Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2016)
179, *Part* 1, *pp.* 293–310

Spatially modelling the association between access to recreational facilities and exercise: the 'Multi-ethnic study of atherosclerosis'

Samuel I. Berchuck,

University of North Carolina at Chapel Hill, USA

Joshua L. Warren,

Yale University, New Haven, USA

Amy H. Herring and Kelly R. Evenson,

University of North Carolina at Chapel Hill, USA

Kari A. B. Moore,

Drexel University, Philadelphia, USA

Yamini K. Ranchod

University of California at Berkeley, USA

and Ana V. Diez-Roux

Drexel University, Philadelphia, USA

[Received September 2014. Revised January 2015]

Summary. Numerous studies have investigated the relationship between the built environment and physical activity. However, these studies assume that these relationships are invariant over space. In this study, we introduce a novel method to analyse the association between access to recreational facilities and exercise allowing for spatial heterogeneity. In addition, this association is studied before and after controlling for crime, which is a variable that could explain spatial heterogeneity of associations. We use data from the Chicago site of the 'Multi-ethnic study of atherosclerosis' of 781 adults aged 46 years and over. A spatially varying coefficient tobit regression model is implemented in the Bayesian setting to allow for the association of interest to vary over space. The relationship is shown to vary over Chicago, being positive in the south but negative or null in the north. Controlling for crime weakens the association in the south with little change observed in northern Chicago. The results of this study indicate that spatial heterogeneity in associations of environmental factors with health may vary over space and deserve further exploration.

Keywords: Environment; Physical activity; Spatially varying coefficients; Tobit regression

1. Introduction

It is estimated that only 3.0% of Americans engage in a fully healthy lifestyle, which entails

Address for correspondence: Samuel I. Berchuck, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 135 Dauer Drive, Chapel Hill, NC 27599-7400, USA.
E-mail: berchuck@live.unc.edu

 0964–1998/16/179293

refraining from smoking, eating five or more fruits and vegetables daily, maintaining a healthy weight and participating in regular exercise (a component of physical activity) (Reeves and Rafferty, 2005). Lack of physical activity is a leading risk factor for chronic disease (US Department of Health and Human Services, 2008) and having access to recreational facilities has been associated with an individual's level of physical activity (Brownson *et al.*, 2000; Troped *et al.*, 2001; Humpel *et al.*, 2002; Huston *et al.*, 2003; Powell *et al.*, 2003; Giles-Corti *et al.*, 2005; Remington *et al.*, 2010) and exercise (Sallis *et al.*, 1990). A perceived presence of such facilities has also been shown to be positively associated with physical activity (Duncan *et al.*, 2005). A number of cross-sectional studies have also shown a positive relationship between features of the built environment potentially conducive to walking (e.g. density of destinations, street network features and proximity to parks) and physical activity (Brownson *et al.*, 2001; Humpel *et al.*, 2002; Hoehner *et al.*, 2005; Gebel *et al.*, 2007; Harris *et al.*, 2013). In particular, the Institute of Medicine (US Institute of Medicine Committee on Accelerating Progress in Obesity Prevention and Glickman, 2012) and the National Prevention Council (National Prevention Council, 2011) both recommended enhancing the built environment to increase physical activity. Recently, longitudinal studies have documented relationships between changes in access to physical activity resources and changes in physical activity over time (Van Cauwenberg *et al.*, 2011; Ranchod *et al.*, 2014).

Most prior work has assumed that associations of the physical activity environment with physical activity are invariant over space. However, it is plausible that associations vary spatially. This may be due to differential distributions and associations of confounders with exposures over space or to spatially varying factors that modify the effects of the physical activity environment on physical activity. In this paper, we analyse spatial heterogeneity in the association of density of physical activity resources with exercise by using buffer level neighbourhood characteristics and individual level data from the Chicago site of the 'Multi-ethnic study of atherosclerosis' (MESA). We allow for the main association of interest (recreational facility access and total exercise) to vary across the city through use of a spatially varying regression coefficient model (Gelfand *et al.*, 2003) for tobit responses. Tobit regression (Tobin, 1958) is used since our outcome variable exercise is zero inflated. A similar model was shown to be effective in capturing the spatial heterogeneities in the relationship of individual level characteristics to physical activity patterns among pregnant women over central North Carolina (Reich *et al.*, 2010). Reich *et al.* (2010) focused on Bayesian variable selection for spatially varying parameters and did not identify any neighbourhood level variables that were spatially varying with respect to activity patterns for pregnant women. Meanwhile, our focus is on better understanding the relationship between exercise and recreational facilities access and how crime impacts this association. Additionally, our zero-inflated model is similar to the model of Reich *et al.* (2010); however, we do not include an extra probability for a zero response, since they had more 0s in their data set due to working only with pregnant women (nearly 80%). Finally, our model allows for location-specific regression slopes and intercepts, leading to increased modelling flexibility and the possibility that the relationship between exercise and recreational facility access varies across Chicago.

To understand further the reasons for spatial heterogeneity in associations of physical activity resources with exercise, we examine the effect of crime. Crime and perceived safety are important environmental predictors of physical activity (Loukaitou-Sideris and Eck, 2007; Foster and Giles-Corti, 2008). It has been suggested that reductions of violence and crime and increased perceptions of neighbourhood safety may contribute to higher population levels of physical activity (Evenson *et al.*, 2012). We hypothesize that the association between access to recreational facilities and individual amounts of exercise will vary spatially because of unaccounted for neighbourhood differences in crime. Crime may be a confounder of the association between physical activity resources with physical activity (and the confounding effects may differ over space if crime is differentially associated with the exposure over space). Crime may also be an effect modifier if, for example, when crime is high the presence of resources has a weaker effect on physical activity because individuals are less likely to utilize these resources.

We first fit the spatially varying regression coefficient tobit model for Chicago without controlling for crime. We then include buffer level crime totals as a covariate and investigate how the spatial relationship between exercise and access to recreational facilities changes. The results are compared with the crime-free model and possible explanations for the observed spatial patterns and changes are discussed. Accounting for spatial dependences in the data is necessary to characterize the association correctly, and ignoring space could result in misleading conclusions regarding the relationship. Therefore, the spatial modelling results are compared with a model that assumes a common association across the city. In Section 2 we introduce the data that are used in the modelling whereas the methods are described in Section 3. We present results from the study in Section 4 and close in Section 5 with conclusions and potential areas of future work.

2. Data

The MESA (`www.mesa-nhlbi.org`) is a longitudinal study of adults ages 45–84 years at enrolment that aims to identify characteristics and risk factors for subclinical atherosclerosis at six study sites in the USA (Bild *et al.*, 2002). These analyses focus on the Chicago site of the MESA because it is the site with the most detailed available crime data. The study was approved by the Institutional Review Boards at Northwestern in Chicago and all participants gave written informed consent. Participants were free of clinical cardio-vascular disease at baseline and were recruited by using a variety of population-based approaches. Among those screened and deemed eligible, the rate of participation was approximately 60%. We use data from examination 2 owing to the completeness of the variables of interest. Of the 1073 MESA Chicago participants at examination 2, 825 had home addresses that had geocoding accuracy of at least ZIP+4 centroid level and had full 1-mile buffers contained entirely within Chicago. (ZIP+4 takes the five-digit zip area and further subdivides the region into more precise geographical sections by using four extra digits.) An additional 44 participants were excluded because of missing covariates, resulting in 781 participants for analysis. Participants who changed locations from examination 1 to examination 2 are included in the analysis group as long as their address at examination 2 fell entirely within Chicago (accounting for the 1-mile buffer). This allows for spatial locations in Chicago that were not originally sampled at the baseline examination to be included in the analysis region. In total, 43 participants out of the 781 changed locations from examination 1 to examination 2.

The primary outcome variable is defined as exercise measured in metabolic equivalents MET minutes per week. A MET is a measure that describes the energy cost that is needed to perform a physical activity, where 1 MET represents the energy cost of a person seated at rest. Participants completed the MESA typical week physical activity survey, adapted from the cross-cultural activity participation study (Ainsworth *et al.*, 1999) that was designed to identify the time spent in and frequency of various physical activities during a typical week in the past month. The survey has 28 items in categories of household chores, lawn, yard, garden or farm, care of children or adults, transportation, walking (not at work), dancing and sport activities, conditioning activities, leisure activities and occupational and volunteer activities. To capture activities that are typically recommended by the US physical activity guidelines (US Department of Health and Human Services, 2008), we used a summary measure for exercise (the sum of walking for exercise, sports or dancing, and conditioning (in MET-hours per week), which was converted into MET-minutes per week (Bertoni *et al.*, 2009). Minutes of activity were summed for each discrete type of activity, converted to hours for ease of presentation and multiplied by MET level (Ainsworth *et al.*, 2012). Additionally, for computational purposes the outcome variable was scaled by 1000.

The primary exposure of interest is defined as the density of recreational facilities within a 1-mile buffer of an individual's residence. Kernel density estimation (Gatrell *et al.*, 1996; Guagliardo, 2004) was used to allow facilities that were closer to a participant's residence to carry more weight than those further away (Ranchod *et al.*, 2014). These densities were estimated by using ArcGIS version 9.2 (Environmental Systems Research Institute, 2006) on the basis of point locations of recreational facilities. The results were not sensitive to the choice of density estimate type (kernel *versus* simple) that was used in the analyses presented. This buffer size was chosen on the basis of past studies including a study on recreational use of active adults (Roux *et al.*, 2007) and an MESA study of physical activity in Chicago that found that the 1-mile buffer was most relevant for typical exercise patterns among participants (Evenson *et al.*, 2012). Recreational facility data were purchased from the national establishment time series database from Walls & Associates (Denver, Colorado). Data were purchased for years 2000–2010 for a total of 133 standard industrial classification codes which were selected on the basis of previous work (Powell *et al.*, 2007). Data for the years 2002–2004 were used and linked to study participants by the year in which examination 2 was administered. In particular, this study included recreational facilities within a 1-mile buffer which include conditioning, recreational, team sports, water activities, water activities conditioning, racquet sports, instructional conditioning, instructional recreational, instructional team sports, instructional water activities and instructional racquet sports. The recreational facilities definition includes both indoor and outdoor activities, where indoor and outdoor are not mutually exclusive categories (i.e. a facility could be categorized as both indoor and outdoor). Densities were expressed as the number of facilities per square mile. Information on outdoor parks is not available in the national environment time series data.

The crime data include the total yearly average of all crimes in a 1-mile buffer surrounding an individual's residence per 1000 people. Police-recorded crime data for years 2001–2012 were obtained from the City of Chicago data portal (City of Chicago, 2011), which houses crime data that occurred within the Chicago city limits. Crimes were excluded from analyses if they were missing any of this information. Types of crime were categorized as assault and battery, criminal offences, incivilities and murder and were coded as indoor or outdoor on the basis of the location of the crime. Crimes with missing location information, or locations listed as automated teller machine, coin-operated machine, and other, were not coded as either indoor or outdoor. However, they were included in the total number of crimes for each category. Crimes in which the location description indicated that it occurred at an airport or airplane were excluded from all analyses, as these were determined not to affect neighbourhood facilities usage or health outcomes significantly. Measures for the total number of incidents within each crime category for buffer sizes of 1 mile around the participants' addresses were created by using ArcGIS and then population-normalized 1-year rates of crime were created. The rates were multiplied by 1000 for a rate of crime per 1000 people.

Additional individual level potential confounders included age, height and weight summarized by body mass index in kilograms per metre squared, race or ethnicity (black or African-American, Chinese-American or Caucasian), gender, arthritis status (yes or no or do not know), marital status (married or living as married, or other), education level (graduate or professional school, college, or high school or less than high school), household income level (more than

Table 1. Summary of continuous variables

Variable	*Mean*	*Standard deviation*	*Minimum*	*2.5% percentile*	*97.5% percentile*	*Maximum*
Exercise (MET-min per week)	1862.14	2310.70	1102.50	0.00	0.00	8688.75
Recreational facilites† (units per mile)	10.80	10.85	4.73	0.36	0.73	33.56
Crime‡	90.09	34.62	80.62	38.29	43.38	156.62
Age (years)	64.08	9.87	64.00	46.00	47.00	82.00
Body mass index (kg m^{-2})	27.18	5.36	26.05	16.00	19.20	40.81

†Kernel density estimate of the number of recreational facilities in a 1-mile buffer.
‡Total yearly average of all crimes in a 1-mile buffer per 1000 people.

Table 2. Summary of categorical variables†

Variable	*Level*	*N*	*Proportion*
Race or ethnicity	Black or African-American	241	0.31
	Chinese-American	98	0.13
	Caucasian	442	0.57
Gender	Female	428	0.55
	Male	353	0.45
Arthritis	Yes	61	0.08
	No or do not know	720	0.92
Marital status	Married or living as married	451	0.58
	Other	330	0.42
Education	Graduate or professional school	291	0.37
	College	378	0.48
	High school or less than high school	112	0.14
Household income	>$100000	275	0.35
	$75000–99999	92	0.12
	$35000–74999	220	0.28
	<$35000	194	0.25
Month of examination 2	January–March	179	0.23
	April–June	171	0.22
	July–September	212	0.27
	October–December	219	0.28
General health	Excellent or very good	508	0.65
	Good	236	0.30
	Fair or poor	37	0.05

†Statistics based on the final sample size of 781 participants from the Chicago MESA study site at examination 2.

$100000, $75000–99999, $35000–74999 or less than $35000), month of examination 2 (January–March, April–June, July–September or October–December) and general health (excellent or very good, good, or fair or poor). Race, gender, marital status, education level and general level of health are all baseline covariates, whereas recreational facilities, crime, exercise, age, body mass index, arthritis status, household income and month of examination 2 were measured at each examination. The summaries of all included continuous variables are displayed in Table 1 and the categorical variable summaries are shown in Table 2.

3. Methods

3.1. Statistical model

To model the spatial relationship between the density of recreational facilities within 1 mile of an individual's residence and exercise, a spatially varying coefficient model was used. First introduced by Gelfand *et al.* (2003), the spatially varying coefficient model is useful because it allows for the spatially varying covariate coefficient to be decomposed into non-spatial and spatial components, thereby increasing modelling flexibility over the domain. We define the observed exercise outcome variable for participant i at location $s(i) \in \{s_1, \ldots, s_q\}$ as $Y_i^*\{s(i)\}$ for $i = 1, \ldots, n$ and $q \leqslant n$, where q is the number of unique locations: 460 in this study. This notation allows for the possibility that multiple participants reside at the same location, because $s(i)$ can map to the same location over different i.

The exercise outcome is zero inflated, with 115 out of 781 individuals reporting no exercise during the study period (15%). To account for this zero inflation, we introduce a first-stage tobit model specification such that

$$Y_i^*\{s(i)\} = \begin{cases} Y_i\{s(i)\} & Y_i\{s(i)\} \geqslant 0, \\ 0 & Y_i\{s(i)\} < 0 \end{cases}$$

where $Y_i\{s(i)\}$ is an unobserved latent variable and our observed outcome variable $Y_i^*\{s(i)\}$ is the realization of the tobit specification. Then, we specify $Y_i\,\{s(i)\}$ by using the spatially varying coefficient model as follows:

$$Y_i\{s(i)\} = \tilde{\beta}_0\{s(i)\} + \tilde{\beta}_1\{s(i)\}x_i + \mathbf{v}_i^{\mathrm{T}}\mathbf{\Lambda} + \varepsilon_i,$$

where $\tilde{\beta}_0\{s(i)\} = \beta_0 + \beta_0\{s(i)\}$, $\tilde{\beta}_1\{s(i)\} = \beta_1 + \beta_1\{s(i)\}$ and $\varepsilon_i \sim^{\mathrm{IID}} N(0, \sigma_\varepsilon^2)$. The covariate x_i is the kernel density estimate of recreational facilities within a 1-mile radius of participant i's location, representing the spatially varying covariate of interest. The vector $\mathbf{v}_i^{\mathrm{T}}$ contains covariates including both the categorical variables and standardized continuous variables, with $\mathbf{\Lambda} = (\beta_2, \ldots, \beta_p)^{\mathrm{T}}$, the corresponding parameters. A complete list of the covariates that were used in the model can be found in Table 1 and Table 2, where $p = 18$. Additionally, in the model in which we control for crime, $\mathbf{v}_i^{\mathrm{T}}$ includes participant i's buffer level crime average and $\mathbf{\Lambda}$ contains a corresponding additional parameter; thus $p = 19$. The spatial relationship is incorporated through use of the spatially referenced intercepts $\beta_0\{s(i)\}$ and slopes $\beta_1\{s(i)\}$, allowing for the main association of interest to change spatially. Note that individuals at the same location have an identical spatial intercept and slope but different random deviation ε_i.

Finally, in addition to the spatially varying coefficient model, a non-spatial tobit (NST) regression model is utilized for comparison. This model is implemented in the Bayesian setting with the same prior distributions and includes the same covariates as the spatially varying coefficient model but does not account for possible spatial heterogeneity in the association. It is used to assess the convergence of the spatial model, as well as to identify geographic areas where a naive (non-spatial) approach may provide misleading results.

3.2. Prior specification

To complete the model specification, we assign prior distributions to the model parameters introduced. In the tobit model, the introduced latent variables $Y_i\{s(i)\}$ are assumed to be conditionally independent and normally distributed with common variance such that

$$Y_i\{s(i)\}|\tilde{\beta}_0\{s(i)\}, \tilde{\beta}_1\{s(i)\}, \mathbf{\Lambda}, \sigma_\varepsilon^2 \overset{\mathrm{ind}}{\sim} N[\tilde{\beta}_0\{s(i)\} + \tilde{\beta}_1\{s(i)\}x_i + \mathbf{v}_i^{\mathrm{T}}\mathbf{\Lambda}, \sigma_\varepsilon^2].$$

Prior distributions for the non-spatial parameters are assigned as follows:

$$\begin{aligned}
\mathbf{\Lambda} &\sim N_{p-1}(\mathbf{0}, \sigma_\beta^2 \mathbf{I}_{p-1}),\\
\beta_0 &\sim N(0, \sigma_\beta^2),\\
\sigma_\varepsilon^2 &\sim \mathrm{IG}(\eta_\varepsilon, \theta_\varepsilon),\\
\beta_1 &\sim N(0, \sigma_\beta^2),
\end{aligned}$$

where $N_d(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_d)$ indicates a multivariate normal distribution of dimension d with mean vector $\boldsymbol{\mu}$ and covariance matrix $\sigma^2 \mathbf{I}_d$ and $\mathbf{I}_d$ is the identity matrix with dimension d. Furthermore, $\mathrm{IG}(\eta, \theta)$ represents an inverse gamma distribution with shape parameter η and scale parameter θ.

The vector of spatial intercepts, $\boldsymbol{\beta}_0^{(s)} = (\beta_0(s_1), \ldots, \beta_0(s_q))^{\mathrm{T}}$, and of spatial slopes, $\boldsymbol{\beta}_1^{(s)} = (\beta_1(s_1), \ldots, \beta_1(s_q))^{\mathrm{T}}$, are each assigned a multivariate normal prior distribution such that $\boldsymbol{\beta}_0^{(s)} | \sigma_0^2, \phi_0 \sim N_q\{0, \sigma_0^2 \boldsymbol{\Sigma}(\phi_0)\}$ and $\boldsymbol{\beta}_1^{(s)} | \sigma_1^2, \phi_1 \sim N_q\{0, \sigma_1^2 \boldsymbol{\Sigma}(\phi_1)\}$, with $\boldsymbol{\Sigma}(\phi)$ the spatial correlation matrix defined as $\boldsymbol{\Sigma}(\phi)_{ij} = \mathrm{corr}\{\beta(s_i), \beta(s_j)|\phi\} = \exp\{-\phi \|s_i - s_j\|\}$, where $\mathrm{corr}(X, Y)$ represents the correlation between random variables X and Y. The covariance matrices of the spatial parameters have an exponential form where σ_0^2 and σ_1^2 are the variances of the spatial processes and ϕ_0 and ϕ_1 represent parameters that control the level of spatial correlation in the data. In addition, a sensitivity analysis to assess the appropriateness of the exponential covariance structure is performed using a spherical structure. Inferences for the hyperparameters are of interest. Thus priors are assigned as follows: $\sigma_0^2 \sim \mathrm{IG}(\eta_0, \theta_0)$, $\sigma_1^2 \sim \mathrm{IG}(\eta_1, \theta_1)$, $\phi_0 \sim U(\alpha_0, \gamma_0)$ and $\phi_1 \sim U(\alpha_1, \gamma_1)$, where $U(a, b)$ represents the uniform distribution with lower bound a and upper bound b.

In particular, relatively uninformative priors are chosen to allow the data to dictate the analysis. Thus, hyperparameters are selected as follows: $\sigma_\beta^2 = 1 \times 10^{10}$, $\eta_\varepsilon = \eta_0 = \eta_1 = 3$, $\theta_\varepsilon = \theta_0 = \theta_1 = 1$, $\alpha_0 = \alpha_1 = 0.001$ and $\gamma_0 = \gamma_1 = 1000$, where the prior bounds for the uniform distribution on ϕ_0 and ϕ_1 are chosen to allow for the maximum and minimum distances between individuals to yield plausible correlation values that range from near 0 (uncorrelated) to near 1 (strong spatial correlation).

3.3. Markov chain Monte Carlo sampling algorithm

To obtain samples from the posterior distributions of the model parameters, we use a data augmentation approach in which we condition on the unobserved latent variables, allowing for the use of Gibbs sampling for a majority of the model parameters (Chib, 1992). We then write the distribution of the latent process in matrix notation:

$$\mathbf{Y} | \boldsymbol{\beta}, \boldsymbol{\beta}_0^{(s)}, \boldsymbol{\beta}_1^{(s)}, \sigma_\varepsilon^2 \sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_0 \boldsymbol{\beta}_0^{(s)} + \mathbf{Z}_1 \boldsymbol{\beta}_1^{(s)}, \sigma_\varepsilon^2 \mathbf{I}_n),$$

where $\mathbf{Y} = (Y_1\{s(1)\}, Y_2\{s(2)\}, \ldots, Y_n\{s(n)\})^{\mathrm{T}}$, each row of the design matrix $\mathbf{X}$ is given by $\mathbf{x}_i = (1, x_i, \mathbf{v}_i^{\mathrm{T}})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \mathbf{\Lambda}^{\mathrm{T}})^{\mathrm{T}}$. Each vector of spatial parameters is multiplied by an $n \times q$ linear transformation matrix, $\mathbf{Z}_0$ and $\mathbf{Z}_1$ for $\boldsymbol{\beta}_0^{(s)}$ and $\boldsymbol{\beta}_1^{(s)}$ respectively, that converts the unique location spatial parameters into parameters for each individual. It is necessary to include the $\mathbf{Z}_0$ and $\mathbf{Z}_1$ to map the spatial parameters to the correct individual and to adjust the dimension of the spatial vectors. Thus, $\mathbf{Z}_0$ and $\mathbf{Z}_1$ include 0s everywhere, except in components z_{ij} where individual i belongs to location s_j. Then, non-zero components of $\mathbf{Z}_0$ take the form $z_{0ij} = 1$ and $\mathbf{Z}_1$ are $z_{1ij} = x_i$, the spatially varying covariate.

The derivations of the full conditional distributions for the parameters can be found in the on-line supporting Web materials. The priors introduced are semiconjugate, with the exception of those for the spatial correlation parameters ϕ_0 and ϕ_1. To obtain samples from the posterior distributions of the parameters, a Gibbs sampler is used with a Metropolis step for ϕ_0 and ϕ_1. An outline of the Markov chain Monte Carlo sampler, demonstrating the sampling of the $(t+1)$th iteration, is given as follows.

(a) Sample

$$Y_i^{(t+1)}\{s(i)\} \sim \begin{cases} \mathrm{TN}(\mu_i^{(t)}, \sigma_\varepsilon^{2(t)}; \leqslant 0) & Y_i^*\{s(i)\}=0, \\ Y_i\{s(i)\} & Y_i^*\{s(i)\}=Y_i\{s(i)\}, \end{cases}$$

where $\mathrm{TN}(\mu_i^{(t)}, \sigma_\varepsilon^{2(t)}; \leqslant 0)$ specifies a truncated normal distribution (truncated above by 0) with $\mu_i^{(t)} = \tilde{\beta}_0^{(t)}\{s(i)\} + \tilde{\beta}_1^{(t)}\{s(i)\}x_i + \mathbf{v}_i^{\mathrm{T}}\mathbf{\Lambda}^{(t)}$.

(b) Sample $\boldsymbol{\beta}^{(t+1)} \sim N_{p+1}(\mathbb{E}_{\boldsymbol{\beta}}^{(t)}, \mathbb{V}_{\boldsymbol{\beta}}^{(t)})$, where $\mathbb{E}_{\boldsymbol{\beta}}^{(t)}$ and $\mathbb{V}_{\boldsymbol{\beta}}^{(t)}$ are as follows:
(i) $\mathbb{V}_{\boldsymbol{\beta}}^{(t)} = \{\mathbf{X}^{\mathrm{T}}(\sigma_\varepsilon^{2(t)}\mathbf{I}_n)^{-1}\mathbf{X} + (\sigma_\beta^2\mathbf{I}_{p+1})^{-1}\}^{-1}$;
(ii) $\mathbb{E}_{\boldsymbol{\beta}}^{(t)} = \mathbb{V}_{\boldsymbol{\beta}}^{(t)}\mathbf{X}^{\mathrm{T}}(\sigma_\varepsilon^{2(t)}\mathbf{I}_n)^{-1}(\mathbf{Y}^{(t+1)} - \mathbf{Z}_0\boldsymbol{\beta}_0^{(s)\,(t)} - \mathbf{Z}_1\boldsymbol{\beta}_1^{(s)\,(t)})$.

(c) Sample $\boldsymbol{\beta}_0^{(s)\,(t+1)} \sim N_q(\mathbb{E}_{\boldsymbol{\beta}_0^{(s)}}^{(t)}, \mathbb{V}_{\boldsymbol{\beta}_0^{(s)}}^{(t)})$, where $\mathbb{E}_{\boldsymbol{\beta}_0^{(s)}}^{(t)}$ and $\mathbb{V}_{\boldsymbol{\beta}_0^{(s)}}^{(t)}$ are as follows:
(i) $\mathbb{V}_{\boldsymbol{\beta}_0^{(s)}}^{(t)} = [\mathbf{Z}_0^{\mathrm{T}}(\sigma_\varepsilon^{2(t)}\mathbf{I}_n)^{-1}\mathbf{Z}_0 + \{\sigma_0^2\,\mathbf{\Sigma}(\phi_0^{(t)})\}^{-1}]^{-1}$;
(ii) $\mathbb{E}_{\boldsymbol{\beta}_0^{(s)}}^{(t)} = \mathbb{V}_{\boldsymbol{\beta}_0^{(s)}}^{(t)}\mathbf{Z}_0^{\mathrm{T}}(\sigma_\varepsilon^{2(t)}\mathbf{I}_n)^{-1}(\mathbf{Y}^{(t+1)} - \mathbf{X}\boldsymbol{\beta}^{(t+1)} - \mathbf{Z}_1\boldsymbol{\beta}_1^{(s)\,(t)})$.

(d) Sample $\boldsymbol{\beta}_1^{(s)\,(t+1)} \sim N_q(\mathbb{E}_{\boldsymbol{\beta}_1^{(s)}}^{(t)}, \mathbb{V}_{\boldsymbol{\beta}_1^{(s)}}^{(t)})$, where $\mathbb{E}_{\boldsymbol{\beta}_1^{(s)}}^{(t)}$ and $\mathbb{V}_{\boldsymbol{\beta}_1^{(s)}}^{(t)}$ are as follows:
(i) $\mathbb{V}_{\boldsymbol{\beta}_1^{(s)}}^{(t)} = [\mathbf{Z}_1^{\mathrm{T}}(\sigma_\varepsilon^{2(t)}\mathbf{I}_n)^{-1}\mathbf{Z}_1 + \{\sigma_1^2\,\mathbf{\Sigma}(\phi_1^{(t)})\}^{-1}]^{-1}$;
(ii) $\mathbb{E}_{\boldsymbol{\beta}_1^{(s)}}^{(t)} = \mathbb{V}_{\boldsymbol{\beta}_1^{(s)}}^{(t)}\mathbf{Z}_1^{\mathrm{T}}(\sigma_\varepsilon^{2(t)}\mathbf{I}_n)^{-1}(\mathbf{Y}^{(t+1)} - \mathbf{X}\boldsymbol{\beta}^{(t+1)} - \mathbf{Z}_0\boldsymbol{\beta}_0^{(s)\,(t+1)})$.

(e) Sample $\sigma_\varepsilon^{2(t+1)} \sim \mathrm{IG}(\eta_\varepsilon + n/2, \xi^{(t)})$, with

$$\xi^{(t)} = \tfrac{1}{2}(\mathbf{Y}^{(t+1)} - \mathbf{X}\boldsymbol{\beta}^{(t+1)} - \mathbf{Z}_0\boldsymbol{\beta}_0^{(s)\,(t+1)} - \mathbf{Z}_1\boldsymbol{\beta}_1^{(s)\,(t+1)})^{\mathrm{T}}(\mathbf{Y}^{(t+1)} - \mathbf{X}\boldsymbol{\beta}^{(t+1)} - \mathbf{Z}_0\boldsymbol{\beta}_0^{(s)\,(t+1)} - \mathbf{Z}_1\boldsymbol{\beta}_1^{(s)\,(t+1)}) + \theta_\varepsilon.$$

(f) Sample

$$\sigma_0^{2(t+1)} \sim \mathrm{IG}\left\{\eta_0 + \frac{q}{2}, \frac{\boldsymbol{\beta}_0^{(s)\,(t+1)^{\mathrm{T}}}\mathbf{\Sigma}(\phi_0^{(t)})^{-1}\boldsymbol{\beta}_0^{(s)\,(t+1)}}{2} + \theta_0\right\}.$$

(g) Sample

$$\sigma_1^{2(t+1)} \sim \mathrm{IG}\left\{\eta_1 + \frac{q}{2}, \frac{\boldsymbol{\beta}_1^{(s)\,(t+1)^{\mathrm{T}}}\mathbf{\Sigma}(\phi_1^{(t)})^{-1}\boldsymbol{\beta}_1^{(s)\,(t+1)}}{2} + \theta_1\right\}.$$

(h) Metropolis step: define a new parameter $\psi_0 = \log\{(\phi_0 - \alpha_0)/(\gamma_0 - \phi_0)\}$, with distribution $f_{\psi_0}(\psi) = \exp(\psi)/\{1+\exp(\psi)\}^2$. Then sample from the proposal distribution, $\psi_0^* \sim N(\psi_0^{(t)}, C_0)$, where C_0 is a tuning parameter for the Metropolis step. Then,

$$\phi_0^* = \gamma_0\exp(\psi_0^* + \alpha_0)/\{1+\exp(\psi_0^*)\}$$

and

$r \propto$

$$\frac{|\boldsymbol{\Sigma}(\phi_0^*)|^{-1/2} \exp(-\frac{1}{2}[\boldsymbol{\beta}_0^{(s)\,(t+1)^{\mathrm{T}}} \{\sigma_0^{2(t+1)}\,\boldsymbol{\Sigma}(\phi_0^*)\}^{-1}\boldsymbol{\beta}_0^{(s)\,(t+1)}])\exp(\psi_0^*)/\{1+\exp(\psi_0^*)\}^2}{|\boldsymbol{\Sigma}(\phi_0^{(t)})|^{-1/2} \exp(-\frac{1}{2}[\boldsymbol{\beta}_0^{(s)\,(t+1)^{\mathrm{T}}} \{\sigma_0^{2(t+1)}\boldsymbol{\Sigma}(\phi_0^{(t)})\}^{-1}\boldsymbol{\beta}_0^{(s)\,(t+1)}])\exp(\psi_0^{(t)})/\{1+\exp(\psi_0^{(t)})\}^2}.$$

Now update ϕ_0 under the condition

$$\phi_0^{(t+1)} = \begin{cases} \phi_0^* & \text{with probability } \min(r,1), \\ \phi_0^{(t)} & \text{with probability } 1-\min(r,1). \end{cases}$$

(i) Repeat the analysis in step (h) for ϕ_1 by using the spatial slope information in place of the spatial intercept information.

The sampler is run for 100000 iterations and the final 50000 samples are kept after burn-in. Additionally, the samples are thinned so that the final number of samples is 10000 and pilot adaptation is used to control the Metropolis acceptance rates. Pilot adaptation is a method that adaptively changes the Metropolis tuning parameters during the burn-in so that the acceptance rates remain stable. An explanation of this technique can be found in Banerjee *et al.* (2003). The analyses are carried out by using R statistical software (R Core Team, 2013). R code for the fully spatial no-crime model (model 1A) can be obtained from

```
http://wileyonlinelibrary.com/journal/rss-datasets
```

3.4. *Spatial prediction*

We have interest in analysing the association between recreational facility access and exercise across Chicago, even in areas where we do not directly observe MESA participants. To do this, we must predict the spatial intercepts and slopes across the domain through use of Bayesian kriging (Handcock and Stein, 1993). Bayesian kriging allows us to interpolate spatially correlated parameters at unobserved locations while properly characterizing the uncertainty in the estimated surface. We choose prediction locations on an equally spaced grid across Chicago, allowing for the predictions to cover the region sufficiently. In Bayesian kriging, our interest is in summarizing the posterior predictive distribution (PPD) of the spatial parameters at unobserved locations. Without loss of generality, we discuss results in terms of the spatial intercepts with the understanding that the spatial slopes are handled similarly. We define $\boldsymbol{\beta}_0^{(s_0)} = (\beta_0(s_{0,1}),\ldots,\beta_0(s_{0,r}))^{\mathrm{T}}$ as the vector of spatial intercepts at unobserved locations $s_{0,1},\ldots,s_{0,r}$ where r is the number of prediction locations included. The PPD is defined as

$$f(\boldsymbol{\beta}_0^{(s_0)}|\mathbf{Y}^*) = \int f(\boldsymbol{\beta}_0^{(s_0)}|\mathbf{Y}^*,\Theta)\, f(\Theta|\mathbf{Y}^*)\,\mathrm{d}\Theta,$$

where $\mathbf{Y}^* = (Y^*\{s(1)\},\ldots,Y^*\{s(n)\})^{\mathrm{T}}$ and Θ represents the vector of all model parameters introduced, including the latent variables. On the basis of the conditional properties of the multivariate normal distribution, we can obtain samples from this PPD by using draws from the posterior distribution of our model parameters. Using composition sampling (Banerjee *et al.*, 2003), we can draw samples jointly from the PPD of $\boldsymbol{\beta}_0^{(s_0)}$ such that for posterior sample t we have

$$\boldsymbol{\beta}_0^{(s_0)\,(t)}|\mathbf{Y}^*,\Theta^{(t)} \sim N_r(\Sigma_{12}^{(t)}\Sigma_{11}^{-1(t)}\boldsymbol{\beta}_0^{(s)\,(t)}, \Sigma_{22}^{(t)} - \Sigma_{12}^{(t)}\Sigma_{11}^{-1(t)}\Sigma_{21}^{(t)}),$$

where

$$\sigma_0^{2(t)} \mathbf{\Sigma}(\phi_0^{(t)}) = \begin{pmatrix} \Sigma_{11_{q\times q}}^{(t)} & \Sigma_{21_{q\times r}}^{(t)} \\ \Sigma_{12_{r\times q}}^{(t)} & \Sigma_{22_{r\times r}}^{(t)} \end{pmatrix}$$

is the full spatial covariance matrix of all locations included (observed and predicted).

4. Results

The study population characteristics for examination 2 in Table 1 and Table 2 suggest that an average participant was 64 years old and exercised 1862 MET-minutes per week. Also, it should be noted that Hispanics were not recruited at the Chicago MESA site. To test our hypothesis that crime explains spatial heterogeneity in the relationship between access to recreational facilities and exercise, we implemented the spatially varying coefficient model. In this analysis we present six models as follows, where fully spatial indicates a model with both spatial intercept and slope and partially spatial indicates a model with spatial slope only. The models are as follows:

(a) a full spatially varying coefficient tobit regression model (slopes and intercepts), not controlling for individual buffer level crime averages (model 1A);
(b) a full spatially varying coefficient tobit regression model (slopes and intercepts), controlling for individual buffer level crime averages (model 1B);
(c) a partial spatially varying coefficient tobit regression model (slopes only), not controlling for individual buffer level crime averages (model 2A);
(d) a partial spatially varying coefficient tobit regression model (slopes only), controlling for individual buffer level crime averages (model 2B);
(e) an NST regression model, not controlling for individual buffer level crime averages (model 3A);
(f) an NST regression model, controlling for individual buffer level crime averages (model 3B).

To assess the validity of the fully spatial models (model 1A and model 1B), diagnostics are performed (Table 3). The purpose of fitting the NST models (model 3A and model 3B) is to have a benchmark to an established model to be used for comparison and to emphasize the need to account for spatial associations. Furthermore, it is desirable to outperform the NST models, since, in the presence of true spatial variation, the spatial models should be more appropriate.

Table 3. Model diagnostics†

Model			*DIC*	p_D	*P*	*G*	D_∞
1A	Fully spatial	No crime	3349.98	70.66	5096.04	3418.65	8514.68
1B		Crime	3349.32	70.53	5102.62	3421.13	8523.75
2A	Partially spatial	No crime	3358.85	28.41	5172.47	3905.22	9077.70
2B		Crime	3359.16	29.08	5179.81	3900.80	9080.61
3A	NST	No crime	3355.24	20.02	5144.62	3982.29	9126.91
3B		Crime	3354.78	20.95	5146.72	3967.01	9113.73

†DIC is a function of the deviance statistic and a model complexity parameter p_D. A smaller value of DIC indicates a better model fit; p_D indicates the number of effective model parameters in the model. G decreases as the goodness of fit increases and P inflates as the model becomes overfitted. Smaller values of $D_\infty = P + G$ are desirable.

The partially spatial models (model 2A and model 2B) are used to verify that the gains in model fit are mainly due to the spatial slope and not the intercept.

To assess model fit, the deviance information criterion DIC and D_∞ are used. DIC is based on the deviance statistic and penalizes for the complexity of a model with an effective number of parameters estimate p_D (Spiegelhalter *et al.*, 2002). The D_∞ posterior predictive measure is an alternative diagnostic tool to DIC, where $D_\infty = P + G$. The G-term decreases as the goodness of fit increases, and P, the penalty term, inflates as the model becomes overfitted, so small values of both of these terms and, thus, small values of D_∞ are desirable (Gelfand and Ghosh, 1998). D_∞ is generally preferred for comparing predictive performance, whereas DIC is preferred for comparing explanatory performance (Banerjee *et al.*, 2003).

The DIC-values for the fully spatial models (model 1A and model 1B) are superior to the partially spatial models (model 2A and model 2B); recall that a smaller DIC is preferred. Furthermore, for the D_∞-values, it can be observed that there is a monotone decreasing trend as more spatial components are included. This suggests that the fully spatial models are better at predicting, which is expected given that each location has its own spatial slope and intercept resulting in increased flexibility. Neither DIC nor D_∞ varies much with the inclusion or exclusion of crime, but the main change in goodness of fit occurs within the spatial structure.

In addition to model diagnostics, a sensitivity analysis for the covariance structure is performed, in which the exponential and spherical structures are compared by using model 1A. Output from this sensitivity analysis is found in Table 1 of the accompanying supporting Web materials, indicating that the exponential and spherical covariance structures are comparable with the exponential model providing improved prediction.

The posterior means for the fixed effects parameters of model 1B are included in Table 4, along with their posterior standard deviations and 95% credible intervals. The model parameters for model 1A have been excluded because the results from both models are virtually identical. The

Table 4. Posterior estimates for the non-spatial parameters in model 1B

Variable	*Coefficient*	*Standard deviation*	*2.5% percentile*	*97.5% percentile*
Intercept	−1.076	0.653	−2.360	0.218
Age (standardized)	0.098	0.103	−0.105	0.301
Body mass index (standardized)	−0.169	0.102	−0.371	0.031
Recreational facilities	−0.250	0.322	−0.812	0.430
Crime (standardized)	−0.191	0.151	−0.488	0.104
Education: college	0.539	0.329	−0.108	1.179
Education: graduate school	0.280	0.373	−0.452	1.010
Income: \$35000–75000	0.363	0.288	−0.205	0.913
Income: \$75000–100000	−0.016	0.375	−0.759	0.713
Income: >\$100000	0.615	0.342	−0.067	1.277
Marital status: yes	−0.140	0.219	−0.572	0.283
Race: Chinese-American	0.738	0.432	−0.115	1.579
Race: black or African-American†	0.733	0.334	0.077	1.398
Gender: male	0.305	0.192	−0.074	0.686
Arthritis: yes†	0.861	0.351	0.161	1.543
Examination period: April–June	0.105	0.284	−0.458	0.665
Examination period: July–September	0.139	0.277	−0.402	0.686
Examination period: October–December	−0.112	0.280	−0.666	0.434
General health: good	0.657	0.494	−0.301	1.622
General health: very good or excellent†	1.489	0.494	0.538	2.461

†95% credible interval does not contain 0.

three most important covariates in terms of explaining individual exercise are the race indicator for black, African-American, the presence of arthritis and having very good or excellent health. The coefficient for arthritis $\{0.861\,(0.161, 1.543)\}$, or scaled $\{861\,(161, 1543)\}$, can be interpreted as the additional MET-minutes per week for individuals with arthritis compared with individuals without arthritis. This is roughly 123 extra MET-minutes per day for individuals with arthritis. Using a jumping rope for 1 min has a MET of 10. Therefore, this is equivalent to somebody using a jumping rope for an extra 12.3 min each day on average. Arthritis patients are often advised by medical professionals to maintain a high level of physical activity including participation in range-of-motion, strengthening and aerobic exercises (Mayo Clinic, 2013). This may help to explain the increase observed in the average exercise amounts in the arthritis group of participants. A similar interpretation can be applied to the indicator of very good or excellent health and the black or African-American indicator. The scaled coefficients for very good or excellent health and black or African-American are given by $\{1489\,(538, 2461)\}$ and $\{733\,(77, 1398)\}$ respectively. Therefore, participants who report very good or excellent health are going to use a jumping rope for 21.3 min more on average each day than an individual with fair or poor health and black or African-Americans will have 10.5 min more exercise each day on average than Caucasians.

The posterior estimate for the fixed effect recreational facilities association in model 1B is given by $\{-0.250\ (-0.812, 0.430)\}$. However, this estimate should not be interpreted as an estimate for the association between recreational facilities and exercise without incorporating the spatial variation for each individual, which is discussed later. We can, however, interpret this estimate in the NST models. The recreational facilities estimate in model 3B is $\{0.020\ (-0.0025, 0.042)\}$ and the results in model 3A are virtually identical. Since the 95% credible interval contains 0, the effect of recreational facilities on exercise is negligible when controlling for covariates. However, these results can be misleading since different areas of Chicago appear to have different coefficient sizes (Fig. 1). Furthermore, through use of the spatially varying coefficient model, the association is shown to vary over Chicago and in certain areas the magnitude of the association is much larger.

Heat maps are created to quantify the spatial variation in the spatial slopes and intercepts. The heat maps are created from the samples from the PPD of the spatial parameters as described in Section 3.4. All the parameters converged in the Markov chain Monte Carlo sampler; however, the parameter corresponding to the spatial smoothness of the spatial intercept, ϕ_0, converged to its prior distribution in both model 1A and model 1B (posterior summaries of spatial covariance parameters and model variances in model 1A and model 1B are displayed in Table 2 of the supporting Web materials). This suggests that a constant intercept is appropriate in this setting and that the random spatial deviations to the intercept are not different from 0 across the domain. Therefore, the heat maps for the spatial intercepts in both models appear as white noise due to the lack of spatial variation. For this reason, only the heat maps for the spatial slopes are presented, though the spatial intercept plots are displayed in Fig. 1 of the supporting Web materials. The displayed estimates in the heat maps represent the sum of the non-spatial and spatial parameters at each new location, $\beta_1 + \beta_1(s_{0,k})$, for $k = 1, \ldots, r$. The 95% credible regions for these location-specific estimates all included 0, but spatial heterogeneity is present. In fact, model 1A yielded significant location-specific slopes with 90% credible intervals not containing 0 in the Near North Side neighbourhood of Chicago, which is the most densely populated study location (seen in Fig. 1 in the yellow part of the standard errors maps). However, after controlling for crime (model 1B) none of the location-specific slopes had 90% credible intervals not containing zero.

In Fig. 1, the heat maps for the spatial slope PPD means and standard deviations for both model 1A and model 1B are displayed. Figs 1(a) and 1(b) show the PPD means of the

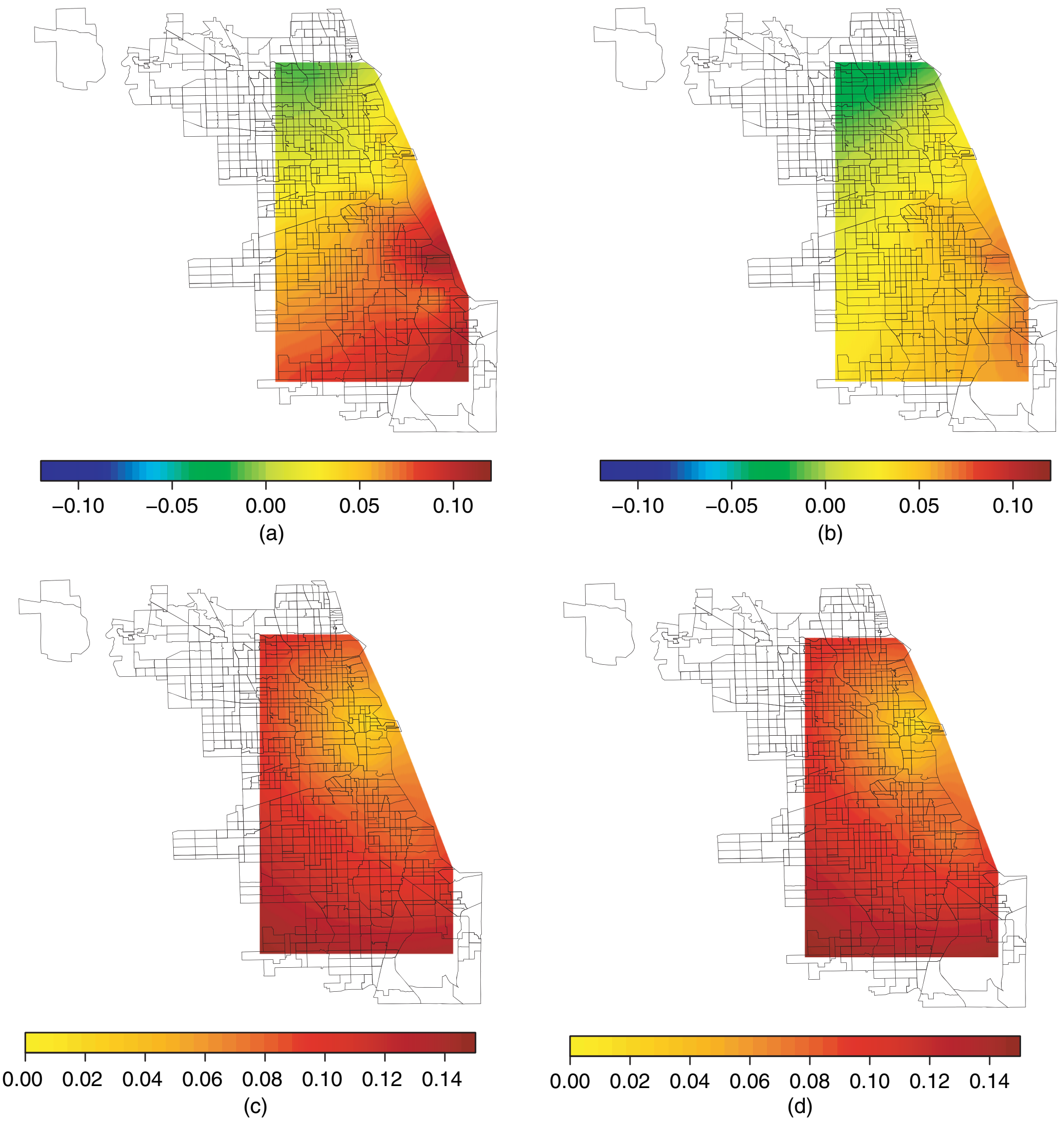

Fig. 1. Spatial slope posterior predictive means and standard deviations (it can be seen that once a crime has been controlled for the PPD means become smaller, the only exception being parts of the north, which become more negative; posterior standard deviations vary over the region of Cook County but do not change over models): (a) model 1A means; (b) model 1B means; (c) model 1A standard deviations; (d) model 1B standard deviations

association between recreational facilities and exercise for models 1A and 1B respectively. The spatial variation is similar over the two models, and it appears that, in locations of extreme magnitude, there is attenuation from the no-crime model to the crime model results. Furthermore, from Fig. 1 it can be seen that the association is slightly negative in the north and positive in the south. Clearly, assuming a constant association is misleading in Chicago. Figs 1(c) and 1(d) contain plots of the spatial variation of the PPD standard deviations. This is a useful indicator of the general location of the participants who were included in the study, since the standard deviation

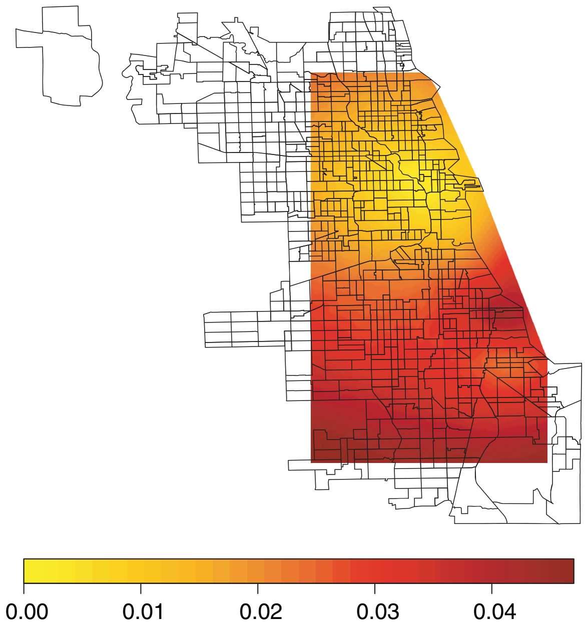

Fig. 2. Mean difference in the posterior predictive slopes from model 1A to model 1B: ■, locations with the greatest mean change after controlling for crime; □, regions in the north representing essentially no change in the association between recreational facilities and exercise after controlling for crime

is lower in locations with numerous participants. There are pockets of low standard deviations in both north and south Chicago and especially along Lake Michigan.

The difference between the PPD means of the slopes in the full no-crime and crime models can be more easily understood in Fig. 2, which shows the mean change in the association between recreational facilities and exercise after controlling for crime; thus a positive value on this plot indicates a weakening association. This is represented by the red and yellow regions in Fig. 2. There is a clear peak in the south of Chicago, indicating that the presence of crime impacts the relationship between access to recreational facilities and exercise, since the association is weakened. In contrast, in the north side of Chicago the yellow region indicates little to no change in the association, after controlling for crime.

5. Discussion

We hypothesized that the relationship between access to recreational facilities and exercise varies spatially over Chicago. Through the implementation of the spatially varying coefficient model, it is clear that there is a spatial structure that underlies this relationship. In fact, there appears to be a difference in this association between the north and south sides of Chicago. The north side of Chicago is characterized as being the most densely populated residential area of Chicago, which is mainly populated by middle and upper class residents, and characterized by having public parkland and residential high-rise buildings (Chicago Tribune Communities, 2014; City of Chicago, 2014). Meanwhile, the south side is known to have a higher proportion of single-family homes, to contain most of the city's remaining industry, to have public parkland, to have large immigrant and African-American populations historically and to have higher rates of poverty and crime (Bell and Jenkins, 1993; Nyden *et al.*, 2006; Sampson, 2012). Additionally, the north side has more recreational facilities in individual buffers than the south side. This picture of Chicago provides context to interpret the heat maps in Fig. 1. In Fig. 1(a), we view the mean slopes for model 1A, where it can be seen that in the south side the association is positive. On the basis of the characteristics of the south side of Chicago, we can conclude that stronger associations of density of recreational facilities with physical activity are observed in areas with a lower density of facilities. Therefore, because of the large variation of access to recreational facilities that exists in the south, the association of interest becomes more detectable. However, in the north side, where the density of recreational facilities is higher, the association is attenuated or suppressed. This may be because everyone has access to recreational facilities and therefore there is little variation to drive the inference.

In addition to exploring the spatial nature of the relationship between access to recreational facilities and exercise, we hypothesized that the spatial heterogeneity was at least partially explained by crime. In particular, we hypothesized that, after adjusting for crime, the spatial heterogeneity that existed in the no-crime model would be reduced. From Fig. 1, we can recognize that the spatial nature of this relationship was not nullified when crime was included in the model; however, the magnitude of the spatial variation is clearly suppressed. This suppression can be observed in Fig. 2 as well. The most drastic of these changes occur in the south side of Chicago, where the mean differences in Fig. 2 reach almost 0.04 recreational facilities in a 1-mile buffer. In the north side, this recreational facilities effect is only slightly increased when controlling for crime, viewed in the yellow areas in Fig. 2. In areas where the slope noticeably decreases after controlling for crime, as in the cluster in the south, recreational facilities and crime are negatively associated. Thus crime operates as a positive confounder resulting in an overestimate of the true causal association between physical activity resources and exercise. Once crime has been controlled for we can detect a more valid (and weaker) estimate of the recreational facility association with exercise.

The previous interpretation is assuming that crime is a confounder. It is also possible that crime is an effect modifier and the varying results can be explained through ways that crime operates in different areas. Additionally, it is possible that crime is not the main factor driving the spatial variation of the association of access to recreational facilities and exercise, but rather is a consequence of other factors such as neighbourhood levels of poverty or affluence. We control for individual level income and education covariates in the analyses, but in future work it would be interesting to investigate the spatial variation before and after controlling for neighbourhood versions of these variables. Finally, since the MESA study includes individuals changing locations between examinations 1 and 2, self-selection may be an underlying issue for the population. In particular, Eid *et al.* (2008) found that obese individuals self-select living in sprawling neighbourhoods (i.e. no built environment) and therefore an association between access to recreational facilities and exercise may not indicate a causal effect but rather is a consequence of this self-selection. However, only 43 out of the 781 participants moved from examination 1 to examination 2.

These distinctions highlight a limitation of this study that could be addressed in future work. In the future it will be beneficial to utilize the longitudinal nature of the MESA study to perform a spatiotemporal analysis of the relationship between access to recreational facilities and exercise. This will allow a more thorough understanding of the potential mediating role of crime. In addition to expanding the model to include the longitudinal framework of the MESA study, it will be useful to incorporate a more complex treatment of the spatially varying covariate. Similarly to Powell *et al.* (2007), we acknowledge that the exclusion of parks from the density of recreational facilities will reduce generalizability of the results. In particular, Humpel *et al.* (2002) highlighted the association between proximity to parks and physical activity. The Lake Michigan parks are a unique recreational setting and accounting for them in future analyses could provide additional insight into the association of interest in this area of Chicago. Additionally, in this study our main interest was to determine whether the relationship between recreational facility access and exercise amounts was spatially varying, but it would be possible to allow the effects of crime to vary over space as well.

Additionally, the results of this study are limited to the region analysed because of the spatially referenced nature of the model. Although we can successfully predict the association of interest in areas surrounding where we observe data by using the spatial model, these predictions are not valid outside that region. This can be viewed in the posterior standard deviation plots of Fig. 1 where uncertainty in parameter estimates is increased because of the lack of observed data and weakening effect of spatial correlation at large distances. This is why we provide predictions only in areas contained by observed data (Fig. 1). Similarly to the findings of Evenson *et al.* (2012), generalizing these results may not be advisable since we worked with a single city and older adult population (ages 45–84 years) with the exclusion of Hispanics.

Finally, we note that alternative techniques exist to model spatial heterogeneity in the regression parameters with the most common method being geographically weighted regression. However, Finley (2011) compared geographically weighted regression with Bayesian spatially varying coefficient models and concluded that spatially varying coefficient models were generally superior to geographically weighted regression because of their increased modelling flexibility and unified inferential Bayesian framework. Geographically weighted regression was found to be less computationally demanding but was ultimately recommended for use as an exploratory analysis tool because of its lack of inferential framework.

In conclusion, there are important differences in the relationship between recreational facility access and total exercise observed across the north side and south side of Chicago that are missed when the commonly applied NST models are implemented. The north side has more recreational facilities, but the associations of facilities with exercise are not as strong as in the south side. These differences may be due to differential associations of crime (an important confounder) with recreational facilities in the north and south sides. Our results suggest that spatial heterogeneity in associations, and the reasons for them, need to be better characterized to develop improved causal inferences regarding neighbourhood health effects.

Acknowledgements

The analyses were funded by the National Institutes of Health–National Heart, Lung, and Blood Institute, grant 2R01 HL071759, the National Institutes of Health (Fuentes, grant 2R01ES014843-04A1) and the National Institute of Environmental Health Sciences (Herring, grant R01ES020619; Herring, grant T32ES007018; Swenberg, grant P30ES010126). The MESA study was supported by contracts N01-HC-95159–N01-HC-95169 from the National Institutes of Health–National Heart, Lung, and Blood Institute and by grants UL1-RR-024156 and UL1-RR-025005 from the National Center for Research Resources. The authors thank Fang Wen for help with data management and the other investigators, staff and participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at `http://www.mesa-nhlbi.org`. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Ainsworth, B., Haskell, W., Herrmann, S., Meckes, N., Bassett, D., Tudor-Locke, C., Greer, J., Vezina, J, Whitt-Glover, M. and Leon, A. (2012) The compendium of physical activities tracking guide. Healthy Lifestyles Research Center, College of Nursing and Health Innovation, Arizona State University, Phoenix.

Ainsworth, B. E., Irwin, M. L., Addy, C. L., Whitt, M. C. and Stolarczyk, L. M. (1999) Moderate physical activity patterns of minority women: the cross-cultural activity participation study. *J. Women's Hlth Gend. Med.*, **8**, 805–813.

Banerjee, S., Gelfand, A. E. and Carlin, B. P. (2003) *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: CRC Press.

Bell, C. C. and Jenkins, E. J. (1993) Community violence and children on Chicago's southside. *Psychiatry*, **56**, 46–54.

Bertoni, A. G., Whitt-Glover, M. C., Chung, H., Le, K. Y., Barr, R. G., Mahesh, M., Jenny, N. S., Burke, G. L. and Jacobs, D. R. (2009) The association between physical activity and subclinical atherosclerosis: the Multi-Ethnic Study of Atherosclerosis. *Am. J. Epidem.*, **169**, 444–454.

Bild, D. E., Bluemke, D. A., Burke, G. L., Detrano, R., Diez-Roux, A. V., Folsom, A. R., Greenland, P., Jacobs, Jr, D. R., Kronmal, R., Liu, K., Clark Nelson, J., O'Leary, D., Saad, M. F., Shea, S., Szklo, M. and Tracy, R. P. (2002) Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am. J. Epidem.*, **156**, 871–881.

Brownson, R. C., Baker, E. A., Housemann, R. A., Brennan, L. K. and Bacak, S. J. (2001) Environmental and policy determinants of physical activity in the United States. *Am. J. Publ. Hlth*, **91**, 1995–2003.

Brownson, R. C., Housemann, R. A., Brown, D. R., Jackson-Thompson, J., King, A. C., Malone, B. R. and Sallis, J. F. (2000) Promoting physical activity in rural communities: walking trail access, use, and effects. *Am. J. Prev. Med.*, **18**, 235–241.

Chib, S. (1992) Bayes inference in the Tobit censored regression model. *J. Econmetr.*, **51**, 79–99.

Chicago Tribune Communities (2014) Chicago neighborhoods. Chicago Tribune Communities, Chicago. (Available from `http://www.chicagotribune.com/classified/realestate/communities/`.)

City of Chicago (2011) City of Chicago data portal. City of Chicago, Chicago (Available from `https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2`.)

City of Chicago (2014) Community area 2000 and 2010 Census population comparisons. City of Chicago, Chicago. (Available from `http://www.cityofchicago.org`.)

Diez-Roux, A. V., Evenson, K. R., McGinn, A. P., Brown, D. G., Moore, L., Brines, S. and Jacobs, Jr, D. R. (2007) Availability of recreational resources and physical activity in adults. *Am. J. Publ. Hlth*, **97**, 493–499.

Duncan, M. J., Spence, J. C. and Mummery, W. K. (2005) Perceived environment and physical activity: a meta-analysis of selected environmental characteristics. *Int. J. Behav. Nutrn Physcl Activ.*, **2**, article 11.

Eid, J., Overman, H. G., Puga, D. and Turner, M. A. (2008) Fat city: questioning the relationship between urban sprawl and obesity. *J. Urb. Econ.*, **63**, 385–404.

Environmental Systems Research Institute (2006) *ARCGIS*, *Version 9.2*. Redlands: Enivronmental Systems Research Institute.

Evenson, K. R., Block, R., Diez-Roux, A. V., McGinn, A. P., Wen, F. and Rodríguez, D. A. (2012) Associations of adult physical activity with perceived safety and police-recorded crime: the Multi-Ethnic Study of Atherosclerosis. *Int. J. Behav. Nutrn Physcl Activ*., **9**, article 146.

Finley, A. O. (2011) Comparing spatially-varying coeffcients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Meth. Ecol. Evoln*, **2**, 143–154.

Foster, S. and Giles-Corti, B. (2008) The built environment, neighborhood crime and constrained physical activity: an exploration of inconsistent findings. *Prev. Med.*, **47**, 241–251.

Gatrell, A. C., Bailey, T. C., Diggle, P. J. and Rowlingson, B. S. (1996) Spatial point pattern analysis and its application in geographical epidemiology. *Trans. Inst. Br. Geog.*, **21**, 256–274.

Gebel, K., Bauman, A. E. and Petticrew, M. (2007) The physical environment and physical activity: a critical appraisal of review articles. *Am. J. Prev. Med.*, **32**, 361–369.

Gelfand, A. E. and Ghosh, S. K. (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika*, **85**, 1–11.

Gelfand, A. E., Kim, H.-J., Sirmans, C. and Banerjee, S. (2003) Spatial modeling with spatially varying coefficient processes. *J. Am. Statist. Ass.*, **98**, 387–396.

Giles-Corti, B., Broomhall, M. H., Knuiman, M., Collins, C., Douglas, K., Ng, K., Lange, A. and Donovan, R. J. (2005) Increasing walking: how important is distance to, attractiveness, and size of public open space? *Am. J. Prev. Med.*, **28**, 169–176.

Guagliardo, M. F. (2004) Spatial accessibility of primary care: concepts, methods and challenges. *Int. J. Hlth Geog.*, **3**, article 3.

Handcock, M. S. and Stein, M. L. (1993) A Bayesian analysis of kriging. *Technometrics*, **35**, 403–410.

Harris, J. K., Lecy, J., Hipp, J. A., Brownson, R. C. and Parra, D. C. (2013) Mapping the development of research on physical activity and the built environment. *Prev. Med.*, **57**, 533–540.

Hoehner, C. M., Brennan Ramirez, L. K., Elliott, M. B., Handy, S. L. and Brownson, R. C. (2005) Perceived and objective environmental measures and physical activity among urban adults. *Am. J. Prev. Med.*, **28**, 105–116.

Humpel, N., Owen, N. and Leslie, E. (2002) Environmental factors associated with adults participation in physical activity: a review. *Am. J. Prev. Med.*, **22**, 188–199.

Huston, S. L., Evenson, K. R., Bors, P. and Gizlice, Z. (2003) Neighborhood environment, access to places for activity, and leisure-time physical activity in a diverse North Carolina population. *Am. J. Hlth Promn*, **18**, 58–69.

Loukaitou-Sideris, A. and Eck, J. E. (2007) Crime prevention and active living. *Am. J. Hlth Promn*, **21**, 380–389.

Mayo Clinic (2013) Exercise helps ease arthritis pain and stiffness. Mayo Clinic, Rochester. (Available from `http://www.mayoclinic.org/diseases-conditions/arthritis/in-depth/arthritis/art-20047971`.)

National Prevention Council (2011) *National Prevention Strategy: America's Plan for Better Health and Wellness.* Washington DC: US Department of Health and Human Services Office of the Surgeon General.

Nyden, P. W., Edlynn, E. and Davis, J. (2006) The differential impact of gentrification on communities in Chicago. Loyola University Chicago Center for Urban Research and Learning, Chicago.

Powell, L. M., Chaloupka, F. J., Slater, S. J., Johnston, L. D. and OMalley, P. M. (2007) The availability of local-area commercial physical activity—related facilities and physical activity among adolescents. *Am. J. Prev. Med.*, **33**, suppl., S292–S300.

Powell, K. E., Martin, L. M. and Chowdhury, P. P. (2003) Places to walk: convenience and regular physical activity. *Am. J. Publ. Hlth*, **93**, 1519–1521.

Ranchod, Y. K., Roux, A. V. D., Evenson, K. R., Sánchez, B. N. and Moore, K. (2014) Longitudinal associations between neighborhood recreational facilities and change in recreational physical activity in the Multi-Ethnic Study of Atherosclerosis, 2000–2007. *Am. J. Epidem.*, **179**, 335–343.

R Core Team (2013) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reeves, M. J. and Rafferty, A. P. (2005) Healthy lifestyle characteristics among adults in the United States, 2000. *Arch. Intern. Med.*, **165**, 854–857.

Reich, B. J., Fuentes, M., Herring, A. H. and Evenson, K. R. (2010) Bayesian variable selection for multivariate spatially varying coefficient regression. *Biometrics*, **66**, 772–782.

Remington, P. L., Brownson, R. C. and Wegner, M. V. (2010) *Chronic Disease Epidemiology and Control*, 3rd edn. Washington DC: American Public Health Association.

Sallis, J. F., Hovell, M. F., Hofstetter, C. R., Elder, J. P., Hackley, M., Caspersen, C. J. and Powell, K. E. (1990) Distance between homes and exercise facilities related to frequency of exercise among San Diego residents. *Publ. Hlth Rep.*, **105**, 179–185.

Sampson, R. J. (2012) *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago: University of Chicago Press.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.

Tobin, J. (1958) Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24–36.

Troped, P. J., Saunders, R. P., Pate, R. R., Reininger, B., Ureda, J. R. and Thompson, S. J. (2001) Associations between self-reported and objective physical environmental factors and use of a community rail-trail. *Prev. Med.*, **32**, 191–200.

US Department of Health and Human Services (2008) *Physical Activity Guidelines for Americans.* Washington DC: US Department of Health and Human Services and Office of Disease Prevention and Health Promotion.

US Institute of Medicine Committee on Accelerating Progress in Obesity Prevention and Glickman, D. (2012) *Accelerating Progress in Obesity Prevention: Solving the Weight of the Nation.* Washington DC: National Academies Press.

Van Cauwenberg, J., De Bourdeaudhuij, I., De Meester, F., Van Dyck, D., Salmon, J., Clarys, P. and Deforche, B. (2011) Relationship between the physical environment and physical activity in older adults: a systematic review. *Hlth Place*, **17**, 458–469.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supporting Web materials for "Spatially modeling the association between access to recreational facilities and exercise: the Multi-Ethnic Study of Atherosclerosis"'.