

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/268388172>

GIS-based analysis of obesity and the built environment in the US

Article in *Cartography and Geographic Information Science* · October 2014

DOI: 10.1080/15230406.2014.965748

CITATIONS

17

READS

705

2 authors:



Yanqing Xu

Wuhan University

32 PUBLICATIONS 798 CITATIONS

[SEE PROFILE](#)



Lei Wang

Louisiana State University

71 PUBLICATIONS 1,629 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



GIS public health [View project](#)



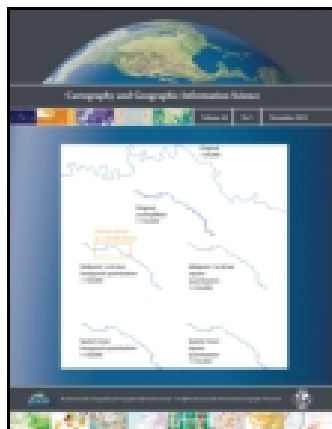
Remote Sensing of Ice [View project](#)

This article was downloaded by: [University Of Maryland]

On: 10 April 2015, At: 09:01

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Cartography and Geographic Information Science

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tcag20>

GIS-based analysis of obesity and the built environment in the US

Yanqing Xu^a & Lei Wang^a

^a Department of Geography and Anthropology, Louisiana State University, Baton Rouge, USA

Published online: 13 Oct 2014.



[Click for updates](#)

To cite this article: Yanqing Xu & Lei Wang (2015) GIS-based analysis of obesity and the built environment in the US, *Cartography and Geographic Information Science*, 42:1, 9-21, DOI: [10.1080/15230406.2014.965748](https://doi.org/10.1080/15230406.2014.965748)

To link to this article: <http://dx.doi.org/10.1080/15230406.2014.965748>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

GIS-based analysis of obesity and the built environment in the US

Yanqing Xu and Lei Wang*

Department of Geography and Anthropology, Louisiana State University, Baton Rouge, USA

(Received 28 January 2014; accepted 18 August 2014)

This paper examines the statewide relationship between built environment and obesity at the county scale by using the Geographically Weighted Regression (GWR) method. The independent variables include three built environment factors – street connectivity, walk score, fast-food/full-service restaurants ratio – and two sociodemographic variables, race heterogeneity and poverty rate. The urban influence is considered as a covariate in the analysis. Through the regression model we found that walk score and street connectivity are negatively related to obesity, poverty rate and metro are positively related, and the fast-food/full-service restaurants ratio is not significant. A regionalization method is used to group US counties to regions based on their GWR coefficients. Qualitative inferences of policies are made available to facilitate better understanding of the obesity problem associated with the built environment in these regions.

Keywords: obesity; built environment; geographically weighted regression; regionalization

1. Introduction

The current obesity epidemic has become a significant contributing factor to several leading causes of morbidity and mortality, including heart disease, stroke, diabetes and some cancers (Zhang, Lu, and Holt 2011). If the prevalence of obesity continues, more than 75% of the US population will be overweight within a few years, and more than 40% will be obese. The cost of treating obesity-related illnesses was estimated to be \$147 billion in 2008. The U.S. Department of Health and Human Services (HHS) awarded more than \$119 million to states and US territories to stimulate obesity research in the US and support public health efforts aimed at increasing physical activity and reducing obesity (e.g., Casagrande et al. 2011; Chi et al. 2013; Flegal et al. 2010; Lopez 2007; Maroko et al. 2009; Rose et al. 2009; Wang, Wen, and Xu. 2013; Yamada et al. 2012). Obesity has also become a worldwide research hotspot, with case studies in the United Kingdom (Edwards et al. 2010; Fraser et al. 2012), Taiwan (Wen, Chen, and Tsai 2010; Chen and Truong 2012), Greece (Chalkias et al. 2013), and the Netherlands (Dijkstra et al. 2013).

Previous research in public health, transportation, and urban planning highlighted the important relationship between environmental factors and people's physical activities at a variety of spatial scales (Feng et al. 2010; Li et al. 2008; Rutt and Coleman 2005; Yamada et al. 2012). For example, researchers have concluded that built-environment attributes, especially walkability, are consistently related to physical activity in general, particularly to 'active transportation' (Casagrande et al. 2011; Smith et al. 2008). Owen et al. (2004) suggested that

accessibility of recreation facilities and opportunities for activities and aesthetics were related to physical activities such as walking. Saelens, Sallis, and Frank (2003) found that ease of pedestrian access to nearby destinations was related to active transportation choices, particularly walking. In addition, the use of Geographic Information Systems (GISs) in public health studies has emerged recently, such as those to measure spatial distribution of accessibility to public resources (e.g., Giles-Corti and Donovan 2002).

Regression models were used to study the relationship between obesity and such environmental factors as fast-food density (Rose et al. 2009), land-use pattern (Heath et al. 2006; Duncan et al. 2010), poverty (Maroko et al. 2009), and walkability (Casagrande et al. 2011). However, it should be noted that in the public health studies in large areas, i.e., the entire US, regression models could be spatially non-stationary, meaning that the coefficients of the regression model are spatially variable (Brunsdon, Fotheringham, and Charlton 1998). In this case, local regression models such as the Geographically Weighted Regression (GWR, Fotheringham, Brunsdon, and Charlton 2002) could be used to avoid the 'ecological fallacy' problem (Holt et al. 1996) and explain the variability of obesity. In addition, we could gain better understanding of the phenomenon by interpreting the spatial pattern of the coefficients (Brunsdon, Fotheringham, and Charlton 1998). Maroko et al. (2009) examined the relationship between park accessibility and social economic status characteristics such as poverty, language barrier, population density and percent of minority ethnic groups in New York City by using the global and GWR models.

*Corresponding author. Email: leiwang@lsu.edu

They found only a weak relationship between park accessibility and physical activity variables and obesity rate. Their results suggest the existence of spatial non-stationarity in the regression models. GWR has been demonstrated to be an effective tool to analyze obesity in a geographical context (Chalkias et al. 2013; Chen and Truong 2012; Chi et al. 2013; Dijkstra et al. 2013; Edwards et al. 2010; Fraser et al. 2012; Wen, Chen, and Tsai 2010). However, only very few studies have addressed the obesity problem in the US continental area. Chi et al. (2013) used GWR and k-mean clustering analysis to examine the association of the food environment and some other socioeconomic variables with obesity in the US. Their work set a basis for a new analysis framework, namely using agglomerates to explain the spatial patterns of the regression coefficients. The built environment factors, however, were not their focus. In public health, built environment is the key to integrating the physical environments of the communities with health and wellness of the residents. The Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO) and other health organizations have recognized the importance of walkability for reducing obesity.

In this research, we focus on the built environment, which is measured by street connectivity and the walk score, in order to conduct a nation-wide analysis of the obesity problem in the US. Street connectivity is a proxy for walkability. Walk score is the measurement of physical activity, i.e., ‘access’ to nearby amenities on foot. We hope to contribute to the existing body of knowledge by answering the following research questions: ‘Are these built environment variables closely and consistently correlated to the obesity problem throughout the US?’ and ‘If not, can geographic analysis based on a local regression model help policy making?’ In this paper, we present procedures of variable selection, model interpretation, and a regionalized analysis of the built environment and obesity problem in the US at the county scale. Then, we provide some insights and discussion of the methodology.

2. Variables and data sources

2.1. Choice of variables

In the regression analysis, the dependent variable is the obesity rate, and the independent variables include four built environment variables – walkability, urbanicity (Vlahov and Galea 2002), street connectivity, and the ratio of fast-food/full-service restaurants (fast-food restaurant ratio) – and two sociodemographic variables: poverty rate and ethnic heterogeneity. These variables have been discussed in previous research. The counties in Alaska and Hawaii were excluded from the data analysis because they

are spatial outliers for the local regression model. Obesity rate is calculated from the obesity data provided by the Diabetes Interactive Atlases (<http://www.cdc.gov/diabetes/atlas/>) along with other data such as diagnosed diabetes and leisure-time physical inactivity at county level. Self-reported weights and heights were calculated to BMI: $BMI = \text{mass (kg)} / (\text{height (m)})^2$. Respondents were considered obese if their BMI values were over 30 (Flegal et al. 2010; CDC 2013). The obesity rate is the ratio of the obese population over the total population in each county. Three built environment variables were used in the analysis: street connectivity, walk score, and ratio of fast-food/full-service restaurants.

2.1.1. Street connectivity

Connectivity is defined by the number of intersections along a certain street network or in an area. Strictly, two-way connections are not intersections. Therefore, only those intersections with three connecting edges and the starting or ending nodes of the street network were included in the connectivity index calculation (Wang, Wen, and Xu. 2013). We used intersection density to measure street connectivity.

$$SC_i = \text{number of intersections} / \text{area} \quad (1)$$

Intersection density corresponds closely to block size – the greater the intersection density, the smaller the blocks. Small blocks make a neighborhood walkable. Street network density and intersection density are highly and positively correlated with each other (Aurbach 2010). Different areas have different patterns of intersection density; the differences will become larger when street network density decreases from urban to suburban and then rural areas. The intersection density measurement is based on a census tract level and then aggregated to county adjusted by population.

2.1.2. Walkability

Walkability is measured by the Walk Score (<http://www.walkscore.com/>) which calculates walking distance from a point of interest to nearby amenities. The Walk Score algorithm has been used in many public health studies (Brewster et al. 2009; Cortright 2009; Duncan et al. 2011; Jones 2010; Kirby et al. 2012; Kumar 2009; Li, Wen, and Henry 2014; Rauterkus, Thrall, and Hangen 2010; Zhu and Lee 2008). Brewster et al. (2009) showed that neighborhood Walk Score is correlated with the level of physical exercise, and hence could predict the levels of obesity, hypertension, and diabetes. Jones (2010) studied the Walk Score and its association with activity levels and she found that the Walk Score is correlated with the GIS-derived walkability index ($r = 0.63$ $p < 0.0001$). Duncan

et al. (2011) concluded that the Walk Score algorithm could produce valid measures of walkability, particularly at the 1600-meter buffer. They suggested that the Walk Score could be used across multiple scales.

The Walk Score algorithm requires user input to locate amenities such as restaurants, grocery stores, schools, parks, and movie theaters, which in this research are sourced from public domain map providers – Google, Education.com, Open Street Map, and Localeze. The algorithm calculates a weighted average of Euclidean distances from a point of interest to the amenities. The weights are determined by facility type priority and a certain distance decay function (Front Seat 2014). Walk Score ranges from 0 (the lowest) to 100 (the highest). Distances with a walk score of 0–49 are car dependent, while 0–24 means almost all errands require a car, and 25–49 means a few amenities are within walking distance. A walk score of 50–69 means some amenities are within walking distance, 70–89 means most errands can be accomplished on foot, and a score between 90 and 100 is walker's paradise where daily errands do not require a car (Front Seat 2014).

Front Seat (2014) provides an application programming interface (API) to query the Walk Score database through URL calls, eliminating the need for manually interacting with the website interface (Front Seat 2014). We developed a Python program which automatically requests walk scores from the server through the Walk Score API. In order to avoid bias caused by population concentration in a limited space within a large area, population-weighted centroids of census tracts are used instead of simple geographic centroids (Wang and Luo 2005). The population data we used are block-level data from the Census 2010. Walk Scores are normalized by population as a weight term and summed to the county level so as to match the scale of obesity data. The weight is the ratio between the population of a lower level unit (e.g., census tract) and a higher level unit (e.g., county). It is calculated as follows:

$$W_k = \sum_{i=1}^{n_k} Pop_i * W_i / Pop_k \quad (2)$$

where W_k is the walkability for the aggregated geographic unit (k), n_k is the number of lower level units to be aggregated into a higher level unit k , Pop_i is the population of the i th lower level unit, and Pop_k is the total population of the higher level unit k .

2.1.3. Urbanicity

According to Lopez and Patricia Hyness (2006), the relationships between built environments and obesity are different for inner city and suburban neighborhoods.

Inner city neighborhoods tend to have greater street connectivity, more sidewalks, and higher walk score; yet, they still have higher obesity rates because inner cities usually have less attractive and less safe environments for physical activity (Weir, Etelson, and Brand 2006). The previous findings suggested that it was necessary to examine the possible variation in the effect of built environments on obesity by an area's urbanicity. We used the 2006 National Center for Health Statistics (Ingram and Franco 2012) Urban-Rural Classification Scheme for the classification of urbanicity. According to Ingram and Franco (2012), there are six urban–rural urbanicity categories, including four levels of metros (large central, large fringe, medium, and small) and two levels of non-metros (micropolitan and non-core). In this research, urbanicity is a dichotomous variable to indicate a county as metro or non-metro.

2.1.4. Food environment

It was hypothesized that accessibility to fast-food restaurants is an indicator of extra calorie intake by the population because reliance on fast-food restaurants may lead to more meals or may increase consumption of high-fat meals, and, consequently, higher calorie intake (Lopez 2007). The food environment is represented by a fast-food/full-service restaurant ratio. The restaurant data is from the 5-year U.S. Economic Census (<http://www.census.gov/econ/>).

2.1.5. Sociodemographic variables

Analysis of built environment should also include its interaction with people's socioeconomic status and race/ethnic composition (Kirby et al. 2012). Zhu and Lee (2008) found that economic and ethnic disparities exist in the environmental support for walking. Similarly, Li, Wen, and Henry (2014) concluded that built environments and socioeconomic conditions were interrelated and that both played an important role in obesity prevention. Therefore, in addition to the built environment variables, we selected two sociodemographic variables: poverty rate (estimated percent of people of all ages in poverty) and ethnic heterogeneity derived from the Census 2010 data. Ethnic heterogeneity reflects the racial/ethnic composition which is defined as $1 - \sum p_i^2$, where p_i is the fraction of the population in a given group (Sampson and Groves 1989). The ethnic heterogeneity index ranges between 0 and 1. If the value is 0, it means that there is only one racial/ethnic group in the unit; while a value approaching 1 reflects maximum heterogeneity. These two sociodemographic variables were also suggested by public health experts (personal communications, 2013, Baton Rouge, Louisiana, USA).

2.2. Statistics and spatial pattern of input data

The mean obesity rate among adults is 27.39% (Table 1). The distribution of obesity patterns is shown in Figure 1. Higher overall obesity is clustered in the south-east central of the U.S. Holmes County, Mississippi, has the highest obesity rate at 42.10%. According to Figure 1, race heterogeneity is higher in the south of the US, whereas the Queens County in New York has the highest race heterogeneity. The average poverty rate is 15.44%, with much lower rates in the urbanized northeast and the West Coast of the US which have high street connectivity and high walk score. The fast-food/full-service restaurants ratio is the lowest in the Midwest. Among the 3109 counties, 1086 (about 35%) are metro while the rest are non-metro.

3. Methods

3.1. Ordinary least squares regression

The regression takes the age-adjusted rates of obesity among adults as the dependent variable; and the independent variables are race heterogeneity, poverty rate, ratio of fast-food/full-service restaurants, street connectivity, walk score, and urbanicity. The relationship was examined on a county-wide basis with cross-sectional analysis by using an Ordinary Least Squares (OLS) regression. The purpose was to test the significance of the variables and potential multicollinearity problems among the variables. The model is:

$$OB = \beta_0 + \beta_1 \text{Race Hetero} + \beta_2 \text{Poverty} + \beta_3 \text{Ratio} + \beta_4 \text{SC} + \beta_5 \text{WS} + \beta_6 \text{Metro} + \varepsilon \quad (3)$$

Table 1. Minimum, maximum, and mean values of dependent and independent variables and statistical difference for these values.

Variables	Min	Max	Mean	SD	No. observations
Dependent variable	—	—	—	—	—
Obesity Rate	12.40	42.10	27.39	3.62	3109
Independent variables	—	—	—	—	—
Race Hetero	0.01	0.78	0.27	0.20	3109
Poverty Rate	2.50	48.50	15.44	6.22	3109
Street Connectivity	0.62	336.08	30.42	37.55	3109
Walk Score	0.00	84.75	11.17	14.99	3109
Ratio of fast-food/full-service restaurant	0.00	18	2.06	1.38	3109
Urbanization	—	—	—	—	—
Metro	—	—	—	—	1086
Non-metro	—	—	—	—	2023

where *OB* stands for obesity; $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$, and β_6 are the regression coefficients; and ε is the random error in the two models.

Moran's *I* is used to test the spatial autocorrelation of the residuals from the regression model:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n W_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

where *n* is the total number of counties in the area, *i* and *j* represent different counties, x_i is the residual of *i*, and \bar{x} is the mean of residuals. W_{ij} is a measure of spatial proximity pairs of *i* and *j* (Wong and Lee 2005). The values of Moran's *I* would be between -1 and $+1$. Negative autocorrelation values mean nearby locations tended to have dissimilar values; positive autocorrelation values mean that similar values tended to occur in adjacent areas. Along with the index, Z-scores are usually reported for the statistical significance test. If *Z* is out of ± 1.96 , the null hypothesis of the randomness test is rejected at the 95% confidence level, which means the pattern is spatially auto-correlated. Otherwise, the spatial arrangement would be completely random (Lin and Wen 2011; Goodchild 1986).

The OLS regression result shows that the significant variables for obesity are race heterogeneity, poverty rate, street connectivity, and walk score (Table 2). The ratio of fast-food/full-service restaurants is not significant. The poverty variable has a positive coefficient (0.31), indicating that the relationship is positive, or, in other words, obesity prevalence is higher in areas with high poverty rate. In addition, the positive sign of the urbanization variable indicates that residents living in more urbanized areas are more likely to be at a higher risk of obesity. This confirms the previous findings that urban areas usually have more disadvantaged populations (i.e., low socioeconomic status or minorities) and less safe environments for people to engage in physical activity (Doyle et al. 2006; Weir, Etelson, and Brand 2006). The negative sign of the race heterogeneity variable suggests that it is more common for minorities to get obese. The coefficient for street connectivity and walk score is negative and significant, confirming that for these two variables, higher values are associated with lower obesity. The Variance Inflation Factor (VIF) values in Table 2 do not suggest any multicollinearity among the independent variables. The coefficient of determination R^2 for obesity is 0.30, a significant amount of variance unexplained. The residual maps (Figure 2) show some spatial autocorrelation in the residuals. The Moran's *I* of the residuals is 0.31 ($p < 0.01$). The spatial autocorrelation in the residuals

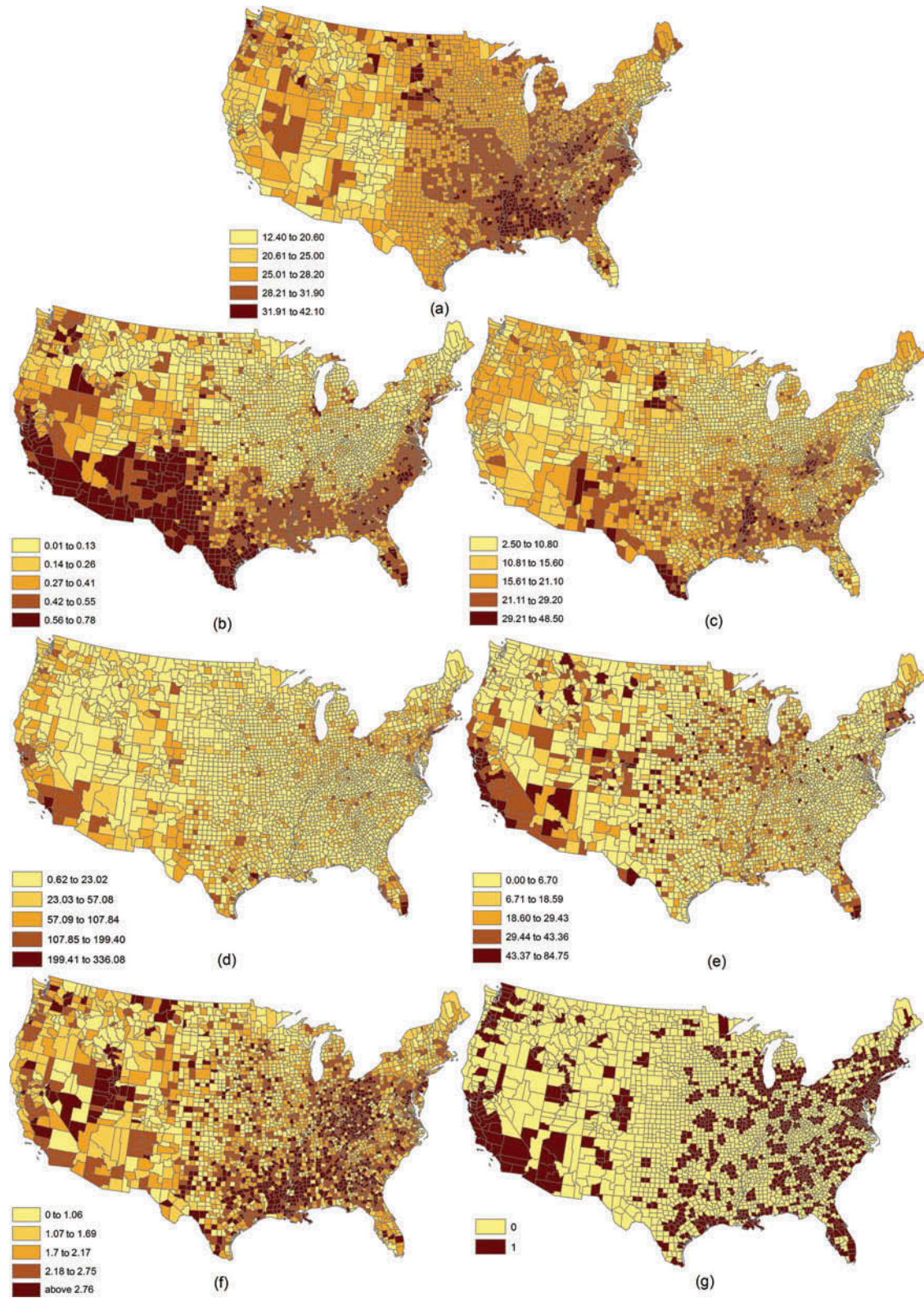


Figure 1. Distribution patterns of variables: (a) Obesity Rate; (b) Race Heterogeneity; (c) Poverty Rate; (d) Street Connectivity; (e) Walk Score; (f) Fast/Full Ratio; (g) Urbanicity.

Table 2. Results of ordinary least squares regression.

Variable	Coefficient	SE	p-Value	VIF
Intercept	23.46	0.17	—	—
Race Hetero	-1.69	-5.09	0.00	1.45
Poverty Rate	0.31	0.01	0.00	1.42
Street Connectivity	-0.02	0.002	0.00	1.54
Walk Score	-0.01	0.004	0.00	1.08
Ratio of fast-food/full-service	-0.00	0.00	0.28	1.01
Metro	1.13	0.00	0.00	1.42
Moran's I	0.31	—	—	—
Adjusted R^2	0.30	—	—	—
AIC	15,701	—	—	—

suggests there is some spatially correlated variability that is unexplained by the global OLS model. Therefore, instead of the global model, we used the local regression model which allows the regression coefficients to vary over the spatial domain.

3.2. Geographically weighted regression

GWR is a localized regression model that allows the parameters of a regression estimation to vary over the spatial domain (Lin and Wen 2011). The model can be expressed as:

$$\begin{aligned}
 OB_i = & \beta_{0i} + \beta_{1i}Race\ Hetero + \beta_{2i}Poverty \\
 & + \beta_{3i}Ratio + \beta_{4i}SC + \beta_{5i}WS + \beta_{6i}Metro \\
 & + \varepsilon_i
 \end{aligned} \quad (5)$$

where β_{ni} is the estimated regression coefficient at county i . The spatial variability of an estimated local regression coefficient was examined to determine whether the underlying process exhibited spatial heterogeneity (Fotheringham, Brunsdon, and Charlton 2000). The optimal solution of the regression equation in GWR is constrained by a geographically weighted matrix W_i (Fotheringham, Brunsdon, and Charlton 2002):

$$\beta_i = (X^T W_i X)^{-1} X W_i Y \quad (6)$$

where W_i is defined by the spatial relationships between neighboring points:

$$W_i = \begin{pmatrix} w_{i1} & 0 & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots & 0 \\ 0 & 0 & w_{i3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & w_{in} \end{pmatrix} \quad (7)$$

where w_{ij} is the strength of association between location i and location j (i and $j = 1 \dots n$) defined by their distance and a kernel function. The kernel function is a distance decay function usually defined as Gaussian with a user specified band width or spatially adaptive band widths. In this research, we used the calibration method that minimizes Akaike Information Criterion (AIC) of regression models to obtain the spatially adaptive band width values.

The analyses were done using ESRI's ArcGIS 10.1 and GWR 4 software packages. The GWR model shows significant improvement over the OLS model (see Table 3). It

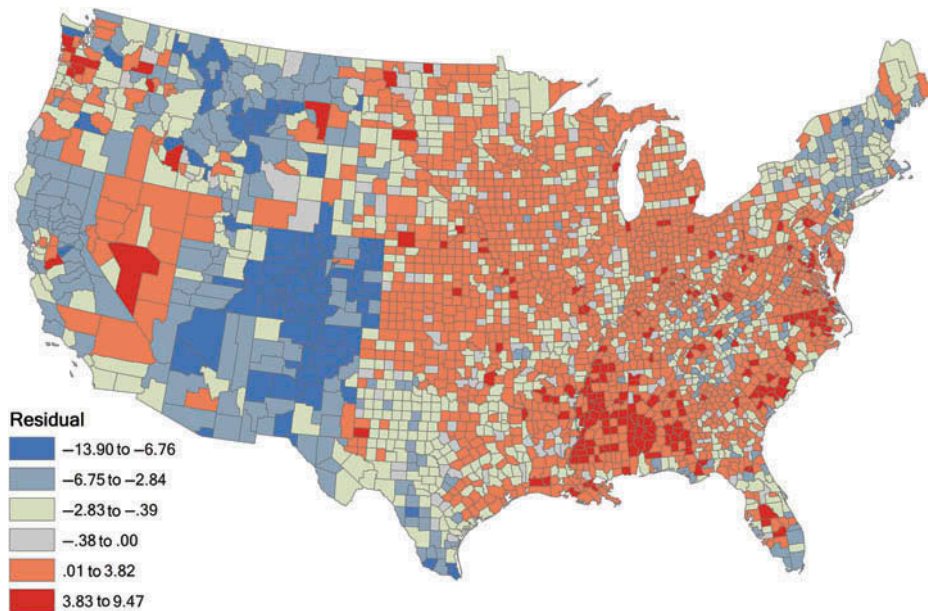


Figure 2. Residuals from the OLS regression. The map shows strong spatial autocorrelation.

Table 3. Results of geographically weighted regression.

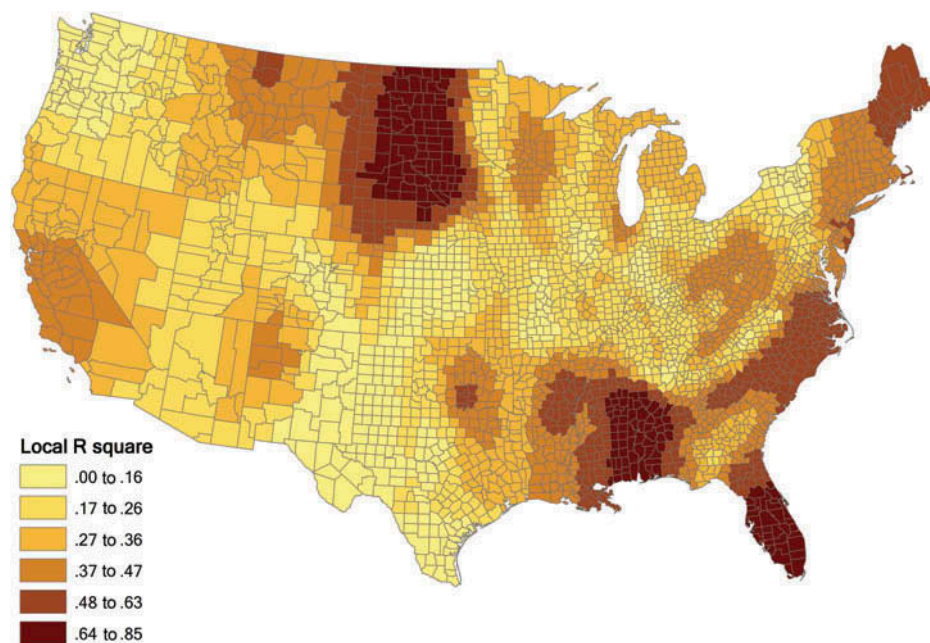
	Min	25% quartile	50% quartile	75% quartile	Max
Intercept	13.66	23.09	24.96	26.29	31.05
Race Hetero	-14.57	-1.30	1.34	3.75	11.43
Poverty Rate	-0.35	0.10	0.18	0.27	0.52
Street Connectivity	-0.05	-0.02	-0.01	-0.004	0.04
Walk Score	-0.13	-0.02	-0.01	0.004	0.07
Ratio of fast-food/ full-service	-0.30	0.00	0.00	0.15	1.75
Metro	-11.55	-0.33	0.09	0.55	2.68
Moran's I	0.01	—	—	—	—
Adjusted R^2	0.72	—	—	—	—
AIC	13,215	—	—	—	—

returns an overall R^2 of 0.72, much better than the OLS model ($R^2 = 0.30$). Figure 3 shows the spatial pattern of the locally weighted R^2 . Furthermore, the residuals of the GWR model only have a slight level of spatial autocorrelation (Moran's $I = 0.01$, $p < 0.001$).

The spatial distribution of R^2 is not even over the study area (Figure 3). Some counties have a high R^2 (of up to 0.85) and some are very low. Generally, the counties in most areas of the north central states and the states of Mississippi, Alabama, and Florida have better regression results than others. Figures 4 and 5 illustrate the intercept and coefficients of race heterogeneity, poverty rate, street connectivity, walk score, the ratio of fast-food/full-service restaurants and urban-rural classification, and the t values

representing the fitting level for each specific variable in GWR. The cartographic method by Mennis (2006) was used to map coefficient values and their significance simultaneously.

Figure 4 shows the spatial patterns of the GWR model coefficients. The intercepts are lower in the mountain areas and the northeast counties, indicating generally lower obesity in those areas (Figure 4a). Figure 4b shows significantly positive coefficients of racial disparity along the West Coast and some inland areas. Poverty rate is strongly associated with obesity in most counties, except for some counties in Colorado (Figure 4c). Further investigation of areas with negative coefficients might be interesting, but this is beyond the scope of the research reported here. The consistency in the poverty coefficients leads us to conclude that socioeconomic disadvantage/poverty is the decisive factor behind the obesity problem in the US. The relationship between street connectivity and obesity is negative in most counties, with outliers of slightly positive values in the mountainous areas (Figures 4d and e). The outliers are mainly in low population density areas. This suggests that in these areas, an increase in street connectivity or walkability may not reduce obesity. The ratio of fast-food/full-service restaurants is strongly and positively correlated with obesity rate in the northeast and in some counties in Washington State (Figure 4f). However, this variable and the urbanicity variable (Figure 4g) do not have much effect on obesity in most of the country. Therefore, they are not included in the discussion of spatial clusters in the following section.

Figure 3. Coefficient of determination (R^2) map from the GWR model.

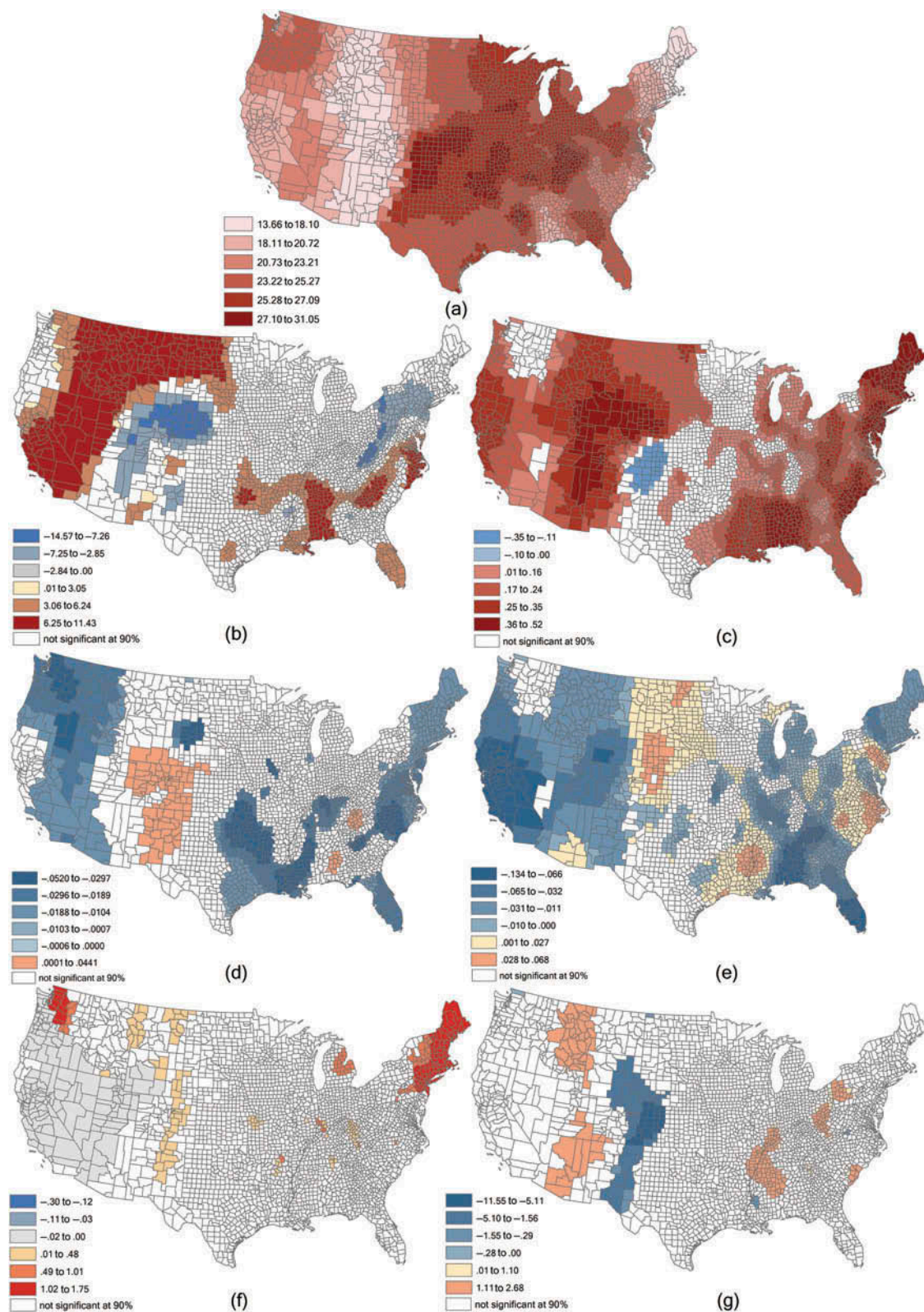


Figure 4. GWR coefficients: (a) Intercept; (b) Race Heterogeneity; (c) Poverty Rate; (d) Street Connectivity; (e) Walk Score; (f) Fast/Full Ratio; (g) Urbanicity.

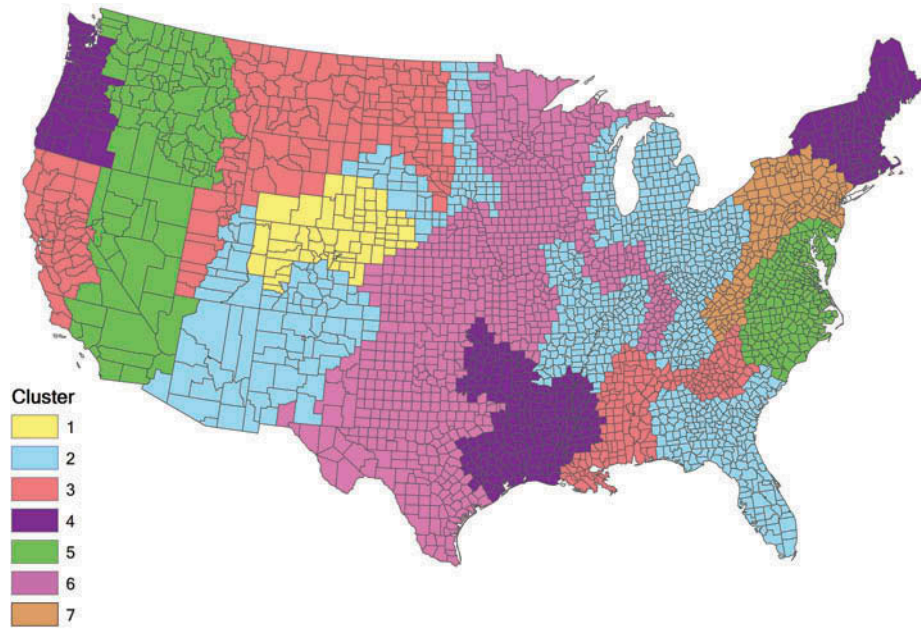


Figure 5. Geographical regions created based on the GWR coefficients.

3.3. Regionalization

The coefficients maps have strong spatial correlation due to the use of local samples in the GWR model. The spatial pattern of coefficients and their t values reflect some underlying physical or social-cultural mechanisms. For example, in Figure 4b we can observe clusters of strong positive coefficient values of racial heterogeneity in the northwest and the West Coast, covering most of California, Nevada, Idaho, Montana, North Dakota, and part of South Dakota. If any future research is to be conducted on the obesity problem of different racial groups, these areas would be interesting. Regionalization of coefficients and their significance level can help to define geographical regions with high and low significant and non-significant coefficients, respectively. The delineation would better reveal the regional heterogeneity of the obesity problem in the US. Coefficients of the three variables – race heterogeneity, poverty rate, and street connectivity – were used in the regionalization analysis. The coefficients of other variables were omitted because they were not significant (Figure 4). The procedure used to delineate the regions is as follows:

- (1) For each variable, group the counties in three groups, based on the sign of their coefficient and significance at 95%: 1 – significantly positive, 2 – significantly negative, and 3 – not significant at 95%.
- (2) Use the ‘dissolve’ algorithm in the GIS spatial analysis tool to eliminate the boundaries of counties that are in the same class and are spatially

adjacent to each other. This will generate regions representing homogeneous area of each variable, meaning, the coefficient values in the region will be all positive or negative or non-significant.

- (3) Generalize the maps by eliminating smaller regions. The goal is to end up with three or four large regions for each variable.
- (4) Intersect the generalized maps of the three variables to create the final regionalization map.

The regionalization map in Figure 5 shows US counties grouped into 16 regions with seven classes. The classification is based on the signs of the coefficients of the selected three variables – poverty, racial heterogeneity, and street connectivity (see Table 4). Class 1 includes New York, Connecticut, Pennsylvania, and Maryland. The two significant variables in this region are poverty, which has a

Table 4. Classification of the regions based on coefficient values of the three variables.

Classes	Race Heterogeneity	Poverty Rate	Street Connectivity
1	–	+	0
2	0	+	0
3	+	+	0
4	0	+	–
5	+	+	–
6	–	+	+
7	0	0	0

Note: ‘+’ means positive and significant, ‘–’ means negative and significant, ‘0’ means not significant.

positive (+) coefficient, and race heterogeneity with a negative (−) coefficient. Class 2 includes areal clusters scattered throughout the US, including the eastern part of the Gulf Coast; the southwest mountain areas of Utah, Arizona, and part of Colorado and New Mexico; the Great Lakes area and its basin, and the area around Memphis, Tennessee. In these areas, the only significant variable is poverty rate. Class 3 includes major parts of California, Montana, Wyoming, North Dakota, western Utah, and a small part of the south US. None of the three variables is significant in these areas, suggesting that the regression model could not explain much of the variability in their obesity problem. Class 4 is located in the states of Washington and New York, as well as the border area between Texas and Louisiana. The two significant variables for this class are poverty (+) and street connectivity (−), which suggests that policies that help the poor or promote walkable environment would help in reducing obesity in these parts of the US. Class 5 includes the coast of Virginia and North Carolina (the so-called the Dominion of Atlantic), Nevada, Idaho, Oregon, Washington, and eastern California. All three variables are significant in this class: race (+), poverty (+), street connectivity (−). Policies dealing with these variables would all be effective. Class 6 includes the border area of Utah, Colorado, Nebraska, and Wyoming. Street connectivity in this class is positively correlated with obesity while race heterogeneity is negatively correlated. Both negate the hypothesis we set out to test with the regression model. A possible explanation is that the population density is generally low in this region. Class 7 is the central zone of the US, extending from Texas all the way to the US border with Canada. None of the three variables was significant for this class. In other words, regression cannot explain the variability of the obesity problem in the areas it covers. More variables would be needed to study their obesity problem.

4. Discussion

This first ever research with Walk Score and street connectivity confirms the previous findings about the role of walkability in reducing obesity at the community level (Frank, Andresen, and Schmid 2004). The global OLS model and the local regression model have both showed that Walk Score is a significant factor in explaining the variability of obesity in the US. The aggregated, county-level Walk Score was significant in modeling obesity by regression. In other words, we have demonstrated a feasible way of measuring walkability at the county level to be used in public health research. The consistency between the global and local regression models suggests the general applicability of this approach to measuring walkability. Nonetheless, some outliers in the local regression model were observed (Figure 4e). For example, some

counties in the center of the US have positive coefficients (the higher the walk score, the higher the obesity) or walk score does not matter (non-significant coefficient). Such outliers may be caused by an inconsistency of the data used in the Walk Score algorithm. The other possible reason might be the use of centroid of census tracts to approximate the population centers in the algorithm. Further revision of the walk score algorithm might help improve the use of the Walk Score algorithm in public health studies.

While the global OLS regression model can measure the relationship between obesity rate and, respectively, race heterogeneity, poverty rate, street connectivity, walk score, ratio of fast-food/full-service, and urban–rural classification, the local regression model (GWR) has its strength in finding geographic heterogeneity among counties by clustering their coefficients. In fact, the spatial patterns of coefficients are more useful than the regression itself to a geographical analysis. General statistic methods used in Human Geography have been criticized for generalizing human objects and neglecting the spatial structure of society. The use of the localized regression model compensates for the weakness of statistical models that neglect spatial heterogeneity. In fact, the GWR is more powerful in explaining the variability in obesity than the OLS model when race heterogeneity, poverty rate, street connectivity, walk score, ratio of fast-food/full-service, and urban–rural classification are used. The spatial pattern of their coefficients is more interesting to Human Geographers than the regression itself. In each of the coefficient maps (Figure 4), one can visually identify clusters of counties that are significantly different from other areas. It is evident that public health policies cannot depend on a global model. For example, the global model identifies poverty rate as a factor significantly contributing to obesity; but from the local model, we are able to identify areas with negative or non-significant coefficients, which means that the global model's conclusion regarding poverty rate does not apply across the board (Figure 4c). Public policies should be flexible and take into account the unique characteristics of each region.

With regard to the methodologies used in this research, note that our regionalization analysis partitions the entire study area into multiple subareas with unique coefficient values and significance levels. The outcome is similar to the Regional Geography paradigm which focuses on the unique combination of characteristics in an area (Peet 1998). Despite the similarity in their form and descriptive nature, however, our approach is fundamentally different from that of traditional Regional Geography, which has been criticized for its lack of scientific justification, and because region identification is subjective and unpredictable. In contrast, our approach is based on quantitative information from the regression models – i.e., region divisions are empirical. We used the sign of the

coefficients and significance level (95%) to define regions. The regions thus created are predictable and scientifically justifiable. Essentially, it is GIS that makes such an approach possible. Hence, one of the objectives of this paper is to illustrate and promote the use of GIS spatial analysis and statistics in public health studies.

The classes defined in Table 4 for the regions shown in Figure 5 could improve policy-making aimed at curbing obesity. They will help to answer such questions as: ‘What measures could be taken to reduce the obesity risk in an area?’ and ‘What are the areas that a policy measure could reduce the obesity rate?’ To answer the first question, find the class of the target area in the map of Figure 5 and then find the variables in Table 4 that are significant for that class. To answer the second question, find in Table 4 the classes for which the specified policy measure (variable) is significant and then refer to Figure 5 to find the relevant counties in those classes. In this way, policy-makers with no expertise in GIS and quantitative methods would be able to use our research report.

Although the ultimate goal of public health research is to thoroughly understand the interaction between obesity and physical and socioeconomic conditions, our research focused on only a few built environment and social status variables. In future research, we would consider land use mix (Talen and Anselin 1998) as one of the built environment variables. Furthermore, age, gender, income, marital status, education level, and employ status were not taken into account because of the scope of this research. To improve the understanding of obesity and built environment associations, it is possible to apply a space-time framework to analyze different years instead of one epidemic year. By doing so we could predict the spatio-temporal pattern of the obesity-built environment relationship. Moreover, we did not include weather as a variable because it was not a common practice in previous research; but as suggested by the GWR model analysis of the counties clustering pattern, weather might be a factor in the spatial distribution of obesity. Furthermore, linear regression cannot handle non-linear relationships. Certain transformations will be necessary if any non-linearity in the variables is identified; although we did not observe any non-linearity in this research, care should be taken if more variables are included in future work.

5. Summary

In this paper, we analyze the obesity problem and related built-environment factors for counties in the contiguous US using regression models with a GIS. We used a global model to analyze the overall relationship and GWR model to identify regional differences. We found that in

most counties, poverty rate, street connectivity, and walk score played a role in obesity rates; however, the contribution of these variables in the counties varies spatially. These findings were extended to region-based qualitative inferences that could help policy-making. GIS made location-sensitive policy-making possible by local regression and regionalization analyses. Such data analysis methodology could enhance our understanding of the obesity problem in the US. We expect similar approaches to be applied to other public health problems in the US or other countries.

References

- Aurbach, L. 2010. “The Power of Intersection Density.” Accessed May 27, 2010. <http://pedshed.net>
- Brewster, M., D. Hurtado, S. Olson, and J. Yen. 2009. “Walkscore.com: A New Methodology to Explore Associations between Neighborhood Resources, Race, and Health.” APHA Poster presentation. Accessed September 27, 2014. https://apha.confex.com/recording/apha/137am/pdf/free/4db77adf5d9ff0d3caf5cafe28f496/paper205082_1.pdf
- Brunsdon, C., S. Fotheringham, and M. Charlton. 1998. “Geographically Weighted Regression – Modelling Spatial Non-Stationarity.” *Journal of the Royal Statistical Society: Series D (The Statistician)* 47: 431–443. doi:10.1111/1467-9884.00145.
- Casagrande, S. S., J. Gittelsohn, A. B. Zonderman, M. K. Evans, and T. L. Gary-Webb. 2011. “Association of Walkability with Obesity in Baltimore City, Maryland.” *American Journal of Public Health* 101 (Suppl S1): SS318–24. doi:10.2105/AJPH.2009.187492.
- CDC. 2013. “Diabetes Interactive Atlases.” Atlanta, GA. <http://www.cdc.gov/diabetes/atlas/>
- Chalkias, C., A. G. Papadopoulos, K. Kalogeropoulos, K. Tambalis, G. Psarra, and L. Sidossis. 2013. “Geographical Heterogeneity of the Relationship between Childhood Obesity and Socio-Environmental Status: Empirical Evidence from Athens, Greece.” *Applied Geography* 37: 34–43. doi:10.1016/j.apgeog.2012.10.007.
- Chen, D.-R., and K. Truong. 2012. “Using Multilevel Modeling and Geographically Weighted Regression to Identify Spatial Variations in the Relationship between Place-Level Disadvantages and Obesity in Taiwan.” *Applied Geography* 32: 737–745. doi:10.1016/j.apgeog.2011.07.018.
- Chi, S.-H., D. S. Grigsby, N. Bradford, and J. Choi. 2013. “Can Geographically Weighted Regression Improve our Contextual Understanding of Obesity in the US? Findings from the USDA Food Atlas.” *Applied Geography* 44: 134–142. doi:10.1016/j.apgeog.2013.07.017.
- Cortright, J. 2009. “Walking the Walk: How Walkability Raises Home Values in U.S. Cities.” Report prepared for CEO of The Cities 2009. Accessed September 27, 2014. http://blog.walkscore.com/wp-content/uploads/2009/08/WalkingTheWalk_CEOsforCities.pdf
- Dijkstra, A., F. Janssen, M. De Bakker, J. Bos, L. J. Rene Lub, G. Van Wissen, and E. Hak. 2013. “Using Spatial Analysis to Predict Health Care Use at the Local Level: A Case Study of Type 2 Diabetes Medication Use and its Association with Demographic Change and Socioeconomic Status.” *PLoS ONE* 8 (8). doi:10.1371/journal.pone.0072730.

- Doyle, S., A. Kelly-Schwartz, M. Schlossberg, and J. Stockard. 2006. "Active Community Environments and Health: The Relationship of Walkable and Safe Communities to Individual Health." *Journal of the American Planning Association* 72 (1): 19–31. doi:10.1080/019443606008976721.
- Duncan, D. T., J. Aldstadt, J. Whalen, S. J. Melly, and S. L. Gortmaker. 2011. "Validation of Walk Score® for Estimating Neighborhood Walkability: An Analysis of Four US Metropolitan Areas." *International Journal of Environmental Research and Public Health* 8: 4160–4179. doi:10.3390/ijerph8114160.
- Duncan, M. J., E. Winkler, T. Sugiyama, E. Cerin, L. Dutoit, E. Leslie, and N. Owen. 2010. "Relationships of Land Use Mix with Walking for Transport: Do Land Uses and Geographical Scale Matter?" *Journal of Urban Health* 87 (5): 782–795. doi:10.1007/s11524-010-9488-7.
- Edwards, K. L., G. P. Clarke, J. K. Ransley, and J. Cade. 2010. "The Neighbourhood Matters: Studying Exposures Relevant to Childhood Obesity and the Policy Implications in Leeds, UK." *Journal of Epidemiology & Community Health* 64: 194–201. doi:10.1136/jech.2009.088906.
- Feng, J., T. A. Glass, F. C. Curriero, W. F. Stewart, and B. S. Schwartz. 2010. "The Built Environment and Obesity: A Systematic Review of the Epidemiologic Evidence." *Health & Place* 16 (2): 175–190. doi:10.1016/j.healthplace.2009.09.008.
- Flegal, K. M., M. D. Carroll, C. L. Ogden, and L. R. Curtin. 2010. "Prevalence and Trends in Obesity Among US Adults, 1999–2008." *The Journal of the American Medical Association* 303 (3): 235–241. doi:10.1001/jama.2009.2014.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*. Newbury Park, CA: Sage.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. New York, NY: Wiley.
- Frank, L. D., M. A. Andresen, and T. L. Schmid. 2004. "Obesity Relationships with Community Design, Physical Activity, and Time Spent in Cars." *American Journal of Preventive Medicine* 27 (2): 87–96. doi:10.1016/j.amepre.2004.04.011.
- Fraser, L. K., G. P. Clarke, J. E. Cade, and K. L. Edwards. 2012. "Fast Food and Obesity: A Spatial Analysis in a Large United Kingdom Population of Children Aged 13–15." *American Journal of Preventive Medicine* 42 (5): e77–e85. doi:10.1016/j.amepre.2012.02.007.
- Front Seat. 2014. "Walk Score API." Accessed April 20. <http://www.walkscore.com/professional/api.php>.
- Giles-Corti, B., and R. J. Donovan. 2002. "Socioeconomic Status Differences in Recreational Physical Activity Levels and Real and Perceived Access to a Supportive Physical Environment." *Preventive Medicine* 35: 601–611. doi:10.1006/pmed.2002.1115.
- Goodchild, M. F. 1986. *Spatial Autocorrelation*. Vol. 47. Norwich: Geo Books.
- Heath, G. W., R. C. Brownson, J. Kruger, R. Miles, K. E. Powell, and L. T. Ramsey. 2006. "The Effectiveness of Urban Design and Land Use and Transport Policies and Practices to Increase Physical Activity: A Systematic Review." *Journal of Physical Activity & Health* 3 (1): S55–S76.
- Holt, D., D. G. Steel, M. Tranmer, and N. Wrigley. 1996. "Aggregation and Ecological Effects in Geographically Based Data." *Geographical Analysis* 28 (3): 244–261. doi:10.1111/j.1538-4632.1996.tb00933.x.
- Ingram, D. D., and S. J. Franco. 2012. "NCHS Urban–Rural Classification Scheme for Counties." *National Center for Health Statistics. Vital Health Statistics* 2 (154): 1–65.
- Jones, L. I. 2010. "Investigation Neighborhood Walkability and its Association with Physical Activity Levels and Body Composition of a Sample of Maryland Adolescent Girls." Master Thesis, University of Maryland.
- Kirby, J. B., L. Liang, H.-J. Chen, and Y. Wang. 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. doi:10.2105/AJPH.2011.300452.
- Kumar, R. 2009. "Walkability of Neighborhoods: A Critical Analysis of Zoning Codes." Master Thesis, University of Cincinnati.
- Li, F., P. A. Harmer, B. J. Cardinal, M. Bosworth, A. Acock, D. Johnson-Shelton, and J. M. Moore. 2008. "Built Environment, Adiposity, and Physical Activity in Adults Aged 50–75." *American Journal of Preventive Medicine* 35 (1): 38–46. doi:10.1016/j.amepre.2008.03.021.
- Li, K., M. Wen, and K. A. Henry. 2014. "Residential Racial Composition and Black-White Obesity Risks: Differential Effects of Neighborhood Social and Built Environment." *International Journal of Environmental Research and Public Health* 11: 626–642. doi:10.3390/ijerph110100626.
- Lin, C.-H., and T.-H. Wen. 2011. "Using Geographically Weighted Regression (GWR) to Explore Spatial Varying Relationships of Immature Mosquitoes and Human Densities with the Incidence of Dengue." *International Journal of Environmental Research and Public Health* 8: 2798–2815. doi:10.3390/ijerph8072798.
- Lopez, R. P., and H. Patricia Hyness. 2006. "Obesity, Physical Activity, and the Urban Environment: Public Health Research Needs." *Environmental Health: A Global Access Science Source* 5 (1). doi:10.1186/1476-069X-5-25.
- Lopez, R. P. 2007. "Neighborhood Risk Factors for Obesity." *Risk Factors and Chronic Disease* 15 (8): 2111–2119.
- Maroko, A. R., J. A. Maantay, N. L. Sohler, K. L. Grady, and P. S. Arno. 2009. "The Complexities of Measuring Access to Parks and Physical Activity Sites in New York City: A Quantitative and Qualitative Approach." *International Journal of Health Geographics* 8 (1). doi:10.1186/1476-072X-8-34.
- Mennis, J. 2006. "Mapping the Results of Geographically Weighted Regression." *The Cartographic Journal* 43 (2): 171–179. doi:10.1179/000870406X114658.
- Owen, N., N. Humpe, E. Leslie, A. Bauman, and J. F. Sallis. 2004. "Understanding Environmental Influences on Walking: Review and Research Agenda." *American Journal of Preventive Medicine* 27 (1). doi:10.1016/j.amepre.2004.03.006.
- Peet, R. 1998. *Modern Geographical Thought*. Oxford: Blackwell.
- Rauterkus, S. Y., G. I. Thrall, and E. Hangen. 2010. "Location Efficiency and Mortgage Default." *The Journal of Sustainable Real Estate* 2 (1): 117–141.
- Rose, D., P. L. Hutchinson, J. N. Bodor, C. M. Swalm, T. A. Farley, D. A. Cohen, and J. C. Rice. 2009. "Neighborhood Food Environments and Body Mass Index: The

- Importance of In-Store Contents.” *American Journal of Preventive Medicine* 37 (3): 214–219. doi:[10.1016/j.amepre.2009.04.024](https://doi.org/10.1016/j.amepre.2009.04.024).
- Rutt, C. D., and K. J. Coleman. 2005. “Examining the Relationships among Built Environment, Physical Activity, and Body Mass Index in El Paso, TX.” *Preventive Medicine* 40 (6): 831–841. doi:[10.1016/j.ypmed.2004.09.035](https://doi.org/10.1016/j.ypmed.2004.09.035).
- Saelens, B. E., J. F. Sallis, and L. D. Frank. 2003. “Environmental Correlates of Walking and Cycling: Findings from the Transportation, Urban Design, and Planning Literatures.” *Annals of Behavioral Medicine* 25 (2): 80–91. doi:[10.1207/S15324796ABM2502_03](https://doi.org/10.1207/S15324796ABM2502_03).
- Sampson, R. J., and W. B. Groves. 1989. “Community Structure and Crime: Testing Social-Disorganization Theory.” *American Journal of Sociology* 94 (4): 774–802. doi:[10.1086/229068](https://doi.org/10.1086/229068).
- Smith, K. R., B. B. Brown, I. Yamada, L. Kowaleski-Jones, C. D. Zick, and J. X. Fan. 2008. “Walkability and Body Mass Index: Density, Design, and New Diversity Measures.” *American Journal of Preventive Medicine* 35 (3): 237–244. doi:[10.1016/j.amepre.2008.05.028](https://doi.org/10.1016/j.amepre.2008.05.028).
- Talen, E., and L. Anselin. 1998. “Assessing Spatial Equity: An Evaluation of Measures of Accessibility to Public Playgrounds.” *Environment and Planning A* 30 (4): 595–613. doi:[10.1068/a300595](https://doi.org/10.1068/a300595).
- Vlahov, D., and S. Galea. 2002. “Urbanization, Urbanicity, and Health.” *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 79: S1S–S12. doi:[10.1093/jurban/79.suppl_1.S1](https://doi.org/10.1093/jurban/79.suppl_1.S1).
- Wang, F., and W. Luo. 2005. “Assessing Spatial and Nonspatial Factors for Healthcare Access: Towards an Integrated Approach to Defining Health Professional Shortage Areas.” *Health & Place* 11: 131–146. doi:[10.1016/j.healthplace.2004.02.003](https://doi.org/10.1016/j.healthplace.2004.02.003).
- Wang, F., M. Wen, and Y. Xu. 2013. “Population-Adjusted Street Connectivity, Urbanicity and Risk of Obesity in the U.S.” *Applied Geography* 41: 1–14. doi:[10.1016/j.apgeog.2013.03.006](https://doi.org/10.1016/j.apgeog.2013.03.006).
- Weir, L. A., D. Etelson, and D. A. Brand. 2006. “Parents’ Perceptions of Neighborhood Safety and Children’s Physical Activity.” *Preventive Medicine* 43 (3): 212–217. doi:[10.1016/j.ypmed.2006.03.024](https://doi.org/10.1016/j.ypmed.2006.03.024).
- Wen, T.-H., D.-R. Chen, and M.-J. Tsai. 2010. “Identifying Geographical Variations in Poverty-Obesity Relationships: Empirical Evidence from Taiwan.” *Geospatial Health* 4 (2): 257–265.
- Wong, D. W.-S., and J. Lee. 2005. *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Hoboken, NJ: John Wiley & Sons.
- Yamada, I., B. B. Brown, K. R. Smith, C. D. Zick, L. Kowaleski-Jones, and J. X. Fan. 2012. “Mixed Land Use and Obesity: An Empirical Comparison of Alternative Land Use Measures and Geographic Scales.” *The Professional Geographer* 64 (2): 157–177. doi:[10.1080/00330124.2011.583592](https://doi.org/10.1080/00330124.2011.583592).
- Zhang, X., H. Lu, and J. B. Holt. 2011. “Modeling Spatial Accessibility to Parks: A National Study.” *International Journal of Health Geographics* 10 (1): 31–14. doi:[10.1186/1476-072X-10-31](https://doi.org/10.1186/1476-072X-10-31).
- Zhu, X., and C. Lee. 2008. “Walkability and Safety around Elementary Schools Economic and Ethnic Disparities.” *American Journal of Preventive Medicine* 34 (4): 282–290. doi:[10.1016/j.amepre.2008.01.024](https://doi.org/10.1016/j.amepre.2008.01.024).