The Census of Social Institutions (CSI): A Public Health Direct Observation Measure of Local Land Use

Katie M. Heinrich, Joseph Hughey, Anthony Randles,
Dustin Wall, N. Andrew Peterson, Nattinee Jitnarin,
LaVerne Berkel, Peter Eaton, Doug Bowles,
C. Keith Haddock, and W. S. Carlos Poston

ABSTRACT *This manuscript describes the development of the Census of Social Institutions (CSI), a reliable direct observation parcel-level built environment measure. The CSI was used to measure all non-residential parcels (n=10,842) in 21 one-mile-radius neighborhoods centered around census block groups of varying income and ethnicity in a large metropolitan area. One year test-retest and inter-rater intra-class correlations showed high reliability for major use type and detail code observations. The CSI accurately captured the presence of about 9,500 uses, including 828 multiple major use and 431 mixed major use parcels that would have been missed in standard commercial databases. CSI data can be utilized to determine the health impacts of environmental settings.*

KEYWORDS *Built environment, Behavior settings, Land use, Health, Parcel, Geographic information systems*

Built environment audits are increasingly used to understand the influence of street-scale variables on health.¹⁻⁴ To advance this research, sound measures of both single institutions and the mix of institutions in communities are needed, but such measures are lacking.^{1,5,6} Measurement of environmental attributes has presented important challenges, including debate over the accuracy of pre-existing databases (e.g., issues include incomplete data, inaccurate classifications, or inaccurately geocoded locations).⁷ In addition, recent reviews noted weaknesses in studies evaluating associations between environmental factors and health attributable to using secondary data or data aggregated at too high a level of analysis (e.g., zip

Heinrich is with the John A. Burns School of Medicine, Department of Public Health Sciences, University of Hawaii, Honolulu, HI, USA; Hughey is with the Departments of Psychology and Architecture, and Urban Planning and Design, University of Missouri-Kansas, Kansas, MO, USA; Randles is with the Department of Health, Nutrition and Exercise Science, North Dakota State University, Fargo, ND, USA; Wall and Berkel are with the Department of Counseling Psychology, University of Missouri-Kansas, Kansas, MO, USA; Peterson is with the School of Social Work, Rutgers University, New Brunswick, NJ, USA; Jitnarin, Haddock, and Poston are with the Institute for Biobehavioral Health Research, National Development and Research Institutes, Leawood, KS, USA; Eaton and Bowles are with the Department of Economics, University of Missouri, Kansas, MO, USA.

Correspondence: Katie M. Heinrich, PhD., John A. Burns School of Medicine, Department of Public Health Sciences, University of Hawaii, 1960 East West Road, Biomed D201, Honolulu, HI 96822, USA. (E-mail: kmhphd@gmail.com)

code or census tract), and emphasized the need for reliable, direct observation instruments.^{6,8}

Direct assessments enable researchers to observe and measure behavior settings first hand. However, previous direct audit instruments are limited because they used street segments (i.e., the portion of a street between two consecutive cross-streets), and not parcels (i.e., legal units into which property is divided for sale and purchase), as their observation unit, resulting in less detailed information about the built environment^{9–11} including restrictions on the number or combination of possible institutional uses per observation. This paper describes a novel direct observation parcel-level measure, the Census of Social Institutions (CSI). The CSI catalogs all uses and conditions of non-residential parcels, and can even note multiple or mixed uses within establishments or parcels.

METHODS

Study Design

The CSI was developed as part of a larger multi-level study examining environmental risk factors of obesity that employed a cross-sectional three-group nested design in a five-county region from 2003–2009. A total of 21 census block groups were randomly selected from >325, stratified by aggregate income tertiles from the 2000 US Census, to ensure adequate sample size and power for the parent study.¹² Inclusion criteria included (1) being within one standard deviation of the tertile group's average median household income, (2) being non-contiguous, (3) $\geq 19\%$ of residents from an ethnic minority group, and (4) >400 residents.

Measure Development

The CSI coding scheme was developed to be compatible with the North American Industry Classification System (NAICS; <http://www.census.gov/epcd/www/naics.html>). The CSI was used to identify broad *Major Use Types*—principal land use (e.g., commercial) and subsequent *Categories* (e.g., retail trade) and *Subcategories* (e.g., financial institutions). However, the CSI differed from NAICS in three important ways. First, the CSI was developed as a parcel-based, GIS-linked observation procedure that could be tailored to any geographic area where parcel files were available. Second, using many micro-level *Detail codes*, the CSI could identify single or multiple detailed uses within parcels. Third, the CSI recorded a mix of major use types within the same parcel (e.g., commercial and public commons: public park with a concession stand) using major use type, category, or subcategory.

CSI Details

After selecting the correct area, street, block, and parcel number, the CSI was used to verify or correct each parcel's street address. The major use type was selected from the options shown in Table 1.

Next, the buildings on each parcel were counted and the structure profile (i.e., number of building stories) was selected. Individual establishment(s) name(s) were noted if available (e.g., Memorial Sports Complex, Central Park) and all relevant detail use codes were assigned. Multiple categories, subcategories, and detail codes also were assigned to a parcel as appropriate, depending on the number of uses and

TABLE 1 CSI code categories: numbers of major types, categories, subcategories, detail codes and illustration for application of CSI codes

Major use type	Category	Subcategory	Detail code
CSI code categories			
1 Residential—built as a residence, but showing non-residential use (e.g., home daycare, office) coded as mixed major use type, due to residential use paired with a non-residential use)	N=2 Categories	N=2 Subcategories	N=7 Detail codes (not assessed unless mixed major use)
2 Commercial—retail or commercial use	N=7 Categories	N=25 Subcategories	N=266 Detail codes
3 Industrial—light/heavy industrial use	N=7 Categories	N=13 Subcategories	N=67 Detail codes
4 Community institutional—Civic/public/private institutions (e.g., schools, places of worship, libraries, social services, government offices)	N=9 Categories	N=9 Subcategories	N=64 Detail codes
5 Public commons—parks, plazas, and other public spaces (e.g., playgrounds)	N=1 Category	N=3 Subcategories	N=14 Detail codes
6 Private residential commons—private park, field, or other green space	N=1 Category	N=1 Subcategory	N=1 Detail code
7 Parking—surface/structured parking	N=2 Categories	N=2 Subcategories	N=4 Detail codes
8 Vacant lot (non-residential)—no principal, classifiable structure exists on the parcel, but its apparent intended use is non-residential	N=1 Category	N=1 Subcategory	N=1 Detail code
9 Vacant building—unable to determine use of building because of vacant state	N=1 Category	N=1 Subcategory	N=1 Detail code
10 Miscellaneous—structure type does not conform to any specified class	N=1 Category	N=1 Subcategory	N=3 Detail codes
11 Under construction—site is under construction	N=1 Category	N=1 Subcategory	N=1 Detail code
12 Mixed major use—a combination of two or more major use types from above	N=1 Category	N=1 Subcategory	N=1 Detail code
Illustration for application of CSI codes			
12 Mixed major use	1290 Mixed use	12990 Mixed use	129991 Mixed use
2 Commercial	270 Services	2711 Legal services	27111 Private law firm
	270 Services	2710 Health services	27111 Medical imaging
4 Community institution	482 Educational institution	4324 Miscellaneous educational institution	48241 Public library

whether they were of a single major use type (i.e., multiple use) or of more than one major use type (i.e., mixed major use).

Procedure

A 1-mile radius circle was drawn around the centroid of each of the 21 census block groups as a proxy for neighborhood and to approximate frequently traveled areas for residents.¹³ Each 1-mile radius circle was divided into seven rectangular parcel maps, which were used by data collectors for the audit.

All data collectors were trained over 3-day periods in use of handheld computer devices, definitions of CSI terms (e.g., parcel ID, field address), and the use of parcel maps. Data collectors completed two trial data collections in one community training area. Each team was provided a cellular telephone, magnetic automobile signs, safety vests, driving maps, and a letter explaining the study.

Teams of three data collectors coded parcels in a systematic fashion. After driving each street segment to assess safety and the presence of any unmapped parcels, observations were conducted on foot. If parcel uses were not fully visible from the exterior of the parcel's building(s), team members entered the building(s) as necessary. One member of the team recorded data for a given parcel and orally confirmed what codes were recorded. Each parcel was coded for major use type; category, subcategory, and detail code(s). Before leaving each parcel, all codes were reviewed. On average, parcel assessment took five minutes. The few larger office buildings with many uses took considerably longer and are not included in this average. Data verification/quality control procedures were implemented including resurveying areas that contained missing or conflicting information (e.g., multiple use parcel with only one detail code) were surveyed a second time, ensuring accurate recording of all visible parcel uses, and that data were consistent with screen options.

Statistical Analyses

We computed intra-class correlations (ICCs) for major use type and detail codes using one-way random effects models as the primary analysis. Because data were polychotomous ordinal, calculating reliabilities for category and subcategory would have been redundant, given the hierarchical nature of the coding scheme. One-way models, which combined variability due to raters, interactions between raters and the objects rated, and measurement error into within-person variability,¹⁴ were the most appropriate method for computing ICCs because we had over 25 different observers evaluating parcels due to staff turnover. We also examined ICCs for two-way mixed and random effects models. All statistical analyses were conducted using SPSS 16.0.

RESULTS

Data were from 10,842 parcels with a non-residential use on 10,211 street segments covering 56.52 square miles. Vacant lots (65.0%, $n=7,046$) were the most frequent major use type followed by commercial ($n=1,949$), community institutional ($n=566$), industrial ($n=271$), parking ($n=218$), public commons ($n=183$), under construction ($n=51$), vacant buildings ($n=37$), and miscellaneous ($n=2$); 88 parcels were missing data. Multiple major use type parcels represented 11.6% ($n=1,259$) of the total (over half were multiple commercial use), including 431 (4.0%) mixed major use type parcels. Per neighborhood, the number of multiple major use type parcels

ranged from 5 to 138 (mean=39.4) and mixed major use type parcels ranged from 2 to 68 (mean=20.5).

Reliability analyses addressed consistency of observations over a 1-year interval. A total of 1,093 observations for repeat assessments were examined, representing a 10% random subsample. ICCs representing 1-year test-retest and inter-rater reliability for major use type, and primary detail code were 0.637 (0.601-0.671) and 0.647 (0.612-0.681), respectively, for the one-way random effects model. For the two-way random and mixed models, the ICCs were 0.638 (0.601-0.672) for major use type and 0.648 (0.612-0.681) for detail code, respectively. All ICCs were significantly different from zero ($p < 0.001$) with F test values ranging from 4.5 to 4.8. Even if our data had met requirements for two-way ICC models (i.e., we had reason to assume ratings could be associated with particular raters), our results demonstrate that the ICCs would essentially be the same as for our one-way models.

DISCUSSION

The CSI demonstrated robust test-retest and inter-rater reliability at the parcel level. Given the high number of staff members involved in the study and the 1-year period between ratings, it is remarkable that major use type and detail code ICCs were so high. In addition, the CSI reliably detected a substantial percentage of multiple uses within given establishments. The 1,259 parcels with multiple and mixed uses in our study would have been miscoded with instruments that capture only a single use per parcel, under-representing their actual uses.

The CSI directly addresses previous concerns¹ by employing direct measurements of institutional presence across jurisdictions.⁵ Although the CSI is not the first direct observation measure, the utility of using handheld computers for data collection helps reduce missing data, improve data quality, reduce data processing time and costs, and enhance research capacity.¹⁵ The CSI also provides primary observation data on spatiophysical variables at the microscale,³ are disaggregated,^{6,8} and can be matched to multiple spatial area boundaries and combined with individual level data or census characteristics.

The CSI is limited by the same factors that limit all direct observation measures. The staffing required to conduct parcel-level observations can be prohibitive from both a cost and personnel perspective. However, using the CSI on random subsets of geographic areas to verify database accuracy would be an appropriate use when adequate funding or staffing is a concern.⁷ Of importance is the extent to which physical environments remain stable over time. We used a 1-year interval, but other intervals might be useful for other geographic scales.

Our findings show that it is reasonable to expect parcel-level accuracy in audits of physical environments. Future built environment studies should employ direct observation measures like the CSI to confirm environmentally based risk and protective factors, particularly those examining the impact of the built environment on obesity risk.

ACKNOWLEDGMENT

This study was supported by a grant from the National Institute of Diabetes and Digestive and Kidney Diseases awarded to Dr. Poston (R01DK064284) and a NIH Minority Supplement to Dr. Berkel (DK064284-S).

We would like to acknowledge L. Carrin Parker, Ph.D., Natalie Heavilin, B.A., Marc Husted, B.A., and the other research assistants who helped collect CSI data for this study.

REFERENCES

1. Booth KM, Pinkston MM, Poston WS. Obesity and the built environment. *J Am Diet Assoc.* 2005; 105(5 Suppl 1): S110–S117.
2. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am J Prev Med.* 2009; 36(4S): S99–S123.
3. Moudon AV, Lee C. Walking and bicycling: an evaluation of environmental audit instruments. *Am J Health Promot.* 2003; 18(1): 21–37.
4. Berrigan D. Built environment and health. *Prev Med.* 2008; 47(3): 239–240.
5. Northridge ME, Sclar ED, Biswas P. Sorting out the connections between the built environment and health: a conceptual framework for navigating pathways and planning healthy cities. *J Urban Health.* 2003; 80(4): 556–568.
6. Papas MA, Alberg AJ, Ewing R, Helzlouer KJ, Gary TL, Klassen AC. The built environment and obesity. *Epidemiol Rev.* 2007; 29: 129–143.
7. Boone JE, Gordon-Larsen P, Stewart JD, Popkin BM. Validation of a GIS facilities database: quantification and implications of error. *Ann Epidemiol.* 2008; 18(5): 371–377.
8. Black JL, Macinko J. Neighborhoods and obesity. *Nutr Rev.* 2008; 66(1): 2–20.
9. 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.
10. Brownson RC, Hoehner CM, Brennan LK, et al. Reliability of two instruments for auditing the environment for physical activity. *J Phys Activ Health.* 2004; 1: 189–207.
11. Zenk SN, Schulz AJ, Mentz G, et al. Inter-rater and test-retest reliability: methods and results for the neighborhood observational checklist. *Health Place.* 2007; 13(2): 452–465.
12. Murray D. *Design and analysis of group-randomized trials.* New York: Oxford University Press; 1998.
13. McGinn AP, Evenson KR, Herring AH, Huston SL, Rodriguez DA. Exploring associations between physical activity and perceived and objective measures of the built environment. *J Urban Health.* 2007; 84(2): 162–184.
14. Nichols D. Choosing an intraclass correlation coefficient. Available at: <http://support.spss.com/ProductsExt/SPSS/Documentation/Statistics/articles/whichicc.htm>. Accessed 1 June 2009.
15. Gravlee CC, Zenk SN, Woods S, Rowe Z, Schulz AJ. Handheld computers for direct observation of the social and physical environments. *Field Methods.* 2006; 18: 382–394.