Journal of Physical Activity and Health 2006, 3, Suppl 1, S241-S257

Standards for Environmental Measurement Using GIS: Toward a Protocol for Protocols

Ann Forsyth, Kathryn H. Schmitz, Michael Oakes, Jason Zimmerman, and Joel Koepp

Background: Interdisciplinary research regarding how the built environment influences physical activity has recently increased. Many research projects conducted jointly by public health and environmental design professionals are using geographic information systems (GIS) to objectively measure the built environment. Numerous methodological issues remain, however, and environmental measurements have not been well documented with accepted, common definitions of valid, reliable variables. *Methods*: This paper proposes how to create and document standardized definitions for measures of environmental variables using GIS with the ultimate goal of developing reliable, valid measures. Inherent problems with software and data that hamper environmental measurement can be offset by protocols combining clear conceptual bases with detailed measurement instructions. *Results*: Examples demonstrate how protocols can more clearly translate concepts into specific measurement. *Conclusions*: This paper provides a model for developing protocols to allow high quality comparative research on relationships between the environment and physical activity and other outcomes of public health interest.

Key Words: GIS, built environment, physical activity

Attempts to explain leisure-time physical activity using individual and social variables have rarely explained greater than 25% of variance.[1] The potential for the built environment to influence some of the variance unexplained by individual and social variables is supported by some theories currently used by behaviorists interested in physical activity—most notably social ecologic theory.[2] This theory, which is broader than the prevalent psychosocial theories, contends that physical activity (of all domains, not just leisure) occurs in a physical space and that the built environment can serve to promote or discourage activity. This theoretical background has recently spurred increased interdisciplinary research into built environmental influences on physical activity, though the focus is rarely just leisure-time activity.

Forsyth, Oakes, Zimmerman, and Koepp are with the Metropolitan Design Center, University of Minnesota, Minneapolis, MN 55455. Schmitz is with the University of Pennsylvania School of Medicine, Philadelphia, PA 19104.

Many research projects conducted jointly by public health and environmental design professionals use geographic information systems (GIS) to objectively measure the built environment. Prior to the advent of GIS technology, short of very laborious manual mapping, public health researchers had to rely on self-reported knowledge and perceptions about environments in studies that attempted to relate environment to health behaviors such as physical activity. Knowledge that can be self-reported and perceptions about the environment undoubtedly influence physical activity. However, the addition of objective GIS measures presents new frontiers for understanding the role of urban design on physical activity and other health behaviors.

The use of GIS for measurement and analysis of research-quality variables is still in its infancy, however, and numerous methodological issues remain. Conceptually, several dimensions of environmental measurement require attention. First, broad categories of variables that may be plausibly associated with physical activity must be selected (e.g. street pattern or mixed use). Second, methods of analyzing those variables need to be weighed against one another, including self-report and objective approaches. Third, each variable must be specifically defined, e.g., specific survey questions or GIS-based formulae. Finally, these definitions must be documented so that measures can be replicated and assessed for reliability and validity. To date, most work on the relationship between physical activity and the built environment has used self-report surveys to measure the environment.[3] As interest in objective measures of the environment has increased, issues of data quality and availability have become apparent,[4] and surprisingly little attention has been given to precisely defining or documenting GIS-based measures. With a few exceptions, environmental measurements have not been well documented with accepted, commonly used definitions of valid, reliable variables, which makes replication a matter of significant guesswork.

This paper proposes an approach to creating and documenting standardized definitions for measures of environmental variables using GIS. These definitions can then be assessed for reliability and validity, using protocols similar to those used to develop self-reported physical activity measures. The Background section outlines why protocols are needed, the inherent limitations of software and data, and the need for clear definitions. The Methods section outlines how protocols are shaped and how they can be used to overcome software and data limitations, enabling better communication within and between research teams. Examples in the Results section illustrate how protocols clarify ambiguities in measurement. Finally, the Summary discusses the contributions of the protocols and other issues that are yet to be resolved. The appendix presents a two protocols that were developed using the six steps described in the Methods section.

From research to date, four types of features in the physical environment have emerged as likely correlates of physical activity, particularly walking: development density, the mix of land uses, street pattern, and pedestrian infrastructure and amenities.[5-11] Many of these features can be measured through GIS. However, not enough is known about which aspects of these variables matter, how much, for whom, and at what scale. Because of these questions, the study team proposed to measure dozens of specific environmental measures using a variety of geographical perspectives (e.g., grid cells, network/street distance buffers, airline buffers, distance to the nearest feature of a certain type) and scales (e.g., 200 m buffer, 1600 m buffer). In this paper, two example variables are used to illustrate protocol development: points of street access, which is detailed in the appendix but only summarized in this text, and the ratio of "4-way" to other intersections.

This protocol work is based on a study of walking and other physical activity conducted in 36 environmentally diverse residential areas in the Twin Cities in Minnesota, involving 718 participants who wore an accelerometer for 7 d, completed a 7-d travel diary, had their height and weight measured, and answered a survey dealing with demographic, environmental perception, attitudinal, and socioeconomic issues. The results of the study will be reported elsewhere. The study area was initially selected because of its high quality GIS data. To perform the study, the team found it necessary to develop protocols to define and operationalize objective (GIS-based) measures of the environment.

This paper, aimed at public health researchers interested in physical activity, seeks to address the lack of precise definition and documentation of GIS-based measures, and spark systematic approaches to such protocol development. In addition, the proposed definitions of objective (GIS-based) environmental variables may be a useful tool for researchers who want to use GIS without developing measures themselves.

Background

Why Protocols Are Necessary

Joint research projects conducted by professionals in the public health and environmental design fields are complicated by differences in research methods used by the two fields. In public health and physical activity research, measures must be shown to be reliable and valid prior to their use. Environmental measurement research is in a different stage of progress for *data* collection, *measurement of variables* based on those data, and *analyses* of those variables.

There are several reasons for the lack of developed and consistent environmental measures. Transportation planning, which has a large number of quantified measurements, has focused on motorized transportation, leaving issues relating to walking in the hands of urban designers. With few exceptions—such as aspects of environmental perception—urban designers have been less interested in quantification than in developing a great sensitivity to the qualitative aspects of place. Even if they were interested in quantification, however, little funding has been available for such work.

In addition, sophisticated computer mapping only emerged into wide scale use in the early 1990s, and computerized mapping databases have taken years to develop. Their continued development will allow environmental measures to eventually become as standardized as measures of physical activity. At present, however, studies reporting environmental variables often fail to explain in a manner that would allow replication by other investigators how variables are derived. For example, "intersection density" may be a variable used to examine the association of walking and built environment, but authors may not indicate whether freeways or other limited-access roads are included in the measures.

There are certainly exceptions to this lack of clear and precise operational definitions of environmental variables. Steiner et al. have an exemplary assessment of a range of measures of street patterns and connectivity.[12] Dill also provides an excellent discussion of street pattern measures.[11] Other researchers provide a high level of detail in their reports, but not enough to be certain one is replicating their measures.[13] This is partly because the reports are focused on the substance of their findings and there are few publication venues for more detailed documentation of methods.[5, 7-9, 14-27]

Inherent Challenges with GIS Due to the Complexity of Software and Data

The new GIS software programs and databases raise a number of issues that make protocols for measurement important as a means of communication between investigators interested in comparing results. Detailed documentation is also useful within teams to identify any problems with the link between the concept investigators wish to measure and the data and techniques available for measuring it.

Software. GIS software is a bundle of programs. There are researchers who work with a simple mapping program, but most add on database management, statistical tools, and scripting languages. The suite of programs most used in the US is ESRI's ArcGIS suite (including ArcMap, ArcCatalog, Spatial Analyst, ArcInfo). Some of these programs are extremely expensive—money matters in this research. For example, our university was not licensed to use ArcSDE, a database management program allowing automatic "versioning" to track changes to data over time and between multiple users. Also, important analytical tools, such as the network analyst that measures street distance buffers (as opposed to crow flies), were not updated for the new GIS programs, which has forced users for several years to switch between an old version of ArcView (an earlier GIS mapping program) and a new version of ArcGIS.

Advanced users often borrow scripts created by others for derived GIS variables such as variations on measures of the distance to the nearest landmark of a certain type, saving time but risking further errors given that scripts are often merely approximations of the concepts being measured and that assumptions are not always clearly articulated or operationalized.

Even within a single program there may be multiple ways to make a similar calculation with different results. For example, when measuring the distance to the nearest feature of type B (e.g., coffee shop) from a starting point A (e.g., a residential address), some methods of analysis use vector data (data created using points, lines, and polygons or shapes) and some use raster data (maps made from cells such as pixels). These differing methods may have slightly different results.

Data. A number of standard concerns challenge all mapping and most geographical analysis (e.g., the earth curves and so straight lines curve). However, issues related to the diverse environmental variables thought to be associated with physical activity as well as the desire for creating methods that can be replicated have created special challenges that are pushing GIS analyses.

- Consistency: Some GIS data are consistently available across a nation (e.g., census data) and some are purely local (e.g., land use categories may be different in every municipality). Although research on physical activity and the built environment is being pursued throughout the world, and many self-report surveys are being used in multiple countries, no environmental data at a scale or on a topic relevant for physical activity research are available internationally. Satellite photos are available, but resolutions are not always high enough to distinguish relevant features and analysis generally involves a time-consuming and error-prone process of converting the raster image data to vector data for analysis. Road centerline and parcel data may come closest to being internationally available and organizations, such as the International Federation of Surveyors, are investigating standards.[28]

- Purpose: Because municipalities collect data for their own purposes (e.g. charging taxes, planning road maintenance), data across geographical areas are inconsistent. Too, the purposes for which data are collected may not involve many topics relevant to physical activity research. Therefore, de novo data collection, which can take the form of time-consuming systematic observations of environments, is often necessary.[29] Such fieldwork data collection approaches are in their infancy. [30, 31, 33-35]
- Geographies: Features or dimensions of the environment are measured using data collected in varying geographic units (parcels, census blocks, street segments). Creating variables for the research often requires assumptions and/or transformation of the data to fit a new geographic unit. For example, the measurement of variables within a specified straight line buffer of an individual will depend on the variable. For example, median block size requires measurement of complete blocks. However, in measuring block size within, say, 400 meters of a person, is it best to count all blocks with a centroid within the buffer or by counting all the blocks fully contained within the buffer? In contrast, when our research team wanted to measure population density within a buffer, it was more important to match population and area, so that for census blocks cut by the edge of the buffer, we apportioned population within the block according to the land area (40% of the land area received 40% of the population). This was a reasonable assumption, but there are other plausible approaches to dealing with this mismatch between measurement geographies and available data.
- Data resolution: What is considered accurate enough for a map with a scale of 1 meter to 1,000,000 meters is unlikely to be accurate enough for a map with a scale of 1 meter to 1,000 meters, but it can be tempting to ignore these issues of resolution.
- Accuracy: One municipality may consistently underestimate home values by 5 to 7%, another by 9 to 10%. These variations make comparative research very difficult and decrease accuracy.
- Completeness: Complete data for one purpose does not mean that complete data are available for an entire study area (e.g., traffic counts of all arterial roads do not include all roads).
- Time: GIS data often incorporate data captured during different years. Because cities are constantly being built and rebuilt, the environment represented by

the data may be a composite or summary picture representing a "reality" that never actually existed. In our study, for example, several roads had been re-aligned between the time of the parcel mapping and the road centerline data.

- Errors: Errors exist in collection and input, even for consistent data. These errors can be minimized but not eliminated. While this is true for all data the complexity and multiple sources of GIS data can render errors difficult to find.

Measurements Are Not Clearly Defined

Even with reliable data as close to perfection as is possible, researchers measuring environmental correlates of physical activity must still decide which variables to measure and how to measure them.

While there is broad theoretical consensus that the environment affects physical activity, no consensus exists regarding which aspects matter, why, how much, and for whom. Nor do we have standard lists of specific variables typically measured in such research. On the contrary, measures developed in urban geography, planning, and transportation may not be relevant to research on physical activity, and public health researchers are not always aware of the problems with physical environment data. In addition, although it would be logical to test a large number of variables and select those that are most reliable, automatable, and correlated with measured physical activity, no standard approaches exist for developing such variables.

There is a mismatch, too, between the broad basic variables used in conceptualizing studies and the messy data technicians encounter. Although discussions addressing this mismatch often occur within the research team, the content of these discussions is rarely published. The appendix has an example of the complexities that can result from this type of mismatch, with over two pages of thumbnails of variations on X (4-way) and T (3-way) intersections needing to be classified.

The complexity of GIS software and data, researchers' errors, and incorrect and inconsistent decisions make standardization of measurement protocols all the more important.

Methods

The GIS protocols we propose are designed to bridge concept and application, enhance communication among those trained in disparate fields, and enable replication. The protocol for each variable has six parts, beginning with a reasonably precise definition of the variable and followed by an explanation of how to operationalize the variable in GIS. Examples are from *Environment and Physical Activity: GIS Protocols*:[36]

1. Basic Concept: A statement of the concept that the variable is intended to represent, with a discussion about its place in the literature and previous use. A comment about the hypothesized relationship between the variable and physical activity might also be useful. While rudimentary, this is often the sole variable description included in journal articles. E.g., Gross Population Density is the overall residential population divided by the land area excluding water area. (In the protocols manual, the Basic Concept section also cites sources explaining the importance of the variable as a measure of the environment.)[a]

2. Basic Formula, or Basic Definition, Basic Procedure: A more specific formula or definition of the variable, but without enough detail to create a GIS-based measure. In many research reports, this is the most detailed level at which variables are reported. E.g., Population per Unit Land Area (without water) = Persons in housing units per unit gross area excluding water area.

3. Detailed Formula or Detailed Definition: An even more specific formula, including data sources and the spatial unit at which the variable is measured, which affects the measurement. Only a few published research reports provide this level of detail and yet it is the level of detail that is *essential* for performing a measure using GIS. One of the greatest barriers to replicating studies reported in the land use and transportation research literature is the absence of a Detailed Formula or Definition.[b] E.g., Population per Unit Land Area (excluding water area) = Persons in housing units as measured in US Census data at the block level per Unit Land Area excluding area of water features as measured in the Ramsey County water layer or the Metropolitan Council water layer for areas outside Ramsey County.

4. Comments and Explanations: The questions likely to occur when operationalizing formulae. In addition, this section warns about common errors. E.g., where census blocks are cut by the end of the measurement geography, the population is apportioned according to the percentage of land area falling inside and outside the line. The Ramsey County water layer includes all water visible in very high resolution digital orthophotos, excluding such features as swimming pools (see more detail in *Environment and Physical Activity*, chapter 2). For the one area outside of Ramsey County, the study team used the Metropolitan Council's 2000 Water Feature Layer.[36]

5. GIS Approach: A description of the measurement in outline, in a form that a GIS expert could use to perform measures, or that someone using a different software program could use to develop their own steps.[36] E.g., calculate the apportioned population and divide it by the total area that has been recalculated after removing the area of all water features.

6. GIS Steps: Detailed GIS instructions using Arc 8 or Arc 9 designed to be comprehensible to infrequent users of GIS. GIS Steps sections range from one-half to almost three pages. These allow easy measurement replication, more complete discussion about how well GIS operations and functions represent concepts being measured, and avoidance of at least some inconsistencies. They also allow users to script some of the measurements using Model Builder, a new feature in ArcGIS 9. As software is updated these are outmoded but provide a record of detailed decisions that can be adapted to new program versions.

The Minnesota protocols were developed by listing and revising the list of variables; creating the basic formulae, based on earlier research whenever possible; troubleshooting details; writing GIS strategies (approach and steps); and assembling and refining.[36] The process of developing protocols was interactive between the PI (environmental measurement lead) and GIS technical staff, which required all participants to clearly communicate the concepts that needed to be measured, the questions raised in doing so, and the steps for actually making the measures in GIS.

Because we were developing measures as we were developing a final list of variables, protocols were revised more often than had we been certain of our measures from the start. Overall, protocols served to increase the capacity to replicate measures both within the team and across research teams. Across-project replication is particularly important to allow later meta-analysis and potential validation.

Examples

Two examples from street pattern measures serve to illustrate the use of protocols. Due to space limitations, however, the GIS steps for these protocols are not included. For the same reason, the protocol for the first street pattern measure, which indicates porosity of a street by measuring access points, appears only in the appendix; in summary, issues of porosity raise questions regarding how to measure roads such as freeways that cross the boundary of the measurement geography (e.g., buffer) but do not actually give access to it. Our team chose criteria for making those decisions that we considered theoretically defensible and which enabled maximum automation. The appendix details our process and criteria.

The second street pattern measure is the ratio of 4-way or X-intersections to all intersections. This is a measure of road connectivity. High ratios of 4-way intersections are thought to be associated with walkable environments.[5,13] Measurement involves calculating the "valence" of each intersection. A T-intersection has a valence of 3 because three road segments converge to the center of the intersection. A 4-way intersection has a valence of 4 because four road segments converge to the center of the intersection. Analyses show, however, that different methods of calculating valence for a simple intersection can yield different results. The appendix illustrates how a GIS-based calculation of valence identifies as 16-way an intersection which, with respect to porosity, is 4-way. This discrepancy results from GIS routines counting each segment converging to the same point and not automatically recognizing when the same road is divided and represented by two sets of segments. Road width is another such example. Although road widths commonly range from 20 to 30 meters, roads in GIS are represented by their center lines regardless of their width. As a result, when wide roads intersect, their center lines often converge not to one but to several points. A simple 4-way intersection of two wide roads (valence of 4) thus can be represented in GIS as two "offset" 3-way intersections, giving a valence of 6. As can be seen in the illustrations in the appendix there are many other variations on the 3- or 4-way intersection, a fact not readily apparent in the current literature.

To address these issues, we examined aerial photos and took field trips. Eventually we selected two buffers of 10 meters and 15 meters and placed them around each GIS-identified intersection point. Using a GIS routine, we dissolved all points falling within the buffer to a single intersection point, with road centerlines intersecting within 20 m or less for the 10 m buffer and 30 m or less for the 15 m buffer becoming one segment intersection point. We propose to conduct a sensitivity analysis to examine whether buffering creates significantly different measures to "raw" intersections; we will also determine which buffer has the highest correlation with walking.

Summary and Remaining Challenges

Protocols are an important strategy for clearly defining variables and documenting their execution. The examples above demonstrate the need for detailed formulae and data descriptions, so that researchers can be certain that their environmental measurements are replicating earlier work or the work of other teams. Consistency may be difficult even then because data are not collected and classified consistently, particularly in the area of land use.

Overall, the protocols respond to the general problem of measuring environmental features thought to be associated with physical activity. The lack of standards for defining and operationalizing key measurement variables is reflected in the literature. Researchers have only rarely provided detailed formulae for the calculations of specific variables or a detailed approach to conceptualization (e.g., whether water was included in the land area calculations). Even though such measures may be able to be tested for reliability and validated against other measures within a study, other teams of researchers cannot be sure they are replicating them.

This paper points to the necessity of establishing consensus about which variables are important and developing reliable and valid approaches to measurement. Teams using GIS need to document their measurements in detail, either within research reports or in separate public documents.[c] Agreement on standard measures may come over time with development of the literature, or through periodic attempts, perhaps fostered by funding agencies, to come to agreement. Without such documentation and agreement, replication and meta-analysis will be virtually impossible. The Minnesota Protocols provide one set of such measures currently available for use.

In addition, this paper provides a model for the development of measures of the built environment which will be of value for research on the causative relationship of the built environment with a variety of public health outcomes, not just physical activity.[37-39] Only through carefully designed and conducted research on these causal relationships will we approach the ultimate goal of intervening to improve public health through urban planning changes. Such interventions will likely take decades and involve legislative action. A solid empirical basis for causal associations between the built environment and physical activity (and other public health outcomes) will be vital toward those efforts.

Acknowledgments

Preparation of this article was assisted by a grant from The Robert Wood Johnson Foundation Active Living Research Program. Our thanks to Anne Vernez Moudon and several anonymous reviewers for extremely helpful comments.

References

1. Sallis JF, Johnson MF, Calfas KJ, Caparosa S, Nichols J. Assessing perceived physical environment variables that may influence physical activity. *Res Quart Exerc Sport.* 1997; 68:345-351.
2. Stokols D. Establishing and maintaining healthy environments. Toward a social ecology of health promotion. *Am Psychol.* 1992; 47:6-22.

3. Saelens BE, Sallis JF, Black JB, Chen D. Neighborhood-based differences in physical activity: an environment scale evaluation. *Am J Public Health.* 2003, 93(9):1552-8.
4. Porter D, Kirtland K, Neet M, Williams J, Ainsworth B. Considerations for using a geographic information system to assess environmental supports for physical activity. *Preventing Chronic Disease: Public Health Research, Practice, and Policy.* 2004; 1, 4:1-5.
5. Cervero R, Kockelman K. Travel demand and the 3 Ds: density, diversity, and design *Transportation Research; D,* 1997; 2, 3:199-219.
6. Crane R. The influence of urban form on travel: an interpretive review. *J Plann Literature.* 2000; 15, 1:3-23.
7. Ewing R, Cervero R. Travel and the built environment: A synthesis. *Transportation Res Rec.* 2001, 1780:87-113.
8. Frank LD, Pivo G. Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit, and walking. *Transportation Res Rec.* 1994; 1466:44-52.
9. Handy S, Boarnet M, Ewing R, Killingsworth R. How the built environment affects physical activity: views from urban planning. *Am J Prev Med.* 2002; 23, 2S:64-79.
10. Handy S. Critical assessment of the literature on the relationships among transportation, land use, and physical activity. Paper prepared for the TRB and the Institute of Medicine Committee on Physical Activity, Health, Transportation, and Land Use, November 24, 2003.
11. Saelens B, Sallis J, Frank L. Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Ann Behav Med.* 2003; 25, 2:80-91.
12. Steiner R, Bond A, Miller D, Sand P. *Future Directions for Multimodal Areawide Level of Service Handbook: Research and Development*. The Florida Dept of Transportation, Office of Systems Planning, Contract BC-345-78; 2004.
13. Dill J. Measuring network connectivity for bicycling and walking. Presented at the Joint Congress of ACSP-AESOP, Leuven, Belgium; 2003.
14. Cervero R, Duncan M. Walking, bicycling, and urban landscapes: evidence from the San Francisco Bay Area. *Am J Public Health.* 2003; 93, 9:1478-1483.
15. Ewing R, Schmid T, Killingsworth R, Zlot A, Raudenbush S. Relationship between urban sprawl and physical activity, obesity, and morbidity. *Am J Health Promot.* 2003; 18(1):47-57.
16. Forsyth A. Measuring density: working definitions for residential density and building intensity. Design Brief 9. Minneapolis: Design Center; 2003.
17. Hess PM, Moudon AV, Snyder MC, Stanilov K. Neighborhood site design and pedestrian travel. *Transportation Res Rec.* 1999, 1674:9-19.
18. Krizek K. Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? *J Am Plann Assoc.* 2003a, 69, 3:265-281.
19. Krizek K. Operationalizing neighborhood accessibility for land use-travel behavior research and regional modeling. *J Plann Educ Res.* 2003b; 22, 3:270-287.
20. McNally MG, Kulkarni A. Assessment of influence of land-use transportation system on travel behavior. *Transportation Res Rec.* 1997; 1607:105-115.
21. McCann BA, Ewing R. *Measuring the Health Effects of Sprawl*. Smart Growth America Surface Transportation Policy Project; 2003.
22. Messenger T, Ewing R 1996. Transit-oriented development in the Sunbelt. *Transportation Res Rec.* 1996, 1518:145-153.
23. Moudon AV, Lee C, Cheadle A, Collier C, Johnson D, Weathers R, Courbois J-I, Hurvitz P. *Walkable and Bikable Communities Project: A Draft Report on Findings from the Walking Models*. Seattle: University of Washington Urban Form Lab; 2004.
24. Song Y, Knaap G-J. Measuring urban form. *J Am Plann Assoc.* 2004; 70, 2:210-225.
25. Southworth M. Walkable communities? An evaluation of neotraditional communities on the urban edge. *J Am Plann Assoc.* 1997; 63, 1:28-44.

26. US Dept of Transportation, Federal Highway Administration. *Guidebook on Methods to Estimate Non-Motorized Travel: Overview of Methods*. FHWA-RD-98-165. McLean, VA: FHWA, 1999a.
27. US Dept of Transportation, Federal Highway Administration. *Guidebook on Methods to Estimate Non-Motorized Travel: Supporting Documentation*. FHWA-RD-98-166. McLean, VA: FHWA, 1999b.
28. International Federation of Surveyors. Web site; 2005. http://www.fig.net.
29. Forsyth A. Analyzing public space at a metropolitan scale. *Urban Geogr.* 2000; 21, 2: 121-147.
30. Moudon AV, Lee C. Walking and bicycling: an evaluation of environmental audit instruments. *Am J Health Promot.* 2003; 18, 1:21-37.
31. Boarnet M, Day K. Assessing built environment features linked to physical activity. Presentation at Active Living Research Annual Conference, San Diego, January; 2004.
32. Day K, Alfonzo A. Measuring urban design features related to walking and bicycling. Presentation at the Association of Collegiate Schools of Planning Conference, Portland, October; 2004.
33. Forsyth A, Oakes M, Schmitz KH. Observational measures of the physical environment: testing and modifying the Boarnet-Day inventory. Paper presented at the Association of Collegiate Schools of Planning Conference, Portland; 2004.
34. Pikora. T. *Systematic Pedestrian and Cycling Environmental Scan (SPACES), Survey of the Physical Environment in Local Neighborhoods: Observer's Manual*. Nedlands, Western Australia, 2002.
35. Pikora T, Giles-Corti B, Bull F, Jamrozik K, Donovan R. Developing a framework for assessment of the environmental determinants of walking and cycling. *Soc Sci Med.* 2003; 56, 1693-1703.
36. Forsyth A ed. *Environment and Physical Activity: GIS Protocols*. Version 3.0. Minneapolis: Metropolitan Design Center, 2005. http://www.designcenter.umn.edu/projects/current/current_research_areas/walkability/twin_cities_walking/epaGISprotocols.html
37. Raudenbush SW. The quantitative assessment of neighborhood social environments. In I. Kawachi I and Berkman LF, eds. *Neighborhoods and Health*. New York: Oxford; 2003: 112-131.
38. Sampson RJ, Raudenbush SW. Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *Am. J. Sociol*. 1999; 105:603-651.
39. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med.* 2004; 58:1929-52.

Notes:

[a] People walk on water in winter, but we did not collect participant data in winter.
[b] Anne Vernez Moudon helped tremendously in clarifying this point.
[c] An anonymous reviewer helped clarify these recommendations.

Appendix

Appendices/Figures

Caption: Appendix, Protocols for Number of Access Points and Ratio of 4-way

Intersections to All Intersections

6.4 Number of Access Points

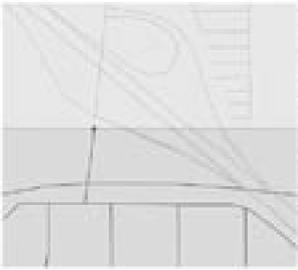
1. Basic Concept
Access points are a measure of the amount of connection an area has to its surroundings. We are not convinced it is a good measure of walkability but we are testing it.

At its most extreme one can imagine a gated neighborhood which inside the neighborhood had small blocks and a highly connected street pattern, but had only one way in and out. In this case walking to destinations outside the development would be made far more complex. Access point measures create a measurement geography, generally a grid cell, and measure how porous its edges are (Southworth 1997; McNally and Kulkarni 1997, 109).

2. Basic Formula
Number of Access Points = Number of times a road crosses the edge of the measurement geography.

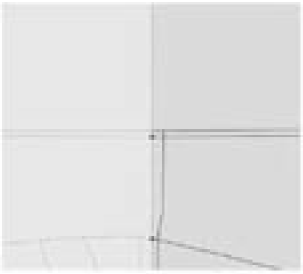
3. Detailed Formula
Number of Access Points = Number of times a road crosses the edge of the measurement geography with divided at-grade roads counted once and limited access highways excluded.

4. Comments and Explanations

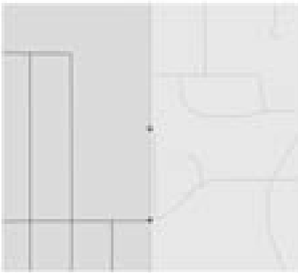


Access points are points of intersection between road centerlines and site boundaries excluding points resulting from the intersections of site boundaries with interstates, on- and off-ramps, and other limited access highways.

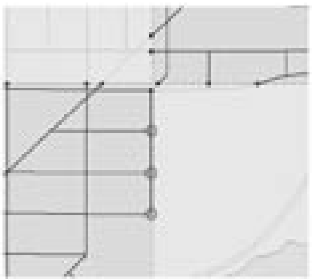
In the diagram to the left, the four green [pale] points are access points created by the intersection of the site boundary and an interstate and its on- and off-ramps, and therefore are excluded from the analysis. The red [dark] point is a valid access point.



Also, site boundary intersections with divided at-grade roads, such as parkways, are counted once (as one access point) instead of twice. The green [pale] point in the diagram to the left is half of a set of paired access points created by the intersection of the site boundary and a divided parkway. Only the red [dark] access points are included in the analysis.



Many access points are somewhat marginal in their quality. That is, these points may exist but not provide significant access to the site, or may fall just inside the site boundary and therefore not be counted (see examples below). The red [dark] point in the middle of the image is technically an access point because the road centerline just crosses the site boundary, even though it provides no significant access to the site. However, it is extremely complex to develop rules to exclude such points—see next image for more examples.



The three purple [dark] circles identify points that should be included in the analysis, since they provide significant access to the site, but are not because the road centerlines lie just inside the site boundary and do not intersect it.

We experimented with methods of buffering these points and derived various rules for exclusion or inclusion, but in the end determined that this opened the door to many subjective decisions. It seemed that the number of access points included that probably should be excluded would roughly offset the number of access points excluded that probably should be included and therefore we decided to set aside the issue of marginal access points in our analysis.

5. GIS Approach
A point file was created that contained all points were road centerlines intersected site boundaries. From this set, Interstate points were deselected. For any divided roads (those with two distinct centerlines, such as parkways), one half of the access points were deselected. The total number of access points for each study site were summed.

6. GIS Steps
The following procedures were completed using ArcMap (ArcInfo) 8.3.

To complete this protocol you will need:

1. Site layer (polygon)
2. Site boundary layer (polyline)
   If you only have a polygon site layer, convert the polygons to polylines using the XTools extension (available from the Downloads section of the ESRI website).
3. Road centerline layer (polyline)
4. AddPointsAtCrossings ArcScript available from the Downloads section of the ESRI website

The basic steps are:

- *Create the Access Point file*
- *Correct for Interstate Access Points*
- *Correct for Parkway Access Points*

- *Remove Unwanted Access Points from View/Table*
- *Summarize Access Point Counts for each Site*

[One and a half pages of detailed steps omitted.]

6.7 Ratio of 4-Way Intersections to All Intersections

1. Basic Concept
Four-way intersections are an indicator of grid street patterns, thought to be more connected. However, one can imagine a high proportion of four way intersections in an area with enormous blocks, so the density of intersections should also be considered.

2. Basic Formula
Ratio of 4-Way Intersections to All Intersections = Raw percent of 4-way intersections versus other intersections (valence 3+).

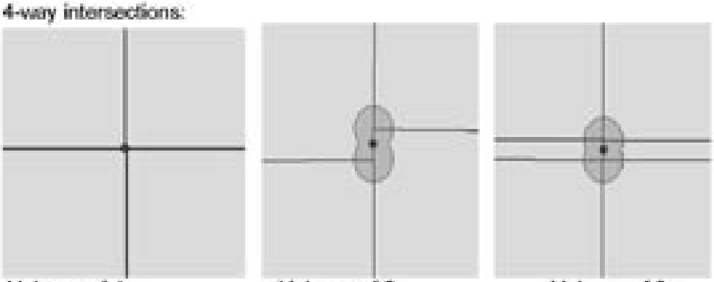
3. Detailed Formula
Ratio of 4-Way Intersections to All Intersections = Same as above although with the added complication that some higher valence intersections are actually 3-way and 4-way intersections.

4. Comments and Explanations
As outlined in Intersections per Area, intersection counts in our study sites (1) did not include intersections formed by the convergence of Interstates and on- or off-ramps and (2) corrected for mis-aligned intersections by creating buffers around each intersection node, merging them where they overlapped, and creating a centroid for each new intersection polygon.

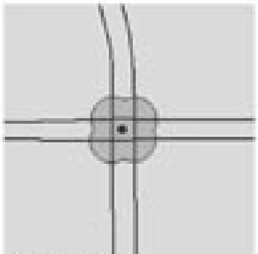
In order to be able to classify these new nodes, we created "rules" to assign types (3-way, 4-way) to the intersections. These assignments were based on the sum of the valences of the original intersections (the valence of an intersection, based on the script that creates the intersection nodes, is the count of the number of road segments that converge to create the node – a valence of 1 is a dead end or cul-de-sac; a valence of 3 is a 3-way intersection; a valence of 4 is a 4-way intersection; etc.).

In particular, we realized that intersections with valences of 6 were typically the result of two intersections that were not quite aligned but that functioned as 4-way intersections. Intersections with valences of 8 were almost always the result of intersections of divided and non-divided roads that functioned as 4-way intersections. Intersections with valences of 16 were the result of intersections of two divided roads that functioned as 4-way intersections.

4-way intersections:



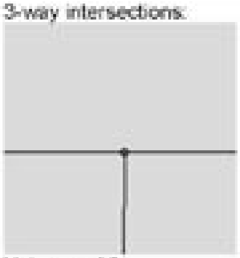
Valence of 4 Valence of 6 Valence of 8



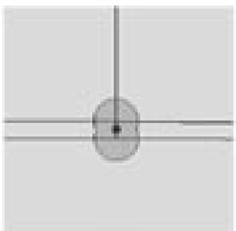
Valence of 16

Likewise, we found that most newly formed intersections with valences of 7 were the result of 3-way intersections of divided and non-divided roads. Similarly, about half of the intersections with valences of 12 were the result of 3-way intersections of two divided roads. These seemed to function as 3-way intersections.

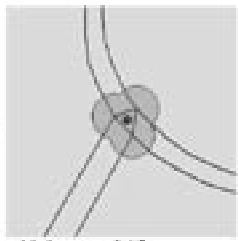
3-way intersections:



Valence of 3

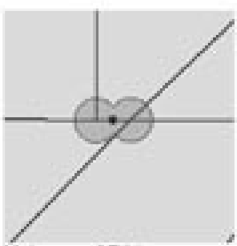


Valence of 7

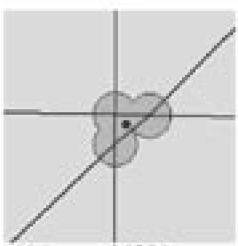


Valence of 12

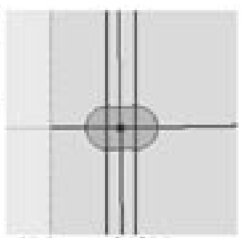
There were some intersections with valences of 7 and 12 that did not conform. Specifically, there were a handful of merged intersections with valences of 7 and 12 that were the formed by the convergence of the gridded street pattern with a diagonal. In addition, there was one intersection with a valence of 12 that seemed to function more like a 4-way intersection. Because there were only three intersections of valence 12 using 10 meter buffers (two 3-way and one 4-way) and five using 15 meter buffers (two 3-way and three 4-way), we chose to apply our 3-way and 4-way intersection rules by hand instead of using a blanket rule. The images below were classed as 4-way.



Valence of 7(b)

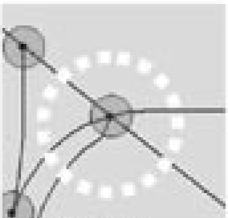


Valence of 12(b)

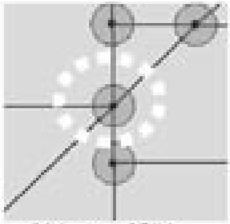


Valence of 12(c)

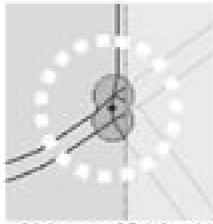
Finally, there were a number of other new intersection types that were more complex than either 3-way or 4-way intersections. These were partly an artifact of our trying to deal with classifying functional 3- and 4- way intersections, through buffering. These are intersections that show a high level of connectivity which is the point of the 4-way intersection count, so they were counted as 4 way intersections. **Note, the intersection being referred to is outlined with a dotted oval.**



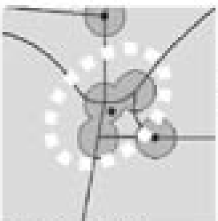
Valence of 5(a)



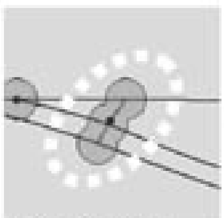
Valence of 5(b)



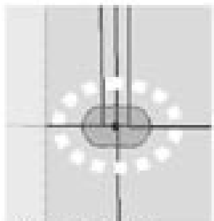
Valence of 9 (should really be 13)



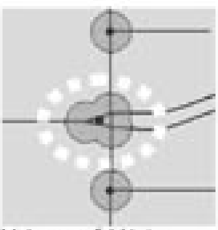
Valence of 10(a)



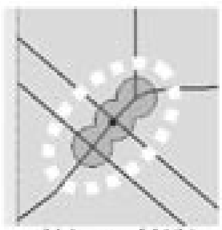
Valence of 10(b)



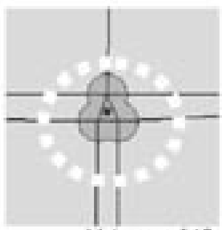
Valence of 10(c)



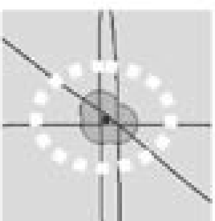
Valence of 11(a)



Valence of 11(b)



Valence of 13



Valence of 14

For all our intersection counts, we ran our numbers three times. During the first run, we counted intersections without any buffers. During the second run, we counted intersections that were buffered 10 meters and adjusted for valences of 6, 7, 8, 12, and 16 and other adjustments as noted above. Finally, in our third run, we counted intersections that were buffered 15 meters and adjusted for valences of 6, 7, 8, 12, and 16 and other adjustments as noted above.

5. GIS Approach

Remove Interstates from the road centerline file. Use the Calculate Fnode Tnode script to create the intersection point file. Remove points with a valence of 1 or 2. Buffer intersections with 10 and 15 meter buffers and create new intersection centroids with reassigned valences. Intersect point file with measurement geography and calculate ratios of 4-way intersection counts to total intersection counts.

6. GIS Steps Omitted—due to length (they are over three pages single spaced)