a discrete-time survival model with spatial frailties to determine which factors are associated with an earlier and/or later than expected closure time (death) of a store.

For each store, we observe (i) the discrete closing date, or if the store did not close during one of the observed study periods, the date in which the study ended and (ii) a censoring variable indicating if the store closed during the study or not. In total, we observe 26 time intervals where a store could potentially fail and denote these as t , where the first study interval ($t = 1$) begins at the start of April 2006 and concludes at the end of June 2006. Each interval includes three calendar months representing quarterly data other than the final interval which includes two months ($t = 26$: July 2012–August 2012). We define the function $q(t)$ as the number of months that separate the end of interval t and the end of interval one where $q(t) = 3(t - 1)$ for $t = 1, \dots, 25$ and $q(26) = q(25) + 2$ to account for the two month separation.

Therefore, for each store we observe $\{\delta_i(s_i), D_i(s_i)\}$, $i = 1, \dots, n$ where $\delta_i(s_i)$ is the censoring indicator for store i at location s_i such that $\delta_i(s_i) = 1$ if the store did not close at some point during the study (censored) and $\delta_i(s_i) = 0$ otherwise; $D_i(s_i)$ is the interval of censoring or closing for store i at location s_i such that $D_i(s_i) = \min\{T_i(s_i), c\}$ where $T_i(s_i)$ is the true (and possibly unobserved) interval of closure for the store and c is final interval of the study ($c = 26$); and n is the number of stores included in the analysis. It is possible for multiple stores to have the same location since a store can open at the location of a

previously failed store. Therefore, $s_i \in \{s_1^*, \dots, s_m^*\}$ for $i = 1, \dots, n$ where s_j^* represents a unique location and m is the total number of unique locations within an MSA with $m \leq n$.

The closing date of a store is only available discretely, at the end of different time periods, resulting in a discrete-time survival modeling framework. For example, a store can first appear in the data during interval $t \in \{0, 1, \dots, 26\}$, meaning that it was opened at some continuous time within the interval. We denote the first interval that store i is represented in the data (birth) as $b_i(s_i)$ from here on out. In this example, $b_i(s_i) = t$. If the store closes during interval $t + 1$, and this represents a time before the conclusion of the study, we define $\delta_i(s_i) = 0$ and $D_i(s_i) = T_i(s_i) = t + 1$. In reality, the store closed at some continuous time within interval $t + 1$. If the store closed during the final interval of the study, we observe $\delta_i(s_i) = 0$ and $D_i(s_i) = T_i(s_i) = c$. If the store remained open at the end of the study (censored), we observe $\delta_i(s_i) = 1$ and $D_i(s_i) = c$ with the true interval of closure, $T_i(s_i)$, unobserved.

In order to model survival data of this type, we work with the discrete-time hazard function which describes the probability that a store closes during interval t given that it survived to that interval such that $\pi_{it}(s_i) = P\{T_i(s_i) = t | T_i(s_i) \geq t\}$, $t = 1, \dots, 26$. Using this definition, the probability that a store closes during some interval t before the final study interval ($t < c$) can be written as

$$P\{D_i(s_i) = t\} = \pi_{it}(s_i) \prod_{j=b_i(s_i)+1}^{t-1} \{1 - \pi_{ij}(s_i)\}.$$

Similarly,

$$P\{D_i(s_i) = c\} = \prod_{j=b_i(s_i)+1}^{c-1} \{1 - \pi_{ij}(s_i)\}$$

represents the probability that the store either closes in the final interval or is censored.

Our interest is in introducing a model for $\pi_{it}(s_i)$ that allows us to identify factors associated with the survival of a store. We model these probabilities using the complementary log-log link function such that

$$\ln[-\ln\{1 - \pi_{it}(s_i)\}] = h\{q(t); \boldsymbol{\gamma}\} + \mathbf{x}_{it}^T(s_i)\boldsymbol{\beta} + w(s_i) \quad (1)$$

where $h\{q(t); \boldsymbol{\gamma}\}$ is a function that describes the baseline risk of closure during interval t , corresponding to $q(t)$ months after the end of the first interval, and $\boldsymbol{\gamma}$ is the vector of unknown parameters associated with the function; $\mathbf{x}_{it}(s_i)$ is the vector of store- and tract-level covariates specific to store i at interval t and location s_i (listed in Section 2); $\boldsymbol{\beta}$ represents the vector of regression parameters that relate the covariates to the probability of closure; and $w(s_i)$ represents a spatial random effect that describes the risk of closure associated with a unique spatial location. The $h\{q(t); \boldsymbol{\gamma}\}$ function is modeled using cubic

splines with five degrees of freedom in order to allow for increased flexibility across time. Use of the complementary log-log link function leads to a hazard ratio (HR) interpretation for the regression parameters.

The vector of spatial random effects are modeled using a multivariate normal distribution (MVN) with an isotropic spatially referenced correlation matrix such that

$$\mathbf{w} = \{w(s_1^*), \dots, w(s_m^*)\}^T | \tau^2, \kappa \sim \text{MVN} \left(0 * \mathbf{1}_m, \frac{1}{2\pi\kappa\tau^2} \Sigma(\kappa) \right) \quad (2)$$

where $\mathbf{1}_m$ is a vector of length m with each entry equal to one and the correlation between random effects from two different spatial locations is defined as a function of the distance between the locations such that

$$\text{Corr} \{w(s_i^*), w(s_j^*) | \kappa\} = \exp \{ -\kappa \|s_i^* - s_j^*\| \}, \quad (3)$$

where $\|\cdot\|$ represents the Euclidean distance function. As the distance between locations increases, the correlation between two parameters decreases and the strength of this spatial decay is controlled by the unknown parameter $\kappa > 0$.

These random effects allow us to investigate the possibility that certain locations within an MSA may be better or worse for the survival of a store, controlling for available covariates. Accounting for spatial correlation is also important in order to make correct inference on the regression parameters of interest. Ignoring spatial correlation can often result in overly optimistic measures of uncertainty and the possibility of drawing incorrect conclusions regarding associations of interest.

3.2. Prior specification

In order to complete the model specification in the Bayesian setting, we assign prior distributions to all model parameters. The regression parameters contained in the $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ vectors are given independent normally distributed prior distributions with a mean of zero and a variance of 1000. A multivariate normal prior distribution is selected for the vector of parameters that in combination define the spatial covariance matrix, $\theta_1 = \ln(\tau)$ and $\theta_2 = \ln(\kappa)$ (Lindgren and Rue, 2015). In our data application, we choose the prior mean and variances/covariances such that $(\theta_1, \theta_2)^T \sim \text{MVN}(0 * \mathbf{1}_2, 10 * \mathbf{I}_2)$ where $\mathbf{I}_2$ is the two by two identity matrix. All priors are chosen to be weakly informative in order to allow the data to drive the inference rather than our prior beliefs.

3.3. Model fitting details

We implement an equivalent data augmentation approach that simplifies the model fitting process. For a given store, we augment $\{\delta_i(s_j), D_i(s_j)\}$ to $(Y_{i,b_f(s_j)+1}(s_j), \dots, Y_{i,D_f(s_j)}(s_j))$ where

$$Y_{it}(s_i) | \pi_{it}(s_i) \stackrel{\text{ind}}{\sim} \text{Bernoulli}(\pi_{it}(s_i)),$$

$D_i(s_i) = \min \{t: Y_{it}(s_i) = 1 \text{ or } t = c\}$, with the Bernoulli random variables defined as

$$Y_{it}(s_i) = \begin{cases} 0 & \text{if } t < D_i(s_i); \\ 1 - \delta_i(s_i) & \text{if } t = D_i(s_i). \end{cases}$$

Therefore, during each interval that the store is open, there is a probability that the store fails during that interval. This probability possibly changes across time, space, and due to store- and tract-level covariates. Augmenting the data in this way is equivalent to working directly with $\{\delta_i(s_i), D_i(s_i)\}$ and this can be shown such that

$$P\{D_i(s_i) = t\} = P\{Y_{it}(s_i) = 1\} \prod_{j=b_i(s_i)+1}^{t-1} P\{Y_{ij}(s_i) = 0\} = \pi_{it}(s_i) \prod_{j=b_i(s_i)+1}^{t-1} \{1 - \pi_{ij}(s_i)\}$$

for any interval occurring before the final study interval and

$$P\{D_i(s_i) = c\} = P\{Y_{ic}(s_i) = 1\} \prod_{j=b_i(s_i)+1}^{c-1} P\{Y_{ij}(s_i) = 0\} + \prod_{j=b_i(s_i)+1}^c P\{Y_{ij}(s_i) = 0\} = \prod_{j=b_i(s_i)+1}^{c-1} \{1 - \pi_{ij}(s_i)\}.$$

Therefore, we can equivalently model the probabilities of interest using a binary regression model while treating $Y_{it}(s_i)$ as the dependent variable.

3.4. Statistical software

Due to the large number of unique spatial locations in each MSA for the different store types (m) (see Section 5), we fit the final models using the R-INLA (integrated nested Laplace approximations) package available in the R Statistical Software program (www.r-inla.org) (Rue et al., 2009; Martins et al., 2013; R Core Team, 2016). Because we are working with point-referenced geostatistical data and modeling framework, the stochastic partial differential equation (SPDE) spatial models were utilized (Lindgren et al., 2011). The SPDE model works by defining the spatially correlated random effects, $w(s)$, that are located continuously across the spatial domain as a Gaussian Markov random field (GMRF) with discrete support over a triangulation of the domain through the use of basis functions (Blangiardo et al., 2013). The resulting GMRF will have a precision matrix whose inverse is close to the original covariance matrix of interest (defined by the spatial correlation function). Working with this GMRF eases the computational burden of having to work directly with dense m by m spatially referenced covariance matrices. See Lindgren et al. (2011) and Blangiardo et al. (2013) for more information.

In general, the INLA methodology represents an alternative model fitting technique to Markov chain Monte Carlo (MCMC) methods which are most commonly used to fit models in the Bayesian setting. INLA approximates marginal posterior distributions of interest and allows for marginal inference that is typically desired in a Bayesian analysis. However, it can often be much faster than MCMC, particularly for large spatial problems like those addressed in this study. This computational convenience allows us to make inference on the parameters of interest in a fraction of the time in our setting where the number of spatial locations is large. In Section 4, we present a simulation study to compare the performance of INLA and full MCMC with respect to marginal posterior inference and computing times.

4. SIMULATION STUDY

We design a simulation study to compare inference obtained from the INLA SPDE approach with that of a full MCMC analysis under different levels of spatial correlation. We also compare the computational performance of each method by monitoring the running times of each fitted model in the study. We begin by simulating data from a spatial process that closely matches the modeling framework introduced in Section 3. The process is defined as $Y_{it}(s_i) | \pi_{it}(s_i) \stackrel{\text{ind}}{\sim} \text{Bernoulli}(\pi_{it}(s_i))$ with

$$\Phi^{-1}(\pi_{it}(s_i)) = -0.10 + 0.50x_{it}(s_i) + w(s_i)$$

for $i = 1, \dots, m = 150$ and $t = 1, \dots, 10$ where the locations are simulated within the unit square and $x_{it}(s_i) \stackrel{\text{iid}}{\sim} N(0, 1)$. The probit link function is used by defining $\Phi^{-1}(\cdot)$ as the inverse cumulative distribution function of the standard normal distribution. We note that this link function results in conjugacy in our MCMC version of the model, allowing for computational benefits due to faster mixing of the Markov chains and therefore requiring shorter runs to make inference on the parameters of interest. This is important since we are fitting the models multiple times in the study. The spatial process is defined based on equations (2) and (3) under two different levels of spatial correlation; (i) strong spatial correlation, where $\kappa = 0.65$ and $\tau = 1$, and (ii) moderate spatial correlation, where $\kappa = 5.00$ and $\tau = 0.36$. Both settings resulting in a total process variance of $\sigma_w^2 = 0.24$ where $\sigma_w^2 = 1/(2\pi\kappa\tau^2)$.

In total, we simulate 50 datasets from this process under each correlation setting, conditional on the spatial locations and covariate values which are constant across all simulated datasets within the same spatial correlation setting. We fit the INLA and MCMC methods to each dataset and compare the performance of each across all datasets. To make these comparisons, we collect multiple pieces of information from each model fit. This includes the posterior means of β_0 , β_1 , κ , and σ_w^2 ; the 95% credible intervals for each parameter; and the run time (in seconds) for each model fit. Using this information, we can estimate the bias, mean squared error (MSE), and coverage for each parameter as well as the average run times for each method. For example, we estimate the bias of the posterior mean of β_0 as

$$\frac{1}{50} \sum_{i=1}^{50} \hat{\beta}_{0i} - \beta_0$$

where $\hat{\beta}_{0i}$ is the posterior mean from the model fit to dataset i and β_0 is the true value of the parameter. Similarly, we estimate the MSE as

$$\frac{1}{50} \sum_{i=1}^{50} (\hat{\beta}_{0i} - \beta_0)^2$$

and the coverage as

$$\frac{1}{50} \sum_{i=1}^{50} I(\beta_0 \in \text{Credible Interval}_i)$$

where $I(\cdot)$ is the indicator function equal to one if the input statement is true and equal to zero otherwise, and “Credible Interval _{i} ” is the 95% quantile-based credible interval obtained from fitting the model to dataset i . We repeat these calculations for each parameter in the study.

For the MCMC method, we base our inference on 20000 posterior samples after removing the first 5000 iterations as a burn-in period. For the MCMC prior choices, we select $\sigma_w^2 \sim \text{Inverse Gamma}(3, 1)$, $\kappa \sim \text{Gamma}(0.10, 0.10)$, and $\beta_0, \beta_1 \stackrel{\text{iid}}{\sim} N(0, 1000)$. The INLA prior choices are the same as those described in Section 3.2. All models are fit separately and on the same computer in order to accurately compare the speed of each method.

In Tables 6 and 7, we display the results from the strong and moderate spatial correlation versions of the simulation study, respectively, averaged across the 50 simulated datasets. With respect to the main regression parameters of interest, INLA and MCMC have very comparable performance in terms of bias, MSE, and coverage across both spatial correlation settings. In the strong spatial correlation setting, INLA results in a larger estimated MSE for κ and σ_w^2 while also having smaller coverage than the 95% credible interval would suggest. However, in the moderate spatial correlation setting the INLA and MCMC results are more similar overall for these parameters. MCMC tends to overestimate σ_w^2 on average in both settings. INLA is estimated to be around 18 (strong spatial correlation) and 3 (moderate spatial correlation) times faster than the MCMC method in this particular simulation study. This is crucial in our data application where the number of unique spatial locations is much larger than 150 (up to 3925 in Chicago). We also note that for a well-behaved simulation study, convergence of the MCMC sampler occurred very quickly for each dataset. However, in practice when the data fail to follow this model exactly, many more posterior samples are required and thinning of the Markov chain may be needed in order to make accurate posterior inference for the parameters of interest. This would further slow the MCMC method in our setting. Overall, INLA has drastically reduced running times while still

providing quality estimation/inference of the main regression parameters of interest in this simulation study.

5. Data application

Over the full period, we found a total of 245, 1750, 439, and 933 unique supermarkets (n) in 244, 1735, 438, and 929 unique spatial locations (m) across Birmingham, AL; Chicago, IL; Minneapolis, MN; and San Francisco, CA; respectively. Similarly, we found a total of 998, 3925, 1363, and 1433 unique convenience stores (n) in 992, 3912, 1348, and 1427 unique spatial locations (m) across the same MSAs over the full period.

The posterior means and 95% quantile-based credible intervals for each of the model parameters are displayed in Tables 8–11 for each MSA. The regression parameters can be interpreted as HRs such that a one unit increase in covariate j leads to a $100(\exp\{\beta_j\} - 1)\%$ increase in the hazard of a store closing. We present these results in parentheses following their discussion and note that because the continuous variables are standardized in our analyses, a one unit increase in these variables corresponds to a one standard deviation increase. These standard deviations are shown in Tables 1–4. Therefore, positive effect estimates suggest an increased probability of store closing while negative effect estimates suggest a decreased probability of store closing.

Increased poverty in a census tract increases the probability that a supermarket closes in Chicago (HR increase: 12.75%). Increased sales volume during a period decreases the probability of store failure across all MSAs (HR increase: –50.84% (Birmingham), –80.80% (Minneapolis), –54.16% (San Francisco), –46.74% (Chicago)). In Chicago, supermarkets in tracts that are racially mixed are more likely to close than supermarkets in tracts that are predominantly White (HR increase: 47.70%). Finally, in Chicago, as the number of competing supermarkets within one km of a store location increases, so does the probability of a supermarket closing (HR increase: 15.03%).

In San Francisco and Birmingham, increased poverty increases the probability of a convenience store closure (HR increase: 20.92% (San Francisco), 69.89% (Birmingham, medium vs. low poverty tertile), 91.55% (Birmingham, high vs. low poverty tertile)). In Minneapolis and Chicago, increased store square footage decreases the probability of store closure (HR increase: –14.79% (Minneapolis), –10.42% (Chicago)). In each MSA other than San Francisco, increased sales volume decreases the probability of store failure (HR increase: –68.65% (Birmingham), –29.53% (Minneapolis), –18.13% (Chicago)). In Chicago, a convenience store in a tract that is predominantly Black is more likely to close than a store in a predominantly White tract (HR increase: 46.23%). In Birmingham, stores in a racially mixed tract are less likely to close than stores in a predominantly White tract (HR increase: –44.57%). In Minneapolis and San Francisco, there appears to be a strong association between competition and store closure, as increasing the number of convenience stores within one km of a store location increases the probability of its failure (HR increase: 27.12% (Minneapolis), 37.71% (San Francisco)). Finally, there was evidence of interaction between square footage and sales volume in Birmingham, Minneapolis, and San Francisco (HR increase: 52.20% (Birmingham), –12.19% (Minneapolis), –14.79% (San Francisco)).

We present posterior mean predictions of the spatial random effects for each MSA in the supermarket (Figure 3) and convenience store (Figure 4) models, allowing us to observe locations where the baseline risk of store closure is higher than expected, after controlling for all available covariates. These regions could inform new hypotheses regarding the causes of the increased risk and/or may suggest relevant spatially-varying covariates that are not currently included in the modeling framework. For the supermarket results in Figure 3, the extreme regions in Birmingham result in an approximately 19.96% increase in the hazard of a store closing based on the spatial location alone. In Minneapolis, the extreme regions result in a 126.05% increase in the hazard of a store closing, a 30.58% increase in San Francisco, and a 61.41% increase in Chicago. Similarly, for the convenience store results in Figure 4, the extreme regions correspond to a 142.41% (Birmingham), 65.93% (Minneapolis), 29.67% (San Francisco), and 36.23% (Chicago) increase in the hazard of a store closing. These pockets of increased risk within the MSAs represent unexplained spatial variability and would not be accounted for without the inclusion of the spatial random effects.

5.1. Model comparisons

We formally compare the spatial version of the models given in (1) with two competing models. First, we fit nonspatial models of the same form that differ only in their removal of the spatial random effects, $w(s_j)$. Second, we fit models that include independent and identically normally distributed census tract random effects that account for similarities in survival outcomes within a census tract. We are interested in determining if there is true spatial variability in the baseline risk of store closure that we are unable to control for given our spatially varying covariates. If the nonspatial or tract random effect versions of the model fit as well or better than the spatial version, then it is likely that this is not the case or that the spatial correlation doesn't extend outside of a tract. We compare the performance of these models using the Watanabe-Akaike information criterion (WAIC) (Watanabe, 2010). WAIC balances model fit and complexity to identify the most appropriate model amongst a group of candidate models, with smaller values of the metric being preferred.

The results in Table 12 suggest that the nonspatial and tract random effect models never meaningfully outperform the spatial models. The effective number of parameters are larger for the tract and spatial models than the nonspatial models as expected, due to the lack of random effects in the nonspatial models.

5.2. Sensitivity analyses

We perform a number of sensitivity analyses to investigate the robustness of our findings to the choice of (i) prior distributions, (ii) spatial scale of our neighborhood covariates, and (iii) method of dealing with tied data. First, we consider two different prior choices for θ_1 and θ_2 . We specify $\theta_1, \theta_2 \stackrel{\text{iid}}{\sim} N(0, 100)$ and $\theta_1, \theta_2 \stackrel{\text{iid}}{\sim} N(0, 1000)$. We rerun all models with these different sets of priors and explore the changes in the main regression parameter estimates of interest. We graphically present the results from these analyses in the supplementary on-line material Figures S1–S8. We compare the estimated posterior means and standard deviations of these regression parameters (intercept excluded) under each prior setting. In the case of low sensitivity to prior choice, we expect these values to be similar across all priors. From

the plots, it is clear that the main associations of interest are not overly sensitive to the choice of prior in the analysis, likely due to the weakly informative prior information being considered in each case.

Next, we explore alternative definitions of neighborhood when defining our covariates. In addition to the original census tract-level covariates, we recreate the covariates at the census block group and census ZIP code tabulation area (ZCTA) levels. Using these covariates, we refit each of the models and present the full results in supplementary on-line material Tables S1–S8 for reference. While differences do exist between results using the different geographies, in our modeling we select census tracts as the most appropriate unit of analysis because by design they are “relatively homogenous units with respect to population characteristics, economic status, and living conditions at the time of establishment” (United States Census Bureau, 2017). Therefore, they may be more representative of a surrounding neighborhood for a specific store than block groups which are generally smaller geographic units and ZCTAs which are larger.

Finally, we explore the use of a different link function in our analyses. The complementary log-log link function is preferable in this setting due to the nature of the discrete ties seen in our data. We know that the closing date of a store isn't actually occurring at the end of a quarterly interval. Instead, the closing dates are happening at some continuous time within an interval and we are only able to view them discretely. By using the complementary log-log link, we are actually obtaining estimates from an underlying continuous-time proportional hazards model. In contrast, use of the more common logit link function assumes that the store closings can only occur at the end of the quarterly intervals. Estimates from that model correspond to a discrete-time proportional odds model (Allison, 2010). However, as a sensitivity analysis, we refit each of the models using the logit link. Because the parameter estimates are no longer directly comparable, we do not present the detailed results from these analyses. However, we note that the results were identical overall with respect to the inclusion/exclusion of zero in the 95% credible intervals for each of the regression parameters (agreement across all fitted models).

6. DISCUSSION

Using longitudinal data and hierarchical spatial modeling methods, we documented spatiotemporal variation in the opening and closure of food stores across four unique MSAs. Our spatial survival model was able to identify a number of factors consistently associated with the survival of different store types. Sales volume across time was typically the strongest indicator of a stores survival probability. In addition, tract-level poverty and racial composition were sometimes observed to be associated with survival probabilities. Stores in high poverty, non-White tracts were often at a disadvantage in terms of length of survival, holding other factors constant. Given the importance of policy decisions that guide store placement and the fact that the current literature is devoid of appropriate modeling techniques to address store opening and closure in a framework that properly accounts for spatial correlation, our statistical approach breaks new ground.

We confirm a greater number of convenience stores relative to supermarkets over time, as we have previously shown (Lamichhane et al., 2014). In contrast to several studies (Morland et al., 2002; Zenk et al., 2005; Moore and Diez Roux, 2006; Morland and Filomena, 2007; Powell et al., 2007), we found greater numbers of supermarkets in high versus low poverty areas and similar to other studies, we found a greater number of convenience stores in high versus low poverty areas (Moore and Diez Roux, 2006; Powell et al., 2007; Sharkey and Horel, 2008; Gustafson et al., 2012). In the current paper, we investigate sociodemographic and internal store characteristics in relation to the survival success of a store over time. Few studies have addressed this issue. One study in Brooklyn showed higher turnover in food stores in lower income areas (Filomena et al., 2013). Another study showed change over a one year period across ZIP codes, with greater decline in supermarkets in high poverty areas (Chen and Wang, 2014). Neither study addressed potential spatial correlation as we did in the current analyses.

Store characteristics were consistently associated with store closure: increased sales volume decreased the probability of supermarket failure in all MSAs and the probability of convenience store failure in all MSAs except San Francisco. There was also a differential effect of sales volume relative to square footage for convenience stores in all MSAs except Chicago. Competition was inconsistently associated with store closure. In contrast, our findings relative to neighborhood sociodemographics were mixed.

In Chicago, we observed a higher probability of store closure in Black vs. White census tracts. Chicago and Birmingham are relatively balanced in terms of predominantly White and non-White populations, whereas Minneapolis has the highest proportion of White population and San Francisco the lowest. Increased poverty was associated with supermarket closure only in Chicago and with convenience store closure in San Francisco and Birmingham. Chicago and Birmingham have a greater proportion of high poverty census tracts, whereas Minneapolis and San Francisco have a greater proportion of low poverty census tracts.

Our analysis of the spatial random effects for each MSA showed geographic areas of each MSA that were associated with increased risk of store closure. We found slightly more spatial variability in convenience stores than in supermarkets. To a certain extent, this is expected as these are smaller stores and there are more of them. However, we were unable to explain the observed patterns with the available covariates. Thus, it would be important for future analyses to understand the factors associated with store closure beyond those we addressed.

While Nielsen TDLinX data are known to be of higher quality (Auchincloss et al., 2012) and are updated monthly, capturing changes in store openings/closings, categorization, ownership, and name (Nielsen, 2013) and thus may provide a more accurate data source of food store availability over time, small and independent food outlets may not be as well captured in TDLinX as larger stores (Rummo et al., 2015). The majority of studies in the literature use Dun & Bradstreet and InfoUSA data to characterize the retail food environment (Liese et al., 2010; Powell et al., 2011; Auchincloss et al., 2012; Fleischhacker et al., 2012; Gustafson et al., 2012; Han et al., 2012); both of which have only moderate

agreement between these data sources and ground-level observations (Liese et al., 2010; Longacre et al., 2011; Powell et al., 2011; Fleischhacker et al., 2012; Gustafson et al., 2012; Han et al., 2012). It is also possible, though unlikely, that stores are opened and closed within the same quarter. In these cases, we would be unaware of their existence given the discrete-time nature of our data and therefore unable to explore the factors associated with their failure.

In conclusion, we found that sales volume was the strongest indicator of store survival across time. In contrast, tract-level sociodemographics were inconsistently associated with store survival probabilities across MSAs, though stores in high poverty, non-White tracts were often at a survival disadvantage, holding other factors constant. In addition, our findings suggest that it is important to address spatial correlation in models that include geographic measures of the retail food environment. The observed patterns of store survival varied by some of the same neighborhood sociodemographic factors associated with lifestyle and health outcomes, which could lead to confusion in interpretation in studies of estimated effects of introduction of food stores into neighborhoods on health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge Marc Peterson of the UNC, Carolina Population Center (CPC); and the CPC Spatial Analysis Unit for the creation of the environmental variables, Tim Monbureau for statistical programming, and Kristin Walls for administrative assistance. This publication was made possible by CTSA Grant Number UL1 TR001863 and KL2 TR001862 from the National Center for Advancing Translational Science (NCATS), components of the National Institutes of Health (NIH), and NIH roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH. This work was also supported by the National Heart, Lung, and Blood Institute at the National Institutes of Health [R01HL114091]. The authors are grateful to the Carolina Population Center, University of North Carolina at Chapel Hill (grant P2C HD050924 from the National Institutes of Health for general support and grant T32 HD007168 for training support); the Nutrition Obesity Research Center, University of North Carolina at Chapel Hill (grant P30DK56350 from the National Institute for Diabetes and Digestive and Kidney Diseases); and the UNC Center for Environmental Health and Susceptibility, University of North Carolina at Chapel Hill (grant P30ES010126 from the National Institute of Environmental Health Sciences).

References

- Allison, PD. Survival analysis using SAS: a practical guide. SAS Institute; 2010.
- Auchincloss AH, Moore KA, Moore LV, Diez Roux AV. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health & Place*. 2012; 18(6):1341–1347. [PubMed: 22883050]
- Blangiardo M, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with r-inla. *Spatial and spatio-temporal epidemiology*. 2013; 7:39–55. [PubMed: 24377114]
- Caspi CE, Sorensen G, Subramanian S, Kawachi I. The local food environment and diet: a systematic review. *Health & Place*. 2012; 18(5):1172–1187. [PubMed: 22717379]
- Chen H-J, Wang Y. The changing food outlet distributions and local contextual factors in the United States. *BMC Public Health*. 2014; 14(1):42. [PubMed: 24433323]
- Cummins S, Petticrew M, Higgins C, Findlay A, Sparks L. Large scale food retailing as an intervention for diet and health: quasi-experimental evaluation of a natural experiment. *Journal of Epidemiology and Community Health*. 2005; 59(12):1035–1040. [PubMed: 16286490]

- Department of Commerce, U. S. Geographic terms and concepts-core based statistical areas and related statistical areas. 2012
- Duffey KJ, Gordon-Larsen P, Steffen LM, Jacobs DR, Popkin BM. Regular consumption from fast food establishments relative to other restaurants is differentially associated with metabolic outcomes in young adults. *The Journal of Nutrition*. 2009; 139(11):2113–2118. [PubMed: 19776183]
- Duran AC, Roux AVD, Maria do Rosario D, Jaime PC. Neighborhood socioeconomic characteristics and differences in the availability of healthy food stores and restaurants in Sao Paulo, Brazil. *Health & Place*. 2013; 23:39–47. [PubMed: 23747923]
- Filomena S, Scanlin K, Morland KB. Brooklyn, New York foodscape 2007–2011: a five-year analysis of stability in food retail environments. *International Journal of Behavioral Nutrition and Physical Activity*. 2013; 10(1):46. [PubMed: 23570574]
- Fleischhacker S, Evenson KR, Rodriguez DA, Ammerman AS. A systematic review of fast food access studies. *Obesity Reviews*. 2011; 12(5):e460–e471. [PubMed: 20149118]
- Fleischhacker SE, Rodriguez DA, Evenson KR, Henley A, Gizlice Z, Soto D, Ramachandran G. Evidence for validity of five secondary data sources for enumerating retail food outlets in seven American Indian communities in North Carolina. *International Journal of Behavioral Nutrition and Physical Activity*. 2012; 9(1):137. [PubMed: 23173781]
- Gordon-Larsen P. Food availability/convenience and obesity. *Advances in Nutrition: An International Review Journal*. 2014; 5(6):809–817.
- Gustafson A, Hankins S, Jilcott S. Measures of the consumer food store environment: a systematic review of the evidence 2000–2011. *Journal of Community Health*. 2012; 37(4):897–911. [PubMed: 22160660]
- Gustafson AA, Lewis S, Wilson C, Jilcott-Pitts S. Validation of food store environment secondary data source and the role of neighborhood deprivation in Appalachia, Kentucky. *BMC Public Health*. 2012; 12(1):688. [PubMed: 22914100]
- Han E, Powell LM, Zenk SN, Rinkus L, Ohri-Vachaspati P, Chaloupka FJ. Classification bias in commercial business lists for retail food stores in the US. *International Journal of Behavioral Nutrition and Physical Activity*. 2012; 9(1):46. [PubMed: 22512874]
- Hollands S, Campbell MK, Gilliland J, Sarma S. Association between neighbourhood fast-food and full-service restaurant density and body mass index: a cross-sectional study of Canadian adults. *Canadian Journal Public Health*. 2014; 105(3):e172–e178.
- Lamichhane AP, Warren JL, Peterson M, Rummo P, Gordon-Larsen P. Spatial-temporal modeling of neighborhood sociodemographic characteristics and food stores. *American Journal of Epidemiology*. 2014; 181:137–150. [PubMed: 25515169]
- Larson NI, Story MT, Nelson MC. Neighborhood environments: disparities in access to healthy foods in the US. *American Journal of Preventive Medicine*. 2009; 36(1):74–81. [PubMed: 18977112]
- Liese AD, Colabianchi N, Lamichhane AP, Barnes TL, Hibbert JD, Porter DE, Nichols MD, Lawson AB. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. *American Journal of Epidemiology*. 2010; 172(11):1324–1333. [PubMed: 20961970]
- Lindgren F, Rue H. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*. 2015; 63(19):1–25.
- Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011; 73(4):423–498.
- Longacre MR, Primack BA, Owens PM, Gibson L, Beauregard S, Mackenzie TA, Dalton MA. Public directory data sources do not accurately characterize the food environment in two predominantly rural states. *Journal of the American Dietetic Association*. 2011; 111(4):577–582. [PubMed: 21443992]
- Martin A, Ogilvie D, Suhrcke M. Evaluating causal relationships between urban built environment characteristics and obesity: a methodological review of observational studies. *International Journal of Behavioral Nutrition and Physical Activity*. 2014; 11(1):142. [PubMed: 25406733]
- Martins TG, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*. 2013; 67:68–83.

- Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health*. 2006; 96(2):325–331. [PubMed: 16380567]
- Morland K, Filomena S. Disparities in the availability of fruits and vegetables between racially segregated urban neighbourhoods. *Public Health Nutrition*. 2007; 10(12):1481–1489. [PubMed: 17582241]
- Morland K, Wing S, Roux AD. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *American Journal of Public Health*. 2002; 92(11):1761–1768. [PubMed: 12406805]
- Morland K, Wing S, Roux AD, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*. 2002; 22(1):23–29. [PubMed: 11777675]
- Nielsen. Service descriptions. 2013
- Nielsen Holdings N.V. Gain a comprehensive view of retail with the leader in location information management with Nielsen TDLinx. 2010
- Pearce J, Hiscock R, Blakely T, Witten K. The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption. *Journal of Epidemiology and Community Health*. 2008; 62(3):198–201. [PubMed: 18272733]
- Pearce J, Hiscock R, Blakely T, Witten K. A national study of the association between neighbourhood access to fast-food outlets and the diet and weight of local residents. *Health & Place*. 2009; 15(1):193–197. [PubMed: 18499502]
- Powell LM, Han E, Chaloupka FJ. Economic contextual factors, food consumption, and obesity among US adolescents. *The Journal of Nutrition*. 2010; 140(6):1175–1180. [PubMed: 20392882]
- Powell LM, Han E, Zenk SN, Khan T, Quinn CM, Gibbs KP, Pugach O, Barker DC, Resnick EA, Myllyluoma J, et al. Field validation of secondary commercial data sources on the retail food outlet environment in the US. *Health & Place*. 2011; 17(5):1122–1131. [PubMed: 21741875]
- Powell LM, Nguyen BT. Fast-food and full-service restaurant consumption among children and adolescents: effect on energy, beverage, and nutrient intake. *JAMA Pediatrics*. 2013; 167(1):14–20. [PubMed: 23128151]
- Powell LM, Slater S, Mirtcheva D, Bao Y, Chaloupka FJ. Food store availability and neighborhood characteristics in the United States. *Preventive Medicine*. 2007; 44(3):189–195. [PubMed: 16997358]
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.
- Richardson AS, Meyer KA, Howard AG, Boone-Heinonen J, Popkin BM, Evenson KR, Kiefe CI, Lewis CE, Gordon-Larsen P. Neighborhood socioeconomic status and food environment: a 20-year longitudinal latent class analysis among CARDIA participants. *Health & Place*. 2014; 30:145–153. [PubMed: 25280107]
- Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009; 71(2):319–392.
- Rummo PE, Gordon-Larsen P, Albrecht SS. Field validation of food outlet databases: the Latino food environment in North Carolina, USA. *Public Health Nutrition*. 2015; 18(6):977–982. [PubMed: 24937758]
- Sharkey JR, Horel S. Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. *The Journal of Nutrition*. 2008; 138(3):620–627. [PubMed: 18287376]
- United States Census Bureau. Decennial management division glossary. 2017
- Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*. 2010; 11(Dec):3571–3594.
- Zenk SN, Schulz AJ, Israel BA, James SA, Bao S, Wilson ML. Neighborhood racial composition, neighborhood poverty, and the spatial accessibility of supermarkets in metropolitan Detroit. *American Journal of Public Health*. 2005; 95(4):660–667. [PubMed: 15798127]

Fig. 1.
Spatial locations of supermarkets that were open at some point during the study period, March 2006 – August 2012. Locations in red indicate the location of a store closing. (Top Left) Birmingham, AL; (Top Right) Chicago, IL; (Bottom Left) Minneapolis, MN; (Bottom Right) San Francisco, CA.

Fig. 2.
Spatial locations of convenience stores that were open at some point during the study period, March 2006 – August 2012. Locations in red indicate the location of a store closing. (Top Left) Birmingham, AL; (Top Right) Chicago, IL; (Bottom Left) Minneapolis, MN; (Bottom Right) San Francisco, CA.

Fig. 3.
Posterior mean predictions from the spatial random field in the supermarket models from
(Top Left) Birmingham, AL; (Top Right) Chicago, IL; (Bottom Left) Minneapolis, MN;
(Bottom Right) San Francisco, CA.

Fig. 4. Posterior mean predictions from the spatial random field in the convenience store models from (Top Left) Birmingham, AL; (Top Right) Chicago, IL; (Bottom Left) Minneapolis, MN; (Bottom Right) San Francisco, CA.

Table 1

Summary of the Birmingham, AL analyses datasets. Summaries are presented averaged over all time periods. Means and proportions are presented with standard deviations shown in parentheses.

Variable	Supermarkets		Convenience Stores	
	Open (4879)	Closed (71)	Open (21757)	Closed (242)
Sq. Foot	25.2 (20.0)	21.1 (16.1)	1.9 (0.7)	1.7 (0.9)
Sales Vol.	238.2 (288.3)	126.8 (90.1)	31.8 (22.8)	25.1 (10.2)
Poverty	0.16 (0.12)	0.18 (0.12)	0.17 (0.12)	0.19 (0.12)
Pops. Dens.	1457 (1425)	1428 (1379)	1309 (1484)	1192 (1501)
CBD Dist.	25.6 (21.4)	21.9 (20.6)	27.8 (22.3)	31.6 (25.3)
# Stores 1 km (Same Type)	0.72 (0.89)	0.76 (0.96)	2.35 (2.18)	2.03 (2.13)
# Stores 1 km (Other Type)	1.85 (2.13)	1.73 (2.15)	0.42 (0.78)	0.39 (0.82)
Tract Race:				
White	0.59	0.42	0.54	0.60
Black	0.20	0.34	0.23	0.24
Mixed	0.21	0.24	0.24	0.16
Hispanic	0.00	0.00	0.00	0.00

Table 2

Summary of the Minneapolis, MN analyses datasets. Summaries are presented averaged over all time periods. Means and proportions are presented with standard deviations shown in parentheses.

Variable	Supermarkets		Convenience Stores	
	Open (8794)	Closed (91)	Open (28885)	Closed (368)
Sq. Foot	28.6 (22.2)	11.1 (12.3)	2.3 (1.0)	2.1 (0.8)
Sales Vol.	325.5 (293.5)	104.0 (110.7)	66.2 (25.2)	58.9 (18.0)
Poverty	0.13 (0.13)	0.18 (0.15)	0.11 (0.11)	0.14 (0.13)
Pops. Dens.	3992 (4087)	5256 (4887)	3458 (3676)	4284 (4237)
CBD Dist.	22.8 (17.5)	20.7 (19.1)	24.7 (18.3)	20.5 (16.5)
# Stores 1 km (Same Type)	0.92 (1.22)	1.22 (1.53)	1.81 (2.09)	2.68 (2.84)
# Stores 1 km (Other Type)	2.05 (2.24)	2.36 (2.34)	0.62 (1.08)	0.82 (1.30)
Tract Race:				
White	0.74	0.60	0.77	0.71
Black	0.00	0.01	0.00	0.00
Mixed	0.26	0.38	0.23	0.29
Hispanic	0.00	0.00	0.00	0.00

Table 3

Summary of the San Francisco, CA analyses datasets. Summaries are presented averaged over all time periods. Means and proportions are presented with standard deviations shown in parentheses.

Variable	Supermarkets		Convenience Stores	
	Open (19901)	Closed (185)	Open (32178)	Closed (214)
Sq. Foot	14.4 (16.8)	10.5 (13.1)	1.8 (0.6)	1.7 (0.6)
Sales Vol.	203.7 (257.1)	95.1 (106.6)	52.2 (26.0)	47.4 (25.6)
Poverty	0.12 (0.09)	0.13 (0.10)	0.12 (0.09)	0.15 (0.12)
Pops. Dens.	17110 (19719)	20576 (23221)	12778 (15728)	19104 (25359)
CBD Dist.	20.8 (12.3)	20.0 (12.1)	23.6 (12.8)	22.4 (14.0)
# Stores 1 km (Same Type)	6.07 (8.68)	6.86 (8.93)	5.76 (7.31)	8.97 (10.57)
# Stores 1 km (Other Type)	1.31 (4.80)	1.64 (5.74)	0.81 (1.89)	1.24 (2.35)
Tract Race:				
White	0.19	0.15	0.14	0.08
Black	0.00	0.01	0.00	0.00
Mixed	0.78	0.82	0.83	0.87
Hispanic	0.03	0.03	0.02	0.04

Table 4

Summary of the Chicago, IL analyses datasets. Summaries are presented averaged over all time periods. Means and proportions are presented with standard deviations shown in parentheses.

Variable	Supermarkets		Convenience Stores	
	Open (34187)	Closed (464)	Open (83500)	Closed (943)
Sq. Foot	22.5 (22.0)	14.8 (17.9)	1.9 (0.8)	1.7 (0.7)
Sales Vol.	243.3 (286.1)	118.0 (154.3)	52.7 (26.8)	46.9 (16.9)
Poverty	0.14 (0.12)	0.19 (0.14)	0.13 (0.11)	0.15 (0.13)
Pops. Dens.	10064 (17475)	12889 (12626)	7486 (9521)	9074 (10899)
CBD Dist.	32.0 (22.9)	25.4 (21.0)	36.8 (23.9)	32.7 (22.9)
# Stores 1 km (Same Type)	2.21 (2.52)	3.09 (2.88)	3.37 (3.74)	3.86 (3.70)
# Stores 1 km (Other Type)	3.22 (3.86)	3.93 (3.94)	1.32 (2.14)	1.55 (2.24)
Tract Race:				
White	0.40	0.23	0.44	0.37
Black	0.14	0.20	0.12	0.19
Mixed	0.38	0.46	0.39	0.38
Hispanic	0.08	0.11	0.05	0.07

Table 5

Average store counts (and standard deviation) across time for the four MSAs and two store types.

MSA	Supermarkets		Convenience Stores	
	Total Open	Closings	Total Open	Closings
Birmingham, AL	190.38 (4.68)	2.73 (2.79)	846.12 (13.15)	9.31 (3.67)
Minneapolis, MN	341.73 (7.72)	3.50 (2.52)	1125.12 (26.94)	14.15 (5.70)
San Francisco, CA	772.54 (18.11)	7.12 (5.60)	1245.85 (67.10)	8.23 (3.79)
Chicago, IL	1332.73 (30.32)	17.85 (10.05)	3247.81 (177.23)	36.27 (12.30)

Table 6
MCMC and INLA simulation study results for the strong spatial correlation setting.

Metric	Parameter				
	Method	β_0	β_1	κ	σ_w^2
Bias (SE)	MCMC	-0.005 (0.005)	0.003 (0.006)	0.327 (0.135)	0.119 (0.007)
	INLA	-0.014 (0.006)	0.001 (0.006)	3.179 (0.529)	0.088 (0.101)
MSE (SE)	MCMC	0.001 (0.000)	0.002 (0.000)	0.999 (0.599)	0.017 (0.002)
	INLA	0.002 (0.000)	0.002 (0.000)	23.800 (7.374)	0.504 (0.303)
Coverage (SE)	MCMC	0.98 (0.02)	0.92 (0.04)	1.00 (0.00)	0.88 (0.05)
	INLA	1.00 (0.00)	0.92 (0.04)	0.68 (0.07)	0.52 (0.07)
Run Time/Seconds (SE)	MCMC	386.16 (1.20)			
	INLA	21.57 (0.28)			

SE: Standard error.

Table 7
MCMC and INLA simulation study results for the moderate spatial correlation setting.

Metric	Parameter				
	Method	β_0	β_1	κ	σ_w^2
Bias (SE)	MCMC	-0.007 (0.005)	0.006 (0.005)	0.543 (0.437)	0.088 (0.011)
	INLA	0.006 (0.012)	0.004 (0.005)	1.508 (0.506)	0.014 (0.015)
MSE (SE)	MCMC	0.001 (0.000)	0.001 (0.000)	9.652 (2.744)	0.014 (0.003)
	INLA	0.007 (0.002)	0.001 (0.000)	14.831 (5.188)	0.011 (0.003)
Coverage (SE)	MCMC	0.94 (0.03)	0.98 (0.02)	0.91 (0.04)	0.98 (0.02)
	INLA	1.00 (0.00)	0.98 (0.02)	0.82 (0.05)	0.90 (0.04)
Run Time/Seconds (SE)	MCMC	297.60 (1.25)			
	INLA	102.45 (1.92)			

SE: Standard error.

Table 8

Birmingham, AL results from the discrete-time spatial survival model fit.

Variable	Supermarkets		Convenience Stores	
	Mean	95% CI	Mean	95% CI
Intercept	-4.46	(-5.44, -3.60)	-4.82	(-5.43, -4.27)
Sq. Foot ^a			0.12	(-0.07, 0.31)
Sales Volume	-0.71	(-1.24, -0.27)	-1.16	(-1.64, -0.70)
(Interaction)			0.42	(0.17, 0.66)
Distance to CBD	-0.23	(-0.62, 0.14)	0.01	(-0.22, 0.25)
# Stores 1 km (Same Type)	-0.02	(-0.28, 0.23)	-0.16	(-0.35, 0.03)
# Stores 1 km (Other Type)	-0.13	(-0.42, 0.15)	0.08	(-0.09, 0.23)
Poverty ^b	-0.10	(-0.45, 0.23)		
Medium vs. Low			0.53	(0.14, 0.93)
High vs. Low			0.65	(0.17, 1.14)
Population Density ^c	-0.26	(-0.64, 0.09)	0.17	(-0.09, 0.44)
Race:				
Black vs. White	0.83	(-0.07, 1.75)	-0.33	(-0.93, 0.26)
Mixed vs. White	0.68	(-0.02, 1.34)	-0.59	(-1.08, -0.12)

^a: log-scaled variable;^b: continuous variable for supermarkets, categorical variable (tertile of census tract poverty level) for convenience stores;^c: log-scaled variable for convenience stores.

Table 9

Minneapolis, MN results from the discrete-time spatial survival model fit.

Variable	Supermarkets		Convenience Stores	
	Mean	95% CI	Mean	95% CI
Intercept	-4.63	(-5.68, -3.70)	-4.28	(-4.71, -3.88)
Sq. Foot			-0.16	(-0.30, -0.02)
Sales Volume	-1.65	(-2.27, -1.12)	-0.35	(-0.48, -0.22)
(Interaction)			-0.13	(-0.26, -0.01)
Distance to CBD	-0.04	(-0.31, 0.22)	-0.20	(-0.41, 0.01)
# Stores 1 km (Same Type)	0.08	(-0.19, 0.34)	0.24	(0.13, 0.34)
# Stores 1 km (Other Type)	-0.11	(-0.40, 0.16)	-0.01	(-0.12, 0.10)
Poverty	-0.10	(-0.42, 0.21)	0.05	(-0.07, 0.17)
Population Density ^a	-0.10	(-0.44, 0.21)	0.02	(-0.19, 0.23)
Race:				
Black vs. White	-0.51	(-1.19, 0.17)	-0.26	(-0.57, 0.04)

^a: log-scaled variable for convenience stores.

Table 10

San Francisco, CA results from the discrete-time spatial survival model fit.

Variable	Supermarkets		Convenience Stores	
	Mean	95% CI	Mean	95% CI
Intercept	-3.88	(-4.39, -3.40)	-4.60	(-5.15, -4.09)
Sq. Foot			-0.13	(-0.28, 0.03)
Sales Volume ^a	-0.78	(-1.09, -0.50)	-0.04	(-0.21, 0.13)
(Interaction)			-0.16	(-0.30, -0.01)
Distance to CBD	0.06	(-0.15, 0.26)	0.15	(-0.01, 0.32)
# Stores 1 km (Same Type)	-0.03	(-0.23, 0.18)	0.32	(0.09, 0.53)
# Stores 1 km (Other Type)	0.03	(-0.12, 0.15)	-0.17	(-0.38, 0.03)
Poverty	0.00	(-0.18, 0.18)	0.19	(0.04, 0.33)
Population Density ^b	-0.01	(-0.21, 0.20)	0.07	(-0.09, 0.23)
Race:				
Black vs. White	-0.09	(-0.56, 0.35)	-0.36	(-0.91, 0.14)

^a: log-scaled for convenience stores;

^b: log-scaled variable for supermarkets and convenience stores.

Table 11

Chicago, IL results from the discrete-time spatial survival model fit.

Variable	Supermarkets		Convenience Stores	
	Mean	95% CI	Mean	95% CI
Intercept	−4.08	(−4.45, −3.73)	−3.90	(−4.16, −3.65)
Sq. Foot			−0.11	(−0.20, −0.03)
Sales Volume	−0.63	(−0.82, −0.46)	−0.20	(−0.29, −0.10)
(Interaction)			−0.09	(−0.18, 0.00)
Distance to CBD ^a	−0.13	(−0.28, 0.02)	−0.09	(−0.19, 0.02)
# Stores 1 km (Same Type)	0.14	(0.01, 0.26)	0.09	(−0.02, 0.19)
# Stores 1 km (Other Type)	−0.08	(−0.22, 0.05)	−0.05	(−0.13, 0.04)
Poverty	0.12	(0.00, 0.23)	0.04	(−0.05, 0.12)
Population Density ^b	−0.05	(−0.20, 0.11)	0.02	(−0.07, 0.10)
Race:				
Black vs. White	0.18	(−0.21, 0.57)	0.38	(0.10, 0.66)
Mixed vs. White	0.39	(0.12, 0.66)	0.02	(−0.15, 0.20)
Hispanic vs. White	0.08	(−0.33, 0.49)	0.27	(−0.07, 0.60)

^a: log-scaled variable for supermarkets;^b: log-scaled variable for supermarkets

Table 12

Watanabe-Akaike information criterion (WAIC) model comparison results. The effective number of parameters is given in parentheses.

Random Effect Specification				
MSA	Store Type	None	Census Tract	Spatial
Bham.	Super.	732.44 (14.10)	732.54 (16.45)	732.91 (19.14)
	Conv.	2618.81 (17.17)	2617.27 (36.61)	2609.24 (83.14)
Minn.	Super.	935.34 (12.20)	934.66 (29.57)	928.82 (43.83)
	Conv.	3836.08 (13.94)	3834.04 (46.90)	3831.96 (72.60)
San Fran.	Super.	2033.29 (12.70)	2033.41 (16.44)	2033.77 (26.92)
	Conv.	2534.61 (15.28)	2533.38 (37.79)	2534.10 (38.61)
Chi.	Super.	4701.51 (14.87)	4701.85 (21.62)	4702.51 (90.65)
	Conv.	10152.00 (17.00)	10151.83 (41.19)	10147.66 (78.35)