Neighborhood Characteristics and Physical Activity

J. Scott Ashwood and Adrian Overton

Abstract
In this paper we will discuss the methods we have used to create a set of neighborhood level measures and how we have used them in our analysis of physical activity (PA) in adolescent girls.

This research is an ancillary study to a multi-centered group (school) randomized controlled intervention trial, called the Trial of Activity for Adolescent Girls (TAAG), that is designed to increase PA among sixth grade girls over a 3 year period.

We will describe the neighborhood level variables that we created using ArcGIS. We have created and used street network measures, measures of access to facilities that support PA (parks, gyms, dance studios, etc.), land use measures, and climate measures. In addition to explaining the measures produced, the paper will include details of how we used ArcGIS to explore how the various measures of neighborhood characteristics were sensitive to scale effects and various neighborhood configurations.

Introduction
We used ArcGIS software to build several neighborhood level measures that we theorized are associated with physical activity of adolescent girls. We did this as part of our ancillary study of the Trial of Activity for Adolescent Girls (TAAG). TAAG is a nationwide, multi-site intervention in middle schools aimed at increasing the physical activity of girls and is funded by the National Heart, Lung, and Blood Institute of the National Institutes of Health. The main TAAG study is evaluating that intervention. Our ancillary study is looking at whether there are neighborhood level influences on physical activity.

At the 2003 ESRI International User Conference we presented our plans for building the various neighborhood measures (Overton and Ashwood, 2003). To study neighborhood effects, we created geodatabases containing GIS derived measures for the neighborhoods around the participating schools as well as around the individual girls' addresses. We created these measures from data that we obtained from the sites using online resources as well as field observation. We investigated several definitions of neighborhood before settling on a definition of neighborhood centered on the girls' addresses.

In this paper, we will describe the data collection process as well as how we created our measures. In addition, we will discuss the different neighborhood definitions we explored.

Data
Treatment and control schools were selected in six sites around the country. Participants at each school were randomly selected to participate in the study (Stevens, et al., 2005). These participants wore Actigraph accelerometers (Manufacturing Technologies Inc.

Health Systems, Model 7164, Shalimar, FL) for six consecutive days during the winter and spring of 2003. Measures of physical activity from these accelerometers served as our primary outcome measure.

At each site, a partner institution assisted with the intervention and with the data collection. The six sites and institutions are: Baltimore/DC (University of Maryland), Columbia, SC (University of South Carolina), Minneapolis, MN (University of Minnesota), New Orleans, LA (Tulane University), San Diego, CA (San Diego State University), and Tucson, AZ (University of Arizona). The University of North Carolina at Chapel Hill serves as the coordinating center. The RAND Corporation has coordinated the ancillary, neighborhoods study.

We collected neighborhood level data for each of the six sites with the assistance of the local institutions and built geodatabases for each site using ArcGIS software. We used US Census Bureau TIGER files for our street network data and census defined boundaries. Much of the data we collected were address level data. We collected data on businesses that support physical activity, parks, schools, food establishments, and crime. We geo-coded the addresses to the TIGER street network data. We also collected locally available feature layers when available. For example, most of our parks location data comes from local, county or city level boundary files. We also have public transportation routes for some sites. In some cases we hand rectified the local data to match our TIGER street network data in order to calculate distances along the street network.

We created several neighborhood feature layers based on different definitions of neighborhood. In general, the neighborhoods fell into two categories: census-based and address based. The census-based neighborhoods we looked at were the census tract and the census block group. In addition, we created 6 neighborhood definitions centered on the girls' addresses and including everything within a fixed distance from the address. We created ¼, ½, and 1-mile circular buffers around each address and created ¼, ½, and 1-mile zones based on distance along the street network from each address.

Once we created the neighborhood boundaries, we then generated several measures from our geo-coded data. We created tools in VBA that used ESRI ArcObjects to combine the information in our database and generate the neighborhood level measures. These included street network measures, neighborhood population measures, and accessibility measures for certain features.

The street network measures we created are based on the number and characteristics of the street segments and intersections that fall within the neighborhood boundaries. We created alpha, beta, and gamma measures that are all based on the ratio of street segments to intersections. These ratio measures give us some idea of how connected the street network is. We also calculated block size and length measures that tell how far people must go to get to intersections.

The neighborhood population measures we have generated are all based on the 2000 US Census and include population density, percent of population that is African American, the percent that is non-white Hispanic, the percent of households with income below poverty, the percent of adults with less than a high school education, and the percent of the workforce that is unemployed. These measures are easy to calculate for the census based neighborhoods and we simply used the values supplied by the US Census Bureau. To get the values of each of these for the address-based neighborhoods, we used area-weighted values from the block group. In those cases where the block group boundary crosses the neighborhood boundary, we took the percentage of the total block group area that falls within the neighborhood boundary and weighted the total population value for the block group by that percentage. For example, Figure 1 shows a single address based neighborhood (in blue) that contains part or all of twelve census block groups (outlined in red). Fifteen percent of the area in block group A is within the address based neighborhood. If the total population in block group A is 200, then we add 15% of that, or 30, to the total population for the address-based neighborhood.

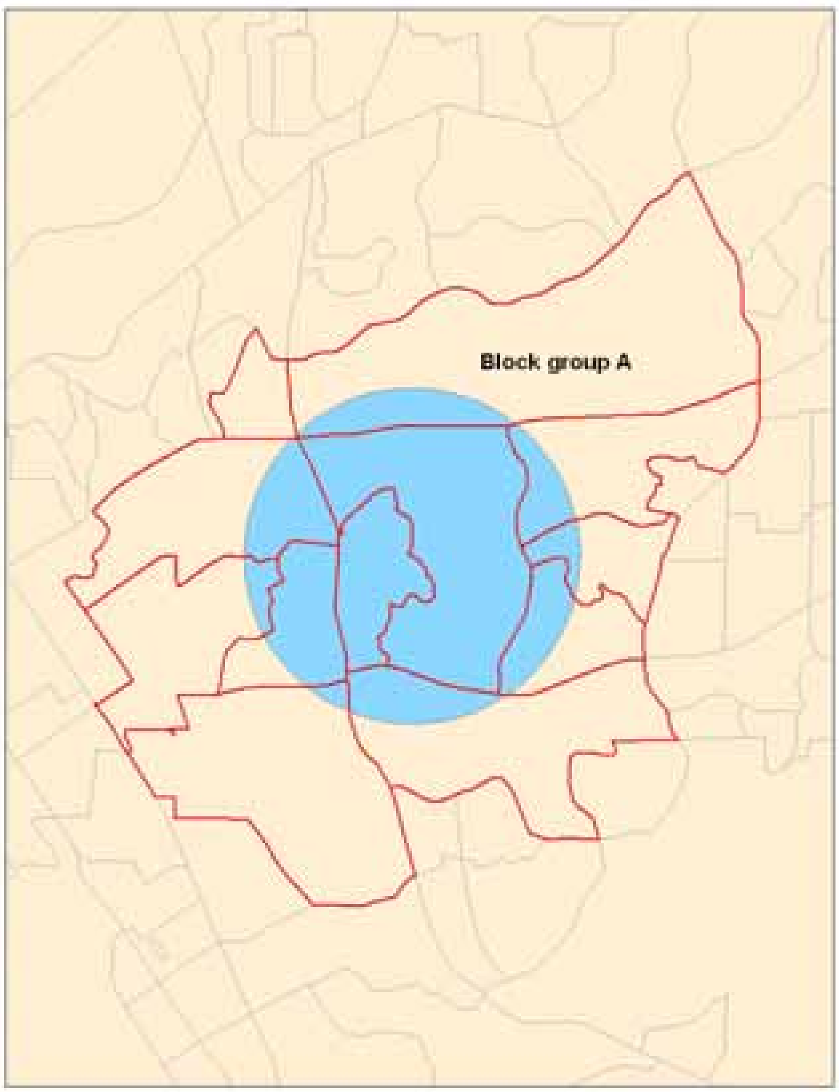

Figure 1: Census block groups that intersect address-based neighborhood

We created two types of accessibility measures: distance-weighted counts and non-weighted counts. To create the non-weighted measures, we simply count up the number of features that fall within the neighborhood boundary. For example, Figure 2 shows businesses that support physical activity near an address. There are 48 businesses within the 1-mile buffer around that address. This is the non-weighted count for the neighborhood.

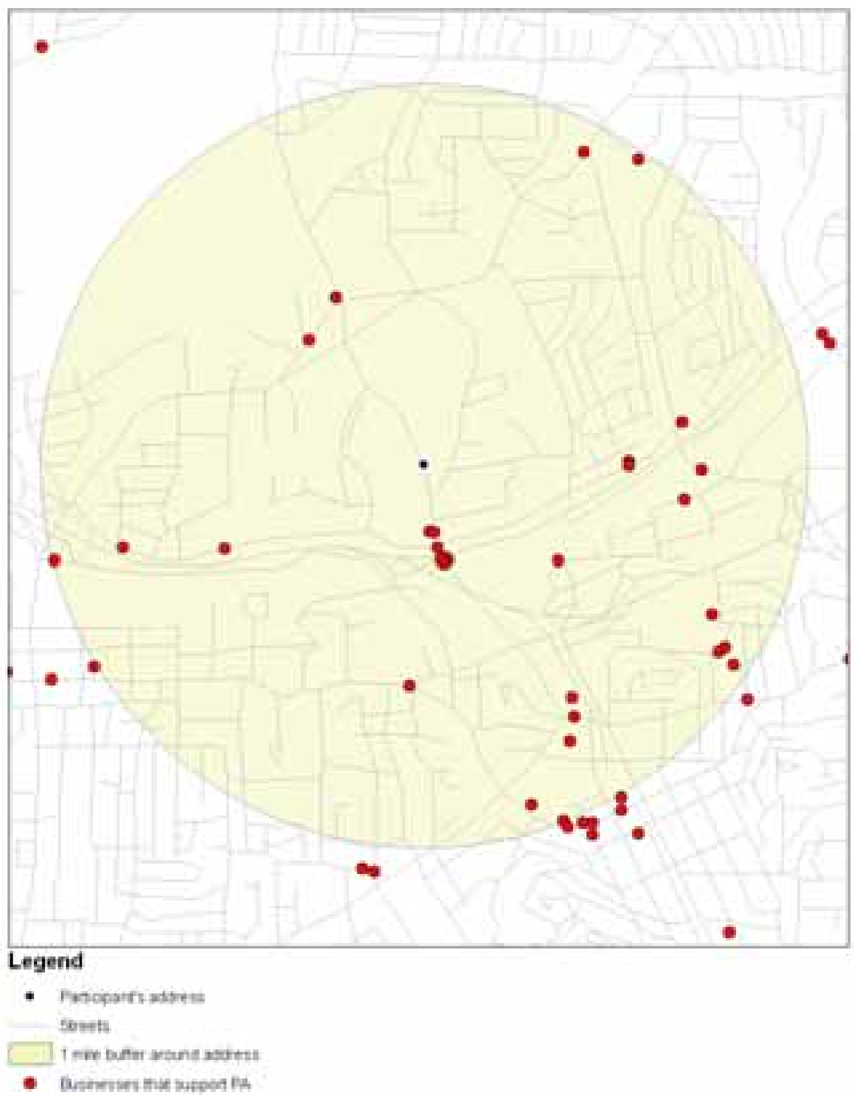

Figure 2: Businesses that support physical activity near a single address

By using the non-weighted count, we are assuming that girls in the study do not care how far away a destination is. They are as likely to go to a martial arts studio that is 1 mile away as they are to go to one that is ¼ mile away. We also created measures that take into account the notion that girls in the study are more likely to choose destinations that are closer to home over places that are farther away. To create the distance weighted counts we calculated the distance from each feature of interest to the girls' addresses, for the address based measures, and the centroid, for the census based neighborhoods. We then penalized, or weighted, each feature for its distance and added the weighted value to the sum for that neighborhood. We explored two weighting functions, negative exponential and negative Gaussian. Figure 3 shows a graph of each of these types of weighting functions. Both functions assign a value of 0.5 to a feature that is 1 mile away. The weight of 0.5 is equivalent to saying that girls are half as likely to go to a destination that is 1 mile away as they are to going next door. The negative Gaussian function does not drop as quickly, giving features that are close by more weight than the negative exponential function. The weighted counts of businesses within 1 mile for the address in Figure 2 are 39.43 for the negative exponential function and 44.42 for the negative Gaussian function.

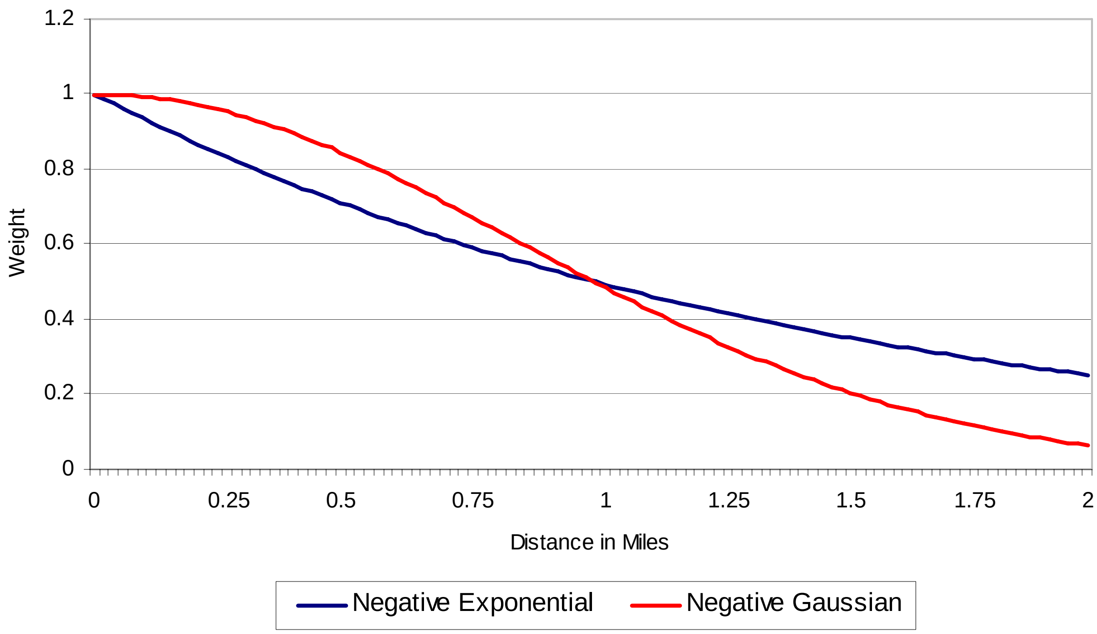

Figure 3: Comparison of negative exponential and negative Gaussian distance weighting functions

We created weighted and unweighted accessibility measures for parks, businesses that support physical activity, and fast food restaurants. In addition to the number of parks and businesses, we also created counts of the types of parks and businesses, and we created counts of the types of facilities within parks. We identified businesses at each of our sites by using a combination of online yellow pages and an InfoUSA database. We identified a set of business types that support physical activity and a list of zip codes that covered the areas. We then pulled the addresses and phone numbers for each of the businesses that were identified by the category/zip combination from each of our sources. Businesses that appeared in both sources were considered valid. We called businesses that appeared in only one source and verified that the business existed, was located where the address placed it, and was the type of business indicated by the yellow pages or InfoUSA category (SIC code).

Classification of parks as well as identifying the types of facilities within the parks involved physically surveying them. We used the parks classification proposed by Mertes and Hall (Mertes and Hall, 1996) to divide our parks into categories. The classification includes the following types: mini-park, neighborhood park, school park, community park, large urban park, sports complex, natural resource area, greenway, special use, and private park/recreation facility. In addition to an overall classification, we surveyed the parks for the presence or absence of several types of facilities. We looked for types of fields and play areas that are available as well as amenities like drinking fountains and lights.

Discussion

We are still in the process of analyzing the relationship between our neighborhood measures and physical activity. So far, it appears that there is a relationship among some of our measures. There are two studies that are either in print or forthcoming as of the writing of this paper. In one study, we identified a relationship between the distance to school and the amount of non-school physical activity (Cohen, et al., 2006). The farther a girl lives from school, the lower the amount of non-school physical activity she has in a week. We have also identified a relationship between number of and access to parks and physical activity (Cohen, et al., forthcoming). In that study, we found that the more parks, closer to home, the higher the amount of non-school physical activity. Furthermore, the types of park are important as well.

Early in our data collection and analysis process we decide to focus on an address-based neighborhood definition. We were able to do this because most of our data is address based or at least we have the coordinates for most of our places. We also wanted a small area of focus in order to facilitate data collection. The address based neighborhoods made sense to us theoretically as well since we are predicting individual level behavior. Using a neighborhood definition and therefore neighborhood measures that are individual level is consistent with our outcomes of interest. If we had to rely on aggregated data instead, then we may have had to choose another neighborhood definition, like census tract.

The choice of a neighborhood definition can be important. The way in which spatial data is aggregated into neighborhoods can affect the values of the neighborhood measures. This is known as the Modifiable Areal Unit Problem (Openshaw, 1984). In general we found that the choice of neighborhood can have an effect on the values for each of our measures.

Table 1 provides a summary of the measures discussed above by neighborhood type. The numbers in Table 1 represent the values for the neighborhoods of the girls who participated in the TAAG study. Girls are assigned to census-based neighborhoods based on where their address places them. In addition to a summary of the measures themselves, Table 1 contains the Pearson Correlation Coefficient estimates for the correlation between each measure and the total hours of non-school physical activity for the week.

	Census Tract				1 Mile Circular Neighborhood			
Measure	**Mean**	**Median**	**Std Dev**	**Corr. Coeff.**	**Mean**	**Median**	**Std Dev**	**Corr. Coeff.**
Gamma Measure	0.44	0.43	0.05	-0.02	0.45	0.45	0.05	-0.02
Street Density	13.87	13.41	6.81	0.04	11.1	10.78	4.37	0.05*
4-way Intersections (%)	0.23	0.2	0.15	0.03	0.23	0.21	0.13	0.03
% Adults Unemployed	0.02	0.02	0.01	0.00	0.03	0.02	0.01	-0.01
% Population Hispanic	0.06	0.03	0.08	-0.02	0.06	0.03	0.08	-0.01
Population Density	3720.35	2477.23	3506.32	0.01	3270.65	2681.13	2452.05	0.01
Total Parks	2.08	1	2.76	-0.01	3.52	2	3.77	0.02

Table 1: Comparison of measures by neighborhood definition

In general the selected measures and correlation coefficients are similar, but there are a few differences. Street density and population density are higher on average for census tracts, while the count of total parks is lower on average. Only one correlation is significant at the 0.10 level, and that is for street density in the 1 mile circular neighborhood. The relationship between parks and physical activity is estimated to be negative for the census tract measure and positive for the address based measure, though neither coefficient estimate is significant. These results suggest that the choice of neighborhood can matter in both describing an area and in analyzing relationships.

Conclusion

We have been able to create neighborhood measures and we have demonstrated that some of these measures are related to physical activity in adolescent girls. The data collection process was a long one, but it yielded data at the address level. This level of data allowed us to create individual level neighborhoods. We did not have to rely on arbitrary neighborhood definitions, which may have changed our ability to identify significant relationships.

References

Cohen D., Ashwood S., Scott M., et al. 2006. Proximity to School and Physical Activity Among Middle School Girls: The Trial of Activity for Adolescent Girls Study. Journal of Physical Activity and Health; 3, Suppl1: S129-138.

Cohen D., Ashwood J.S., Scott M., et al. Forthcoming. Public Parks and Physical Activity Among Adolescent Girls. Pediatrics.

Mertes J, Hall J. 1996. *Park, Recreation, Open Space and Greenway Guidelines*. Ashburn, VA: National Recreation and Park Association.

Openshaw S. 1984. *TheModifiable Areal Unit Problem*. Concepts and Techniques in Modern Geography 38. Norwich, United Kingdom: GeoBooks.

Overton A., Ashwood S. 2003. Using ArcGIS to Analyze the Effects of Community Characteristics on Physical Activity in Adolescent Girls. 23rd Annual ESRI International User Convention: Professional Paper.

Stevens J, Murray D, Catellier D, et al. 2005. Design of the Trial of Activity in Adolescent Girls (TAAG). Contemporary Clinical Trials;26:223-233