A Report from the University of Vermont Transportation Research Center

Exploration of Walking Behavior in Vermont Using Spatial Regression

TRC Report 14-011

June 15

Liu, Lee

Liu, Lee

EXPLORATION OF WALKING BEHAVIOR IN VERMONT USING SPATIAL REGRESSION

ABSTRACT

This report focuses on the relationship between walking and its contributing factors by applying spatial regression methods. Using the Vermont data from the New England Transportation Survey (NETS), walking variables as well as 170 independent variables are derived including some through spatial analysis with Geographic Information Systems (GIS). Among those independent variables, people's lifestyle and perception of the built environment variables are included. A linear regression model is first established to serve as a base model for comparisons with spatial regression models. The results reveal that people's lifestyle and perception of the built environment are significant variables explaining Vermonters' walking behavior. Methodologically, the results reveal that no spatial effect is found and that there are no significant differences between the linear and spatial regression models. Therefore, the study concludes that it may be appropriate to apply traditional non-spatial statistical tools to analyze the relationship between walking and its contributing factors. However, the study suggests that researchers examine whether spatial effect exists in these inherently spatial behaviors before using only traditional statistics. This caution is particularly relevant as methods to estimate spatial models become more commonplace and easily available. The two spatial methods used in this report both reveal small but different challenges.

INTRODUCTION

The important role that walking plays in public health has been well established. Walking is one of the most common forms of physical activity (Eyler et al., 2003), in part because it has fewer physical, social and psychological barriers than most traditional forms of exercise (Allender et al., 2006), and it provides a variety of health and societal benefits. Walking has been associated with decreased body mass index (Kahn et al., 1997), reduced coronary incidents among women (Manson et al. 1999), reduced cardiovascular events among diabetic women (Hu et al. 2001) and reduced health costs (Stokes et al. 2007). Recently Murphy et al. (2012) found that time walking at a brisk pace for personal transport had a strong positive association with being lean.

Given the benefits associated with walking, recent research has focused on the factors that contribute to increased walking rates. A study by Cao et al. (2007) examined four urban and four suburban neighborhoods and found that policies designed to decrease the distance between residences and destinations with alternative transportation led to decreases in driving and increases in walking. Boarnet et al. (2008) also investigated the relationship between walking and urban design using cost-benefit analysis in urban Portland, Oregon. While these and other studies provide valuable insights into the factors that contribute to increased prevalence of walking, several important research gaps remain, notably, the use of spatially explicit modeling techniques and the consideration of a broader set of predictor variables across a wider spectrum of urban to rural communities. Methodologically, broader sets of predictor variables and spatially explicit modeling need to be incorporated into walking related research. People's lifestyles, perceptions of the built environment, and attitudes toward walking, which been found to be crucial to understanding behaviors like walking (Livi et al., 2004), have not been sufficiently studied. The existing literature also relies mostly on traditional linear regression methods which may not be appropriate if spatial effects are at play. In this case, auto-correlated residuals would violate the independence assumption for errors in the classical multiple linear regression model, which could lead to inaccurate degrees of freedom and inflated t-statistics that increase the chance of type 1 error (Fox, 1997, Greene 2000).

In addition, most walking studies have taken place in urban or suburban areas. Walking, however, may have a particularly important role to play in rural area as rural adults tend to have higher levels of obesity and to be less activity in their leisure time than urban and suburban residents (Eberhardt et al., 2001; Parks et al., 2003; Patterson et al., 2004). Due to differences in the built environment characteristics in urban and rural environments, walking-related factors found in urban studies may not be important or relevant in rural environments. With respect to either residential or commercial density, for example, a place that is regarded as highly dense in rural areas may be considered highly sparse in urban areas. What's more, walking normally happens for short distances and often in combination with other transportation modes such as bus or train and the lack of public transportation in rural areas may create another obstacle for rural residents to walk.

Finally, walking research should distinguish between recreational and utilitarian walking. Boarnet et al. (2011) defined recreational walking as walking for pleasure or exercise (e.g. walking a dog) and utilitarian walking as walking to reach a destination, not just for the

sake of walking (e.g., walking to work or school). Since walking for these two purposes has different motivations, it is correlated with different predictor variables. Owen et al. (2007), for example, found a strong independent positive association between walking for transport and an objectively derived neighborhood walkability index but no significant association between environmental factors and recreational walking. Thus, it is essential to differentiate between these the two types of walking and examine them separately.
This study aims to fill the gaps introduced above. Specifically, the study explores the relationship between recreational walking and a wide range of predictor variables in mostly rural Vermont using spatially explicit regression models. Predictor variables from five different categories, social demographics, lifestyle, physical built environment, perceived built environment and attitude, are examined in the study. Equally significantly, the study applies spatial regression models to account for spatial non-independence among points. Finally, the study concludes by pointing out significant findings about spatial regression exploration and the relationship between walking and its contributing variables.

REVIEW OF LITERATURE

Walking can be considered to be the result of a stepwise, three-level, decision-making process (Coogan et al., 2011) that reflects the "volition to [walk]," "factors [that] either facilitate or impede volition" and prevailing "social norm[s]." In practical terms, the likelihood that an individual will walk can be understood by asking the following three questions, "whether one is willing to walk," "whether one is able to walk" and "whether one is satisfied with walking." Willingness to walk is a necessary precursor to walking. Willingness to walk is reflective of people's attitudes and is likely correlated with lifestyles and socio-demographic variables. By itself, however, willing is not sufficient to make people walk. As Jopson (2000) found, even individuals with positive attitudes toward walking and an expressed desire to walk may not do so. Real or perceived physical constraints and environmental limitation that might prevent walking provide a secondary hurdle to action. Finally, the satisfaction derived from the walking behavior itself and influences from peers or family play significant roles in the decision to walk. This framework suggests that social demographics, lifestyle and attitude variables, the built environment and perceptions of that environment all play a role in the decision to walk, as has been supported to varying degrees in the current literature.

Factors the influence walking rates

Extensive research has linked socio-demographic and built environment characteristics to walking (e.g. Handy, 1996; Aultman-Hall et al., 1997; Sallis et al., 1999; Cervero et al., 2003; Frank et al., 2003). As discussed by Livi and Clifton (2004), these data are relatively easy to acquire and analyze while psychological and social factors that are also crucial to understanding behavior are more difficult to monitor and evaluate objectively and consequently are less well understood.
Since 1990s, numerous studies have investigated the relationship between the built environment and travel behavior. These studies have found that residents living in traditional neighborhoods, characterized as high density, high accessibility, mixed land uses and rectangular street networks, drive less and walk more than those living in suburban neighborhoods (e.g., Cervero et al., 2003; Crane et al., 1998). An analysis of

household travel diaries from Portland, Oregon showed that narrow roads, street connectivity, continuous sidewalks, and zonal household density, as well as proximity to commercial uses and transit reduce trip length and increase the prevalence of walking (1000 Friends of Oregon. 1994). Shriver and Katherine (1997) found that neighborhood transportation, land use, and design characteristics influence walk distance, duration, purpose, and number of secondary activities. Handy (1996) and Shriver (1997) concluded that well-designed facilities could encourage walking without compromising safety and convenience. Dwelling density, street connectivity, land-use mix, and net retail area have all been correlated with walking decision (Frank et al., 2003; Aultman-Hall et al., 1997; Handy, 1996). In 2011, Boarnet et al studied possible influence variables of walking for travel, and found that characteristics of the sidewalk infrastructure, street crossings and traffic speeds, and land use are reliably associated with walking.

People's perceptions of the built environment are also important and may not be consistent from person to person or with the actual built environment. Consequently, data availability is a major impediment to conducting studies on these factors. For example, a study conducted by Moudon et al. (2006) found that individuals who self-reported that they did not have any grocery stores in their neighborhood had an average of 2.46 such stores within a 1-km airline buffer of their home, indicating that built environment variables are not an accurate reflection of people's perceptions. Another study was conducted through the Surface Transportation Policy Project (Belden et al., 2003) assessed perceptions of walkability through phone interviews but did not compared the perceptions recorded in the interviews with actual walking behavior. Objective measures refer to the physical existence of infrastructures and facilities, while subjective measures describe how people perceive the objective existence, which is closely tied to individuals' understanding and knowledge. Thus, they are both essential in influencing walking and need to be differentiated.

In addition to physical built environment characteristics, social and psychological influences are essential determinates of walking (Livi et al., 2004). Kitamura et al. (1997) concluded that the variation in travel demand for their San Francisco Bay Area sample owed more to attitudinal factors than to land use characteristics. Bagley and Mokhtarian (2002) employed a structural equations model to investigate the relationships between explanatory variables and travel demand and found that with respect to direct and total effects, attitudinal and lifestyle variables had the greatest impact among all explanatory variables.

Spatial auto-correlation among predictors of walking

Based on the literature, factors that contribute to walking include socio-demographic variables, physical built environment variables, as well as people's perception of and attitudes toward the built environment. These factors are likely to exhibit spatially auto-correlated patterns, meaning that these factors are more likely to be similar for residents living close to one another than for those who live far apart from each other. This reflects Tobler's first law of geography, "everything is related to everything else, but near things are more related than distant things" (1970). Spatial auto-correlation can be either "inherent," meaning that the variable has an intrinsic spatial relationship, or "induced," meaning that the variable is influenced by an external variable that is inherently auto-correlated. Many built environment characteristics, such as sidewalk coverage or the distance to the nearest park, are inherently spatially auto-correlated while perceptions of the built environment, which are influenced by the built environment, are likely to exhibit induced spatial auto-correlation. Walking and its predictors, therefore, are likely to have both inherent and induced spatial auto-correlation.

Methodologically, however, most of the walking studies have used traditional non-spatial regression methods to interpret to quantify factors that contribute to walking. For instance, Krizek (2003) applied linear regression models to test whether changes in travel behavior could be attributed to changes in neighborhood accessibility, controlling for changes in socio-demographic characteristics, workplace accessibility, and regional accessibility. He found that changes in neighborhood accessibility were statistically significant to all models of travel behavior, including walking. However, that study neglected the possibility that the relationship could be spatial. As Greene (2000) and Fox (1997) put it, regression models with auto-correlated residuals violate the independence assumption for errors in the classical multiple linear regression model – an assumption embodied in the Gauss Markov Theorem.

The field of spatial econometrics has developed techniques to explicitly account for spatial variables related to location topography and distance in the model specification process (Anselin, 2006). These model specification techniques are now included in several software packages including R, Geoda, and ESRI's ArcGIS software. Consequently, spatially explicit research methods are gaining traction in a number of research fields. For example, Voss et al. (2006) explored the inter-county variations in child poverty rates in the U.S. using spatial regression techniques and concluded that the explicit treatment of spatial effects in an explanatory regression model improved considerably on the results of linear regression models that do not account for spatial effects. Celebioglu and Dall'erba (2009) performed an exploratory spatial data analysis on regional growth and development levels in Turkey from 1995-2001 and detected the presence of spatial dependence across the provinces. Messner et al. (2011) applied techniques of exploratory spatial data analysis and spatial regression modeling to explain variation in robbery and assault rates across 413 districts Germany. The possibility of spatial auto-correlation among factors related to walking suggests that these techniques need to be evaluated for their appropriateness for walking and other transportation related research.

DATA

The study uses three major datasets, the New England Transportation Survey (NETS), business location data from Nielsen and ESRI's road network data.

The NETS was designed to create a "portrait" of rural transportation patterns in Vermont, New Hampshire, and Maine. With support from the University of Vermont Transportation Research Center, the New England Transportation Institute conducted the survey during 2008 and 2009 and collected 3,630 valid responses that included the geocoded location of the respondents' residences. The survey instrument incorporated questions regarding the respondents' current travel behaviors, attitudes toward the availability of various transportation services, and perceptions of their access to important destinations.

The Nielsen Business database contains a variety of business information for over 14 million establishments in the U.S., including company name, city, zip code, latitude/longitude, business category codes and descriptions, counts of employees, and annual sales. In total, 187,216 businesses are listed in the three northern New England states. By examining and cross-referencing two different business categorization schemes, supermarkets, convenience stores, and specialty food stores were identified. For food outlets that were not labeled, further cross-referencing was performed by using Google and Yellow Pages to decide which of the three categories of interest to use. In the end, 4,137 retail food establishments are considered in this study for analysis. This data is mostly used to calculate commercial density and residents' accessibility to different types of stores.

A road network from ESRI was used to conduct network-based spatial analysis in ArcGIS using the Network Analyst tool. This geodatabase provides complete road network information, including how various routes connect to each other and their speed limits, which enables the calculation of realistic driving times. The closest facilities to the travel survey respondents were used in the analysis to calculate network distances to convenience stores and supermarkets based on the road network data.

Model variables

The dependent variable in this study is the time spent on recreational walking and is derived from the NETS data. The original survey question asks the respondents to report the "number of hours per week spent walking, jogging, running for exercise/pleasure, or walking the dog."

In total, 170 independent variables derived from the NETS dataset or calculated from census and business data were considered. The independent variables fell into five major categories: social demographics, lifestyle, built environment, perceptions of the built environment, and attitudes toward general transportation issues. The 21 social demographic variables included the respondents' age, gender, marital status, education level and income. Lifestyle variables included those that describe people's non-walking, physical activities and the amount of time that people spend doing sedentary activities. Built environment variables provide objective measures the physical environment around the respondents' neighborhood, such as nearby building types and distances to various destinations. Many of these variables, including commercial and residential density, and the distance to closest supermarkets and convenience stores were calculated in ArcGIS. Perceptions of the built environment were measured by people's level of agreement or disagreement with certain statements regarding neighborhood characteristics, and specific aspects that people considered when moving into their current neighborhood or would consider if they moved in the future. The last category captured about people's attitudes toward general transportation issues, such as their agreement with the statement, "I need to drive my car to get where I need to go." A detailed list of the independent variables is in Appendix A, from Table 1 to 5.

The GIS variables which measure density (either commercial or residential) and distances (to supermarkets or convenience stores) are all continuous. The NETS data included both categorical and continuous variables. Categorical variables with seven or more ordinal categories were considered to be continuous since their categories are specific enough to represent the variable meanings. Variables with fewer than three categories were recoded into binary variables. For instance, the gender variable has been converted into female (indicating whether the respondent is female or not). It was not practical to convert variables with three to seven categories into binary variables but it may also be problematic to treat these variables as continuous. Ultimately, 19 variables of this kind were retained and treated as continuous variables. Interpreting the results related to these variables should be done with caution.

Lastly, when recoding variables, duplicates remain in all independent variables. For instance, the employment variable has eight categories (i.e., employed full-time, employed part-time, self-employed, student, student and employed, retired, homemaker or stay-at-home-parent, and not currently employed), represented by continuous values from 1 to 8. If considering the variable in the regression model, changes of values across different categories would not be very meaningful. So this variable was converted into three different binary variables, employment without schooling, employment with any kind of schooling, and employment including employed students. They overlap in some way, but keeping them all in the model helped to identify which variables related more with the dependent variable.

METHODOLOGY

The analysis started with the development of a linear regression model for walking and its predictors. A stepwise method was taken to incorporate the 170 variables to achieve the best model fit. Based on the basic model, two software packages (S-Plus and Geoda) were used to detect spatial effect and establish spatial regression models. When developing spatial regression models, three matrix weighting strategies were explored: one nearest neighbor, three nearest neighbors and a distance-threshold matrix. The construction of each model is described below.

Linear regression model

A linear regression model was developed to serve as the base model for the analysis. With 170 independent variables, the first step was to conduct bivariate regression analysis between the dependent and each independent variable to get a basic evaluation of how these factors correlate to walking. This step filtered out insignificant variables and revealed 75 variables that were significant at the 0.10 level. These significant variables were ranked based on the absolute value of their correlation with walking and this ranking was used to determine the sequence that they entered into the linear regression model. The criteria for finding the best fit model included making sure all the factors in the model were significant at 0.10 level or higher and that the model fit increased when the new variable entered the model. Note that since we have “duplicate” variables (see above), another controlling factor is to remove duplicates and keep the most significant one.

Spatial regression modeling

Building weight matrices is a significant step when building spatial regression models. As Celebioglu and Dall’erba (2009) state, the spatial weight matrix is necessary to specify the neighborhood structure for a spatial dataset. According to Griffith (1995), an incorrect choice of weights inflates the standard error of the model and biases the correlation estimate. It is better to under-specify the weight matrix (have too few neighbors) than to over-specify it (Griffith, 1995). Multiple neighbor weight matrices were analyzed to ensure the model was properly specified.

Though spatial econometrics theory has been evolving for years, its application for spatial regression models is still under-developed (Anselin, 2012). Anselin et al. (2006) identified the R spdep package and Geoda as the existing software with the functionality for spatial regression modeling. S-Plus work, the commercial counterpart to R, is also cable of spatial regression modeling (Quene, 2006). Since S-Plus and Geoda have differing limitations, analyses was conducted using both of these packages.

a. Spatial weight matrix

The first requirement of both S-Plus and Geoda is the specification of weight matrices. In order to compare model results, three weight matrices were established for both software models. Every pair of data points must either be labeled as a "neighbor" or "non-neighbor" with respect to one another. Neighbors have a non-zero weight value, while non-neighbors have weight of zero. Normally, neighbors are assigned with 1 and non-neighbors are assigned with 0.

The k-nearest neighbors method and the distance-band neighbor method are two common criteria for determining whether or not a data pair are neighbors. The k-nearest neighbors method defines neighbors by comparing the distances between data points for all data pairs and chooses the k nearest ones. The distance-band neighbor method defines all data points within distance threshold of one another as neighbors. Distance-band weights are always symmetrical since when A is the within a given radius of B, B must also be within that radius of A. The k-nearest neighbors approach, however, can produce asymmetric results as A may be the nearest neighbors to B but a different location, C, may be closer to A than B is, resulting in an asymmetric weight matrix.

In this report, we decided to create three weight matrices to investigate the distribution of our variables of interest: k_1 nearest neighbor matrix, k_3 nearest neighbor matrix, and D_13 miles matrix which defines as neighbors located within a great circle distance with a cutoff of 13 miles. We chose one (k_1) and three (k_3) nearest neighbors because it is better to under-specify the weights matrix (have too few neighbors) than to over-specify it (Griffith, 1995). The distance threshold of 13 miles was based on semi-variogram analysis which showed a 13 mile extent to spatial auto-correlation for walking in Vermont.

b. Spatial regression models

Using the best-fit model from linear regression analysis, the study conducted spatial regressions using S-Plus and Geoda. The first step in this process was to examine the linear regression residuals for SA based on the three weight matrices (i.e., k_1, k_3 and D_13 miles).

Two common spatial regression models are the spatial error model and the spatial lag model (Voss et al. 2006). The spatial error model is commonly specified according to equations 5 and 6, and the spatial lag model is specified according to equation 7.

$$y = X\beta + u \tag{5}$$

$$u = \rho W u + \varepsilon \tag{6}$$

$$y = \lambda W y + X\beta + \varepsilon \tag{7}$$

In these equations, y is the vector representing the dependent variable, X is the matrix representing independent variables, β is the vector of regression parameters to be estimated, u is the vector of error terms presumed to have a covariance structure as given in equation 6, ρ is a spatial lag parameter to be estimated, and W is the weight matrix defining neighborhood structure in the spatial process.

The spatial error model is specified by using the residuals from the linear model as a proxy for u. In equation 7, Wy is the vector of spatial lags of the dependent variable$y$ and λ is a spatial lag parameter to be estimated. Wy represents the weighted average of the dependent variable for neighboring locations. For greater details on model specifications for both types of models, the reader can refer to Voss et al. (2006).

S-Plus only calculates spatial error models, while Geoda can calculate both spatial error and lag models. However, S-Plus supports distance-band and k-nearest neighbor weight matrices, while Geoda only supports the distance-band weight matrices since it requires a symmetric weight matrix.

RESULTS

Linear regression model

Table 1 shows the results for the final linear regression created using the bivariate regression and stepwise model specification described above. This model included 17 independent variables and had an overall model fit, as measured by R-square, of 0.1251, indicating that the independent variables account for about thirteen percent of the total variance in recreational walking. The lifestyle and perceived built environment categories contributed the largest number of variables to the model.

Two social demographic variables show significant correlation with walking, “walklimit” and “employed.studentemploy,” referring to whether people have limited physical ability to walk and whether they are employed, respectively. Both limitation in walking and being employed negatively influences recreational walking rates. Among lifestyle variables, the more time people spend on other types of physical activities, such as biking, exercising at gym or going hiking, the more time people spend walking recreationally. This relationship among differing types of physical activity is not unexpected. Another lifestyle variable that correlates positively with walking is whether the person worked or volunteered for a candidate or party in the last presidential election, implying people who actively engage in political activities are more likely to walk. Another contributing variable is the time people spend eating in sit-in restaurants, which shows a positive relationship between eating in those restaurants and walking. With respect to physical built environment variables, the distance to the closest convenience store is significantly and positively linked to walking while the length of the longest household commute is significantly and negatively related to recreational walking. This makes sense for Vermonters because when they are far away from work or school, they may tend to drive. Many of the perceived built environment variables are positively correlated with walking. People who value exercise and health issues more also tend to walk more. Social ties with neighbors serve as an influencing factor as well. The more people agree that they know their neighbors well, the more they tend to walk. Proximity to outdoor recreation areas and commercial activities are positively related to walking. Only one attitudinal variable was included in the final model. The belief that reducing automobile mileage was difficult was significantly and negatively correlated with walking.

Table 1 Regression model results for walking

Independent variables	OLS (S-Plus & Geoda)	K_1 spatial error model (S-Plus)	K_3 spatial error model (S-Plus)	D-13 spatial error model (S-plus)	D-13 spatial error model (Geoda)	D-13 spatial lag model (Geoda)
Social demographic variables						
walklimit *[whether limited in walking]*	-0.825(0.06)	-0.825(0.06)	-0.823(0.07)	-0.817(0.07)	-0.812(0.07)	-0.823(0.06)
employed.studentemploy *[whether employed]*	-0.638(0.02)	-0.634(0.02)	-0.636(0.02)	-0.641(0.02)	-0.646(0.02)	-0.641(0.02)
Life style variables						
weeklyact2t *[time to bike for exercise/pleasure]*	0.374(0.00)	0.375(0.00)	0.375(0.00)	0.373(0.00)	0.372(0.00)	0.373(0.00)
weeklyact3t *[time to exercise at a gym/fitness club/health club]*	0.154(0.01)	0.152(0.01)	0.154(0.01)	0.155(0.00)	0.155(0.00)	0.154(0.00)
weeklyact4t *[time for other physical activity]*	0.105(0.00)	0.104(0.00)	0.105(0.00)	0.104(0.00)	0.104(0.00)	0.104(0.00)
quick9.rec *[whether work for candidate in presidential election]*	0.869(0.04)	0.878(0.04)	0.871(0.04)	0.860(0.04)	0.857(0.04)	0.852(0.04)
weeklyact6t *[time to eat at sit-in restaurants]*	0.197(0.01)	0.201(0.00)	0.197(0.01)	0.199(0.00)	0.199(0.00)	0.198(0.00)
bikedest3.rec* *[Frequency of biking to go food shopping]*	0.521(0.02)	0.517(0.02)	0.519(0.02)	0.524(0.02)	0.526(0.02)	0.521(0.02)
Physical built environment						
distance_convenience *[closest distance to convenience stores]*	0.070(0.04)	0.069(0.04)	0.070(0.04)	0.070(0.04)	0.070(0.04)	0.070(0.04)
distances3t *[longest commute to work/school for household]*	-0.009(0.07)	-0.009(0.07)	-0.009(0.07)	-0.009(0.08)	-0.009(0.08)	-0.009(0.07)
Perceived built environment						
nextneigh1 *[importance of living walk-/bikeable neighborhood]*	0.175(0.05)	0.177(0.04)	0.175(0.05)	0.177(0.04)	0.176(0.04)	0.174(0.04)
move14 *[health reasons]*	0.149(0.01)	0.146(0.01)	0.148(0.01)	0.150(0.01)	0.152(0.01)	0.151(0.01)
neighborhood18.rec *[I feel I know my neighbors extremely well]*	0.248(0.00)	0.246(0.00)	0.248(0.00)	0.249(0.00)	0.250(0.00)	0.249(0.00)
move11 *[near outdoor recreation]*	0.137(0.03)	0.139(0.03)	0.137(0.03)	0.136(0.03)	0.138(0.03)	0.136(0.03)
neighborhood7.rec *[near commercial activity]*	0.109(0.03)	0.107(0.03)	0.109(0.03)	0.109(0.03)	0.110(0.03)	0.109(0.03)
neighborhood4.rec *[easy to buy groceries near home]*	-0.246(0.00)	-0.255(0.00)	-0.247(0.00)	-0.241(0.00)	-0.244(0.00)	-0.245(0.00)

Attitudinal variables						
otherconsid1 *[difficult to reduce auto mileage and gasoline use]*	-0.214(0.00)	-0.214(0.00)	-0.214(0.00)	-0.215(0.00)	-0.214(0.00)	-0.214(0.00)
(Intercept)	1.607(0.06)	1.678(0.05)	1.613(0.06)	1.586(0.06)	1.574(0.06)	1.317(0.16)
R-squared	0.1251	M	M	M	0.1374	0.1372
Residual standard error	3.710	3.707	3.710	3.710	3.683	3.683
Lag parameter (rho)	NA	-0.026	-0.002	8.846e-4	0.113	0.090
Log-likelihood	M	-6161	-6162	-6162	-3441.447	-3441.53
AIC	M	M	M	M	6918.89	6921.06

NA: Not applicable, M: missing

R-squared is adjusted R-squared for linear regression; otherwise pseudo R-squared.

Moran's I on OLS residuals

Moran's *I* is one method to check for spatial auto-correlation among OLS residuals. Moran's *I* results based on *k*_1, *k*_3 and *D*_13 weight matrices are shown in Table 2. These results do not suggest significant spatial auto-correlation and, therefore, that OLS regression may be sufficient.

Table 2 Moran's I on linear regression residuals

k_1 neighbor matrix	k_3 neighbor matrix	D_13 matrix
-0.044 (0.187)	-0.003 (0.923)	0.004 (0.328)

In addition to Moran's *I*, Geoda also conducts a number of other diagnostics based on the distance weight matrix. Table 3 shows the results of these diagnostics tests using the *D_13* matrix. The Lagrange Multiplier test for the spatial lag and error models pre-tests whether those two alternative models would improve on OLS. The results do not suggest the lag or error models would improve on the OLS results. However, it is advisable to actually try out spatial regression models, and then compare them with linear regression to determine whether spatial effects really makes a difference.

Table 3 Geoda spatial diagnosis results

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.004	0.977	0.329
Lagrange Multiplier (lag)	1	0.397	0.528
Robust LM (lag)	1	0.037	0.847
Lagrange Multiplier (error)	1	0.584	0.445
Robust LM (error)	1	0.224	0.636
Lagrange Multiplier (SARMA)	2	0.621	0.733

Spatial regression models

Spatial regression models based on *k*_1, *k*_3 and *D*_13 weight matrices were developed using both S-Plus and Geoda software using the same variables as the baseline OLS model. The results of these regressions are shown in Table 1. The second, third and fourth columns in this table are regression models from S-Plus, while the last two columns are models from Geoda. As explained previously, S-Plus only supports spatial error modeling but is capable of performing spatial models based on asymmetric k-nearest neighbors weight matrices. Geoda, in contrast, supports both spatial lag and error models but cannot use asymmetric k-nearest neighbor weights. Consequently, the analyses conducted in each of the packages are complementary.

For S-Plus models, the results using the three different weight matrices produce very similar results both in terms of coefficients and their significance levels. S-Plus does not provide a pseudo R-squared measure but the log-likelihood does not vary substantially across the models, indicating comparable levels of model fit. As would be expected from the results of the Moran's I test, the residual standard error does not show significant differences between three models and OLS model, indicating that the spatial error models did not improve on the OLS model. The lag parameters (rho) for the three models vary significantly, indicating the spatial weights matrices influence the extent that neighboring effect is accounted. In particular, for the k_1 regression model that has the highest lag parameter, it suggests that stronger neighbor effects are accounted than the other two models.

The spatial error model and the spatial lag model created in Geoda using the D_13 weighting matrix are shown in the final two columns of Table 1. Most model parameters for two models are similar, including the coefficients of the independent variables, the pseudo R-square values, the residual standard error and log-likelihood terms. The lag parameter is slightly bigger for the spatial error model while AIC is slight larger for the spatial lag model. When comparing both models with the OLS results, the coefficients of independent variables are very similar. The only difference compared to OLS is that the R-square for spatial models increases to 0.137 from 0.1251 but since the R-square in the spatial model is a pseudo value, the increase in value does not necessarily suggest a better model.

In Table 1, when comparing the D_13 spatial error model from S-Plus with the same model from Geoda, most variables' parameters are very close, but the lag parameter and log-likelihood values differ substantially. The lag parameter for the Geoda model is greater, indicating it estimates a larger spatial neighbor effect. The log-likelihood value for the Geoda model is also greater, meaning it is preferable to the S-Plus model.

CONCLUSIONS

This research examines walking and its predictors for Vermonters. Starting from one hundred and seventy independent variables across five categories, most variables from people's lifestyle and perceived built environment stand out as significantly correlated with walking. A few variables worth noting include physically active individuals who engage in activities such as biking, exercising at a gym or hiking, which impacts walking behavior. Aspects of the neighborhood that residents care about include whether the neighborhood allows exercise by walking or biking, health concerns, proximity to recreation areas and commercial activities, which all indicate people's preference for physical activity. Interestingly enough, the model finds that people's engagement in political activity is positively correlated with their walking levels.

The spatial regression analysis suggests it is not necessary to apply spatial techniques to the walking behaviors considered in this study. The first evidence is that the Moran's I test on OLS residuals does not find any SA based on three different weight matrices, which indicates the linear model is sufficient. To further examine the need for spatial regression analysis, the study proceeds to build spatial

error and lag models using S-Plus and Geoda. The spatial error models based on three weight matrices from S-Plus show no significant difference from the OLS model and among each other. This indicates that different weight matrices do not affect the outcome of the spatial error models significantly in S-Plus. With respect to the Geoda models, the spatial error and lag models are constructed based on symmetric distance-band weight matrix. Both models are diagnosed based on OLS residuals. Comparison of the two models does not reveal any major differences, except for the lag parameter. The parameter for the spatial error model is greater than that for the other model, which would indicate that the spatial error model is better if their diagnoses were significant. This was not found. Thus, the spatial regression analysis performed in this study suggests that it is sufficient to perform the ordinary linear regression models because no spatial effect is detected.
Lastly, as Voss, P.R., et al. (2006) mentioned, this is still an emerging area where software developments have not kept pace with conceptual and theoretical advances. Analyses in this project concur with this finding and suggest a need for further development of estimation techniques to advance more widespread use of spatial techniques in modeling of transportation databases.

ACKNOWLEDGEMENTS

The author is grateful for the support of Professors Brian Lee, Donna Rizzo and Austin Troy. The team acknowledges funding from the US DOT through the University Transportation Center program at the University of Vermont.

DISCLAIMER

REFERENCES

1000 Friends of Oregon. (1994) The Pedestrian Environment: Making the Land Use Transportation Air Quality Connection. *Parsons Brinkerhoff Quade and Douglas, Inc., with Cambridge Systematics, Inc., and Calthorpe Associates*. 4.

Anselin, L. (2012) *Spatial regression* [E-slides]. Retrieved from: https://geodacenter.asu.edu/eslides.

Anselin, L. (2010) Thirty years of spatial econometrics. *Papers in Regional Science*, 89 (1), pp. 3-25.

Anselin, L., Syabi, I. & Kho, Y. (2006) GeoDa: an introduction to spatial data analysis. *Geographical Analysis*, 38, 5–22.

Anselin, L. (1992) *SpaceStat, a Software Program for Analysis of Spatial Data*. Santa Barbara, CA: National Center for Geographic Information and Analysis *(NCGIA)*, University of California.

Audirac, I. (1999) Stated preference for pedestrian proximity: an assessment of new urbanist sense of community. *Journal of Planning Education and Research*, 19(1), pp. 53–66.

Aultman-Hall, L., M. Roorda, BW Baetz. (1997) Using GIS for evaluation of neighborhood pedestrian accessibility. *J Urban Planning Dev*, 123, pp 10–17.

Belden Russonollo & Stewart. (2003) *Americans' attitudes toward walking and creating better walking communities*. Surface transportation policy project. Retrieved from: www.transact.org/library/reports_pdfs/pedpoll.pdf.

Berrigan, David, Richard P. Troiano. (2002) The Association Between Urban Form and Physical Activity in U.S. Adults. *American Journal of Preventative Medicine*, 23 (2S).

Boarnet MG, A. Forsyth, K. Day and et al. (2011) The street level built environment and physical activity and walking: results of a predictive validity study for the Irvine Minnesota Inventory. *Environment and Behavior*. 43(6), pp 735-775.

Boarnet, Marlon G, Greenwald M, McMillan TE. (2008) Walking, urban design, and health, toward a cost-benefit analysis framework. *Journal of Planning Education and Research*. 27, pp 341-358.

Cao, Xinyu, Mokhtarian PL, Handy SL. (2007) Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation*. 34, pp 535-556.

Carsten, OMJ. DJ Sherborne, JA Rothengatter. (1998) Intelligent traffic signals for pedestrians: evaluation of trials in three countries. *Transportation Research Part C*. 6, pp 213–229.

Celebioglu, F. and Dall'erba, S. (2009) Spatial Disparities across the regions of Turkey: an exploratory spatial data analysis. *The Annals of Regional Sciences*, Published online: 05 July 2009.

Cervero, R., Duncan, M. (2003) Walking, bicycling, and urban landscapes: evidence from San Francisco Bay Area. *Am. J. Public Health*. 93(9), pp 1478–1483.

Coogan, MA, Adler T and Karash KH. (2011) The paths from walk preference to walk behavior: applying latent factors in structural equation modeling. *World Symposium on Transport and Land Use Research*. (presentation)

Crane, R. (1996) On Form versus Function: Will the New Urbanism Reduce Traffic, or Increase It? *Journal of Planning Education and Research*. 15, pp 117-126.

Crane, R. (2000) The influence of urban form on travel: an interpretive review. *J. Plann. Lit*. 15(1), pp 3–23.

Crane, R, Crepeau, R. (1998) Does neighborhood design influence travel? A behavioral analysis of travel diary and GIS data. *Transport. Res. D*. 3(4), pp 225–238.

Eberhardt, MS.; Ingram, DD.; Makuc, DM., et al. (2001) *Urban and Rural Health Chartbook*. Health, United States. Hyattsville, MD: National Center for Health Statistics.

Ewing, R., Cervero, R. (2001) Travel and the built environment: a synthesis. *Transport. Res. Rec*. 1780, pp 87–113.

Forsythe, M. J., & Berger, W. G. (1973). *Urban pedestrian accident countermeasures experimental evaluation*. Vol. 1, Appendix C: Behavioral Evaluation Summary Data, Biotechnology, Inc. Falls Church, VA; US Department of Transportation, Washington, DC.

Fox, J. (1997) *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage Publications.

Frank, LD. (2004) Economic determinants of urban form: Resulting trade-offs between active and sedentary forms of travel. *Am J Prev Med*. 27 (suppl 3), pp. 146–153.

Frank, LD, PO Engelke, TL Schmid. (2003) *Health and community design: The impact of the built environment on physical activity*. Island Press, Washington DC.

Gårder, P. (1989) Pedestrian safety at traffic signals: a study carried out with the help of a conflict technique. *Accident Analysis and Prevention*. 21, pp. 435–444.

Greene, W. (2000) *Econometric Analysis*. Upper Saddle River, NJ: Prentice–Hall.

Griffith, D.A. (1995) *Some guidelines for specifying the geographic weights matrix contained in spatial statistical models*. In Arlinghaus, S.L., editor, Practical Handbook of Spatial Statistics, pp 65-82. CRC press, Inc., Boca Raton.

Handy, SL. (1996) Urban form and pedestrian choices: Study of Austin neighborhoods. *Transp Res Rec*. 1552, pp 135–144.

Handy SL, Boarnet MG, Ewing R, et al. (2002) How the built environment affects physical activity: views from urban planning. *American Journal of Preventive Medicine*. 23, pp 64–73.

Hanson, S., and M. Schwab. (1987) Accessibility and Intraurban Travel. *Environment and Planning A*. 19(6), pp 735–748.

Hodgson, F., May, T., and Tight, M. (1997) Evaluation of the MIST travel awareness campaign. *Traffic Engineering and Control*.pp 655-659.

Hu FB, Stampfer MJ, Solomon C, et al. (2001) Physical activity and risk for cardiovascular events in diabetic women. *Annal of Internal Medicine*. 134, pp 96–105.

Jopson, A. F. (2000) *Can psychology help to reduce car use*? Institute for Transport Studies, University Leeds, Leeds, UK.

Kahn, H.S., L.M. Tatham, C. Rodríguez, et al. (1997) Heath Jr. Stable behaviours associated with adults' 10 year change in body mass index and likelihood of gain at the waist. *Am. J. Public Health*. 87, pp 747–754.

Kaluzny, S. P., Vega, S. C., Cardoso, T. P., et al. (1996) *S-Plus spatialstats user's manual*. Seattle: Insightful Corporation. ISBN: 0-387-998226-4. 327 pp.

Kitamura, R., Mokhtarian, P.L., Laidet, L. (1997) A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation.* 24, pp 125–158.

Krizek, K.(2003) Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? *J. Am. Plann. Assoc*. 69(3), pp 265–281.

Kwon, Young-In, Shigeru Morichi, and Tetsuo Yai. (1998) Analysis of Pedestrian Behavior and Planning Guidelines with Mixed Traffic for Narrow Urban Streets, *Transportation Research Record 1636*, Paper No. 98-0541.

Leslie, E., N. Coffee, L. Frank, and et al. (2007) Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health Place*. 13, pp 111–122.

Livi AD, Clifton KJ. (2004) Issues and methods in capturing pedestrian behaviors, attitudes and perceptions: experiences with a community-based walkability survey. *TRB 2004 Annual Meeting* CD-ROM.

Manson JE, Hu FB, Rich-Edwards JW, and et al. (1999) A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. *New England Journal of Medicine*. 341, pp 650–658.

Messner, S. F., Teske, R. H. C., Jr., Baller, R. D., and et al. (2011) Structural covariates of violent crime rates in Germany: Exploratory spatial analysis in Kreise. Justice Quarterly, online first.

Moudon, AV, Lee C, Cheadle AD, et al. (2006) Operational definitions of walkable neighborhood: theoretical and empirical insights. *Journal of physical activity and health*, 3(suppl 1), pp s99-117.

Owen N., Cerin E., Leslie E., et al. (2007) Neighborhood walkability and the walking behavior of Austrian Adults. *American Journal of Preventive Medicine*. 33(5), pp 387-395.

Parks SE, Housemann RA, Brownson RC. (2003) Differential correlates of physical activity in urban and rural adults of various socioeconomic backgrounds in the United States. *J Epidemiol Community Health*. 57, pp 29–35.

Patterson PD, Moore CD, Probst JC, et al. (2004) Obesity and physical activity in rural America. *J Rural Health*. 20(2), pp 151-9.

Pikora, Terri. (2000) *Systematic Pedestrian and Cycling Environmental Scan (SPACES), Survey of the Physical Environment in Local Neighbourhoods: Observer's Manual*. Nedlands, Western Australia: University of Western Australia.

Quene H. (2006) Statistics with R and S-Plus. Retrieved from: http://cran.r-project.org/doc/contrib/Quene.pdf.

Randall, TA, BW Baetz. (2001) Evaluating pedestrian connectivity for surburban sustainability. *J Urban Planning Dev*. 127, pp 1–15.

Saelens, B.E., J.F. Sallis, L.D. Frank. (2003) Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Ann Behav Med*. 25, pp 80–91.

Sallis, James, Marilyn Johnson, Karen Calfas, and et al. (1997) Assessing perceived physical environmental variables that may influence physical activity. *Research Quarterly for Exercise and Sport*. 68(4).

Sallis JF, Owen N. (1999) *Physical activity and behavioral medicine*. Thousand Oaks, CA: Sage.

Shriver, K. (1997) Influence of environmental design on pedestrian travel behavior in four Austin (TX, USA) neighborhoods. *Transportation Research Record*, 1578, pp 65–73.

Sisiopiku V.P., Akin D. (2003) Pedestrian behaviors at and perceptions towards various pedestrian facilities: an examination based on observation and survey data. *Transportation Research Part F*. 6, pp 249-274.

Stokes, Robert J., John MacDonald, and et al. (2004) Estimating the effects of light rail transit on health care costs. *Health and Place*. 10.1016/j.healthplace.2007.04.002. 20, pp 151–159.

Voss, P.R., et al. (2006) County child poverty rates in the US: a spatial regression approach. *Popul Res Policy Rev.* 25, pp 369-91.

Zahran S, Brody SD, Maghelal P, et al. (2008) Cycling and walking: explaining the spatial distribution of healthy modes of transportation in the United States. *Transp Res Part D Transp Environ*. 13, pp 462–470.

APPENDIX A INDEPENDENT VARIABLES

1.1. APPENDIX A.1 SOCIO-DEMOGRAPHIC VARIABLES

No	Variable name	Variable meaning	Variable type
1	age1t	Age in years	continuous
2	female	Whether the respondent is female	binary
3	married	Whether the respondent is married	binary
4	loweduc	Highest completed education lower than bachelor	binary
5	employed	Whether employed, including full-time, part-time, and self-employed	binary
6	employed_studentemploy	Whether employed, including full-time, part-time, self-employed, and employed student	binary
7	employed_students	Whether employed, including full-time, part-time, self-employed, student, and employed student	binary
8	numveh1t	Number of vehicles in the household	continuous
9	housechild	At least one child under 18 live in home	binary
10	income	Household income	continuous
11	income_value	Household income, recode income into its average wages based on the range, then divide by 1000, just to easily represent	continuous
12	lowincome	Household income lower than 25,000 are low income	binary
13	openspacehouse	whether the residence is house on working farmland or major open space	binary
14	multifamilylive	whether the residence is apartment, townhouse, condominium, multi-family house/dormitory or other institutional	binary
15	own	Whether the house is owned	binary
16	long	how long you lived at current home(primary residence) location	continuous
17	long_value	the years lived at current home are converted into number	continuous
18	quick1_rec	are you a licensed driver	binary
1	quick7_rec	does your household own a 'second house'	binary

9			
20	drivelimit	Whether limited in driving	binary
21	walklimit	Whether limited in walking	binary

1.2. APPENDIX A.2 LIFE STYLE VARIABLES

No	Variable name	Variable meaning	Variable type
1	weeklyact2t	number of hours per week spent biking for exercise or pleasure	continuous
2	weeklyact3t	number of hours per week spent exercising at a gym, fitness club, or health club	continuous
3	weeklyact4t	number of hours per week spent doing other physical activity, such as hiking, climbing or kayaking	continuous
4	weeklyact5t	Number of hours per week spent eating at fast food restaurants	continuous
5	weeklyact6t	Number of hours per week spent eating at sit-in restaurants	continuous
6	weeklyact7t	Number of hours per week spent attending non-work meetings, movies, plays, or concerts	continuous
7	weeklyeatout	combine weekly5t+weekly6t, number of hours per week spent eating at fast food restaurant/sit-in restaurants	continuous
8	dailyact1t	number of hours per day spent watching TV	continuous
9	dailyact2t	number of hours per day spent playing video games	continuous
10	dailyact3t	number of hours per day spent using the internet or email	continuous
11	dailyact	Sedentary duration	continuous
12	tranbike1	whether most often they bike to work	binary
13	tranbike2	whether most often they bike to school	binary
14	tranbike3	whether most often they bike to food shopping	binary
15	tranbike4	whether most often they bike to go shopping for non-food items	binary
16	tranbike5	whether most often they bike to go to doctor	binary

17	tranbike6	whether most often they bike to a restaurant, bar or out for entertainment	binary
18	tranbike7	whether most often they bike to park or recreation area	binary
19	tranbike8	whether most often they bike to see family	binary
20	tranbike9	whether most often they bike to see friends	binary
21	tranbike10	whether most often they bike to attend church/worship	binary
22	bikedest1*	Frequency of biking to go to work in the last month	continuous
23	bikedest2*	Frequency of biking to go to school in the last month	continuous
24	bikedest3*	Frequency of biking to go food shopping in the last month	continuous
25	bikedest4*	Frequency of biking to go shopping for non-food items in the last month	continuous
26	bikedest5*	Frequency of biking to go to the doctor in the last month	continuous
27	bikedest6*	Frequency of biking to go to a restaurant, bar, or out for entertainment	continuous
28	bikedest7*	Frequency of biking to go to a park or recreation area in the last month	continuous
29	bikedest8*	Frequency of biking to see family in the last month	continuous
30	bikedest9*	Frequency of biking to see friends in the last month	continuous
31	bikedest10*	Frequency of biking to attend church/worship in the last month	continuous
32	bikedest1_rec*	Frequency of biking to go to work in the last month	continuous
33	bikedest2_rec*	Frequency of biking to go to school in the last month	continuous
34	bikedest3_rec*	Frequency of biking to go food shopping in the last month	continuous
35	bikedest4_rec*	Frequency of biking to go shopping for non-food items in the last month	continuous
36	bikedest5_rec*	Frequency of biking to go to the doctor in the last month	continuous
37	bikedest6_rec*	Frequency of biking to go to a restaurant, bar, or out for entertainment	continuous
38	bikedest7_rec*	Frequency of biking to go to a park or recreation area in the last month	continuous
3	bikedest8_re	Frequency of biking to see family in the last month	continuous

9	c*		
40	bikedest9_rec*	Frequency of biking to see friends in the last month	continuous
41	bikedest10_rec*	Frequency of biking to attend church/worship in the last month	continuous
42	quick2_rec	do you drive less than you used to	binary
43	quick4_rec	do you belong to any groups or social clubs in or near your community	binary
44	quick5_rec	do you belong to a gym, health club or fitness class	binary
45	quick6_rec	have you had a hunting or fishing license in the last two years	binary
46	quick8_rec	did you vote in the last presidential election	binary
47	quick9_rec	did you work or volunteer for a candidate or party in the last presidential election	binary

Note: variables with asterisks (*) are variables that are categorical with three to six categories. Reservations should be made when interpreting these variables.

1.3. APPENDIX A.3 PHYSICAL BUILT ENVIRONMENT VARIABLES

No	Variable name	Variable meaning	Variable type
1	transit5	No public transit available in my neighborhood	binary
2	nearby1	detached single-family homes	binary
3	nearby2	apartment buildings, townhouses, condominiums, multi-family houses(dulexes)	binary
4	nearby3	other types of homes	binary
5	nearby4	non of these-only farmland or major open space	binary
6	distances1t	one-way commute to work	continuous
7	distances2t	one-way commute to school	continuous
8	distances3t	largest commute distance to work or school for anyone in your household	continuous
9	distances4t	distance from home to nearest store for basic needs	continuous
10	distances5t	distance from home to place where you buy groceries	continuous
11	distances6t	distance from home to place where you buy major retail items	continuous
12	distances7t	distance from home to medical facility/hospital you would use in an emergency	continuous
1	distances8t	distance from home to a place where you eat or drink	continuous

13		and have an enjoyable evening	
14	distances9t	distance from home to a place where you can get a bus or train to Boston or NYC	continuous
15	commercial density	commercial density	continuous
16	residential density	residential density	continuous
17	distance_supermarket	cloest distance to supermarkets in miles	continuous
18	time_supermarket	shortest time to supermarkets in minutes	continuous
19	distance_convenience	cloest distance to supermarkets in miles	continuous
20	time_convenience	shortest time to supermarkets in minutes	continuous

1.4. APPENDIX A.4 PERCEIVED BUILT ENVIRONMENT VARIABLES

No	Variable name	Variable meaning	Variable type
1	move6	close to family, friends/other family reasons	binary
2	move7	close to church or other place of worship	binary
3	move8	close to job or school	binary
4	move11	near major outdoor recreation areas	binary
5	move12	walkable neighborhood, near local activities	binary
6	move13	farming or gardening	binary
7	move14	health reasons	binary
8	move15	value having space and separation from others	binary
9	move16	get away from urban life/value being rural	binary
10	move17	concerns about crime or unpleasant disturbances	binary
11	feel1*	feelings on distance to work	continuous
12	feel2*	feelings on distance to school	continuous
13	feel3*	feelings on distance to get basic food	continuous
14	feel4*	feelings on distance to get groceries	continuous
15	feel5*	feelings on distance to large retail stores	continuous

16	feel6*	feelings on distance to hospital	continuous
17	quick3_rec	are you comtemplating moving within the next five years	binary
18	quick10_rec	in the past year, did you ever decide not to address a medical concern or keep an appoitment because it was too difficult to get to the doctor or medical center	binary
19	neighborhood1_rec	My neighborhood has an adequate number of good sidewalks or walking paths.	continuous
20	neighborhood2_rec	It is easy to get to a town center or other place of activity.	continuous
21	neighborhood3_rec	I worry that it would be difficult to get help in case of an auto accident on my local roads.	continuous
22	neighborhood4_rec	It is easy to get to a place to buy groceries near my home.	continuous
23	neighborhood5_rec	I can easily get to places where people gather, like community centers, libraries, or social clubs.	continuous
24	neighborhood6_rec	My home has adequate room for parking two or more cars.	continuous
25	neighborhood7_rec	I live within walking distance of commercial activity, like stores and places where I can get coffee or other casual meals.	continuous
26	neighborhood8_rec	Biking in my neighborhood is safe and enjoyable.	continuous
27	neighborhood9_rec	I have friends and relatives who could help me get where I need to go.	continuous
28	neighborhood10_rec	My home is conveniently located near to where I work or go to school.	continuous
29	neighborhood11_rec	Other people think my home and neighborhood are very nice.	continuous
30	neighborhood12_rec	To get to my home, I rely on dirt roads, or narrow, winding two lane roads.	continuous
31	neighborhood13_rec	I like the feeling that I am physically isolated from other residents.	continuous
32	neighborhood14_rec	I worry about how long it would take police and fire to get to my home.	continuous
33	neighborhood15_rec	I worry about how long it would take to get from my home to the hospital in an emergency.	continuous
34	neighborhood16_rec	I worry that, in the future, I will not be able to get to medical services from where I live now.	continuous
35	neighborhood17_rec	I feel safe in my home.	continuous
36	neighborhood18_rec	I feel I know my neighbors extremely well.	continuous

37	neighborhood19_rec	Having less money in my retirement account would tend to make it harder for me to move to a more densely settled area.	continuous
38	neighborhood20_rec	Having less money in my retirement account might make it more important for me to move to a more densely settled area.	continuous
39	neighborhood21_rec	I feel safe when outside in my neighborhood.	continuous
40	neighborhood22_rec	It is the government's job to get me to the hospital, so I don't worry about it.	continuous
41	area1	Overall, how satisfied are you with the area where you live?	continuous
42	area2	How satisfied are you with the potential for economic advancement (good jobs) in this area?	continuous
43	area3	Overall, how satisfied are you with staying in your area, compared to moving to a more urban area?	continuous
44	area4	How satisfied are you with your ability to get where you need to go in a reasonable amount of time?	continuous
45	nexthome1	Importance of having an adequate number of sidewalks or walking paths in good, safe condition at next home	continuous
46	nexthome2	Importance of having a place to do my shopping reasonably near my home at next home	continuous
47	nexthome3	Importance of having a large lot with plenty of space at next home	continuous
48	nexthome4	Importance of having a feeling of privacy from other people at next home	continuous
49	nexthome5	Importance of having adequate room for parking two or more cars at next home	continuous
50	nexthome6	Importance of having a safe and enjoyable place to ride a bike at next home	continuous
51	nexthome7	Importance of being close to outdoor recreational areas at next home	continuous
52	nextneigh1	For me, living in a neighborhood where I could exercise by walking or biking would be…	continuous
53	nextneigh2	For me, having neighbors close by and making friends with neighbors would be…	continuous
54	nextneigh3	For me, to live within walking distance of a town or village center with basic stores would be…	continuous
55	nextneigh4	For me, to live in a place where it was easier to get to essential medical services would be…	continuous
56	nextneigh5	For me to live closer to my job, and drive less, would be…	continuous
57	nextneigh6	For me, to be able to take public transportation or carpool to work or for other trips would be…	continuous
5	nextneigh7	For me, to always have friends and relatives who can	continuous

8		take me places would be…	
59	nextneigh8	For my household to get along with fewer cars would be…	continuous
60	nextneigh9	For me, to live in less living space would be…	continuous
61	nextneigh10	For me, having access to places where people meet and gather would be…	continuous
62	nextneigh11	For me, the idea of moving away to a less rural state is…	continuous
63	nextneigh12	For me, to live with a smaller lot would be…	continuous

Note: variables with asterisks (*) are variables that are categorical with three to six categories. Reservations should be made when interpreting these variables.

1.5. Appendix A.5 Attitudinal variables

No	Variable name	Variable meaning	Variable type
1	otherconsid1	It would be hard for me to reduce my auto mileage and use of gasoline.	continuous
2	otherconsid2	I'd be willing to drive less to improve the environment and reduce my use of foreign oil.	continuous
3	otherconsid3	I love the freedom and independence that owning several cars provides for my household.	continuous
4	otherconsid4	I think I am wasting too much time driving.	continuous
5	otherconsid5	I think I should spend more time walking, just to be healthier.	continuous
6	otherconsid6	I need to drive my car to get where I need to go.	continuous
7	otherconsid7	I feel that biking is too dangerous.	continuous
8	otherconsid8	I really enjoy driving and don't want to reduce the amount I drive.	continuous
9	otherconsid9	I think there is more chance for economic advancement in a more urban state.	continuous
10	otherconsid10	I can solve most of the problems facing me if I invest the necessary effort.	continuous
11	otherconsid11	I can usually handle whatever comes my way.	continuous
12	otherconsid12	It is easy for me to stick to my aims and accomplish my goals.	continuous
13	otherconsid13	I feel there is not enough time to do what I have to do.	continuous

14	otherconsid 14	I think that people are fair, helpful and can be trusted.	continuous
15	otherconsid 15	Carbon emissions from driving my vehicle contribute to climate change.	continuous
16	otherconsid 16	As gas prices increase, I am more conscious of how many trips I take each day.	continuous
17	otherconsid 17	Sometimes I feel that I am trapped in this place and cannot move away.	continuous
18	Prefer*	If all else were equal, where would you prefer to live?	continuous
19	Prefer_rural	Whether respondents prefer rural dwelling	binary

Note: variables with asterisks (*) are variables that are categorical with three to six categories. Reservations should be made when interpreting these variables.