



SHORT PAPER

Open Access

# Neighborhood sampling: how many streets must an auditor walk?

Tracy E McMillan<sup>1</sup>, Catherine Cubbin<sup>2</sup>, Barbara Parmenter<sup>3</sup>, Ashley V Medina<sup>4</sup>, Rebecca E Lee<sup>4\*</sup>

## Abstract

This study tested the representativeness of four street segment sampling protocols using the Pedestrian Environment Data Scan (PEDS) in eleven neighborhoods surrounding public housing developments in Houston, TX. The following four street segment sampling protocols were used (1) all segments, both residential and arterial, contained within the 400 meter radius buffer from the center point of the housing development (the core) were compared with all segments contained between the 400 meter radius buffer and the 800 meter radius buffer (the ring); all residential segments in the core were compared with (2) 75% (3) 50% and (4) 25% samples of randomly selected residential street segments in the core. Analyses were conducted on five key variables: sidewalk presence; ratings of attractiveness and safety for walking; connectivity; and number of traffic lanes. Some differences were found when comparing all street segments, both residential and arterial, in the core to the ring. Findings suggested that sampling 25% of residential street segments within the 400 m radius of a residence sufficiently represents the pedestrian built environment. Conclusions support more cost effective environmental data collection for physical activity research.

## Findings

Neighborhood context has been associated with health and physical activity (PA) [1-13]. Studies of specific neighborhood characteristics, including pedestrian pathways, reduced automobile traffic, and aesthetic appeal, and their association with PA [14], have yielded rigorous instrument development, validation and implementation to increase understanding of the role the built environment plays in PA [15]. Nevertheless, there remain unresolved issues concerning specific sampling and data collection protocols that have implications for future research, promotion and policy.

Although some municipalities collect and compile GIS data about the built environment that aid PA research, most do not. Very few have detailed data such as sidewalk condition or pathway obstructions, and the quality and consistency of the GIS data vary widely [16]. Many environmental audit instruments have been developed to address these limitations. Four frequently used instruments include the Systematic Pedestrian and Cycling Environmental Scan (SPACES) [17]; the Irvine

Minnesota Inventory (I-M) [18]; the Analytic Audit Tool and Checklist Audit Tool (SLU) [19]; and the Pedestrian Environment Data Scan (PEDS) [20]. Each has adequate reliability and provides a rich assortment of micro-scale environment data. The principal limitation to these instruments is the time and cost involved in data collection. PEDS has the lowest data collection time of these audit instruments [20], averaging 3-5 minutes per segment, compared to 10 to 20 minutes for the I-M and SLU. A full inventory of street segments within a quarter mile radius of a selected address can exceed 100 segments – 17-34 hours for data collection per neighborhood.

Environmental audits on the complete census of streets in a neighborhood may be unnecessary, as there is likely substantial homogeneity within street types in a neighborhood, particularly residential streets. Pikora and colleagues noted that the lack of variation of built environment characteristics among assessed segments “resulted in skewed distributions of responses to some items,” which led to a high level of chance agreement and low kappa scores for these items [17]. To address this, Agrawal, Schlossberg, and Irvin suggest that data collection be both streamlined and customized within street types [21]. For example, characteristics such as

\* Correspondence: releephd@yahoo.com

<sup>4</sup>Texas Obesity Research Center, Department of Health and Human Performance, University of Houston, 3855 Holman St, Garrison Gymnasium Rm 104, Houston, TX 77004, USA

sidewalk, traffic volumes, posted speeds, number of lanes of traffic and safe crossings will vary more significantly across a sample of arterial streets than of residential streets. Therefore, although a full sample of arterials may be necessary to capture this variability, a reduced sample of residential street segments may be possible due to less variation from one street segment to the next.

This study tested four different street segment sampling protocols using a gently modified version of the PEDS to determine whether abbreviated data collection protocols can sufficiently and accurately represent the pedestrian built environment. Analyses were conducted to determine whether (1) the immediately proximal neighborhood (within 400 meters of the residence) is notably different from nearby neighborhoods (by adding an additional 400 meters to the radius) that might impact daily physical activity and (2) it is necessary to sample the complete census of all residential street segments within the 400 meter radius buffer. Our hypothesis in both cases was that no variation in the tested characteristics would exist, thereby supporting the scientific integrity of sampling strategies using abbreviated data collection.

Eleven neighborhoods in the City of Houston were selected for this study. Neighborhoods were defined as the area within an 800 meter radius circumscribed around a public housing development managed by the Houston Housing Authority in Houston, Texas [Eugeni, Baxter, Lee: Disconnections of African American public housing residents: Connections to Physical Activity, Dietary Habits and Obesity, submitted; [22]]. Housing developments are affordable rental housing for families, seniors, and persons with disabilities, federally-funded and managed by the Houston Housing Authority. Defining the neighborhood as the area within the boundaries of the circle has several advantages [23,24]. First, it captures all areas to which a resident may be exposed on a daily basis during both foot and automobile travels. Second, the straight line distance allows for capture of distance traveled on footpaths and other "short cut" routes that may not be captured by using a street network or aerial satellite photography strategy. Third, it may reduce the effect of spatial correlation that arises from using census boundaries where points near the boundary of the census area are influenced by factors in adjacent census areas, as housing developments were selected to be at least 1600 meters apart. All housing development neighborhoods were located in urban areas that were predominantly lower income, with higher proportions of ethnic minorities.

The neighborhoods varied by socioeconomic status (SES) index, racial ethnic concentration and street node density. To characterize neighborhood-level SES, we

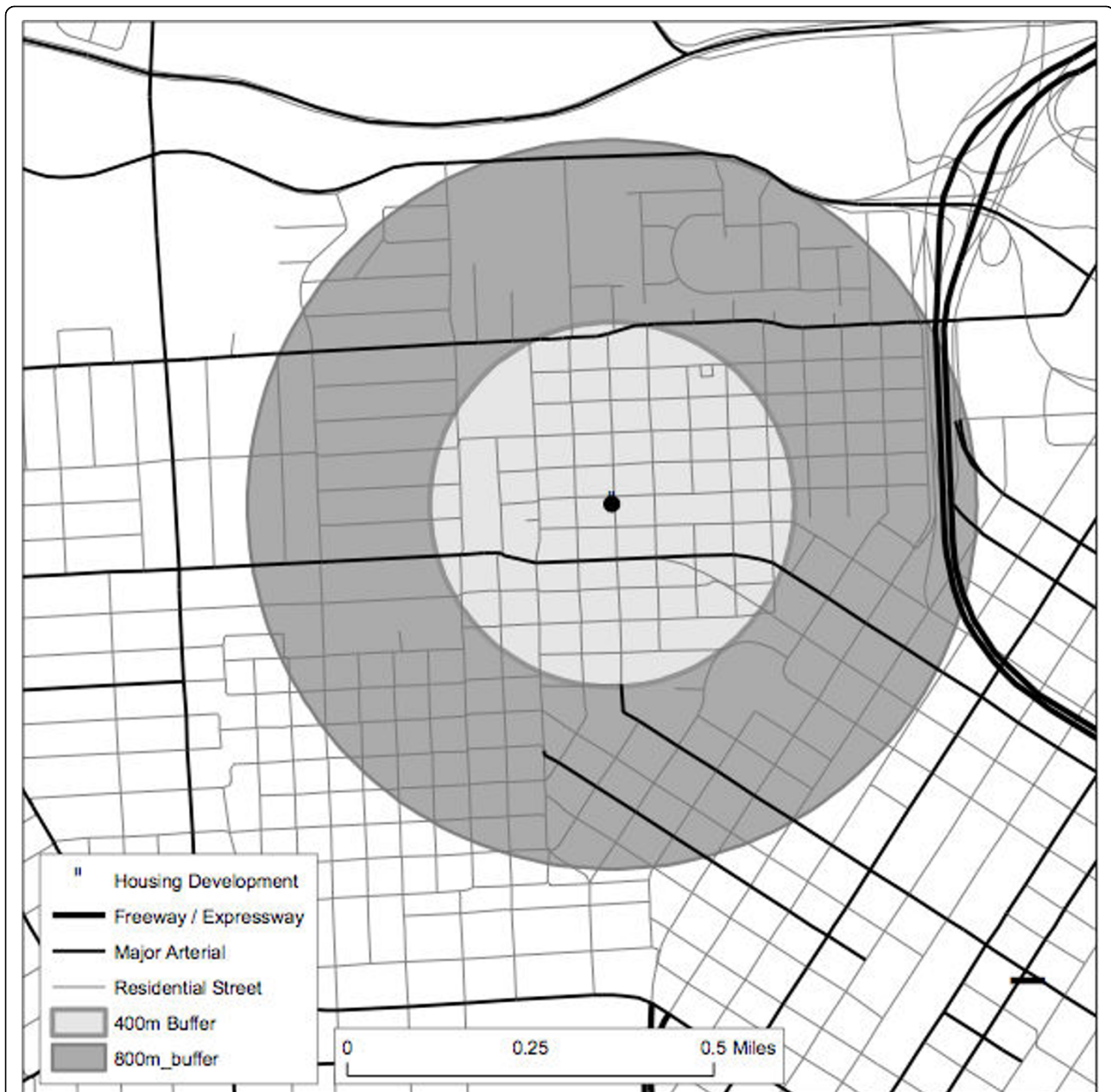
constructed an index based on one of the author's (CC) previous work [25]. Five variables from U.S. Census block-group level data from the year 2000 for Harris County were standardized and then summed with equal weights to compute the index: percentage aged 25 and older with less than a high school education, median annual family income, percentage blue collar workers, percentage unemployed, and median housing value. Correlations among the five variables ranged from 0.35 to 0.89 and principal components analysis revealed that the five variables explained 69% of the total variance. Harris County and HD neighborhood socio-demographic characteristics are presented in Additional File 1: Table S1 [Harris County and HD Neighborhood socio-demographic characteristics]. All street segments in each neighborhood were identified using ArcVIEW. Prior to going into the field, neighborhoods were mapped using GIS technology. Street segments were numbered in preparation for field assessment and assessed using the PEDS [20]. This instrument assesses a number of street characteristics associated with physical activity, in particular, pedestrian activities and bicycling. Trained field assessors were deployed to neighborhoods in teams of two following established safety protocols [24]. Each segment was carefully assessed and rated using the operational definitions from the PEDS instrument. Since a full sample of street segments was initially collected for this project, street segments were not identified as residential or arterial prior to conducting data collection. The road type of the street segment was recorded at the time of the assessment. Residential streets were defined as moderate to low volume roads that carry less than 5,000 cars per 24 hour period. Arterial streets were defined as high volume, main roads that carry approximately 5,000-10,000 cars per 24 hour period. On and off ramps were not included in the sample of residential street segments. However, it is important to note that although the functional classification of roadways (e.g. arterial vs. residential) may be available in GIS datasets obtained from municipalities, this should be verified in the field prior to sampling procedures and before completing environmental audits.

Analyses were conducted on five key variables associated with walking: sidewalk presence; observer ratings of attractiveness and safety for walking; connectivity; and number of traffic lanes (proxy for speed/volume). HD neighborhood pedestrian built environment characteristics are described in Additional File 2: Table S2 [HD Neighborhood pedestrian built environment characteristics]. The rating of attractiveness pertained to finding the area aesthetically pleasing and to the existence of destinations. It answered the question: "would you want to walk/bike this segment?" The rating of safety for walking took

into consideration not only walking along the sidewalk but crossing the street. "Would a child be safe walking the segment?" Response to safety for cycling considered road attributes such as speed limits and presence of bicycle facilities. All data were collected, entered into an Access database and proofed by two research team members using established protocols [22,24].

Chi square analyses were used to compare similarity in the five key variables for different geographic areas

and different sampling strategies. First, we compared all segments contained within the 400 meter radius buffer from the center point of the housing development (the core) with the segments contained between the 400 meter radius buffer and the 800 meter radius buffer (the ring) as illustrated in Figure 1. Second, we compared all residential segments in the core with three different percentages of randomly selected residential street segments in the core: 75%, 50% and 25%—to



**Figure 1 Map of Housing Development Neighborhood.** Figure 1 graphically displays the geographic comparisons between all segments contained within the 400 meter radius buffer from the center point of the housing development (the core) with the segments contained between the 400 meter radius buffer and the 800 meter radius buffer (the ring).

determine whether a sample of segments fewer than 100% would still be representative of the pedestrian environment. Analyses were conducted for all neighborhoods combined and separately. Figure 1 graphically displays the geographic comparisons between all roads within the 400 m and 800 m buffers.

As presented in Additional File 3: Table S3 [All HD and HD Neighborhood core vs. ring comparison of pedestrian built environment characteristics], a few significant differences were seen between the core and the ring. There was significant variability in sidewalk presence in two neighborhoods, connectivity varied in four neighborhoods, and traffic lanes varied in three. Attractiveness for walking varied significantly between core and ring in three neighborhoods and feelings of walking safety significantly varied in four neighborhoods. General patterns indicated that ratings of attractiveness and safety were lower in the core than in the ring.

There were no significant differences in the comparison of the residential street segments in the core compared to a random sample of residential segments (75%, 50% and 25%) as presented in Additional File 4: Table S4 [Comparisons of core vs. random sample of core residential segments (75%, 50% and 25%)].

The goal of this study was to determine the extent of sampling necessary to provide a representative sample of the built environment in residential neighborhoods. Some differences were found when comparing the core to the ring, likely due to the increased variability of arterials in this broader street sample covering a larger geographic area. No differences existed when residential segment samples of 75%, 50% and 25% were compared to the census of all residential streets in the core. These findings suggest that sampling as few as 25% of residential street segments within the 400 m radius of a residence may sufficiently represent the pedestrian environment and provides support for abbreviated data collection schemes of homogeneous street data for efficiency and cost-savings.

Although this study only included 11 neighborhoods, the systematic protocols and considerable detail provided comprehensive data with good variability. Strengths included detailed data collection, trained data collectors, and a complete census of street segments. A limitation of this study is that findings may not be generalizable to other Houston areas not surrounding a housing development or to older cities that have greater historical diversity in micro neighborhood design; thus, this study should be replicated.

Findings suggest that future studies may reduce the burden of exhaustive neighborhood data collection as a relatively small sample of the neighborhood residential street segments may appropriately represent the residential built environment. Arterial streets and streets with

more mixed use introduce much greater variability and richness in datasets, and future studies are needed to capture the depth and influence of arterial streets in the pedestrian environment.

**Additional file 1: Table S1. Harris County and HD Neighborhood socio-demographic characteristics.** Table S1 describes the neighborhood socio-demographic characteristics of Harris county and of each housing development neighborhood.

**Additional file 2: Table S2. HD Neighborhood pedestrian built environment characteristics.** Table S2 describes the pedestrian built environment characteristics of each housing development neighborhood.

**Additional file 3: Table S3. All HD and HD Neighborhood core vs. ring comparison of pedestrian built environment characteristics.** Table S3 describes the core vs. ring comparisons of pedestrian built environment characteristics of all housing development neighborhoods and of each housing development neighborhood.

**Additional file 4: Table S4. Comparisons of core vs. random sample of core residential segments (75%, 50% and 25%).** Table S4 describes the comparisons of residential street segments in the core and a random sample of 75%, 50%, 25% residential segments in the core, respectively.

#### Acknowledgements

This work was funded by a grant to Dr. Lee from the Active Living Research program of the Robert Wood Johnson Foundation. The authors gratefully acknowledge Scherezade Mama for assistance in the formatting of this manuscript and Dr. Daniel O'Connor for assistance with analyses.

#### Author details

<sup>1</sup>PPH Partners, 7710 W Sweetwater Tr, Flagstaff, AZ 86001, USA. <sup>2</sup>School of Social Work, Population Research Center, University of Texas at Austin, 1 University Station G1800, Austin, TX 78712, USA. <sup>3</sup>Department of Urban and Environmental Policy and Planning, Tufts University, 97 Talbot Avenue, Medford, MA 02155, USA. <sup>4</sup>Texas Obesity Research Center, Department of Health and Human Performance, University of Houston, 3855 Holman St, Garrison Gymnasium Rm 104, Houston, TX 77004, USA.

#### Authors' contributions

REL led the study and assisted with writing. TM assisted with the study and writing. CC completed analyses and assisted with the study and writing. BP led the cartography and assisted with the analyses and writing. AM assisted with the study and writing. All authors read and approved the final manuscript.

#### Competing interests

The authors declare that they have no competing interests.

Received: 17 August 2009 Accepted: 12 March 2010

Published: 12 March 2010

#### References

1. Cubbin C, Hadden WC, Winkleby MA: Neighborhood context and cardiovascular disease risk factors: the contribution of material deprivation. *Ethn Dis* 2001, **11**(4):687-700.
2. Cubbin C, Sundquist K, Ahlen H, Johansson SE, Winkleby MA, Sundquist J: Neighborhood deprivation and cardiovascular disease risk factors: protective and harmful effects. *Scand J Public Health* 2006, **34**(3):228-237.
3. Cubbin C, Winkleby MA: A multilevel analysis examining the influence of neighborhood-level deprivation on health knowledge, behavior changes, and risk of coronary heart disease: findings from four cities in northern California. *American Journal of Epidemiology* 2005, **162**(559-568).
4. Lee RE, Cubbin C: Neighborhood context and youth cardiovascular health behaviors. *Am J Public Health* 2002, **92**(3):428-436.
5. Lee RE, Cubbin C, Winkleby M: Contribution of neighbourhood socioeconomic status and physical activity resources to physical activity among women. *J Epidemiol Community Health* 2007, **61**(10):882-890.

6. Ross CE: **Walking exercising and smoking: Does neighborhood matter?** *Soc Sci Med* 2000, **51**:265-274.
7. Yen IH, Kaplan GA: **Poverty area residence and changes in physical activity level: Evidence from the Alameda County Study.** *Am J Public Health* 1998, **88**:1709-1712.
8. Casagrande S, Whitt-Glover M, Lancaster K, Odoms-Young A, Gary T: **Built Environment and Health Behaviors Among African Americans: A Systematic Review.** *American Journal of Preventive Medicine* 2009, **36**(2):174-181.
9. Van Dyck D, Deforche B, Cardon G, De Bourdeaudhuij I: **Neighbourhood walkability and its particular importance for adults with a preference for passive transport.** *Health & Place* 2009, **15**(2):496-504.
10. Gordon-Larsen P, Nelson MC, Page P, Popkin BM: **Inequality in the Built Environment Underlies Key Health Disparities in Physical Activity and Obesity.** *Pediatrics* 2006, **117**(2):417-424.
11. Cutts BB, Darby KJ, Boone CG, Brewis A: **City structure, obesity, and environmental justice: An integrated analysis of physical and social barriers to walkable streets and park access.** *Social Science & Medicine* 2009, **69**(9):1314-1322.
12. Lee RE, Mama SK, Banda JA, Bryant LG, McAlexander KP: **Physical Activity Opportunities in low Socioeconomic Status Neighborhoods.** *Journal of Epidemiology and Community Health* 2009, **63**:1021.
13. Lee RE, Heinrich KM, Medina AV, Maddock JE, Regan GR, Reese-Smith JY, Jokura Y: **A Picture of the Healthful Food Environment in Two Diverse Urban Cities.** *Environmental Health Insights* .
14. Institute of Medicine of the National Academies: **Does the built environment influence physical activity? Examining the evidence.** Committee on Physical Activity, Health, Transportation, and Land Use. Washington D.C 2005.
15. **Research Active Living Tools and Resources.** [<http://www.activelivingresearch.org/resourcesearch>].
16. Parmenter B, McMillan T, Cubbin C, Lee RE: **Developing Geospatial Data Management, Recruitment, and Analysis Techniques for Physical Activity Research.** *Journal of the Urban and Regional Information Systems Association* 2008, **20**(2):13-19.
17. Pikora TJ, Bull FC, Jamrozik K, Knuiman M, Giles-Corti B, Donovan RJ: **Developing a reliable audit instrument to measure the physical environment for physical activity.** *Am J Prev Med* 2002, **23**(3):187-194.
18. Boarnet MG, Day K, Alfonzo M, Forsyth A, Oakes M: **The Irvine-Minnesota inventory to measure built environments: reliability tests.** *Am J Prev Med* 2006, **30**(2):153-159.
19. Hoehner CM, Brennan Ramirez LK, Elliott MB, Handy SL, Brownson RC: **Perceived and objective environmental measures and physical activity among urban adults.** *Am J Prev Med* 2005, **28**(2 Suppl 2):105-116.
20. Clifton KJ, Livi Smith A, Rodriguez D: **The Development and Testing of an Audit for the Pedestrian Environment.** *Landscape and Urban Planning* 2007, **80**(1-2):95-110.
21. Agrawal AW, Schlossberg M, Irvin K: **How Far, by Which Route and Why? A Spatial Analysis of Pedestrian Preference.** *Journal of Urban Design* 2008, **13**(1):81-98.
22. McAlexander KM, Banda JA, McAlexander JW, Lee RE: **Physical Activity Resource Attributes and Obesity in Low-Income African Americans.** *Journal of Urban Health* 2009, **86**(5):696-707.
23. Lee RE, Reese-Smith J, Regan G, Booth K, Howard H: **Applying GIS Technology to Assess the Obesogenic Structure of Neighborhoods Surrounding Public Housing Developments.** *Medicine and Science in Sports and Exercise* 2003, **35**(5 Suppl).
24. Lee RE, Booth KM, Reese-Smith JY, Regan G, Howard HH: **The Physical Activity Resource Assessment (PARA) instrument: Evaluating features, amenities and incivilities of physical activity resources in urban neighborhoods.** *Int J Behav Nutr Phys Act* 2005, **2**(1):13.
25. Winkleby MA, Cubbin C, Ahn D: **Effect of cross-level interaction between individual and neighborhood socioeconomic status on adult mortality rates.** *American Journal of Public Health* 2006, **96**:2145-2153.

doi:10.1186/1479-5868-7-20

**Cite this article as:** McMillan et al.: **Neighborhood sampling: how many streets must an auditor walk?.** *International Journal of Behavioral Nutrition and Physical Activity* 2010 **7**:20.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

