Using ArcGIS to Analyze the Effects of Community Characteristics on Physical Activity in Adolescent Girls

Adrian Overton and Scott Ashwood

Abstract:

Our paper and presentation will describe how we used ArcGIS to manage multiple sources data and to derive quantitative measures of neighborhood level characteristics for six study sites around the US. Some of the neighborhood characteristics include proximity to recreational facilities, street connectivity, population density, crime, mass transit, neighborhood SES, neighborhood terrain, and land use mix. The study will analyze community impacts on individual physical activity using multilevel modeling techniques combined with spatial auto-regressive methods.

Introduction

This paper will focus on a new approach to analyzing the influence that the built environment has on physical activity (PA) of middle school-aged girls. We will use ArcGIS 8.3 (GIS software) to operationalize the neighborhood contexts of the girls participating in a randomized controlled intervention aimed at curtailing the decline in physical activity among young females. The indicators of the built form of neighborhoods will be developed from spatial data sources that are readily available from national and/or local government sources. These neighborhood indicators of the built environment will be integrated with the measurements taken of physical activity levels of each girl participating in the intervention using a CSA (Computer Science and Applications, Inc.) accelerometer device. The CSA accelerometer is worn on the waist for seven days, and records the quantity of motion for every minute, to calculate the average daily minutes of intensity adjusted moderate-to-vigorous physical activity (MVPA). These measurements will be repeated several times over the study period of 3 years.

There have been several studies in the urban planning literature that look at the relationship between the built environment and travel behaviors that have health benefits such as walking and bicycling (Handy, et al. 2002), and a few in the public health literature that use an ecological approach to analyzing the influence the physical environment has on physical activity in individuals (Corti and Donovan, 2002; Sallis, et al. 1990). However, these studies tend to rely on survey instruments to obtain data on the level and extent of exercise and not on observational and empirical measure(s) of physical activity or the built environment. In addition, many studies in the urban planning literature tend to focus on the travel behavior of individuals or aggregates in going from origins to destinations, or in economic terms as a derived demand for travel. These studies and surveys do not capture the physical exertion of individuals or the fact that many walking and bicycling trips are made just for exercise or for simply "getting outdoors" for unstructured or unplanned activity. In other words, these studies assume that all travel demands are derived demands to partake in some activity such as work, shopping, appointments, etc. We do not attempt to provide an alternative framework to

the derived demand approach here, but the use of an objective measure of physical activity in adolescent girls, coupled with survey data, does present new possibilities that may suggest an alternative framework for studying the relationship between the built environment and physical activity of a non-working segment of the population.

Project Description

This project is an ancillary study to an intervention study funded by the National Heart Lung and Blood Institute (NHBLI) called the multi-centered Trial of Activity for Adolescent Girls (TAAG); which is a group (school)-randomized controlled intervention trial to increase PA among a cohort of sixth grade girls over a 3 year period. By focusing on empirically-based measures of the built environment of communities and the PA levels of adolescent girls who reside in six communities spread over the U.S., we hope to add to the debate surrounding the effect that the physical environment has on PA.

In this paper we will present our methodology for using an ArcGIS-based approach to determine whether the physical form of neighborhoods influences the likelihood and extent that young females residing in six U.S. cities will engage in PA. Our hypothesis is that the physical environment of communities will have an impact on the PA levels of adolescent girls. To determine these effects we need to develop a contextual database that allows us to consider the spatial locations of amenities, the physical form of neighborhoods, and socioeconomic status of neighborhoods in relation to the residential location of the girls. We do not track the actual spatial movement of the girls, but the level of physical exertion using an accelerometer device to be worn by each girl participating in the study, administered and monitored by the six TAAG field sites. A questionnaire will also be administered to ascertain the locations at which the girls were active.

Population surveys consistently demonstrate that youth activity levels fail to meet recommended guidelines for PA. Although PA participation is important for both boys and girls, girls are at higher risk for inactivity than boys according to the Centers for Disease Control (CDC). The importance of reversing the decline in activity among adolescents was clearly recognized in the Healthy People 2010 Objectives (DHHS November, 2000). While adolescents spend most of the hours of the week during the school year, and nearly all of those hours during vacation times in their neighborhoods, homes, or other environments, the role of these environments in influencing PA is not well understood at the current time. Neighborhoods are particularly relevant for youth, as they are unable to drive and for which activity is often limited to the immediate distance they are able to walk or ride a bike. For this reason, adolescents as a group are highly sensitive to their environments and represent an excellent subsection of the population through which to understand the influence of the physical environment on exercise levels.

How does the built environment impact physical activity?

The following diagram illustrates the TAAG intervention conceptual model of how the built environment may potentially influence PA in adolescent girls. The diagram in Figure 1 below presents a social-ecological model of the influence of the TAAG intervention and the theorized contextual effects of the social and physical environment of the neighborhoods in which the participating girls live.

Figure 1.
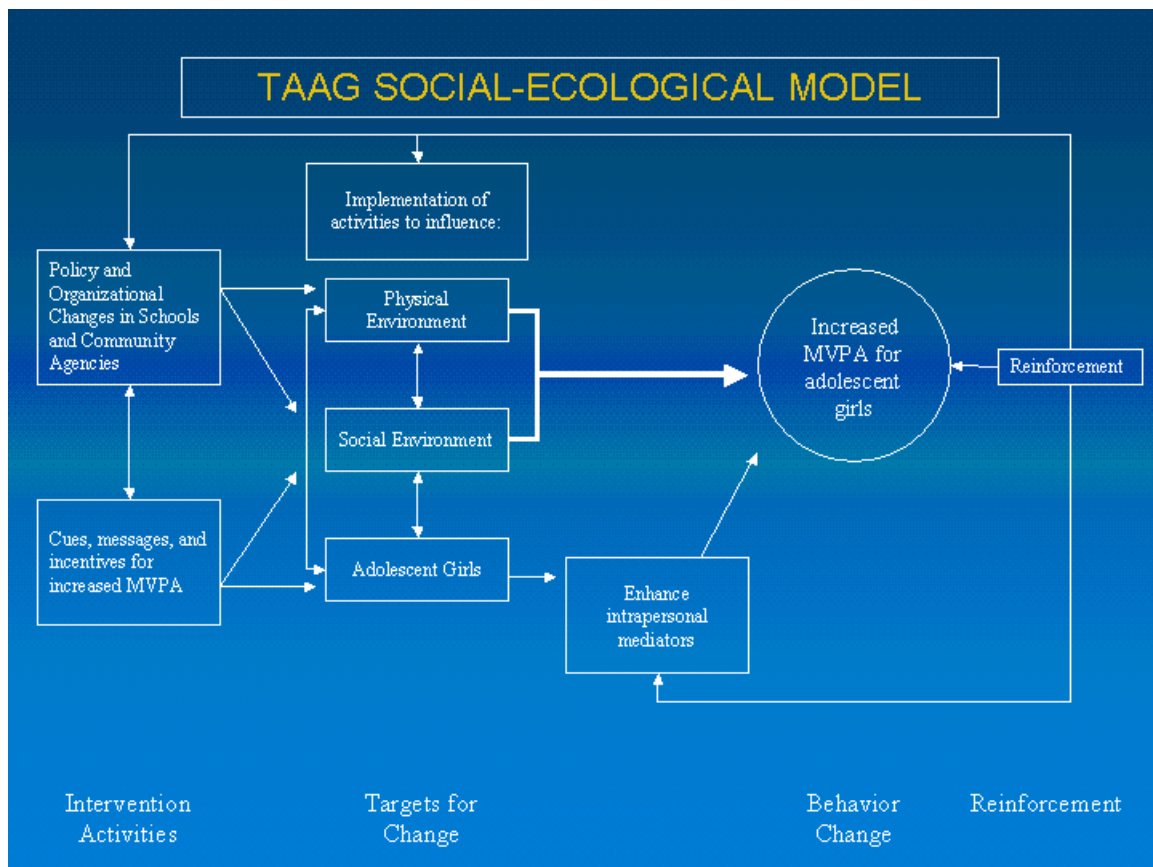
Source: TAAG Presentation, July 2002

This paper will focus on 1) the hypothesized relationship between the physical environment and increased MVPA for adolescent girls that the TAAG intervention is expected to demonstrate, and 2) how we plan to use ArcGIS to develop the contextual database for the study.

Conceptual GIS Database Design

To analyze the neighborhood level effects of the physical environment on PA, we need a model that captures the significant elements or features of the spatial context of neighborhoods. ArcGIS provides the capability to model the physical environment at the appropriate scale for assessing neighborhood effects on individuals. However, the problem of defining what a neighborhood is, and in a definition that satisfactorily captures all relevant features from the perspective of residents is very challenging given

existing data sources and other available resources. We will attempt to assess the effects of various neighborhood definitions by using several arbitrary definitions of neighborhood, such as census tracts/blockgroups, 1-, 2-, and 5-mile radius areas, and areas formed by the intersection of tertiary/residential streets or t-communities (Grannis, 1998). The ArcGIS database will be used to capture the hierarchical nesting of girls within neighborhoods in six study sites consisting of:

- Baltimore, Maryland metro area and areas in Montgomery County, Maryland
- The counties of Greenville, York, and Richland South Carolina
- Minneapolis-St. Paul, Minnesota metro area
- New Orleans, Louisiana metro area
- San Diego, California
- Tucson, Arizona

Each site has recruited 3 intervention and 3 control middle schools for a total of 36 schools. Each site has recruited approximately 120 sixth-grade girls per school and will monitor them over the next 3 years. An ArcGIS geodatabase will store and manage the data for each site. This section will briefly describe the data requirements, conceptual database design, database implementation, and data collection procedures for the study.

Data Requirements
The first step in designing the geodatabase was to identify 1) the geographic data needed to develop measures of the built environment and 2) potential data sources. The data requirements were driven by our conceptual model of how neighborhood characteristics might influence the PA of adolescent girls. Thus, the physical environment box depicted in Figure 1 is composed of the following aspects of a neighborhoods built environment:

- **Land use**
- **Street design**
- **Proximity to amenities such as parks, public transit, convenience stores**
- **Terrain, e.g. steep hills, etc.**
- **Climate**

The next step was to determine how to operationalize these aspects of the built environment as empirical and quantitative measures or indicators. We consulted several sources in the urban planning and design literature to identify measures that we thought were 1) substantive measures of urban form 2) detailed enough to capture the built and material qualities of neighborhoods, and 3) could be quantitatively measured (Song and Knaap (2002); Cervero and Kockelman (1997); Krizek (2003); Talen (2002); Handy et. al (2002); Allen (2001); Galster (2000); Ewing et al (2002); Greenwald and Boarnet (2001); Handy and Clifton (2000); Kockelman (1996)). Also, the measures would need to allow for analysis at different levels of spatial resolution to accommodate our need to assess various definitions of the neighborhood area. The operationalizing the physical environment section of this paper will briefly describe the empirical measures/indicators we selected to operationalize the built environment, and more importantly, how each measure theoretically relates to PA in our cohort of adolescent girls.

Based on the selected measures of the built environment, we developed a list of the geographic data layers required to compile the measures. These were data layers that we thought would be available from most state or local government sources. To allow for variation in the data availability across sites and within sites (some of our buffer-based neighborhood areas crossed county jurisdictional boundaries), we included measures that would be computed from federal sources such as Census TIGER/Line files when more detailed local data wasn't available.

The diagram in Figure 2 shows the data layers identified for the contextual database and the proposed hierarchical structure.

Figure 2
DRAFT
TAAG PA Study Contextual Database Design
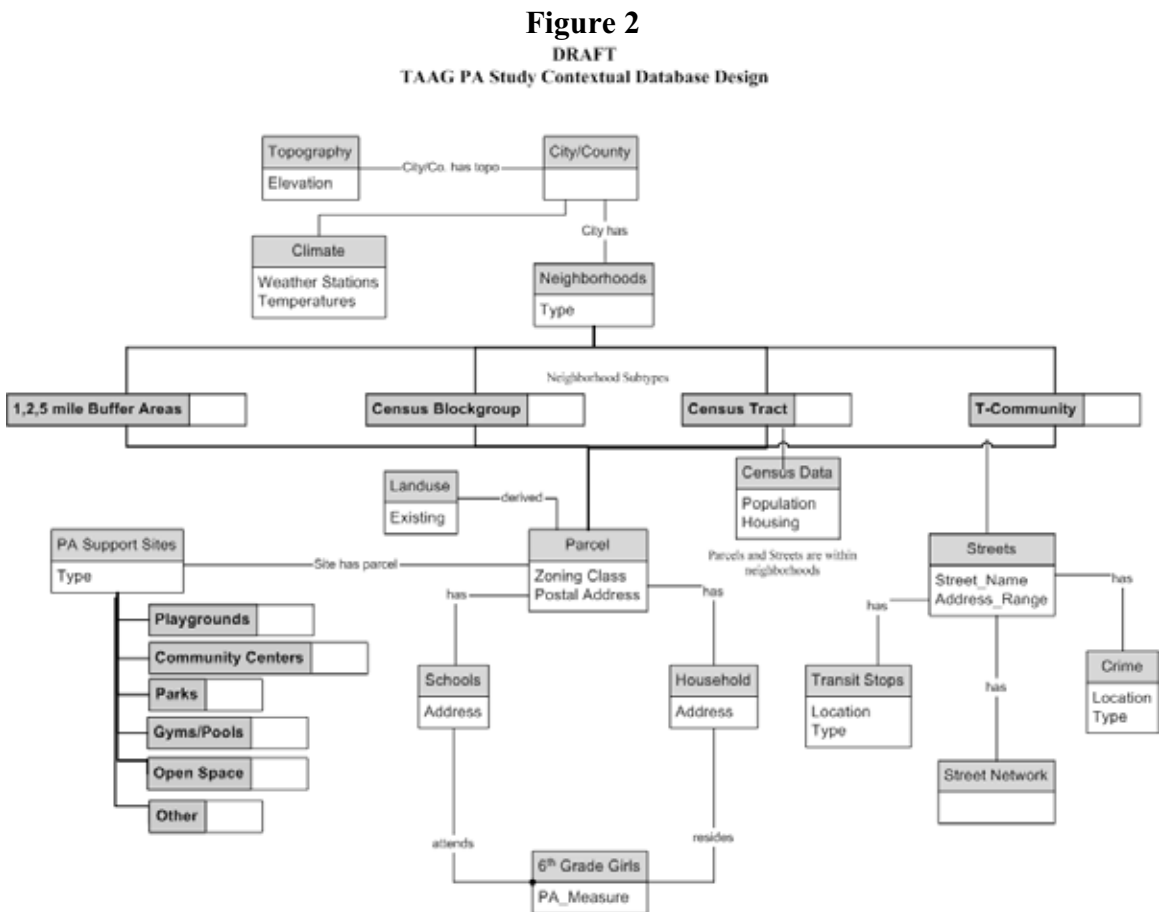
There are two base layers, parcels and street centerlines, for each site because of data availability issues. The parcel layer will serve as the base layer, and all definitions of neighborhoods where such data is available or can be easily created from existing paper or digital maps. To accommodate sites where parcel level data is not available, we will use the Census TIGER/Line street centerline layer to geocode addresses and neighborhood boundaries. The diagram in figure 2 shows the parcel layer as the base map. Since it is not likely that we will receive parcel level data from all, or a major

portion of our study areas, the streets layer will replace parcels as the base layer where this is the case.

Data Collection
We have begun collecting data from local sources and compiling the contextual database of layers. Specifically, the addresses of the participating middle schools and girls have been collected, and we have begun geocoding addresses of schools and girls to the 2000 TIGER/Line files and parcels where available. We have not received data from the CSA accelerometer devices as of this writing, but the baseline measurements for the study have been conducted. The girls were monitored for a one-week period using the CSA monitors and questionnaires were administered to ascertain the types and locations of activity.

Currently, we are developing our programming specifications of the indicators of the built environment. These programs will be implemented using VBA and ESRI's ArcObjects component-based GIS model. The programs are being developed to take into account various definitions of neighborhoods so that measurements can be taken at various scales and mapped for further analysis. Since all of the physical environment indicators will be compiled at the neighborhood level, the indicators will be stored as attributes of each neighborhood subtype for analyzing their impact upon the PA measures of individual study participants.

Operationalizing the Physical Environment
This section will describe how we plan to operationalize some of the characteristics of the built environment and their relevance to PA of adolescent girls. Interest in the influence of the physical environment on PA is a relatively new area of inquiry, and the evidence collected to date has been fairly sparse. Thus, it is unclear which specific features or characteristics of the built environment are likely to be important and how they influence PA. To describe how the indicators we will develop relate to PA, we will use a conceptual framework developed by Pikora et al., to explicate our characterization of the physical environment of neighborhoods that may influence PA. This framework was chosen based on the rigorous consensus-building methodology used to construct it. The framework was developed through the use of semi-structured interviews with experts from a cross-section of relevant disciplines, and a Delphi study with a diverse panel of recognized experts to identify and rank a set of indicators of the built environment based on their relevance to PA (Pikora et al, 2003). The conceptual framework developed by Pikora et al., identified four features of the physical environment that may influence PA at the local neighborhood level: functional, safety, aesthetic, and destination. These features are useful for our purposes because they are focused on features of the local neighborhood environment that will be captured in our study. The conceptual framework also provides a concise way to present the indicators we will develop for each neighborhood in the study.

According to Pikora et al. the *functional* feature in the framework relates to the physical attributes of streets and neighborhood structures primarily used for travel. These are built features that may influence the ability of girls to access PA support sites and thereby

influence the amount of PA. Also, the functional features of neighborhood streets can influence unstructured PA such as walking to friends' houses, stores and other local sites. In other words, the overall accessibility of destinations within and outside of the local neighborhood may influence individual decisions and behaviors toward outdoor activity. Thus, the indicators of functional features focus on the street network and other relevant pathways.

Figure 3

Neighborhood Physical Environmental Factors

Functional	Safety	Aesthetic	Destination
Route Directness to nearest PA Support Sites	Pedestrian Crossing Distance	Land Use Mix	Land Use Mix
Internal Street Connectivity	Proximity to Crime	Climate	Land Use Balance
External Street Connectivity	Average block size	Terrain/Hill Steepness	Park Adjacency
Street Network Density	Non-residential Street Intersections/Crossings	Park Space Supply	Park Proximity
Street Network Extent	Pedestrian Intersection Safety	Park Adjacency	Transit Adjacency to Housing
Pedestrian Network Coverage		Single Family Dwelling Density	Transit Proximity to Housing
Pedestrian Intersection Safety		Multi-Family Dwelling Density	PA Support Sites
Bicycle Network Coverage		Park and School Parcel Size	Jobs to Housing Balance
		Average Single Family Parcel Size	Amenities Proximity

The *safety* feature of the framework reflects the need to provide safe physical environments in which people can engage in structured or unstructured outdoor activities. The indicators of neighborhood safety reflect the need to quantify the potential for interaction with vehicle traffic. Thus, the indicators we will construct measure the width of street crossings within the neighborhoods; the size of neighborhood blocks where larger block sizes will mean longer travel distances away from home; and increased potential for interaction with automobile traffic. The measures of crime will be based on the location of reported crime occurrence by crime type and severity. Although it is known that perceptions of neighborhood crime may have a larger impact on individual and parental behavior than actual crime occurrence, for our purposes the proximity to crime locations provides a more quantifiable and objective characterization of neighborhood crime. Additionally, we will control for neighborhood crime perceptions of study participants using data collected from the follow-up surveys. We do not expect the safety characteristic of neighborhoods to have a direct influence on PA but rather to

provide individual cognitive cues and influences that may positively or negatively affect PA levels for unstructured activities such as walking and biking for exercise or for travel to an activity.

The *aesthetic* feature includes characteristics that may influence PA by providing cognitive cues and opportunities to be active in unstructured activities such as exploring nearby parks and bodies of water such as rivers, lakes, beaches, or streams. The land use mix measure is listed high on the list as a proxy for the presence of interesting or varying sights and architecture. Additionally measures, such as weather and presence of steep hills in the neighborhood, are likely to directly influence PA of the girls in our study in either a positive or negative direction.

The *destination* feature relates to the presence and availability of community and commercial facilities that would be significant to adolescent girls within or near the neighborhood area. Several transportation planning studies have shown that where there are appropriate local destinations, there is an increased chance that people will walk (Burden et al., 1999; Atash, 1994; Federal Highway Administration, 1993; Untermann, 1994). The presence of other people walking to various destinations in the neighborhood will provide social cues and incentives to walk or bike to nearby activities.

Data Analysis and Multi-level Model Development Plan
The goal of the analysis phase of the study is to determine the relative influence (not causal) of the physical environmental measure we will construct on the individual measures of PA of the adolescent girls participating in the TAAG intervention. To accommodate our objective of conducting an ecological analysis without the attendant problem of the so-called ecological fallacy, we plan to develop a multi-level model to analyze the influence that the variation in physical characteristics of neighborhoods will have on the variation in individual PA measures.

The multi-level model approach is appropriate for our study design because we wish to separate the effects of personal characteristics, (PA measure obtained from the CSA monitors and other immutable factors like race/ethnicity, age, household SES, etc.), from the contextual effects such as the indicators of the design and structure of neighborhoods in which the girls reside (Greenland 2001; Raudenbush and Bryk 2002). The multi-level model approach also avoids other common problems of ecologic regression such as the atomistic fallacy of modeling behavior solely at the individual level and excluding the context in which the behavior occurs (Greenland, 2001). Equivalently, modeling only at the aggregate or contextual level is prone to the ecological fallacy of making inferences about individual behavior that might not be applicable (Robinson, 1950). The multi-level model avoids both problems by combining an individual-level model representing disaggregate behavior while controlling for individual confounding factors with a macro-level model representing the relationship between PA and the variation in neighborhood contexts.

Before specifying the multi-level model, we will conduct an analysis to determine the extent of data redundancy resulting from spatial autocorrelation and the use of multiple

variables or attributes. Our use of several indicators to capture different aspects of the four features (functional, safety, aesthetic, and destination), is based on the use of discrete and arbitrary spatial definitions of neighborhoods. To describe a spatial process such as the physical design or form of neighborhood areas is problematic because it assumes that the physical characteristics of the neighborhood end and change form once the boundaries are reached. Like most spatial processes, this is an unrealistic assumption. There is very likely to be some spatial spillover or spatial autocorrelation latent in the indicators of neighborhood physical design. Our approach to dealing with this problem is to conduct a multivariate spatial autocorrelation analysis using GIS-based tools to determine the extent of the spatial spillover for each indicator and to adjust our analysis accordingly by computing spatially lagged variables for indicators exhibiting significant spatial autocorrelation.

Another concern is that, redundant information will be introduced into our analysis through the inclusion of multiple variables, as the attributes themselves are repetitious at least to some degree. For instance, our study design will use a relatively large number of indicators of the physical environment. To address the concerns that arise from this type of redundancy, we will use factor analysis techniques to reduce the number of variables under consideration in our model. Using factor analysis methods we can create four synthetic indicators representing the four features of the physical environment (function, safety, aesthetic, and destination) that are believed to have an influence on PA. The result of the factor analysis will be four indicators constructed from the indicators of each feature listed in Figure 3.

Conclusions
Our approach to analyzing the influence of the physical design and form of neighborhoods on PA would not be possible without the current tool sets and techniques provided by GIS. The presentation to accompany this paper will demonstrate specifically how we are applying ArcGIS to the management and development of a neighborhood level contextual database, and some of the issues involved in it's construction, for researching an important public health issue. A future paper will present the results of our analysis for the baseline period. In conclusion, the analysis is expected to provide a new source of objective analysis and insight to the current debate surrounding the issue of the role of the design/structure of the built environment in providing support for, or creating barriers to, PA at the neighborhood level. We hope that this information will be of use to policy-makers, planners, public health analysts, and citizens in making decisions and possibly considering new policy instruments or actions that could be beneficial to the public health and well-being of our communities.

References

Allen, E. 2001. INDEX: Software for community indicators. In Brail, R. K. and Klosterman, R. E. (Eds.), *Planning Support Systems: Integrating geographic systems, models, and visualization tools* (pp. 229-261). Redlands, CA: ESRI Press.

Atash, F. 1994. Redesigning suburbia for walking and transit: Emerging concepts. *Journal of Urban Planning and Development*, 120, 48-57.

Burden, D., Wallwork, M., Sides, K., Trias, R., & Rue, H. 1999. *Street design guidelines for healthy neighborhoods*. Center for Liveable Communities, Sacramento, CA.

Cervero, R., and Kara Kockelman. 1997. Travel demand and the three Ds: density, diversity, and design. *Transportation Research, Part D* 2 (2):199-219.

Corti, B., Donovan, R., and Holman, C. 1996. Factors influencing the use of physical activity facilities: results from qualitative research. *Health Promotion Journal of Australia*. 6:16-21.

Ewing, R., Pendall, R., and Chen, D. 2002. *Measuring Sprawl and Its Impact*. Smart Growth America. www.smartgrowthamerica.org.

Federal Highway Administration 1993. *The FHWA National Bicycling and Walking Study case study No. 4: Measures to overcome impediments to bicycling and walking.* Federal Highway Administration, Washington, DC.

Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S. and Freihage, J. 2001. Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing Policy Debate* Vol. 12, Issue 4, pp. 681 – 717.

Grannis, Rick. 1998. The Importance of Trivial Streets: Residential Streets and Residental Segregation. *American Journal of Sociology* 103(6):1530-64.

Greenland, Sander. 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology*, 30:1343-1350.

Greenwald, M. J. and Marlon G. Boarnet. 2001. The Built Environment as a Determinant of Walking Behavior: Analyzing Non-Work Pedestrian travel in Portland, Oregon. Irvine, CA: Institute of Transportation Studies at Univ. of California, Irvine. Working paper UCI-ITS-WP-01-1, July 2001.

Griffith, D. A. and Amrhein, C. G. 1997. *Multivariate Statistical Analysis for Geographers.* Upper Saddle River, NJ: Prentice Hall.

Handy, Susan and Kelly Clifton. 2000. Evaluating Neighborhood Accessibility: Issues and Methods Using Geographic Information Systems. Austin, TX: Center for Transportation Research. Report # SWUTC/00/167202-1. November 2000.

Handy, S., Boarnet, M. G., Ewing, R., Killingsworth. 2002. How the Built Environment Affects Physical Activity: Views from Urban Planning. *American Journal of Preventive Medicine* 23(2S):64-73.

Kockelman, K. M. 1996. Travel Behavior as a Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from the San Francisco Bay Area. In *City and Regional Planning*. Berkeley: University of California Press.

Krizek, Kevin. 2003. Operationalizing neighborhood accessibility for land use-travel behavior research and regional modeling. *Journal of Planning Education and Research* 22:270-287.

Pikora, T., Corti, B., Bull, F., Jamrozik, K. and Donovan, R. 2003. Developing a framework for assessment of the environmental determinants of walking and cycling. *Social Science & Medicine*. 56:1694-1703.

Raudenbush, S. W. and Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd Edition*. Thousand Oaks, CA: Sage Publications.

Sallis, J. et al. 1990. Distance between homes and exercise facilities related to frequency of exercise among San Diego residents. *Public Health Reports*., 105:179-186.

Song, Yan and Gerrit-Jan Knaap. 2002. Measuring Urban Form: Is Portland Winning the War on Sprawl? National Center for Smart Growth Research and Education, University of Maryland. A Working Paper.

Talen, Emily 2002. "Measurement Issues in Smart Growth Research." Paper presented at the Smart Growth and New Urbanism Conference, University of Maryland. May 4, 2002.

Untermann, R. 1984. *Accomodating the Pedestrian: Adapting Towns and Neighbourhoods for Walking and Bicycling*. New York: Van Nostrand Reinhold.