Explaining Variations in Obesity and Inactivity between US Metropolitan Areas

Peter Congdon, Queen Mary University of London; p.congdon@qmul.ac.uk

Abstract This paper discusses measurement of the main dimensions of the urban environment that have been proposed as relevant to explaining geographic variations in obesity and inactivity. It considers urban sprawl, food access and exercise access as latent constructs, defined by sets of observed indicators for areas. In an application to 993 US metropolitan counties, the paper shows how these latent constructs may be incorporated in an ecological (area-scale) model, which recognizes spatial aspects in the patterning of both outcomes and environmental factors. Urban sprawl and area socioeconomic status emerge from regression modelling as leading influences on obesity and inactivity.

Keywords Obesity; Inactivity; Sprawl; Food Access; Income; Spatial correlation.

Introduction

It is increasingly recognised that, in addition to individual health behaviours and genetic factors, aspects of the urban environment may affect levels of obesity and physical activity (Hill and Peters, 1998; Jeffery and Utter, 2003; Feng et al, 2012; Black et al, 2008). In the USA, significant geographic variation in age standardised obesity and inactivity has been reported (Ford et al, 2005), and the nature of their link to urban environments is therefore an issue (Levine, 2007). This paper discusses measurement of the main dimensions of the urban environment (environmental constructs) proposed as relevant to explaining obesity and inactivity. In an application to 993 US metropolitan counties, it then shows how they may be incorporated in an ecological (area-scale) model. The model recognizes spatial aspects in the patterning of obesity, inactivity and the environmental constructs. By contrast, other studies consider spatial clustering or spatial heterogeneity in regression (Koh, 2011; Schuurman et al, 2009; von Hippel and Benson, 2014; Black, 2014; Chi et al, 2013).

Environmental influences on obesity and inactivity

In attempts to explain ecological variations in obesity and physical activity, much recent research has highlighted the potential influence of dispersed urban development (urban sprawl) (Ewing et al, 2006; Lopez, 2004; Garden and Jalaludin, 2009; Cho et al, 2006). Sprawl has been characterised as low density or leapfrog development, with segregated land uses, low walkability, and high automobile dependence (Frumkin, 2002; Ewing & Hamidi, 2010; Saelens et al, 2003; Lopez and Hynes, 2003). Automobile dependence and other aspects of recently developed urban environments may act to discourage activity and increase obesity risk (Lopez and Hynes, 2006; Berrigan and Troiano, 2002).

Also implicated in the rise in obesity rates is the urban food environment (Bodor et al, 2010; Chi et al, 2013). Thus easy access to fast-food restaurants and convenience stores, combined with low access to supermarkets and large grocery stores, is seen as a potential source for unhealthy diets and overconsumption of processed food (French et al, 2001).

Similarly while physical activity may be discouraged by sprawl, also relevant is access to opportunities for such activity (Owen et al, 2004; Parks et al, 2003; Jilcott et al, 2011; Berrigan and

Troiano, 2002; Sallis et al, 2012; Gladwell et al, 2013), though the exercise environment and sprawl may be intercorrelated to some degree.

Obesity and inactivity are also correlated with poverty and low income levels (i.e. measures of socioeconomic status, SES), both at individual and area scales, although trends in these effects are subject to debate (McLaren, 2007). Restricted access to healthy food outlets for lower income groups, the relative costs of different food types, and location behaviour of food retailers, may partly explain relationships between area SES and obesity (Ver Ploeg, 2010; Drewnowski and Specter, 2004; Levine, 2011; Moore and Diez Roux, 2006; Bennett et al, 2011). Effects of area socio-economic status on obesity or inactivity may be subject to a deprivation amplification effect (Macintyre, 2007), for example, due to effects on food access patterns of concentrations of low income or high income groups. Similarly underlying causes of poverty-inactivity associations may be lesser access to recreation facilities, parks, and natural/greenspace environments for lower income groups.

Measurement issues

In order to carry out an ecological analysis of environmental influences on variations in obesity and inactivity, measurement issues are important. It is necessary to operationalize what are essentially multidimensional latent constructs, such as sprawl or food access, which are imperfectly recognized in any single observed index (Nardo et al, 2005). Ideally the constructs are designed to be optimal in terms of explaining variations in obesity and inactivity.

Indicators of various aspects of urban sprawl have been proposed, and the complexities of measuring urban sprawl noted in several studies (Lopez, 2004; Lopez, 2014). A recent analysis (Ewing and Hamidi, 2014) proposes four indices of sprawl (and conversely compactness): density, mixed use (neighborhood mix of homes, jobs and services); centering (strength of activity centers and downtowns); and street accessibility. The study of Ewing and Hamidi (2014) proposes a composite score of compactness (i.e. a negative index of spread), obtained by summing the four indices. However, simple summation may not provide optimal weighting of the different components in terms of explaining variations in obesity and inactivity. Appendix 1 contains definitions of these four indices, and other input indices used in the subsequent analysis in the current paper.

Measurement of food access in a summary index or indices has attracted both research and policy interest (Neckerman et al, 2009; Jilcott et al, 2010a). Policy interest is apparent in the development of the Food Environment Atlas (USDA, 2014a) which includes small area measures of supermarket, grocery and convenience store availability and access, as well as comparative numbers of fast-food and full service restaurants.

Difficulties in measuring food access have been noted. For example, the modified retail food environment index (MRFEI) (healthy food retailers as a percentage of all food retailers) may dilute specific components of the food environment associated with diet and weight (Jilcott et al, 2010a). Gustafson et al (2012) note that supermarket availability does not necessarily indicate an abundant resource of healthy, high quality foods. The literature review of food access by Health Canada (2013) mentions problematic questions such as appropriate distance thresholds for travelling to food stores, whether to use administrative boundaries, buffer zones, or activity spaces to define food access, and whether access differs by socio-demographic group, vehicle ownership, etc. The latter issue is examined in detail by Bader et al (2010).

There are similar issues in measuring access to opportunities for physical activity (Petrella et al, 2008; Edwards et al, 2011). These include which exercise activities (walking, cycling, etc) should be included, what activity threshold should be applied (e.g. moderate vs. vigorous exercise), what facilities (e.g. recreational sports facilities, parks, cycle paths, pedestrian trails) should be included, and whether exercise should be restricted to recreational activity.

Reviews of access indicators mention the predictive validity of such measures in terms of explaining relevant outcomes such as obesity (Jilcott et al, 2010a, p 433). Ensuring predictive validity is addressed naturally in the adaptive method for deriving latent access constructs (see below): the higher loadings will be for those measured indices or scales that best explain obesity or activity. Alternatives such as summing standardized scores of measured indices are not adaptive in the same way.

Methods

This paper considers urban sprawl and healthy food access as latent constructs (Lopez, 2014; Sturm and Cohen, 2004), with the relative weighting of measured indicators in defining the constructs governed by actual variation in county obesity and inactivity rates. The counties considered are 993 metropolitan counties, as in Ewing and Hamidi (2014), for which the four sub-components of sprawl mentioned above are available. These counties are a subset of 1167 metropolitan counties under the 2013 definition developed by the National Center for Health Statistics (NCHS) (Ingram and Franco, 2012), since counties were included only if they had at least one Census tract with population density exceeding 100 people per square mile. Table 1 compares selected indicators used in the analysis (and discussed further below) between the subset of 993 counties, all metropolitan counties (NCHS definition), and all 3141 counties. Differences between the subset and all metro counties are relatively small, though obesity and inactivity are slightly lower than the averages across 3141 counties.

The obesity and inactivity rates used in the model in the current paper are for adults (ages 18 and over) at county level in the year 2011, and are estimates based on analysis by the Centre for Disease Control and Prevention (CDC) (http://www.cdc.gov/diabetes/atlas/countydata/County_Methods.html#countylevelestimates). These are modelled estimates (using a Bayesian methodology) based on aggregated individual level responses to the annual US health survey, the Behavioral Risk Factor Surveillance System (BRFSS), and data from the U.S. Census Bureau's Population Estimates Program (Barker et al, 2013). They were developed using modern small area estimation techniques (Rao, 2003), using a statistical model that "borrows strength" in making an estimate for one county from data collected for other counties. Obesity is defined as a BMI of 30 or greater, while respondents are considered physically inactive if they answered "no" to the question, "During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?"

The CDC estimates of obesity and inactivity rates are provided in terms of estimated central rates and 95% intervals, with the latter defining the precision of the estimate of the central rate. The wider the gap between the upper and lower limit, the less precisely is the mean obesity rate estimated, i.e. it has higher variance. To incorporate both aspects of the CDC rates (i.e. both the mean rate and its precision) a hierarchical model, namely a form of meta-analysis, is adopted (e.g.

Welton et al, 2012; Pigott, 2012). Appendix 2 describes such models in general, while Appendix 3 describes modelling assumptions in the hierarchical analysis here. This includes a regression model to predict obesity and inactivity rates.

The analysis uses counties as units of geographic analysis, while recognizing that this choice of scale may affect results obtained, in line with the modifiable area unit problem (Sexton, 2008). The choice of county units is motivated in part by the much wider range of relevant indices developed for counties, partly because of policy relevance (Edwards et al, 2011), including the Food Environment Atlas, and the CDC estimates of obesity and inactivity. A number of studies related to the present one have used counties as their analysis units (e.g. Salois, 2012; von Hippel and Benson, 2014; Jilcott et al, 2010a; Jilcott et al, 2011; Edwards et al, 2011; Black, 2014; Chi et al, 2013; Shoultz et al, 2007) so that refining an existing evidence base is a relevant consideration in further research.

Admittedly, counties are not necessarily ideal for measuring impacts of some aspects of the urban environment considered here. For example, Williamson (2010, page 28) analyses the impact of sprawl at various geographic scales (at both neighbourhood and metropolis-wide levels). Disaggregated analysis (e.g. at a census tract rather than county level) might at first sight be considered; there are 65 thousand census tracts with an average population of 4000. However, the form of the outcome variable is also relevant. The obesity and inactivity outcomes are based on the BRFSS health survey data, and there are no inclusive population registers of such outcomes in the US. Developing stable estimates of these outcomes (from the BRFSS) at a scale below counties would be problematic, and they could only be obtained subject to a considerable loss of precision as compared to the county estimates.

Using Latent Constructs

Variations in obesity and inactivity for the 993 metropolitan counties are explained both by latent constructs (sprawl, food access, and exercise access), and by directly observable indicators of area socioeconomic status. With regard to sprawl, food and exercise access, we are seeking to summarise overlapping correlations in manifest (observed) indicators by an underlying construct (factor). Specifically, the three latent constructs are respectively defined by four, seven and two measured indices (see following section). In the factor analysis method used in the current paper, such constructs are also defined in such a way as to optimally explain variation in obesity and/or inactivity.

Factors are latent variables posited to explain correlations or covariances between observed indicators. To quote Brown (2006, p.13) "a factor is an unobservable variable that influences more than one observed measure and that accounts for the correlations among these observed measures". Generally the number of factors is less than the number of measured variables, which assists in reducing the influence of multicollinearity on regression estimates.

The latent constructs are obtained from weighted combinations of observed indicators. The analysis is confirmatory in the sense that observed indicators are taken to represent a particular construct, and that construct only. Observed indicators are linked to latent constructs via loadings (sometimes called item-construct loadings). However, a joint likelihood (see Appendix 3) ensures that the definition of the constructs (e.g. the relative weighting of constituent indices expressed by the loadings) is adaptive - in the sense that providing optimal weightings in terms of explaining obesity and inactivity.

This differs from approaches such as simple summation of constituent indicators, summation of standardized values, or principal components analysis, which do not adapt the latent constructs to explain the outcomes (e.g. obesity) that are postulated to depend on them. A Bayesian estimation strategy is implemented via the OPENBUGS package (Lunn et al, 2009) using Markov Chain Monte Carlo (MCMC) sampling. This strategy has advantages, for example, if some indicators of a particular construct are available but some are missing data, then imputation of missing data is simply carried out. Imputation of the missing indicators is relevant for a few counties (under 2%) where some of the sprawl indices (density and centering) are not defined. Missing data values are treated as an additional unknown quantity in the model and imputed at each iteration in the MCMC sampling (Cowles, 2013, p 141).

Observed indicators for three latent constructs

The three latent constructs (sprawl, food access, and exercise access) are respectively defined by four, seven and two measured indices (see Appendix 1 for definitions of the input indicators defining the constructs). As potential indicators of compactness, we consider the four indices (based on the 2010 US Census) set out by Ewing and Hamidi (2014). These are expected to be negative measures of sprawl (i.e. they take lower values as sprawl increases), and the loadings linking the indicators to the construct are constrained to maintain this feature. In this way, a correlation can be obtained between the compactness score derived here and the composite compactness index of Ewing and Hamidi (2014).

Seven potential indicators are taken to measure the food environment, and the loadings are set in such a way that the construct is a negative index of healthy food access. The first three indicators are measures of low access, based on distance to a supermarket or large grocery store (Ver Ploeg et al, 2012), and with low access defined by distances over 1 mile (for urban Census tracts) and over 10 miles (for rural Census tracts) (USDA, 2014a). Specifically the first indicator is the percentage (in 2011) of the total population with low access, the second is the percentage of the population who have low access combined with low income, and the third is the percentage with low access combined with no car. The fourth and fifth indicators are measures of healthy food access, namely the modified retail food environment index (CDC, 2014), and the ratio of grocery stores to population (USDA, 2014a). The sixth and seventh indicators compare food outlets (Chi et al, 2013; Jilcott et al, 2010a), and are respectively the ratio of convenience stores to grocery stores, and the ratio of fast-food to full service restaurants (USDA, 2014a).

There are two indicators of exercise access: recreation and fitness facilities per head of population in 2011 (USDA, 2014a), and the USDA natural amenities scale (USDA, 2014b). The latter index has been used in a number of studies of activity/recreation and obesity (Jilcott et al, 2013; von Hippel and Benson, 2014; Zahran et al, 2008). In practice all indicators except the amenities scale are log transformed, to reduce skewness; see Appendix 1 for the transformations used.

Spatial Dependencies

The analysis allows for the latent constructs to be spatially correlated (cf. Wang and Wall, 2003; Hogan and Tchernis, 2004). This feature is intended to reproduce observed patterns. Spatial clustering in urban sprawl is illustrated by high sprawl in the Southeast of the US (Ewing and Hamidi, 2010; Terando et al, 2014), while physical inactivity is highest in the South east and the Plains states (Robert Wood Johnson Foundation, 2014). Spatial pattering in environmental influences is expected to induce spatial correlation in the obesity and activity outcomes. However, not all risk factors for ecological variations in obesity and inactivity are measurable. We use a spatially correlated random effect (analogous to a residual in conventional regression) to represent unmeasured influences on county level obesity and inactivity (Wakefield et al, 2000, p 108-113). This random effect also represents unexplained spatial correlation in obesity and inactivity rates.

Spatial dependence in construct scores and in the residuals is accommodated using a conditional autoregressive approach, specifically the scheme of Leroux et al (1999). This scheme includes a measure of global spatial dependence, denoted λ in Leroux et al (1999), and varying between 0 and 1 (see Appendix 3). In this way, one allows for absence of spatial dependence: if λ=0, the construct scores show no spatial dependence. Because the spatial units are not necessarily contiguous, spatial interactions are specified in distance decay terms for the ten nearest neighbours to each county: if d_{ij} is the distance between counties i and j, then the interactions are $w_{ij}=1/d_{ij}^{\eta}$ where η is a positive parameter.

Regression Analysis

Variations in obesity and inactivity for the 993 metropolitan counties are explained, using a regression model, by the latent constructs (sprawl, food access, exercise access), and by observed indices of area socioeconomic status (county SES). Loadings linking indicators to constructs are constrained so that the first construct (COMPACT) is a measure of compactness (i.e. a negative measure of sprawl), the second (LOWHLTHFD) is a measure of low access to healthy food, and the third (EXERACC) is a positive index of exercise access.

Obesity rates are taken as depending on sprawl, food access and area SES, while inactivity is taken as depending on sprawl, exercise access and area SES. Let CSES denote county SES. Two indicators of area SES are compared in the analysis below: the county poverty rate in 2011 and the estimated county median household income in 2011. These are from the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) (US Census Bureau, 2014).

Then the regression model to predict the means (μ_1,μ_2) of the obesity and inactivity rates (means of can be summarised as

$\mu_1 = \beta_0+\beta_1\text{COMPACT}+ \beta_2\text{LOWHLTHFD}+ \beta_3\text{CSES}+s_1$

$\mu_2= \gamma_0+\gamma_1\text{COMPACT}+ \gamma_2\text{EXERACC}+ \gamma_3\text{CSES}+s_2$

where s_1 and s_2 are spatially clustered residual terms. As in other regression, the coefficients (β and γ parameters) from the regression will take account of correlations between the constructs and county SES. We do not specify a direct link of inactivity on obesity as the purpose of the model is to achieve conditional independence between these outcomes given the latent environmental constructs and observed county SES (Skrondal and Rabe-Hesketh, 2007). That is, the correlation between the sets of residuals $e_1=\theta_1-\mu_1$ and $e_2=\theta_2-\mu_2$ (see Appendix 3) should be effectively zero. Analysis showed that inclusion of s_1 and s_2 was necessary to achieve conditional independence.

Results

As mentioned above, two measures of area SES are compared: county poverty rates and county median household incomes. A measure of predictive loss (Gelfand and Ghosh, 1998) is used as a measure of fit, and shows that median household income (log transformed in the regression) produces a better fit than if county poverty is used as a measure of area SES.

Table 2 presents standardised β and γ coefficients (with 95% estimation intervals) for the county income model. Also shown are estimated standardised loadings linking indicators to constructs. From the latter one may ascertain those indicators most central to defining the construct, and relevant to the issue of predictive validity mentioned by Jilcott et al (2010a). Spatial dependence parameters for the three constructs are also shown. There are five such parameters: for the three constructs (COMPAC, LOWHLTHFD, EXERACC), and for the two spatial residual effects (s_1,s_2).

Model performance is satisfactory in that predictions, obtained using the mixed predictive approach of Marshall and Spiegelhalter (2003), are in line with the observed data. The goal of conditional independence is achieved: the correlation between residuals e_1 and e_2, as defined above, is 0.04 with a 95% interval (-0.03,0.12) straddling zero.

Table 2 shows that all regression effects (the β and γ parameters) on obesity and physical inactivity are significant and in anticipated directions. However, effects of compactness and area SES on obesity outweigh those due to food access. Similarly, effects of compactness and area SES on inactivity outweigh those due to exercise access.

All three constructs exhibit spatial dependence, with significant λ parameters. This is also apparent from maps of the scores (standardised values for continental US counties) in Figures 1 to 3. These Figures contain the area scores on the latent constructs that are derived by the Bayesian spatial factor model (COMPAC, LOWHLTHFD, EXERACC), as set out in Appendix 3. In precise terms these are posterior mean scores. Values are only mapped for the 993 counties included in the analysis.

Figures 1 and 2 show low compactness and relatively poor healthy food access in the south east USA, while Figure 3 shows high exercise access in the Pacific states. Scores for food access in particular parts of the south east USA, such as Texas, concord with other studies (Cole, 2012). The pattern fits in with a recurrent finding (e.g. Larson et al, 2009; Block et al, 2004) that residents of low-income minority neighbourhoods (which are disproportionately located in the south east) are most affected by poor access to supermarkets and healthful food.

The four indicators of compactness have broadly similar relevance to the overall construct, as assessed by standardised loadings (δ parameters), except that centering has a slightly lower impact. In defining low healthy food access, the most important influences are the ratio of convenience stores to grocery stores, and grocery stores per head. The relatively low salience of distance based access to supermarkets has some affinity with findings of Gustafson et al (2012), though higher weights attach to low distance based access combined with SES measures (namely low income or no car). The natural amenity score is the main influence on exercise access, providing some support for the "green exercise" hypothesis, namely the benefits of a "synergistic combination of exercise and exposure to nature" (Gladwell et al, 2013).

Table 3 lists out counties with the highest and lowest scores on the three constructs. There is a correlation of 0.93 between the scores for compactness obtained here, and the composite score provided in Ewing and Hamidi (2014). Hence counties with high scores for compactness in Table 3 are similar to those reported by that study. Highest scores for low healthy food access are for counties with low grocery store provision and/or high ratios of convenience stores to grocery stores. Compactness and low healthy food access are negatively associated, with a correlation between them of -0.54. There is a correlation of 0.23 between compactness and exercise access.

Table 4 summarises relationships between constructs and outcomes using Pearson correlations between mean values on the constructs (posterior county means), and CDC obesity and inactivity rates for 2011. It can be seen that compactness has similar correlations with obesity and inactivity outcomes as does 2011 median county income, and higher correlations with these outcomes than county poverty has. Low healthy food access has a higher correlation with obesity than poverty does, although its impact is less in the full regression setting (which takes account of uncertainty in obesity and inactivity rates, correlations between predictors, etc.).

Table 5 summarises associations in terms of the nine US Census Divisions (Figure 4). Highest obesity (over 30% of adults), lowest activity (over 25% inactivity), and adverse environments (low compactness, low healthy food access, and relatively low incomes) coincide in the East South Central and West South Central divisions. Relatively high compactness and healthy food access characterise metropolitan counties in the Pacific and Mid-Atlantic divisions.

Discussion

It is increasingly recognised that aspects of the urban environment may affect levels of obesity and physical activity. Previous ecological (area-scale) studies have generally considered only one or other of these outcomes, and considered the impacts of either sprawl or of food access. The present study is distinctive in considering both environmental factors together, and in evaluating their relative importance. It is also distinctive in considering obesity and inactivity as joint outcomes, and examining the way latent environmental constructs can be defined in relation to them, and in a way that best explains their variation.

Sprawl, food access and exercise access are considered as latent multidimensional constructs, imperfectly represented by any single measured index. The latent constructs are taken to be potentially spatially dependent (e.g. to allow for concentration of high values in sub-regions), and the analysis confirms this. Highest obesity, lowest activity, and highest concentrations of sprawling metropolitan counties with poor food access, are in the Southeast divisions of the US.

The centrality of area socio-economic status in explaining geographic variations in obesity and inactivity is also confirmed in the analysis here, an effect remaining after allowing for sprawl and food access (Levine, 2011). The importance of area deprivation in accounting for ethnic differentials in obesity has recently been established (Rossen, 2013). However, the current analysis has shown that median household income has a better predictive value for explaining obesity and inactivity variations than area poverty. It may be that taking account of concentrations of high income groups, or of low income groups, reflects unmeasured aspects of food and exercise access (McLaren, 2007, p 35; Swinburn et al, 2004, p. 133), or indirect effects of area income. A methodological extension of the work here could allow for indirect effects of area SES (e.g. via food access) on obesity and inactivity, as well as direct effects.

The analysis has shown that sprawl, typically in newer suburban areas, is an important influence on obesity and inactivity, and outweighs the impact of healthy food access on obesity. There may be scope to improve measures of food access (e.g. Gustafson et al, 2012). Whether and how far this would enhance the impact of food access (e.g. sufficient to match the impacts of compactness apparent in Table 2) is uncertain. There is some evidence that suburban sprawl and diminished food access are interrelated, as illustrated by the correlation of 0.54 obtained above.

It has been noted that inner city residents may also have elevated obesity and restricted activity levels, despite greater street connectivity and land use mix (i.e. greater compactness) than in suburban areas (Lopez and Hynes, 2006). However, the analysis here does suggest a distinct "suburbs effect" as can be seen from Table 6. This sets out (for the n=993 counties in the metropolitan subset) averages on the three constructs, and obesity/activity levels, according to NCHS urban category and poverty level. Counties are characterised as above or below average poverty according to whether they exceed the US wide rate. It can be seen that large central metro areas (inner cities) have lower obesity, higher activity and generally more favourable environments (in terms of obesogenicity) than large fringe metro (suburban) counties with similar poverty levels.

Although not considered in the present paper, health outcomes related to obesity and activity may also be linked to environmental influences such as sprawl and food access (Sturm and Cohen, 2004; Riediker and Koren, 2004). For example, the CDC estimates mentioned above include diabetes prevalence, allowing a continuation of the comparison in Table 6. Thus a "suburbs effect" shows in a diabetes rate of 11.2% in high poverty large fringe metro counties, in contrast to a rate of 9.7% in high poverty large central counties. Specific inclusion of health outcomes, as well as modifiable health behaviours, is possible in an extended analysis using the methods in the present paper.

Limitations of the analysis here may be mentioned. Firstly the response outcomes (obesity, inactivity) are based not on full population coverage (e.g. population disease registers), but on CDC estimates using survey data, and hence subject to sampling variability and measurement errors. Imprecision (variability) is expected to be greater for smaller counties, though the Bayesian borrowing strength method used by CDC can in fact provide estimates with greater precision than classical estimates (Rao, 2003; Schirm et al, 1999). Secondly, the notion of sprawl is primarily considered in the literature in relation to metropolitan areas, and this study follows that practice. However, certain attributed consequences of sprawl (e.g. variations in walkability and auto dependence) may be relevant to all US counties, of which there are over 3100 (Shoultz et al, 2007). For example, some studies (von Hippel and Benson, 2014; Jilcott et al, 2010b; Berrigan and Troiano, 2002; Zahran et al, 2008) use commuting distances or age of housing stock as indices of sprawl.

Another caveat is that the ideal analysis is multilevel, with interactions between contextual and individual level influences considered. The present study focuses on ecological analysis, as the interplay between the major environmental and contextual factors has been relatively little considered. The present paper addresses this while taking account of complexities such as spatial correlation in their pattern, and defining the constructs adaptively to best explain the health outcomes. As mentioned by Black (2014), it is it is “imperative to understand the role of place on health, especially in a context of stark spatial inequalities.”

The BRFSS, on which the CDC estimated county obesity and activity rates are based, would at first sight be the most relevant dataset for carrying out a multilevel analysis. Unfortunately, the BRFSS does not include potentially important individual level modifiers of environmental influences, such as vehicle ownership.

Additionally, extending the analysis to a multilevel setting would involve rather complex and computationally demanding modelling, unless simplifications were made in the form of the environmental (area) variables. For example, in a realistic multi-level model it is likely that responsiveness to (or impacts of) sprawl, food access or exercise access would differ by household income group or ethnic group. Health Canada (2013, p 13) mentions that “eating behaviours of people who are socially or economically disadvantaged would be more strongly associated with the quality of their food environment”. Such considerations mean that it would be difficult to allow the area factor scores (and other parameters of the spatial factor model) to remain as unknown parameters in a multilevel model with large numbers of individual subjects. The advantages of the latent construct methodology of the present paper (e.g. adaptiveness to the outcomes being explained, and allowing factor scores to be spatially correlated) would therefore most likely have to be dropped, and simple summary scores (e.g. principal component scores, sums of standardised area indicators) used instead.

Acknowledgements The author acknowledges comments on an earlier draft by Stephanie Jilcott Pitts

Appendix 1 Indicator Definitions and Sources

Density A principal component score obtained from five variables: overall population density (persons per square mile, ppsm); percentage of population living at low densities (under 1500 ppsm); percentage of population living at medium to high urban densities (over 12500 ppsm); urban density based on the National Land Cover Database; and employment density obtained from the Local Employment Dynamics database (source: Smart Growth America Project; Ewing and Hamidi, 2014). In Appendix 3, Z_1 is the log of the density score.

Mixed use A principal component score obtained from three variables: average job-population balance; degree of job mixing (using five employment sectors); a walkability score measuring proximity to amenities (source: Smart Growth America Project; Ewing and Hamidi, 2014). In Appendix 3, Z_2 is the log of the mixed use score.

Centering A principal component score obtained from four variables: coefficient of variation in census block group population densities, namely standard deviation of block group densities divided by average density of block groups; coefficient of variation in census block group employment densities; percentage of county population in central business district or sub-centers; percentage of county employment in central business district or sub-centers (source: Smart Growth America Project; Ewing and Hamidi, 2014). In Appendix 3, Z_3 is the log of the centering score.

Street Accessibility A principal component score obtained from four variables: average block size; percentage of blocks with areas less than 1/100 square mile; intersection density for urban and suburban census tracts within the county, excluding rural tracts with gross densities of less than 100 persons per square mile; percentage of 4-or-more-way intersections, again excluding rural tracts (source: Smart Growth America Project; Ewing and Hamidi, 2014). In Appendix 3, Z_4 is the log of the street access score.

Low food store access Percentage of county population with low distance access to supermarket or large grocery store in 2010 (source: Food Environment Atlas; USDA, 2014a). Low access is defined as greater than one mile from a supermarket or grocery store in urban areas, or greater than ten miles from a supermarket or grocery store in rural areas. In Appendix 3, Z_5 is the log of 1 plus this percentage.

Low food store access and low income Percentage of county population with low-income and low access to a supermarket or large grocery store in 2010 (source: Food Environment Atlas; USDA, 2014a). In Appendix 3, Z_6 is the log of 1 plus this percentage.

Low food store access and no car Percentage of county population without a car and with low access to a supermarket or large grocery store in 2010 (source: Food Environment Atlas; USDA, 2014a). In Appendix 3, Z_7 is the log of 1 plus this percentage.

Modified retail food environment index (MRFEI) Healthy food retailers as a percentage of all food retailers, population weighted average of census tract values within each county (source: CDC; CDC, 2014). Healthy food retailers include supermarkets, larger grocery stores, supercenters, and produce stores within census tracts of residence or 0.5 mile from the tract boundary. Less healthy food retailers include fast-food restaurants, small grocery stores, and convenience stores within census tracts or 0.5 mile from the tract boundary. In Appendix 3, Z_8 is the log of 1 plus this percentage.

Grocery stores access. Grocery stores per thousand county population in 2011 (source: Food Environment Atlas; USDA, 2014a). In Appendix 3, Z_9= log(1+100*R), where R is grocery stores per thousand population.

Ratio of convenience to grocery stores in 2011 (source: Food Environment Atlas; USDA, 2014a). With CNVS being the number of convenience stores in a county, and GRCS the number of grocery stores, Z_{10}= log[(1+CNVS)/(1+GRCS)] in Appendix 3.

Ratio of fast-food to full service restaurants in 2011 (source: Food Environment Atlas; USDA, 2014a). With FFR being the number of fast-food restaurants in a county, and FSR the number of full service restaurants, Z_{11}= log[(1+FFR)/(1+FSR)] in Appendix 3.

Recreation access. Recreation and fitness facilities per thousand population in 2011 (source: Food Environment Atlas; USDA, 2014c). These facilities are defined as those providing fitness and recreational sports activities, such as swimming, skating, or racquet sports. In Appendix 3, Z_{12}= log(1+100*R), where R is facilities per thousand population.

Natural amenity scale. Constructed by combining six measures of climate, topography, and water area (source: USDA; USDA, 2014b). In Appendix 3, Z_{13} is this index, without transformation.

Appendix 2 Hierarchical (Meta-Analysis) Models

There is an extensive statistical technology (often called meta-analysis) applied to various summary statistics (e.g. means/variances, proportions, odds ratios, risk differences) from sets of interrelated units (schools, hospitals, counties, etc). Goals of such analysis include "borrowing strength" to provide more precise estimates for each unit using information from other units, and to make comparative inferences such as rankings of units on exam rates, hospital mortality, disease rates, etc. These techniques are most commonly applied in health and education (Deely and Smith, 1998), but also relevant to geographic studies of disease variation.

The idea is to use the totality of the summary statistics (information on means and precisions for each unit) to provide information about the distribution of underlying rates over all units. The first stage in a hierarchical model specifies the density of the observed summary statistics, including known precision information, and referring to underlying rates for each unit. The second stage specifies the smooth density (e.g. normal) for the underlying rates. The second stage can also involve regression, an approach often called meta-regression.

Appendix 3 Model Specification

Bayesian estimation is applied to estimate parameters and derive construct scores. Estimation is based on the second halves of two chain runs of 10,000 iterations, with convergence assessed using Brooks-Gelman-Rubin statistics (Brooks and Gelman, 1998). Let y_{1i} and y_{2i} denote logit transforms of the CDC obesity and inactivity rates for US metropolitan counties (i=1,..,993), with the logit transform being one possible transformation to improve approximation to normality (Atkinson, 1985, chapter 7). The CDC estimates include 2.5% and 97.5% quantiles, which define the precision of the estimates of the central obesity and inactivity rates. The wider the gap between the upper and lower quantiles, the less precisely is the central rate estimated, i.e. it has higher variance.

To incorporate both aspects of the CDC estimates, a hierarchical model is used (see Appendix 2): the likelihood for the observed outcomes (y_{1i}, y_{2i}) at stage 1, the model for latent means $(\theta_{1i}, \theta_{2i})$ at stage 2, and the specification (e.g. assumed prior densities) of hyperparameters at stage 3. Specifically at stage 1, the two outcomes (logit transformed rates) are assumed normally distributed with known county specific variances V_{1i} and V_{2i} (variances in the logit transformed scale are obtained from the 2.5% and 97.5% limits in the CDC data). Thus

$$y_{1i} \sim N(\theta_{1i}, V_{1i})$$
$$y_{2i} \sim N(\theta_{2i}, V_{2i}).$$

To define θ_{1i} and θ_{2i}, a meta-regression is adopted involving the three constructs and median county income, MEDINC. There are 13 indicators (Z_1 to Z_{13}) of the three constructs (F_1, F_2, F_3) (respectively denoted COMPACT, LOWHLTHFD, and EXERACC in the main text). Except for Z_{13} (the natural amenity scale), which is retained without transformation, the indicators are log transforms of originally percentage or ratio indices. The transformations are designed to avoid taking logarithms of zero values.

The confirmatory measurement model for the three constructs is then

$$Z_{ki} \sim N(\alpha_k + \delta_k F_{1i}, \tau_k) \quad k = 1,..,4$$
$$Z_{ki} \sim N(\alpha_k + \delta_k F_{2i}, \tau_k) \quad k = 5,..,11$$
$$Z_{ki} \sim N(\alpha_k + \delta_k F_{3i}, \tau_k) \quad k = 12,..,13.$$

To ensure consistent labelling of the constructs (i.e. avoid label switching), the first loading for each construct (i.e. δ_1, δ_5, and δ_{12}) is assigned a positive prior, namely an exponential with mean 1. Remaining loadings are assigned N(0,10) priors.

Obesity rates are taken to depend on sprawl (F_1), food access (F_2) and area SES, while inactivity is taken as depending on sprawl (F_1), exercise access (F_3) and area SES. As mentioned in the main text, county median income (MEDINC) provided better fit than county poverty. The meta-regression model for the latent mean obesity and inactivity rates is then

$$\theta_{1i} \sim N(\mu_{1i}, \phi_1),$$
$$\theta_{2i} \sim N(\mu_{2i}, \phi_2),$$
$$\mu_{1i}=\beta_0 +\beta_1 F_{1i} + \beta_2 F_{2i} + \beta_3 MEDINC_i +s_{1i},$$
$$\mu_{2i}=\gamma_0 +\gamma_1 F_{1i} + \gamma_2 F_{3i} + \gamma_3 MEDINC_i +s_{2i},$$

where the factor scores F_{ki} $(k = 1,..,3)$, and the residuals s_{1i} and s_{2i} are spatially dependent according to the conditional autoregressive (CAR) scheme of Leroux et al (1999). The α, β and γ parameters are assigned N(0,1000) priors, and inverse variances ($1/\tau_k$ and $1/\phi_m$) are assigned gamma priors with shape 1 and rate 0.001.

The counties included in the analysis are not necessarily contiguous, so spatial interactions w_{ij} are defined in terms of distances d_{ij} to the ten nearest neighbouring counties. So for the first construct, spatial interactions are obtained as $w_{ij} = 1/d_{ij}^{\eta_1}$, where $\eta_1 > 0$ represents distance decay, while the construct scores have conditional densities

$$F_{1i} \mid F_{1[i]} \sim N\left(\frac{\lambda_1}{1-\lambda_1+\lambda_1 \sum_{j \neq i} w_{ij}} \sum_{j \neq i} w_{ij} F_{1j}, \frac{\sigma^2_{F1}}{1-\lambda_1+\lambda_1 \sum_{j \neq i} w_{ij}}\right),$$

where $F_{1[i]}$ denotes all counties apart from county i. As described in the main text, λ_1 is a global index of spatial dependence, whereas η_1 represents a local decay effect. For identification, the variance σ^2_{F1} is set to 1, corresponding to factor standardization (Skrondal and Rabe-Hesketh (2007, p 718). Specifications for the other constructs, and for the spatial effects, (s_1, s_2), follow this scheme also, with spatial correlations denoted λ_{s1} and λ_{s2}.

To assess predictive fit, the mixed predictive scheme of Marsall and Spiegelhalter (2003) is used. Thus let $\theta_{new,1i}$ and $\theta_{new,2i}$ be replicates of the latent means for obesity and inactivity, and $y_{new,1i}$ and $y_{new,2i}$ be the corresponding predictions of the outcomes. Then the indicators

$$d_{1i} = I(y_{1i} - y_{new,1i})$$
$$d_{2i} = I(y_{2i} - y_{new,2i})$$

are monitored through MCMC iterations, where $I(C) = 1$ if condition C is true, and $I(C) = 0$ otherwise. For example, underprediction of obesity rates would be apparent if at most iterations $y_{new,1i}$ was less than y_{1i}. Similarly, overprediction would be apparent if at most iterations y_{1i} was less than $y_{new,1i}$. The resulting posterior predictive p-values for county i, namely $\Pr(d_{1i} = 1 | y)$, would be respectively high (e.g. over 0.95) or low (e.g. under 0.05). In fact, these p-values indicate satisfactory fit: there are respectively 2 and 8 counties (of 993) with underpredicted and overpredicted obesity rates, and 0 and 4 counties with underpredicted and overpredicted inactivity rates.

References

Atkinson, A. (1985) *Plots, Transformations, and Regression*. New York: Oxford University Press. 5.

Bader M., Purciel M., Yousefzadeh P., Neckerman K. (2010) Disparities in neighborhood food environments: implications of measurement strategies. *Economic Geography,* 86(4):409-30.

Barker L., Thompson T., Kirtland K., Boyle J., Geiss L., McCauley M., & Albright A. (2013) Bayesian small area estimates of diabetes incidence by United States county. *Journal of Data Science,* 11:249–269.

Bennett K., Probst J., Pumkam C (2011) Obesity among working age adults: the role of county-level persistent poverty in rural disparities. *Health and Place*, 17(5):1174-81.

Berrigan D., & Troiano, R. (2002) The association between urban form and physical activity in U.S. adults. *American Journal of Preventive Medicine*, 23(2S):74-79.

Black, N. (2014) An ecological approach to understanding adult obesity prevalence in the United States: a county-level analysis using geographically weighted regression. *Applied Spatial Analysis and Policy*, 7(3), 283-299.

Black J. & Macinko J. (2008) Neighborhoods and obesity. *Nutrition Reviews*, 66(1): 2–20.

Block J., Scribner R., DeSalvo K. (2004) Fast-food, race/ethnicity, and income: a geographic analysis. *American Journal of Preventive Medicine,* 27(3):211-7.

Bodor J., Rice J., Farley T., Swalm C., & Rose D. (2010) The association between obesity and urban food environments. *Journal of Urban Health,* 87(5):771-81.

Brooks S, & Gelman A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics,* 7(4), 434-45

Brown, T. (2006) *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.

Centers for Disease Control and Prevention (CDC) (2014) *Healthier Food Retail: Beginning the Assessment Process in Your State or Community.* Atlanta: U.S. Department of Health and Human Services.

Chi, S., Grigsby-Toussaint, D., Bradford, N., & Choi, J. (2013) Can geographically weighted regression improve our contextual understanding of obesity in the US? Findings from the USDA Food Atlas. *Applied Geography,* 44(1), 134-142.

Cho, S., Chen, Z., Yen, S., & Eastwood, D. (2006) The effects of urban sprawl on body mass index: where people live does matter? *Consumer Interests Annual*, 52 (1), 159-169.

Cole, C. (2012*) Access to Healthy and Affordable Food Is Critical to Good Nutrition*. Austin, Texas: Center for Public Policy Priorities.

Cowles, M. (2013) *Applied Bayesian Statistics: with R and OpenBUGS Examples* New York: Springer.
Deely, J. & Smith, A. (1998) Quantitative refinements for comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161(1), 5-12.
Drewnowski A., & Specter S. (2004) Poverty and obesity: the role of energy density and energy costs. *American Journal Clinical Nutr*ition, 79(1):6-16.
Edwards, M., Jilcott, S., Floyd, M., & Moore, J. (2011) County-level disparities in access to recreational resources and associations with adult obesity. *Journal of Park and Recreation Administration,* 29(2), 39-54
Ewing R., Brownson R., & Berrigan D. (2006) Relationship between urban sprawl and weight of United States youth. *American Journal of Preventive Medicine*, 31(6):464-74.
Ewing, R., & Hamidi, S. (2010) *Measuring Urban Sprawl and Validating Sprawl Measures*. Metropolitan Research Center, University of Utah.
Ewing, R., & Hamidi, S. (2014) *Measuring Sprawl 2014. Smart Growth America and Metropolitan* Research Center, University of Utah.
Feng J, Glass T., Curriero F., Stewart W., & Schwartz B. (2012) The built environment and obesity: a systematic review of the epidemiologic evidence. *Health & Place,* 16 (2): 175–190.
Ford, P., & Dzewaltowski, D. (2008) Disparities in obesity prevalence due to variation in the retail food environment: three testable hypotheses. *Nutrition Reviews*, 66(4), 216-228.
Ford E., Mokdad A., Giles W., Galuska D., & Serdula, M. (2005) Geographic variation in the prevalence of obesity, diabetes, and obesity-related behaviors. *Obesity Research,* 13(1),118-22.
French S., Story M., Jeffery R. (2001) Environmental influences on eating and physical activity. *Annual Review of Public Health*, 22:309-35.
Frumkin H. (2002) Urban sprawl and public health. *Public Health Reports,* 117(3):201-17
Garden F., Jalaludin B. (2009) Impact of Urban Sprawl on Overweight, Obesity, and Physical Activity in Sydney, Australia. *Journal of Urban Health,* 86(1): 19–30
Gelfand, A., & Ghosh, S. (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85(1), 1-11.
Gladwell V., Brown D., Wood C., Sandercock G., Barton J. (2013) The great outdoors: how a green exercise environment can benefit all. *Extreme Physiology and Medicine*, 2013, 2(1):3.
Gustafson A., Lewis S., Wilson C., & Jilcott-Pitts S. (2012) Validation of food store environment secondary data source and the role of neighborhood deprivation in Appalachia, Kentucky. *BMC Public Health,* 2012 Aug 22;12:688.
Health Canada (2013) *Measuring the Food Environment in Canada.* Ottawa: Health Canada
Hill J., & Peters J. (1998) Environmental contributions to the obesity epidemic. *Science*, 280 (5368):1371–1374
Hogan J., & Tchernis R. (2004) Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association,* 99(466): 314-324.
Ingram, D., & Franco, S. (2012) 2013 NCHS Urban–Rural Classification Scheme for Counties. *Vital Health Statistics,* 2(154), 1-73
Jeffery R., & Utter J. (2003) The changing environment and population obesity in the United States. *Obesity Research*, 11, Suppl:12S-22S.
Jilcott S., McGuirt J., Imai S, Evenson K. (2010a) Measuring the retail food environment in rural and urban North Carolina counties. *Journal of Public Health Management and Practice,* 16(5):432-40.

Jilcott S., Liu H., Moore J., Bethel J., Wilson J., & Ammerman A. (2010b) Commute times, food retail gaps, and body mass index in North Carolina counties. *Preventing Chronic Disease* 2010 Sep;7(5):A107.

Jilcott S., Moore J., Shores K., Imai S., & McGranahan D. (2011) Associations between natural amenities, physical activity, and body mass index in 100 North Carolina counties. *American Journal of Health Promotion,* 26(1):52-5.

Jilcott Pitts S., Edwards M., Moore J., Shores K., DuBose K., McGranahan D. (2013) Obesity is inversely associated with natural amenities and recreation facilities per capita. *Journal of Physical Activity and Health*, 10(7):1032-8.

Koh, K (2011) A spatial analysis of county level obesity prevalence in Michigan. Presented to 2011 Mid-Continent Regional Science Association Conference. Available at http://www.mcrsa.org/Assets/Documents/Proceedings/A Spatial Analysis of County-level Obesity Prevalence in Michigan.pdf

Larson N., Story M., Nelson M. (2009) Neighborhood environments: disparities in access to healthy foods in the U.S. *American Journal of Preventive Medicine,* 36(1):74-81.

Leroux, B., Lei X., & Breslow, N. (1999) Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: *Statistical Models in Epidemiology, the Environment and Clinical Trials*, Halloran, M and Berry, D (eds), pp. 135-178. New York: Springer-Verlag

Levine, J. (2011) Poverty and obesity in the U.S. *Diabetes,* 60(11), 2667-2668

Lopez, R. (2004) Urban sprawl and risk for being overweight or obese. *American Journal of Public Health,* 94(9):1574-9.

Lopez, R (2014) Urban Sprawl in the United States: 1970-2010. *Cities and the Environment,* 7(1). http://digitalcommons.lmu.edu/cate/

Lopez, R., & Hynes, P. (2003) Sprawl in the 1990s: measurement, distribution, and trends. *Urban Affairs Review,* 38(3): 325–35

Lopez R., & Hynes, P. (2006) Obesity, physical activity, and the urban environment: public health research needs. *Environmental Health,* 2006 Sep 18;5:25.

Lunn D, Spiegelhalter D, Thomas A, & Best N. (2009) The BUGS project: Evolution, critique and future directions. *Statistics in Medicine,* 28(25):3049-67

Macintyre S. (2007) Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *International Journal of Behavioral Nutrition and Physical Activity*, 2007, 4:32.

McLaren L (2007) Socioeconomic status and obesity. *Epidemiologic Reviews,* 29(1):29-48

Marshall, E., & Spiegelhalter, D. (2003) Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine,* 22(10), 1649-1660.

Moore, L., & Diez Roux, A. (2006) Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health,* 96(2), 325-331.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD.

Neckerman, K., Bader, M., Purciel M., & Yousefzadeh, P. (2009) *Measuring Food Access in Urban Areas*. National Poverty Center, University of Michigan.

Owen N., Humpel N., Leslie E., Bauman A., & Sallis J. (2004) Understanding environmental influences on walking; review and research agenda. *American Journal of Preventive Medicine,* 27(1):67-76.

Parks, S, Housemann, R., & Brownson, R. (2003) Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. *Journal of Epidemiology and Community Health,* 57(1), 29-35

Petrella R., Kennedy E., Overend T. (2008) Geographic determinants of healthy lifestyle change in a community-based exercise prescription delivered in family practice. *Environmental Health Insights*, 2008:1; 51-62.

Pigott, T (2012) *Advances in Meta-Analysis*. New York: Springer

Rao J (2003) *Small Area Estimation*. Hoboken, New Jersey: Wiley.

Riediker M., & Koren H. (2004) The importance of environmental exposures to physical, mental and social well-being. *International Journal of Hygiene and Environmental Health,* 207(3):193-20

Robert Wood Johnson Foundation (2014) *2014 County Health Rankings Key Findings*. http://www.countyhealthrankings.org/file/2014-county-health-rankings-key-findingspdf

Rossen L. (2013) Neighbourhood economic deprivation explains racial/ethnic disparities in overweight and obesity among children and adolescents in the USA. *Journal of Epidemiology and Community Health,* 68(2), 123-129

Saelens, B., Sallis, J., & Frank, L. (2003) Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine,* 25(2), 80-91

Sallis J., Floyd M., Rodríguez D., & Saelens B. (2012) Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation,* 125(5):729-37

Salois, M. (2012) Obesity and diabetes, the built environment, and the ‘local’ food economy in the United States. *Economics & Human Biology*. 10(1): 35-42.

Schirm A, Czajka J, Zaslavsky A (1999) Large numbers of estimates for small areas. *Federal Committee on Statistical Methodology Research Conference,* 1999, 77–83 (https://fcsm.sites.usa.gov/files/2014/05/IV-A_Schirm_FCSM1999.pdf)

Schuurman N., Peters P., & Oliver L. (2009) Are obesity and physical activity clustered? A spatial analysis linked to residential density. *Obesity,* 17(12):2202-9.

Sexton K. (2008) Modifiable Areal Unit Problem (MAUP). In: *Encyclopedia of Quantitative Risk Analysis and Assessment*, Volume 3. Melnick E., and Everitt B. (eds.) Chichester: Wiley

Shoultz, G., Givens, J., Drane, J. (2007) Urban form, heart disease, and geography: A case study in composite index formation and Bayesian spatial modeling. *Population Research and Policy Review,* 26(5-6), 661-685.

Skrondal, A., & Rabe-Hesketh, S. (2007) Latent variable modelling: a survey. *Scandinavian Journal of Statistics,* 34(4), 712-745.

Sturm R., & Cohen D. (2004) Suburban sprawl and physical and mental health. *Public Health,* 118(7):488-96.

Swinburn B., Caterson I., Seidell J., & James W. (2004) Diet, nutrition and the prevention of excess weight gain and obesity. *Public Health Nutrition*, 7(1A):123-46.

Terando, A., Costanza, J., Belyea, C., Dunn, R., McKerrow, A., & Collazo, J. (2014) The southern megalopolis: using the past to predict the future of urban sprawl in the southeast US. *PloS One,* 9(7), e102261.

U.S. Census Bureau (2014) *Small Area Income and Poverty Estimates* http://www.census.gov/did/www/saipe/about/index.html

U.S. Department of Agriculture (USDA) (2014a) *Food Environment Atlas.* http://www.ers.usda.gov/data-products/food-environment-atlas.aspx. Economic Research Service, USDA

U.S. Department of Agriculture (USDA) (2014b) *Natural Amenities Scale.* http://www.ers.usda.gov/data-products/natural-amenities-scale.aspx. Economic Research Service, USDA

U.S. Department of Agriculture (USDA) (2014c) *Food Environment Atlas Data Documentation.* http://ers.usda.gov/datafiles/Food_Environment_Atlas/Data_Access_and_Documentation_Downloads/Current_Version/documentation.pdf. Economic Research Service, USDA

Ver Ploeg, M. (2010) Access to affordable, nutritious food is limited in "food deserts". *Amber Waves,* 8(1), 20-7.

Ver Ploeg M., Breneman V., Dutko P., Williams R., Snyder S., Dicken C., & Kaufman P. (2012) *Access to Affordable and Nutritious Food: Updated Estimates of Distance to Supermarkets Using 2010 Data.* US Department of Agriculture, Economic Research Report No. (ERR-143)

von Hippel P., & Benson R. (2014) Obesity and the natural environment across US counties. *American Journal of Public Health,* 104(7):1287-93.

Wakefield, J., Best N., & Waller L. (2000) Bayesian approaches to disease mapping. In Elliott, P., Wakefield, J., Best, N., & Briggs, D. (Eds.), *Spatial Epidemiology: Methods and Applications* (pp 104-127). Oxford: Oxford University Press

Wang, F., and Wall, M. (2003) Generalized Common Spatial Factor Model. *Biostatistics.* 4(4), 569-582.

Welton, N., Sutton, A., Cooper, N., Abrams, K., & Ades, A. (2012) *Evidence Synthesis for Decision Making in Healthcare.* Chichester: Wiley

Williamson, T. (2010) *Sprawl, Justice, and Citizenship: The Civic Costs of the American Way of Life.* New York: Oxford University Press

Zahran, S., Brody, S. D., Maghelal, P., Prelog, A., & Lacy, M. (2008) Cycling and walking: Explaining the spatial distribution of healthy modes of transportation in the United States. *Transportation Research D,* 13(7), 462-470.

Table 1 Indicator Comparisons between Case Study Counties, All Metropolitan Counties, and All counties

		Metropolitan Counties in Analysis (n=993)	All Metrop-olitan (n=1167)	All Counties (n=3141)
Outcomes	Obesity	29.5	29.8	30.7
	Inactivity	24.2	24.6	25.8
Food Accesss	Low distance-based food store access	21.4	21.4	23.5
	Low distance-based food store access & low income	6.1	6.4	8.4
	Low distance based food store access & no car	2.2	2.4	3.1
	Modified retail food environment index	12.0	12.2	13.7
	Grocery stores access	0.174	0.183	0.259
Exercise/Environment	Recreation access	0.090	0.085	0.071
	Natural amenity scale	0.27	0.27	0.05
Area SES	Median County Income ($000s)	51.3	50.0	43.8

Table 2 Estimated Regression and Factor Model Parameters

Regression coefficients					
Explaining Obesity	**Standardised Effect of:**	**Estimate**	**LL**	**UL**	**Parameter**
	COMPACT	**-0.37**	**-0.42**	**-0.31**	β_1
	LOWHLTHFD	**0.14**	**0.08**	**0.21**	β_2
	MEDIAN INCOME	**-0.56**	**-0.60**	**-0.51**	β_3
Explaining Inactivity	**Standardised Effect of:**				
	COMPACT	**-0.29**	**-0.33**	**-0.24**	γ_1
	EXERACC	**-0.14**	**-0.22**	**-0.06**	γ_2
	MEDIAN INCOME	**-0.50**	**-0.54**	**-0.46**	γ_3
Standardised factor loadings	**Indicator**				
Loadings on COMPACT	**Density**	**0.75**	**0.71**	**0.80**	δ_1
	Mix	**0.74**	**0.69**	**0.79**	δ_2
	Centering	**0.55**	**0.49**	**0.61**	δ_3
	Street Accessibility	**0.78**	**0.73**	**0.82**	δ_4
Loadings on LOWHLTHFD	**Population, low access to supermarket**	**0.16**	**0.10**	**0.23**	δ_5
	Low income low access to supermarket	**0.31**	**0.25**	**0.37**	δ_6
	No car & low access to supermarket	**0.41**	**0.35**	**0.46**	δ_7
	Proportion of healthy food retailers (mRFEI)	**-0.18**	**-0.24**	**-0.12**	δ_8
	Ratio of grocery stores to population.	**-0.70**	**-0.74**	**-0.65**	δ_9
	Ratio of convenience stores to grocery stores	**1.00**	**0.99**	**1.00**	δ_{10}
	Ratio of fast food to full service restaurants.	**0.15**	**0.09**	**0.21**	δ_{11}
Loadings on EXERACC	**Recreation and fitness facilities per head**	**0.05**	**0.01**	**0.11**	δ_{12}
	Natural amenities scale	**1.00**	**0.99**	**1.01**	δ_{13}
Spatial Dependence	**Construct or Residual Effect**				
	COMPACT	**0.63**	**0.33**	**0.95**	λ_1
	LOWHLTHFD	**0.90**	**0.79**	**0.99**	λ_2
	EXERACC	**0.96**	**0.88**	**1.00**	λ_3
	s_1	**0.94**	**0.79**	**1.00**	λ_{s1}
	s_2	**0.94**	**0.80**	**1.00**	λ_{s2}

TABLE 3 High and Low Scores, Environmental Constructs

	COMPACTNESS			LOW ACCESS TO HEALTHY FOOD			EXERCISE ACCESS	
	County (State abbrev)	Score		County (State abbrev)	Score		County (State abbrev)	Score
Highest Sprawl	Oglethorpe County, GA	-1.13	High	New York County, NY	-3.12	Low Exer-	Tipton County, IN	-1.47
(low compact-	Harris County, GA	-1.01	food	Kings County, NY	-2.98	cise	Grand Forks County, ND	-1.47
ness)	Greene County, NC	-1.01	access	Bronx County, NY	-2.88	Access	Dodge County, MN	-1.45
	Spencer County, KY	-0.98	(low	Queens County, NY	-2.22		Cass County, ND	-1.44
	Macon County, TN	-0.98	scores)	San Francisco County, CA	-2.01		Champaign County, IL	-1.30
	Grant Parish, LA	-0.91		Richmond County, NY	-1.93		Piatt County, IL	-1.27
	Fayette County, TN	-0.91		Essex County, NJ	-1.90		Clay County, MN	-1.25
	Elbert County, CO	-0.90		Passaic County, NJ	-1.74		Ford County, IL	-1.21
	Morrow County, OH	-0.90		Westchester County, NY	-1.69		Olmsted County, MN	-1.15
	Anson County, NC	-0.89		Spencer County, KY	-1.42		Lincoln County, SD	-1.14
	Grainger County, TN	-0.88		Hudson County, NJ	-1.38		Polk County, MN	-1.13
	Blount County, AL	-0.86		Philadelphia County, PA	-1.36		Wells County, IN	-1.12
	Edgefield County, SC	-0.85		Asotin County, WA	-1.32		Howard County, IN	-1.11
	Lawrence County	-0.84		Douglas County, WA	-1.31		Benton County, IA	-1.08
	Jackson County, KS	-0.84		San Benito County, CA	-1.31		Hancock County, IN	-1.08
	Hampden County, MA	-0.82		Marin County, CA	-1.30		Shelby County, IN	-1.07
	Brown County, IN	-0.82		Santa Cruz County, CA	-1.23		McLean County, IL	-1.07
	Monroe County, GA	-0.82		Franklin County, IN	-1.20		Madison County, IN	-1.05
	Franklin County, NC	-0.81		Alameda County, CA	-1.19		Jones County, IA	-1.05
	Wakulla County, FL	-0.80		Nassau County, NY	-1.19		Story County, IA	-1.05
	Orleans Parish, LA	0.73	Low	San Juan County, NM	0.98	High	Clear Creek County, CO	1.76
Lowest Sprawl	Nassau County, NY	0.77	Food	Butts County, GA	0.99	Exercise	Imperial County, CA	1.78
(high compact-	Essex County, NJ	0.79	Access	Bartow County, GA	0.99	Access	Riverside County, CA	1.82
ness)	Passaic County, NJ	0.80	(high	Liberty County, GA	1.00		Washoe County, NV	1.87
	Baltimore city, MD	0.82	scores)	Tate County, MS	1.02		Stanislaus County, CA	1.96
	Cook County, IL	0.85		Fairfield County, SC	1.02		Carson City, NV	1.99
	Richmond County, NY	0.87		Pittsylvania County, VA	1.03		Napa County, CA	2.12
	Denver County, CO	0.88		McLennan County, TX	1.06		Sonoma County, CA	2.20
	Westchester County, NY	0.91		Rapides Parish, LA	1.07		San Luis Obispo County, CA	2.20
	Multnomah County, OR	0.93		Bell County, TX	1.09		Contra Costa County, CA	2.29
	Alexandria city, VA	1.00		Grant County, KY	1.11		San Mateo County, CA	2.31
	Hudson County, NJ	1.02		Wagoner County, OK	1.16		Marin County, CA	2.33
	Arlington County, VA	1.06		Elmore County, AL	1.19		Santa Cruz County, CA	2.36
	Philadelphia County, PA	1.09		George County, MS	1.20		Orange County, CA	2.42
	Queens County, NY	1.23		Randall County, TX	1.22		Monterey County, CA	2.52
	District of Columbia, DC	1.39		Kendall County, TX	1.23		San Diego County, CA	2.71
	Bronx County, NY	1.55		Victoria County, TX	1.24		San Francisco County, CA	2.80
	San Francisco County, CA	1.57		Nueces County, TX	1.27		Los Angeles County, CA	2.85
	Kings County, NY	1.82		McClain County, OK	1.41		Santa Barbara County, CA	3.01
	New York County, NY	2.79		Sequatchie County, TN	1.61		Ventura County, CA	3.03

Table 4 Correlations between Constructs and Outcomes

	Obesity	Inactivity
Compactness	-0.51	-0.45
Low Healthy Food Access	0.43	0.35
Exercise Access	-0.39	-0.36
County Median Household Income	-0.53	-0.51
County Poverty	0.36	0.37

Table 5 Profile of Outcomes and Environmental Factors, US Census Divisions

	Compact-ness	Low Food Access	Exercise Access	Median Household Income ($000s)	Obesity (%)	Inactivity (%)
New England	-0.14	-0.14	0.17	60.1	24.5	20.0
Mid-Atlantic	0.17	-0.58	-0.10	57.0	27.0	23.2
East North Central	-0.16	-0.02	-0.62	51.3	30.2	24.4
West North Central	-0.16	0.12	-0.52	53.2	29.6	23.3
South Atlantic	-0.20	0.15	0.10	50.7	29.6	24.6
East South Central	-0.37	0.38	-0.10	44.6	32.3	29.7
West South Central	-0.20	0.43	0.16	47.4	31.5	27.4
Mountain	0.00	0.05	0.99	52.1	24.0	18.2
Pacific	0.17	-0.60	1.47	54.5	25.5	17.0
US Average	-0.14	0.05	0.02	51.3	29.2	24.2

Table 6 Profile of Outcomes and Environmental Factors, by NCHS Category and Poverty Level

	Compact-ness	Low Food Access	Exercise Access	Poverty (%)	Obesity (%)	Inactivity (%)	% Diabetes
Below National Poverty Rate							
Large central	0.49	-0.64	0.66	12.2	23.4	18.7	7.9
Large fringe metro	-0.18	0.00	-0.04	10.3	28.0	22.9	8.9
Medium metro	-0.19	0.08	0.05	12.2	28.5	23.1	8.9
Small metro	-0.17	0.02	-0.17	12.3	28.9	22.6	8.5
All low poverty counties	-0.16	0.01	-0.02	11.3	28.2	22.8	8.8
Above National Poverty Rate							
Large central	0.56	-0.53	0.22	20.0	27.0	22.9	9.7
Large fringe metro	-0.29	0.16	0.04	19.5	32.5	28.0	11.2
Medium metro	-0.17	0.18	0.14	20.1	31.1	26.2	10.3
Small metro	-0.14	0.18	0.01	20.1	30.6	26.0	10.2
All high poverty counties	-0.10	0.10	0.08	20.1	30.7	26.0	10.3

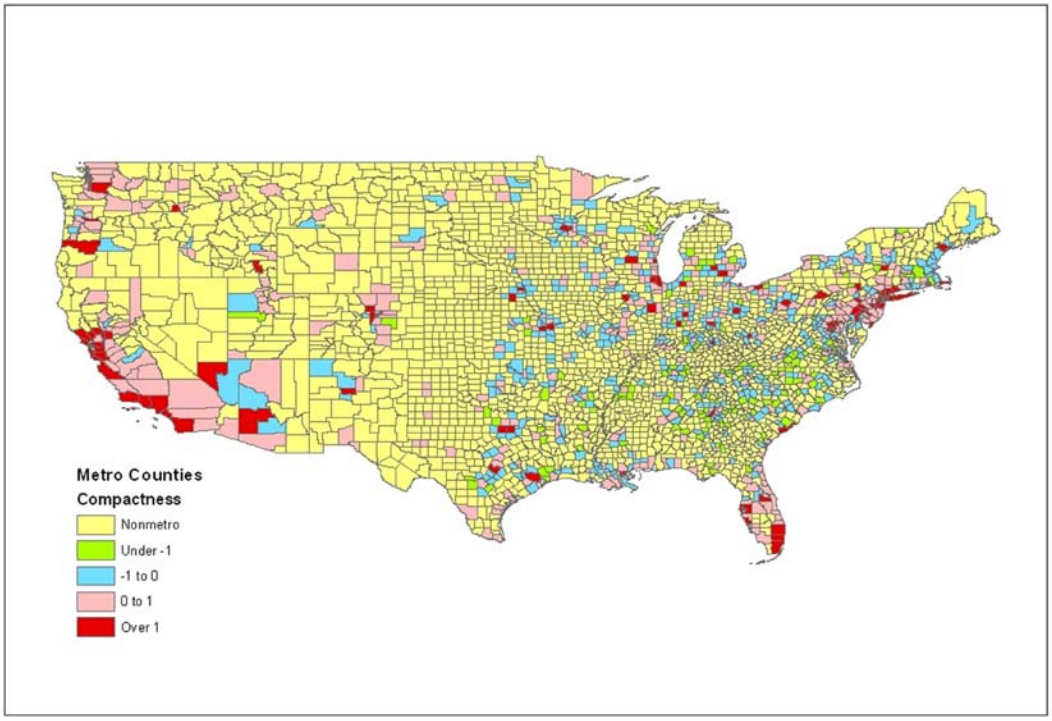

Figure 1 Compactness Scores (Standardized Values)

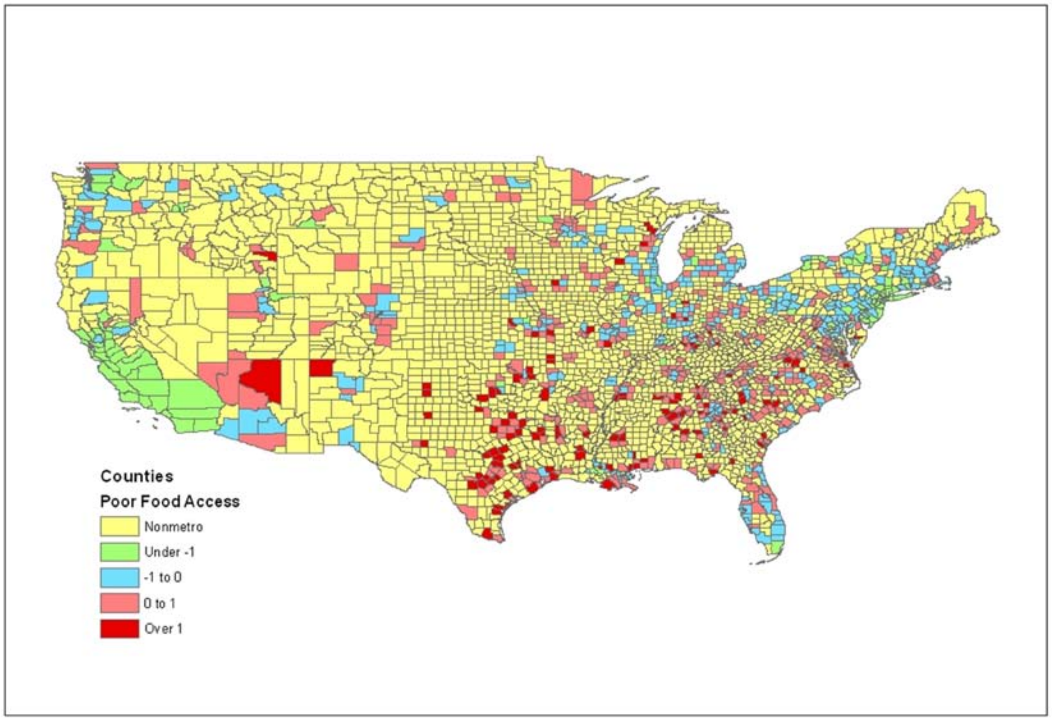

Figure 2 Scores for Low Healthy Food Access (Standardized Values)

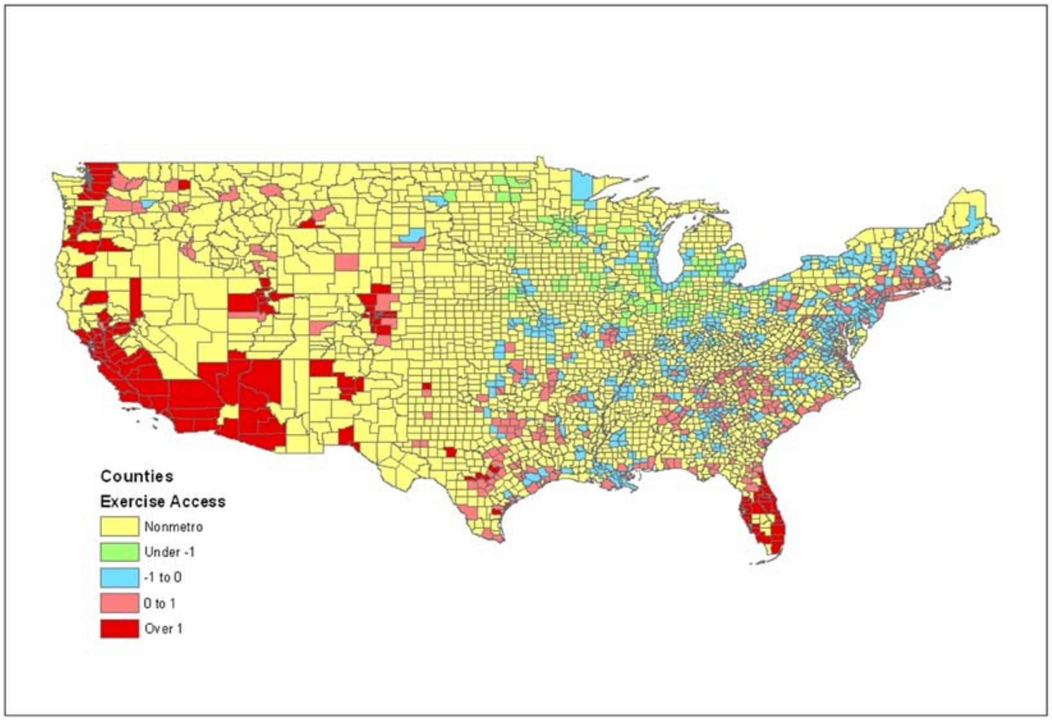

Figure 3 Scores for Exercise Access (Standardized Values)

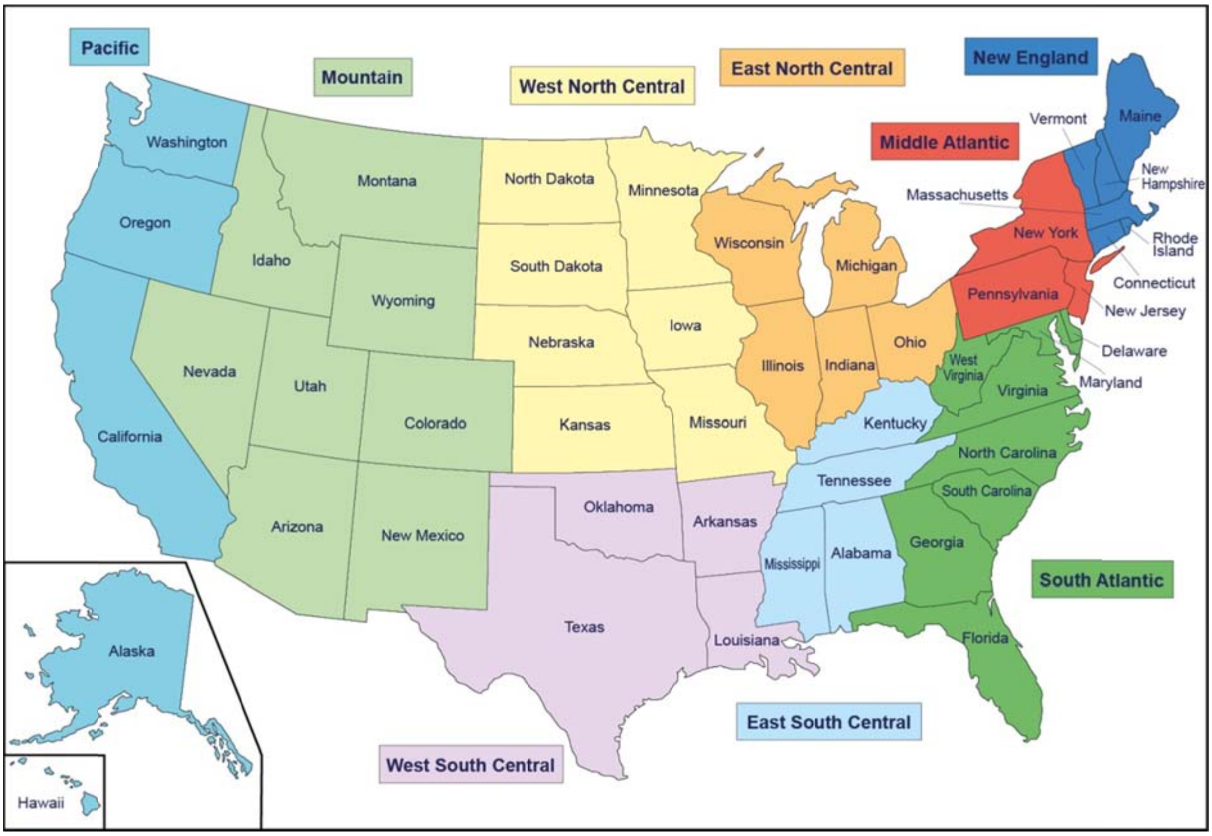

Figure 4 US Census Divisions