Applied Geography 44 (2013) 134–142

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Can Geographically Weighted Regression improve our contextual understanding of obesity in the US? Findings from the USDA Food Atlas

Sang-Hyun Chi [a], Diana S. Grigsby-Toussaint [b], Natalie Bradford [b], Jinmu Choi [a,*]

[a] Department of Geography, Kyung Hee University, Seoul, Republic of Korea
[b] Department of Kinesiology and Community Health, Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Keywords:
Obesity
Food environment
Geographically Weighted Regression
Spatial analysis

A B S T R A C T

There is growing interest in the role the food environment as well as demographic and socio-economic factors play in the prevalence of obesity in the US. Existing empirical evidence examining the association between the food environment and obesity risk, however, remains equivocal. We hypothesized that spatial heterogeneity may account for the conflicting results. Using Geographically Weighted Regression, we examined how the associations between the food environment, and demographic and socio-economic variables associated with obesity vary over space at the county level in the US. The analysis shows that higher ratios of convenience-to-grocery stores, poverty rate, and urban environments were positively associated with obesity risk in the US. Conversely, areas with better physical environments were negatively associated with obesity risk. Most importantly, the association between obesity and all major explanatory variables in our analysis significantly varied over space.

Introduction

Globally, obesity has evolved into a major public health concern. In the United States specifically, obesity is a recognized epidemic, prevalent amongst 35.7% of adults and 16.9% of children (Ogden, Carroll, Kit, & Flegal, 2010). The increased prevalence of obesity accounts for billions of dollars in medical and prescription costs to the health care industry (i.e., private insurers, Medicare, and Medicaid), further highlighting the need to prevent the onset of this epidemic (Finkelstein, Trogdon, Cohen, & Dietz, 2009). Given the association between obesity and several chronic conditions such as diabetes and cardiovascular disease, research to disentangle the complexity of obesity risk is particularly important (Khan et al., 2009; World Health Organization, 2003). In efforts to impede the growing obesity epidemic and its burdens, it is crucial to identify and understand the impact of risk factors associated with the condition.

It is well documented that obesity is often the result of modifiable behaviors, namely unhealthy eating patterns and physical inactivity. While important to the study and prevention of obesity, individual behavioral risk factors are difficult to target and extrapolate to the population at large (Summerbell et al., 2005). Furthermore, no significant individual psychological, biological, or metabolic phenomena have been observed in recent decades that could explain the trends in obesity rates (Sallis & Glanz, 2009). Thus, research focused solely on individual risk factors fail to address the full scope of the obesity epidemic. Instead, environmental risk factors such as socio-economic status, access to healthy and affordable food options, and physical activity resources are essential to the study of obesity (Hendrickson, Smith, & Eikenberry, 2006; Larsen & Gilliland, 2009). Hence, there is growing interest in the study of the food environment and the relationship with obesity risk in the US (Ford & Dzewaltowski, 2010; Gordon-Larsen, Nelson, Page, & Popkin, 2006; Holsten, 2009; Morland, Diez Roux, & Wing, 2006; Russell & Heidkamp, 2011). The food environment refers to public places in a community like grocery and convenience stores, restaurants, and farmers' markets where food can be obtained (Sallis & Glanz, 2009). Dramatic changes in the prevalence of obesity have been accompanied by dramatic changes in the food environment. For example, the number of states with obesity rates above 15% rose from zero to 30 between 1986 and 1996 (Centers for Disease Control and Prevention, 2010). During this same period the number of fast-food restaurants increased by 85 percent while the number of food stores decreased by 15 percent (Cohen, 2008).

* Corresponding author.
E-mail addresses: hyungeo@khu.ac.kr (S.-H. Chi), dgrigs1@illinois.edu (D.S. Grigsby-Toussaint), bradfor3@illinois.edu (N. Bradford), cjm89@khu.ac.kr, jinmuchoi@gmail.com (J. Choi).

0143-6228/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apgeog.2013.07.017

Many studies support a positive correlation between unhealthy food environments and obesity (Bodor, Rice, Farley, Swalm, & Rose, 2010; Mehta & Chang, 2008; Morland et al., 2006; Spence, Cutumisu, Edwards, Raine, & Smoyer-Tomic, 2009). These studies operate on the assumption that obesogenic environments develop in areas where there is limited access to healthy food resources or an abundance of energy dense foods (Block, Scribner, & DeSalvo, 2004; Morland & Evenson, 2009; Pearce, Blakely, Witten, & Bartie, 2007; Spence et al., 2009). Similarly, previous research identifies certain characteristics of the food environment that promote healthy dietary behaviors and are associated with a reduced risk for obesity. For example, the availability of more supermarkets in neighborhoods has been linked to increased fruit and vegetable consumption as well as a decrease in the risk for obesity (Ford & Dzewaltowski, 2008; Larson, Story, & Nelson, 2009). Some studies, however, refute the correlation between food environments and obesity (Burdette & Whitaker, 2004; Macintyre, McKay, Cummins & Burns, 2005; Simmons et al., 2005). A cross-sectional study of low-income children in Ohio, for example, showed no association between overweight and proximity to fast-food restaurants and other elements of the built environment (Burdette & Whitaker, 2004).

These mixed results may be explained by the absence of studies that accurately account for regional differences amongst food environments. Most studies employ a universal model or focus on a single explanatory variable of the food environment – methods that neglect to adjust for spatial variation in the food environment (Caspi, Sorensen, Subramanian, & Kawachi, 2012). Glanz, Sallis, Saelens, and Frank (2005) propose a conceptualization of the food environment using an ecological-based model that accounts for external variables such as industry policies, type and location of food outlets, price, and food advertisements. Ecological models like this are extremely useful for research and intervention purposes because they examine multiple risk factors at the population-level in addition to individual variables.

Thus, it is imperative for future research on obesity and the food environment to adjust for regional differences in government policies, community demographics, allocation of resources, and other external variables. This paper attempts to further explain the complexity in understanding risk factors for obesity. In particular, we are concerned with the impact of food environments on obesity risk, and how that impact varies across space. We argue that the geographic scale of a study has a major influence on measurements of accessibility and implications about the association between food environments and obesity. Multivariate and spatial (Geographically Weighted Regression) analyses are used to demonstrate the contextual nature of the food environment and its impact on obesity risk in the US. Ultimately such research may lead to the implementation of community-specific interventions to combat the growing obesity epidemic.

Methods

Data source

The United States Department of Agriculture (USDA) Economic Research Service (ERS) Food Environment Atlas (hereafter the Atlas), was used as the primary data source for this study [http://www.ers.usda.gov/data-products/food-environment-atlas.aspx, accessed July 1, 2011]. The Atlas is a publically available database with more than 160 indicators on the food environment primarily at the county level in the US, including availability of food stores, expenditures on food, as well as other socio-economic characteristics that may influence food access and consumption. The information used for this study in the Atlas was compiled from *Access to Affordable and Nutritious Food — Measuring and Understanding Food Deserts and Their Consequences: Report to Congress* of 2006 and the census of 2000. Data on 3108 counties were used for the analysis while counties in Alaska and Hawaii were excluded due to the non-contiguity with the rest of the US States and the presence of many missing variables.

Measures

Food and physical activity environment

Several variables were selected to capture contextual influences of the food environment on obesity risk. In order to account for *food availability*, the number of grocery stores, convenience stores, and fast-food restaurants were selected from the database. Instead of using the conventional method of calculating food and restaurant density per capita, we opted to use a measure developed by Mehta and Chang (2008) that involves estimating the ratio of convenience vs. grocery stores and fast-food vs. full-service restaurants in an area of interest. The assumption behind this method is that grocery stores are typically the main source of healthy food, such as fresh fruits, vegetables and high fiber-containing food, while convenience stores primarily provide unhealthy, energy dense processed foods (Michimi & Wimberly, 2010). Additionally, there is some evidence to suggest that the utilization of fast-food restaurants may result in higher consumption of energy dense foods compared to full-service restaurants (French, Harnack, & Jeffery, 2000; Jeffery & French, 1998; Kim & Leigh, 2011).

To account for levels of *food accessibility* (Bodor et al., 2010; Spence et al., 2009), specifically healthy foods across the US, we used the percentage of households without cars and the absence of grocery stores within a 1-mile radius. Again, our goal here was to reflect access to healthier foods that could attenuate obesity risk. Additionally, the role of the *physical environment* on obesity risk was considered by including the USDA's natural amenity index as a proxy for the influence of climate and the natural landscape (e.g., greenspace) on physical activity behaviors (Tucker & Gilliland, 2007). The index is a composite score of varied topography, such as lakes, ponds, oceanfront and climate to encompass all four seasons in the US (i.e., winter, spring, summer, fall). As several studies have demonstrated an association between access to an aesthetically pleasing environment (e.g., greenspace) and higher levels of physical activity among children and adults (Coombes, Jones, & Hillsdon, 2010; Grigsby-Toussaint, Chi, & Fiese, 2011; Jilcott, Keyserling, Crawford, McGuirt, & Ammerman, 2011; Jilcott, Moore, Shores, Imai, & McGranahan, 2011; Wang, Wen, & Xu, 2013), we hypothesized that an inverse relationship would exist between increasing natural amenity index scores and increasing obesity risk in the US.

Socio-demographic factors

Due to the well-documented racial and economic disparities in the food environment across the US (Grigsby-Toussaint, Zenk, Odoms-Young, Ruggiero, & Moise, 2010; Kirby, Liang, Chen, & Wang, 2012; Morland, Wing, & Roux, 2002; Powell, Slater, Mirtcheva, Bao, & Chaloupka, 2007; Zenk et al., 2005), we incorporated measures of economic deprivation and racial composition in our analysis. Specifically, the poverty rate of a county was included as a proxy for the economic deprivation (Bodor et al., 2010; Pearce et al., 2007; Wen, Chen, & Tsai, 2010), and the percentage of Whites in each county was used to account for racial disparities in obesity risk (Bloc, Scribner, & DeSalvo, 2004; Powell et al., 2007). Due to the high inverse correlation between racial groups at the county level (e.g., higher percentages of Blacks mirroring lower percentages of Whites), we included only the percentage of Whites in each county since many of the counties are

White dominant (Kim, Subramanian, Gortmaker, & Kawachi, 2006).[1] Finally, due to observed differences in obesity risk between urban and non-urban areas (Murray et al., 2006; Sallis & Glanz, 2006), urban influence was considered in the analysis as a dichotomous variable of metro and non-metro counties based on the US census divisions.

Analysis

We used Geographically Weighted Regression (GWR) to examine the spatially heterogeneous association between obesity risk and 7 explanatory variables (see Table 1) across the US (Chalkias et al., 2013; Procter, Clarke, Ransley, & Cade, 2008; Wen et al., 2010). The GWR model can be expressed by using matrix notation.

$$y_i = x_i\beta_i + \varepsilon_i$$

where β_i refers to the column vector of estimated regression coefficients and X_i indicates the row vector of explanatory variables at location i (Wheeler & Páez, 2010; 462). As denoted by i, observations close enough to the point of interest (W_i) were included in the process of estimation of the regression coefficient. Therefore the estimated local coefficient ($\widehat{\beta}$) is expressed as

$$\widehat{\beta} = \left[X^T W_i X\right]^{-1} X^T W_i Y$$

where $[X^T W_i X]^{-1}$ is the inverse matrix of independent variables that are in the estimation process while $W_i Y$ is the dependent variable that is also used for the regression calculation by the design of the weight matrix.

In GWR, parameters are estimated at all observations with a dependent variable and independent variables. Hence, the regression coefficients are not static but show spatial variation over the study area (Charlton, Fotheringham, & Brunsdon, 2009). In other words, the relationship between a certain food environment variable and obesity prevalence may change over space. Since the parameters estimated in GWR are location specific, the influence and explanatory power of independent variables can be examined at every point. The estimation of local parameters in GWR is accomplished by placing more weight on observations in closer geographic proximity and excluding observations over the distance of indifference. The inclusion/exclusion of observations varies with a spatial kernel that is used in the estimation process. Since the observations in the estimation process are decided by the spatial kernel, the choice of the kernel is crucial. There are two frequently used spatial kernels, one is a fixed kernel and the other is an adaptive kernel. The former keeps a one-sized kernel over the whole study area, whereas the size of kernel varies according to the spatial distribution of observations in the latter. In the adaptive method, the size of kernel becomes large to secure a sufficient number of observations when the observations are sparsely distributed. In this study, the adaptive method was used to account for the difference in the size of counties across the US. Consequently, we eliminate the challenge of having too few observations at the county level in the West and too many in the East.

The Akaike Information Criterion (AIC) minimizing method was used to estimate the size of the kernel (Fotheringham, Brunsdon, & Charlton, 2002). The adaptive kernel size estimated by the AIC method in our study varied to secure 160 observations in the kernel. The spatial weights (W_{ij}) can be expressed as:

$$w_{ij} = \begin{cases} \left[1-\left(\frac{d_{ij}}{d_{ik}}\right)^2\right]^2 & \text{if } d_{ij} \le d_{ik} \\ 0 & \text{if } d_{ij} > d_{ik} \end{cases}$$

where d_{ij} is the distance between observation i and j, and d_{ik} refers to the distance from observation i to k that is located at the rim of the spatial kernel.

In GWR, the regression model is expressed as:

$$\text{Obesity rate at location } i(u) \sim \beta_{1i}(u) \text{ Ratio of convenience stores vs. grocery stores}_i + \beta_{2i}(u) \text{ Ratio of fast}-\text{food restaurant vs. full}-\text{service restaurants}_i + \beta_{3i}(u) \text{ Ratio of household with no car and more than one mile to grocery store}_i + \beta_{4i}(u) \text{ Natural Amenity}_i + \beta_{5i}(u)\%\text{ White}_i + \beta_{6i}(u) \text{ Poverty Rate}_i + \beta_{7i}(u) \text{ metro}_i\text{:}$$

where β_{ki} indicates the parameter that explains the association around location i with the geographic coordinate (u). In the GWR model, goodness of fit, coefficients and statistical significance are estimated at each point whereas the OLS model produces one fixed value. Also, Monte Carlo simulation is employed to test whether the spatial variation of local parameters of GWR is statistically significant (Charlton et al., 2009). The Monte Carlo simulation is performed to test the null hypothesis that β_{ki} does not vary with i.

The variability (V_j) of β_{ki} can be expressed as the sum of squares of the difference between estimated coefficients and the mean of the coefficients on explanatory variable j (Brunsdon, Fotheringham, & Charlton, 1996).

$$V_j = \sum_i \left(\beta_{ij} - \beta_{\bullet j}\right)^2 / N$$

Table 1
Descriptive summary of the study population (mean value if not specified).

Variable	National ($N = 3108$)		Regions (N for each region below)			
	Mean	Std. dev.	Northeast	Midwest	South	West
Outcome variable						
Obesity rate, %	28.93	3.70	27.53	29.18	30.32	24.55
Explanatory variables						
Ratio of convenience-to-grocery stores	2.90	1.84	2.36	2.51	3.50	2.27
Ratio of fast-food-to-full-service restaurants	1.45	1.20	1.43	1.81	1.13	1.63
Households with no car living more than 1 mile from a grocery store, %	3.98	2.60	3.74	3.13	5.06	2.74
Natural Amenity Score [a]	3.49	1.04	3.47	2.70	3.64	5.03
White, %	79.54	19.04	81.45	90.66	71.83	75.55
Poverty rate, %	15.27	6.05	12.99	12.84	17.91	14.14
Urbanization [b]						
Metro	1085 [c]		146	528	285	126
Non-metro	2023		120	846	770	287
Valid N	3108		266	1374	1055	412

[a] Lowest = 1, Highest = 7.
[b] Dichotomous variable, Metro = 1, Non-metro = 0 (reference category).
[c] Number of counties in the categories.

[1] The Pearson correlation coefficient between the percentage of White and Black is −0.647 (statistically significant at a 0.001 level). The proportion of Whites is more than 70% in 2291 counties (73.7%) out of 3108.

The Monte Carlo simulation shuffles the location of i and calculates the values of V. Then the simulation computes the rank of V_j in the continuum of V to see if the value of V_j is due to chance or not. The GWR3 software (Fotheringham et al., 2002) was used to perform the analysis.

Results

Summary characteristics of the study population

Table 1 shows the descriptive statistics of the 3108 counties across the continental US included in our analysis with the maps of key variables in Fig. 1. Our sample consisted of 2023 non-metro counties, and 1085 metro counties. The mean adult obesity prevalence across the US was 28.93%, with the highest overall rates found in the South (30.32%), and the lowest rates found in the Northeast (27.53%). The highest obesity rate is 43.5% in Greene County, Alabama, and the lowest is 12.5% in Boulder, Colorado. On average, Whites accounted for 79.54% of each county (SD, 19.04), and the average poverty rate was 15.27% (SD, 6.05). Approximately 4% of individuals lived in homes more than 1 mile away from a grocery store without a car, but this rate went up to 5.06% for individuals residing in the South. The ratio of convenience-to-grocery stores is highest in the South (3.50) and lowest in the West (2.27), while the ratio of fast-food-to-full-service restaurants shows the opposite pattern. In the South the number of fast-food restaurants is 1.13 times higher than full-service restaurants but in the Midwest the ratio goes up to 1.81 (Table 1). The spatial variations of obesity and the main food environment variables are visualized in Fig. 1.

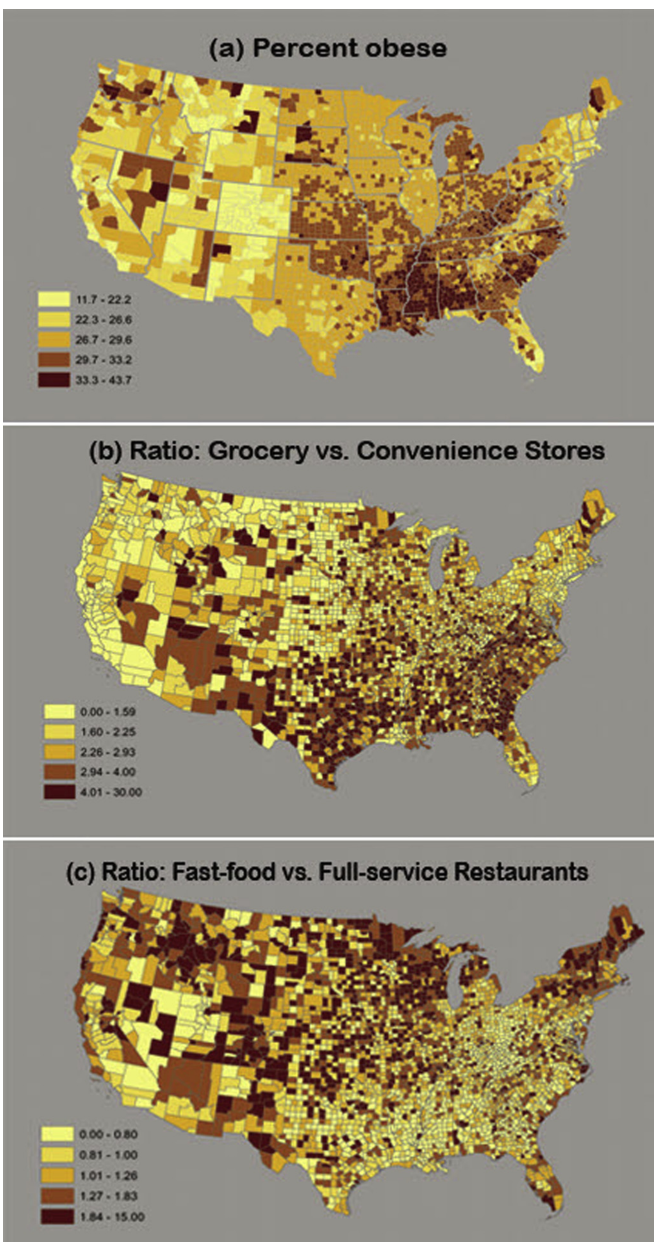

Fig. 1. Distribution of major variables.

OLS and GWR regression analysis

Table 2 shows the results of the OLS regression analysis with the comparison GWR model. Many variables in the model are statistically significant in the OLS model. The ratio of convenience stores-to-grocery stores shows a positive association with the obesity rate at the national level. Contrary to our expectation, however, the ratio of fast-food to full-service restaurants shows a negative relationship. Areas with higher levels of natural amenity, i.e., better physical environments, were associated with a lower risk of obesity, while higher poverty rates were highly associated with obesity risk. In addition, people living in more urbanized environments were more likely to be at higher risk for obesity.

Compared to the OLS, in the GWR model, the r-square significantly increased from 0.44 (OLS) to 0.73 (GWR), showing that the GWR outperformed OLS. The result confirms that the differences in obesity prevalence in the US can be better explained by examining spatially heterogeneous processes. In other words, the regional model indicates better explanatory power than the national scale analysis. The advantage of GWR, however, should be examined with a caveat. Fig. 2 shows where the local r-squares of the GWR are larger/smaller than that of OLS. As seen in the map, GWR does not always outperform the OLS. In some regions, such as Washington State and its vicinities, and part of the Midwest, the explanatory power of the GWR is not satisfactory. This means that the GWR may not be the best alternative to OLS when there is a high level of heterogeneity and GWR and OLS should be used as complementary methods. Consistent with the OLS model, the GWR model also found a positive association between higher ratios of convenience stores-to-grocery stores, and obesity risk. The association between higher ratios of fast-food-to-convenience stores, however, was positively associated with increased obesity risk. Interestingly, in the GWR model, a negative association was found between obesity and households with limited access to grocery stores based on car-ownership, concentration of Whites in an area, and areas with higher indices of natural amenity. However, the car-ownership variable was not significant at the $p < 0.05$ level.

Visualization and clustering techniques

Although the directions of coefficients in the GWR are similar to those of OLS at the national scale (Table 2), the Monte Carlo test on local coefficients of variables shows that the spatial variation is statistically significant, which means that the impacts of these variables change over space. This result indicates that the assumption of a universal simple linear relationship cannot be used to explain the relationship between food environment and obesity prevalence at the US county level. Although GWR provides useful information, it is sometimes challenging to interpret the statistical results because GWR produces locally specific parameters, which means there are as many regression coefficients as the number of

Table 2
Comparison of parameters (OLS vs. GWR).

Explanatory variables	OLS		GWR							
	B value (S.E)	*p*-value	Median value (S.D)	Min	1st quartile	2nd quartile	3rd quartile	4th quartile	Max	Spatial variation of local estimates [a]
Constant	29.41 (0.47)	0.00	29.9 (7.00)	0.00	25.96	30.05	34.49	49.25	49.25	0.00
Ratio (convenience vs. grocery store)	0.18 (0.03)	0.00	0.14 (0.23)	−0.45	0.00	0.00	0.00	1.72	1.72	0.00
Ratio (fast-food vs. full-service restaurants)	−0.2 (0.04)	0.00	0.1 (0.48)	−2.52	0.00	0.00	0.00	0.67	0.67	0.00
Percent of households with no car and more than 1 mile to grocery store	0.33 (0.03)	0.00	−0.14 (0.21)	−0.39	0.00	0.00	0.28	1.36	1.36	0.07
Natural Amenity	−1.44 (0.05)	0.00	−0.28 (0.73)	−3.35	−0.61	0.00	0.00	1.19	1.19	0.00
Percent White	0.00 (0.00)	0.13	−0.02 (0.06)	−0.17	−0.07	0.00	0.00	0.22	0.22	0.00
Poverty rate	0.21 (0.01)	0.00	0.09 (0.12)	−0.26	0.00	0.00	0.17	0.54	0.54	0.00
metro	0.47 (0.12)	0.00	0.09 (0.63)	−6.22	0.00	0.00	0.00	1.96	1.96	0.14
Residual sum of squares	23,647		10,113							
Akaike Information Criterion	15,145		13,342							
Adjusted *r*-square	0.443		0.730							
Moran's I of residuals [b]	0.443		0.091							

[a] Monte Carlo test for the spatial variability of regression coefficients.
[b] Contiguity-based weight matrix was used to model the spatial relationships.

counties. Therefore, we used visualization techniques and cluster analysis to summarize the spatial variation of regression parameters. Fig. 3 shows the variation of local parameters of each variable. As seen in the map, there is significant spatial variation across the US for all of the variables included in our analysis. For example, the variation of local intercepts (Fig. 3(a)) shows that the residents in some counties in the Deep South and Dakotas are more likely to be obese even after we control for other explanatory variables. Also the statistical significance of independent variables shows a high level of regional variation. In other words, if we control for food environment, socio-economic status, and built environment, a single factor may be important for explaining the level of obesity in one region but not in other regions.

In order to further summarize the information generated by the GWR, we used *k*-means cluster analysis to classify groups of regions according to the different associations between dependent and independent variables (Windle, Rose, Devillers, & Fortini, 2010). Although clustering analyses have been widely used in various studies, a few unsolved problems have existed since the methods were introduced (Everitt, 1979). Such problems include 1) the number of predefined clusters or even numbers, 2) the lack of reliable validation techniques to determine whether if the classification is successful, and 3) variations in the measurement of distance among groups, among others. With the consideration of these limitations, we used cluster analysis for descriptive analysis. In other words, the cluster analysis was used to summarize the massive amount of information produced by the GWR. We also used *k*-means cluster analysis rather than resorting to hierarchical clustering since we did not assume the existence of a definite number of clusters. We tried to examine regional patterns by increasing the depth of classification.

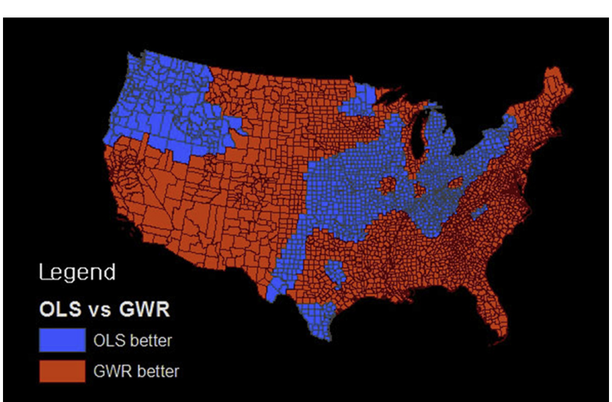

Fig. 2. Comparison of OLS and GWR.

There are some benefits in the use of cluster analysis to summarize the results of GWR. First, cluster analysis separates GWR parameters into different zones (Windle et al., 2010). By grouping counties into several zones, it is possible to determine a more relevant unit of analysis to investigate the role of explanatory variables. For example, it is well known that poverty rate or urbanized physical environment is positively associated with high obesity rates at the national scale. However, the cluster analysis using the *t*-values of the GWR, may provide an opportunity to look at the association between obesity and poverty rate or physical environment variables between designated clusters after controlling for other variables. Second, the cluster analysis shows the relative importance of risk factors according to the regions in the US. The OLS regression at the national scale only provides the average of the influence of risk factors on obesity. Compared to the result of OLS, the cluster analysis of GWR suggests the relative importance of risk factors according to the zones. Some risk factors may have crucial role in one region but may not in other regions. The relative importance of food environment, physical environment, and socio-economic factors can be used for more effective policy engagement.

Fig. 4 shows the groups of regions identified by the *k*-means cluster analysis. To make sense of the maps, the descriptive statistics of clusters in each round are provided in Table 3. Table 3 summarizes the characteristics of each cluster based on the associations between obesity and the explanatory variables we used. The result of the first round ($k = 2$) classified counties into two groups. Cluster 1 (red) includes the Northeast, a part of the Midwest, California, Texas, Florida, Arizona and Colorado. Counties in cluster 1 are separated from cluster 2 in that the ratio of convenience-to-grocery stores has a larger influence on the obesity rate. In addition, the obesity rate in cluster 1 is less influenced by natural amenity and more influenced by poverty rate compared to cluster 2. In the next depth ($k = 3$), cases in the Northeast (cluster 1: red) are spatially limited to the Northeast and the Chicago area. Also, cluster 3 (blue) is centered on the Dakotas, Montana, and the Deep South. In cluster 1 (red), the obesity rate is positively correlated with households without a car, percent white, poverty rate, and metro variables. In other words, the obesity rates in urbanized regions and their vicinities are more likely correlated with mobility in buying food and socio-economic status. In addition, metro counties are more likely to show a high level of obesity prevalence

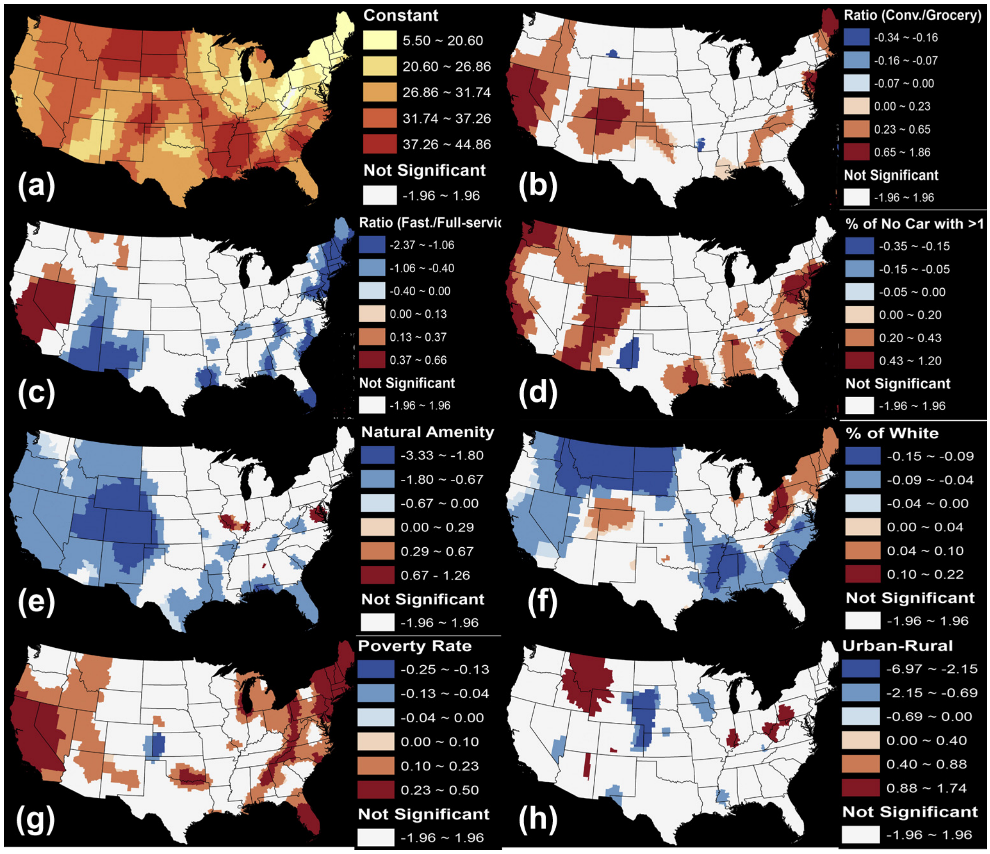

Fig. 3. Local parameters from GWR modeling. ((a) Constant, (b) Ratio (convenience vs. grocery), (c) Ratio (fast-food vs. full-service), (d) Percent of no car with more than 1 mile to grocery store, (e) Natural Amenity, (f) Percent White, (g) Poverty Rate, (h) Metro vs. non-metro.)

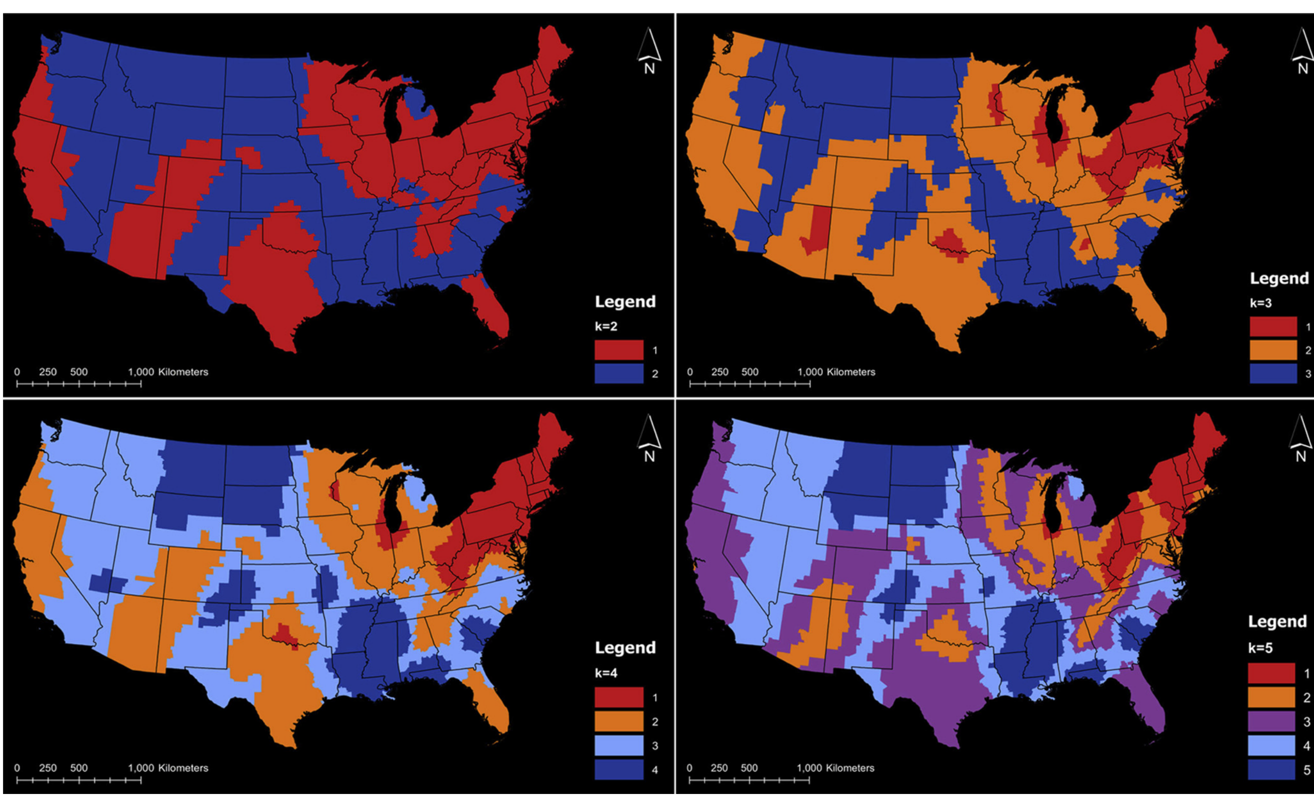

Fig. 4. Regions classified by *k*-means cluster analysis of local coefficient from the GWR.

Table 3
Characteristics of clusters identified by the *k*-means cluster analysis.

K	Clusters	*n*	Mean GWR parameter estimates							
			Constant	Ratio (convenience vs. grocery store)	Ratio (fast-food vs. full-service restaurant)	Percent of homes without cars and more than 1 mile to grocery store	Natural Amenity	Percent White	Poverty rate	Metro
2	1	1661	25.12	0.18	−0.34	0.18	−0.26	0.02	0.15	0.09
	2	1447	35.57	0.08	−0.25	0.17	−0.65	−0.06	0.06	−0.15
	F-statistics [a]		4808*** [b]	125.1***	29.62***	0.825	240.6***	2697***	607.5***	78.11***
3	1	526	19.59	0.15	−0.62	0.22	−0.06	0.06	0.23	0.34
	2	1588	28.82	0.17	−0.22	0.16	−0.37	−0.01	0.10	−0.06
	3	994	37.36	0.07	−0.25	0.16	−0.74	−0.08	0.04	−0.16
	F-statistics		1013***	116.5***	10.35***	0.216	128.2***	672.8***	118***	5.049**
4	1	436	18.77	0.14	−0.70	0.24	−0.09	0.07	0.25	0.39
	2	1148	27.19	0.20	−0.21	0.15	−0.30	0.00	0.12	−0.02
	3	970	32.84	0.11	−0.21	0.18	−0.55	−0.05	0.08	−0.05
	4	554	39.61	0.04	−0.30	0.15	−0.79	−0.09	0.02	−0.30
	F-statistics		410.8***	2.75*	301.2***	13.74***	28.75***	385.1***	251.6***	29.28***
5	1	274	16.85	0.16	−0.70	0.19	−0.20	0.09	0.28	0.49
	2	624	24.43	0.16	−0.36	0.16	0.01	0.02	0.13	0.05
	3	1006	29.22	0.18	−0.20	0.19	−0.47	−0.02	0.11	−0.04
	4	756	34.25	0.11	−0.22	0.17	−0.65	−0.06	0.08	−0.06
	5	448	40.29	0.03	−0.30	0.15	−0.78	−0.10	0.01	−0.34
	F-statistics		24.99***	35.8***	18.02***	12.1***	43.84***	26.46***	39.88***	6.081**

[a] ANOVA test on the differences between mean of local coefficients in clusters.
[b] * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

compared to surrounding non-metro counties. Compared to cluster 1, cluster 3 (blue) is the group of counties where natural amenity plays a pivotal role in explaining obesity rate in a county while influences of socio-economic variables and the food environment seem to have less influence. Also, non-metro counties show higher obesity rates than metro counties in cluster 3. The cluster 3 is unique in that the cluster is distinguished from the others mainly because natural amenity is a key explanatory variable in cluster 3 rather than other socio-economic variables. The results should be interpreted with caution because it does not mean that socio-economic factors are not important in cluster 3. Rather it does mean that natural amenity is important since there is a high level of homogeneity in socio-economic status in cluster 3 compared to other clusters, which suggests that general OLS and local scale analysis (GWR) are complementary to each other. Cluster 2 (orange) shows intermediate characteristics between clusters 1 and 3. In the later stages of cluster analysis ($k = 4$ and 5), the basic structure of the first two stages hold. In the last round ($k = 5$), clusters 1–2, centered on cluster 1 can be regarded as urban style clusters that are composed of big cities and their surrounding counties. In clusters 1–3, the food environment, specifically convenience vs. grocery store ratio has a larger influence on obesity rate than in clusters 4 and 5. The natural amenity index has limited explanatory power while metro counties show higher obesity rate than non-metro counties in clusters 1 and 2. While clusters 1 and 2 are composed of counties in the Northeast, clusters 4 and 5 are centered on the Great Plains and Deep South where good natural amenity seems to have a preventive impact on obesity while food environment and poverty rate have limited explanatory power among counties in clusters 4 and 5. Also metro counties in these counties are more likely to show lower obesity rate than non-metro counties.

Discussion and conclusion

The results from this multivariate and GWR analysis demonstrate several implications for improving our contextual understanding of obesity risk in the US. Similar to previous studies, high density of convenience stores over grocery stores (Bodor et al., 2010; Mehta & Chang, 2008; Spence et al., 2009), poverty rate (Bodor et al., 2010; Pearce, Hiscock, Blakely, & Witten, 2008; Pearce et al., 2007), and urbanized built environments (Murray et al., 2006; Sallis & Glanz, 2006) are positively associated with obesity whereas better natural amenity (Coombes, Jones, & Hillsdon, 2010; Grigsby-Toussaint et al., 2011; Tucker & Gilliland, 2007), such as mild weather and greenspace access is negatively associated with obesity. Second, the correlation of the food environment with obesity strongly suggests the interpretation and the analysis of the Atlas should be carefully implemented. For instance, the ratio of fast-food/full-service restaurant shows a negative association with obesity rate in many counties. The seemingly odd result may be due to the definition of fast-food restaurants that is defined as "pay before eat" establishments. As Austin (Austin et al., 2005) discussed, the definition does not accurately represent the traditional notion of fast-food establishments that provide high-energy dense food. Another possible explanation is that the net effect of fast-food is not influential if we controlled for other factors since eating-out accounts for a small portion of energy intake compared to the meals at home (Austin et al., 2005). When it comes to the energy density of food, the definition does not capture the food quality since there are many restaurants classified as fast-food restaurants that serve healthy food. The same problem can be found in the definition of grocery stores as well. In fact, the distribution of chain supermarkets, non-chain supermarkets, and grocery stores is significantly different at zip code level (Powell et al., 2007). In the same vein, grocery stores can be further classified into chain supermarkets, independently owned grocery stores, and specialty food stores (Morland et al., 2006). Because the size and type of grocery stores are closely related to the price of food, there is a need to subdivide the type of stores and restaurants for better representation of the food environment. Finally and the most importantly, the GWR model demonstrates that the association between food environment and obesity varies across space. This result is a partial answer for the reason why there is controversial empirical evidence about the relationship between variables of food environment and obesity. In other words, the result of the GWR model suggests a caveat for the generalization of the study conducted at the local or neighborhood scale. The local level analysis focusing on spatial heterogeneous processes cannot be standalone methodology since the combination of local level analysis and global scale analysis may provide a better understanding of obesity by "separating the forest from the trees" (Wang et al., 2013). Examining the association among obesity and per capita famers' markets, grocery stores/super markets, and supercenters in US counties, Jilcott et al. found that the high density of food venue is positively associated with obesity (Jilcott, Keyserling, et al., 2011). However, they pointed out that the association is not linear and may vary based on region of the US. Our study is based on the question left by the research of Jilcott, Keyserling, et al. (2011). We tried to investigate how the association between food environment and obesity vary over space. Also this study may provide a partial answer about why there are mixed results in empirical studies on the influence of food outlets and obesity risk. For instance, while several studies have confirmed the positive association between high density of fast-food restaurants and convenience stories with obesity (Bodor et al., 2010; Morland & Evenson, 2009; Morland et al., 2006; Spence et al., 2009), there are contradictory studies showing no correlation between the density of food venues and obesity (Burdette & Whitaker, 2004; Macintyre et al., 2005; Simmons et al., 2005). Despite the implication, however, the result of the GWR at county level leaves many questions and potential research avenues in the study of food environment. The spatial variation of the food environment influence on obesity suggests the need to investigate the mediating processes between food environment and obesity as well as the necessity of improving food environment data.

References

Austin, S. B., Melly, S. J., Sanchez, B. N., Patel, A., Buka, S., & Gortmaker, S. L. (2005). Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments. *American Journal of Public Health, 95*(9), 1575–1581.

Block, J. P., Scribner, R. A., & DeSalvo, K. B. (2004). Fast food, race/ethnicity, and income. *American Journal of Preventive Medicine, 27*(3), 211–217.

Bodor, J. N., Rice, J. C., Farley, T. A., Swalm, C. M., & Rose, D. (2010). The association between obesity and urban food environments. *Journal of Urban Health, 87*(5), 771–781.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis, 28*(4), 281–298.

Burdette, H. L., & Whitaker, R. C. (2004). Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children. *Preventive Medicine, 38*(1), 57–63.

Caspi, C. E., Sorensen, G., Subramanian, S. V., & Kawachi, I. (2012). The local food environment and diet: a systematic review. *Health & Place, 18*(5), 1172–1187. http://dx.doi.org/10.1016/j.healthplace.2012.05.006.

Centers for Disease Control and Prevention. (2010). *Obesity trends among U.S. adults between 1985 and 2010*. Retrieved from: http://www.cdc.gov/obesity/downloads/obesity_trends_2010.pdf.

Chalkias, C., Papadopoulos, A. G., Kalogeropoulos, K., Tambalis, K., Psarra, G., & Sidossis, L. (2013). Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: empirical evidence from Athens, Greece. *Applied Geography, 37*, 34–43.

Charlton, M., Fotheringham, S., & Brunsdon, C. (2009). *Geographically weighted regression*. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.

Cohen, D. A. (2008). Obesity and the built environment: changes in environmental cues cause energy imbalances. *International Journal of Obesity, 32*, S137–S142. http://dx.doi.org/10.1038/ijo.2008.250.

Coombes, E., Jones, A. P., & Hillsdon, M. (2010). The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social Science & Medicine, 70*(6), 816–822.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. *Biometrics*, 169–181.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: payer- and service-specific estimates 2009. *Health Affairs, 28*(5), w822–w831. http://dx.doi.org/10.1377/hlthaff.28.5.w822.

Ford, P. B., & Dzewaltowski, D. A. (2008). Disparities in obesity prevalence due to variation in the retail food environment: three testable hypotheses. *Nutrition Reviews, 66*(4), 216–228. http://dx.doi.org/10.1111/j.1753-4887.2008.00026.x.

Ford, P. B., & Dzewaltowski, D. A. (2010). Limited supermarket availability is not associated with obesity risk among participants in the Kansas WIC program. *Obesity, 18*(10), 1944–1951.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression*. Chichester: Wiley.

French, S. A., Harnack, L., & Jeffery, R. W. (2000). Fast food restaurant use among women in the Pound of Prevention study: dietary, behavioral and demographic correlates. *International Journal of Obesity and Related Metabolic Disorders, 24*(10), 1353–1359.

Glanz, K., Sallis, J. F., Saelens, B. E., & Frank, L. D. (2005). Healthy nutrition environments: concepts and measures. *American Journal of Health Promotion, 19*(5), 330–333.

Gordon-Larsen, P., Nelson, M. C., Page, P., & Popkin, B. M. (2006). Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics, 117*(2), 417–424.

Grigsby-Toussaint, D. S., Zenk, S. N., Odoms-Young, A., Ruggiero, L., & Moise, I. (2010). Availability of commonly consumed and culturally specific fruits and vegetables in African-American and Latino neighborhoods. *Journal of the American Dietetic Association, 110*(5), 746–752.

Grigsby-Toussaint, D., Chi, S.-H., & Fiese, B. (2011). Where they live, how they play: neighborhood greenness and outdoor physical activity among preschoolers. *International Journal of Health Geographics, 10*, 66.

Hendrickson, D., Smith, C., & Eikenberry, N. (2006). Fruit and vegetable access in four low-income food deserts communities in Minnesota. *Agriculture and Human Values, 23*(3), 371–383.

Holsten, J. E. (2009). Obesity and the community food environment: a systematic. *Public Health Nutrition, 12*(3), 397–405.

Jeffery, R. W., & French, S. A. (1998). Epidemic obesity in the United States: are fast foods and television viewing contributing? *American Journal of Public Health, 88*(2), 277–280.

Jilcott, S. B., Keyserling, T., Crawford, T., McGuirt, J. T., & Ammerman, A. S. (2011). Examining associations among obesity and per capita farmers' markets, grocery stores/supermarkets, and supercenters in US counties. *Journal of the American Dietetic Association, 111*(4), 567–572.

Jilcott, S. B., Moore, J. B., Shores, K. A., Imai, S., & McGranahan, D. A. (2011). Associations between natural amenities, physical activity, and body mass index in 100 North Carolina counties. *American Journal of Health Promotion, 26*(1), 52–55.

Khan, L. K., Sobush, K., Keener, D., Goodman, K., Lowry, A., Kakietek, J., et al. (2009). *Recommended community strategies and measurements to prevent obesity in the United States*. US Department of Health & Human Services, Centers for Disease Control and Prevention.

Kim, D., & Leigh, P. (2011). Are meals at full-service restaurants "normal" or "inferior"? *Population Health Management, 14*(6), 307–315.

Kim, D., Subramanian, S. V., Gortmaker, S. L., & Kawachi, I. (2006). US state-and county-level social capital in relation to obesity and physical inactivity: a multilevel, multivariable analysis. *Social Science & Medicine, 63*(4), 1045–1059.

Kirby, J. B., Liang, L., Chen, H. J., & Wang, Y. (2012). Race, place, and obesity: the complex relationships among community racial/ethnic composition, individual race/ethnicity, and obesity in the United States. *American Journal of Public Health, 102*(8), 1572–1578.

Larsen, K., & Gilliland, J. (2009). A farmers' market in a food desert: evaluating impacts on the price and availability of healthy food. *Health & Place, 15*(4), 1158–1162.

Larson, N. I., Story, M. T., & Nelson, M. C. (2009). Neighborhood environments: disparities in access to healthy foods in the U.S. *American Journal of Preventive Medicine, 36*(1), 74–81. http://dx.doi.org/10.1016/j.amepre.2008.09.025.

Macintyre, S., McKay, L., Cummins, S., & Burns, C. (2005). Out-of-home food outlets and area deprivation: case study in Glasgow, UK. *International Journal of Behavioral Nutrition and Physical Activity, 2*(1), 16.

Mehta, N. K., & Chang, V. W. (2008). Weight status and restaurant availability: a multilevel analysis. *American Journal of Preventive Medicine, 34*(2), 127–133.

Michimi, A., & Wimberly, M. C. (2010). Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States. *International Journal of Health Geographics, 9*(49), 94–112.

Morland, K. B., & Evenson, K. R. (2009). Obesity prevalence and the local food environment. *Health & Place, 15*(2), 491–495.

Morland, K., Wing, S., & Roux, A. D. (2002). The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *American Journal of Public Health, 92*(11), 1761–1767.

Morland, K., Diez Roux, A. V., & Wing, S. (2006). Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *American Journal of Preventive Medicine, 30*(4), 333–339.

Murray, C. J., Kulkarni, S. C., Michaud, C., Tomijima, N., Bulzacchelli, M. T., Iandiorio, T. J., et al. (2006). Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Medicine, 3*(9), e260.

Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2012). *Prevalence of obesity in the United States, 2009–2010*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Pearce, J., Blakely, T., Witten, K., & Bartie, P. (2007). Neighborhood deprivation and access to fast-food retailing. *American Journal of Preventive Medicine, 32*(5), 375–382.

Pearce, J., Hiscock, R., Blakely, T., & Witten, K. (2008). The contextual effects of neighbourhood access to supermarkets and convenience stores on individual fruit and vegetable consumption. *Journal of Epidemiology and Community Health, 62*(3), 198–201.

Powell, L. M., Slater, S., Mirtcheva, D., Bao, Y., & Chaloupka, F. J. (2007). Food store availability and neighborhood characteristics in the United States. *Preventive Medicine, 44*(3), 189–195.

Procter, K. L., Clarke, G. P., Ransley, J. K., & Cade, J. (2008). Micro-level analysis of childhood obesity, diet, physical activity, residential socioeconomic and social capital variables: where are the obesogenic environments in Leeds? *Area, 40*(3), 323–340.

Russell, S. E., & Heidkamp, P. C. (2011). 'Food desertification': the loss of a major supermarket in New Haven, Connecticut. *Applied Geography, 31*, 1197–1209.
Sallis, J. F., & Glanz, K. (2006). The role of built environments in physical activity, eating, and obesity in childhood. *Future Child, 16*(1), 89–108.
Sallis, J. F., & Glanz, K. (2009). Physical activity and food environments: solutions to the obesity epidemic. *Milbank Quarterly, 87*(1), 123–154. http://dx.doi.org/10.1111/j.1468-0009.2009.00550.x.
Simmons, D., McKenzie, A., Eaton, S., Cox, N., Khan, M. A., Shaw, J., et al. (2005). Choice and availability of takeaway and restaurant food is not related to the prevalence of adult obesity in rural communities in Australia. *International Journal of Obesity, 29*(6), 703–710.
Spence, J. C., Cutumisu, N., Edwards, J., Raine, K. D., & Smoyer-Tomic, K. (2009). Relation between local food environments and obesity among adults. *BMC Public Health, 9*(1), 192.
Summerbell, C. D., Waters, E., Edmunds, L. D., Kelly, S., Brown, T., & Campbell, K. J. (2005). Interventions for preventing obesity in children. *Cochrane Database of Systematic Reviews, 3*(3). http://dx.doi.org/10.1002/14651858.CD001871.pub2.
Tucker, P., & Gilliland, J. (2007). The effect of season and weather on physical activity: a systematic review. *Public Health, 121*(12), 909–922.
Wang, F., Wen, M., & Xu, Y. (2013). Population-adjusted street connectivity, urbanicity and risk of obesity in the U.S. *Applied Geography, 41*, 1–14.
Wen, T. H., Chen, D. R., & Tsai, M. J. (2010). Identifying geographical variations in poverty-obesity relationships: empirical evidence from Taiwan. *Geospatial Health, 4*(2), 257–265.
Wheeler, D. C., & Páez, A. (2010). Geographically weighted regression. In *Handbook of applied spatial analysis* (pp. 461–486). Berlin, Heidelberg: Springer.
Windle, M. J., Rose, G. A., Devillers, R., & Fortin, M. J. (2010). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES Journal of Marine Science, 67*(1), 145–154.
World Health Organization. (2003). *World health report 2003: A vision for global health shaping the future*. World Health Organization.
Zenk, S. N., Schulz, A. J., Hollis-Neely, T., Campbell, R. T., Holmes, N., Watkins, G., et al. (2005). Fruit and vegetable intake in African Americans income and store characteristics. *American Journal of Preventive Medicine, 29*(1), 1–9.