



NIH PUBLIC ACCESS

Author Manuscript

Ann Epidemiol. Author manuscript; available in PMC 2009 May 1.

Published in final edited form as:

Ann Epidemiol. 2008 May ; 18(5): 371–377.

Validation of a GIS facilities database: quantification and implications of error

Janne E. Boone, MPH^{*,†}, Penny Gordon-Larsen, PhD^{*,†}, James D. Stewart, MA[†], and Barry M. Popkin, PhD^{*,†}

^{*} Department of Nutrition, Schools of Public Health & Medicine, University of North Carolina at Chapel Hill; Chapel Hill, NC, USA

[†] Carolina Population Center, University of North Carolina at Chapel Hill; Chapel Hill, NC, USA

Abstract

Purpose—To validate a commercial database of community-level physical activity facilities that can be used in future research examining associations between access to physical activity facilities and individual-level physical activity and obesity.

Methods—Physical activity facility characteristics and locations obtained from a commercial database were compared to a field census conducted in 80 Census block groups within two U.S. communities. Agreement statistics, agreement of administratively-defined neighborhoods, and distance between locations were used to quantify count, attribute, and positional error.

Results—There was moderate agreement (concordance: non-urban: 0.39; urban: 0.46) of presence of any physical activity facility and poor to moderate agreement (kappa range: 0.14 to 0.76) of physical activity facility type. The mean Euclidean distance between commercial database versus field census locations was 757 and 35 meters in the non-urban and urban communities, respectively. However, 94 and 100% of non-urban and urban physical activity facilities, respectively, fell into the same 5-digit ZIP code, dropping to 92 and 98% in the same block group and 71% along the same street.

Conclusions—Our findings suggest that the commercial database of physical activity facilities may contain appreciable error, but patterns of error suggest that built environment-health associations are likely biased downward.

Keywords

Geographic Information Systems; Validation Studies; Environment Design

INTRODUCTION

Research demonstrating statistical relationships, particularly in adolescents (1–3), between obesity and physical activity and aspects of the built environment, such as nearby retail destinations or parks and open space (4–7), has facilitated a growing interest in spatial analyses using geographic information system (GIS) technology.

Please Address Correspondence & Reprint Requests To: Penny Gordon-Larsen, Ph.D., University of North Carolina at Chapel Hill, Carolina Population Center, University Square, 123 West Franklin St., Chapel Hill, NC 27516-3997, Phone: (W) 919-843-9966; (H) 919-968-6594; Fax: 919-966-9159, Email: gordon_larsen@unc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

GIS-derived measures of the built environment enable examination of built environment effects in large population studies because such measures do not rely on resource-intensive neighborhood audits or other forms of direct observation (8) and avoid limitations of perceived measures of the environment (6,9,10). However, large-scale GIS's rely on existing databases of built environment components such as retail food outlets and recreation facilities.

Improved understanding of the implications of such analyses requires quantification of the degree of error in GIS data and ultimately the potential bias that such error introduces to environment-health associations. Errors may include incomplete data (count error), inaccurate classification of facilities or characteristics (attribute error), or inaccurate geocoded location (positional error) (Figure 1).

Despite the recognition of these potential errors and considerable growth in the use of GIS technology in health research, existing validation studies generally focus on the positional error of geocoded residential addresses (11–16). Count and attribute error in databases of community resources, as opposed to residential addresses that are managed in GIS programs have not been assessed.

In this study, we quantify the count, attribute, and positional error in a commercial database of physical activity facilities in two communities to be used in future research examining adolescent physical activity in relation to the built environment. While this validation study focuses on physical activity resources, it will inform future work with the analogous diet environment as well as broader built environment research using commercial databases.

METHODS

Study Sample

The sample included 40 census block groups within each of two distinct geographic locations selected on the basis of urbanicity, representativeness, and safety concerns. The non-urban location was described by field staff as a “rural town,” with a population density of <2,000 persons/mi². In contrast, the urban location was extremely dense (average >50,000 persons/mi²), comprised of block groups small in area, with a smaller proportion of racial/ethnic minorities and relatively low median income (Table 1).

Commercial Database

The commercial database contained information on licensed businesses in the U.S. as of August 18, 2005, coded to eight-digit Standard Industrial Classification (SIC) codes. The vendor provided data (n=384 facilities) including businesses (1) with any of 6 SIC codes corresponding to a list of 169 SIC codes representing physical activity facilities or with facility name containing YMCA or YWCA text phrases; and (2) that fell within ZIP codes containing the 80 block groups described above. The list of SIC codes was generated collaboratively by the investigators and a vendor representative to ensure coverage of all physical activity resources and facilities. The vendor queries data by ZIP code, so facilities in ZIP codes corresponding to the 80 selected block groups were requested. Because park addresses in the commercial database correspond to their usually off-site administrative office or commercial entity contained in the park, parks data are more appropriately obtained from other databases and were excluded from this analysis.

Street address format was standardized by the commercial vendor, which facilitated the geocoding process. Preference was given to facility locations geocoded as a street segment match using ArcView 3.3 with default settings (spelling sensitivity=80, minimum match score=60, minimum score to be considered a candidate=30), 10-foot road offset, and Tele Atlas Dynamap/2000 for Download v15.1 (April 2005) street data (17), which is updated quarterly.

An initial batch match was followed by an interactive session to resolve unmatched addresses, yielding 92.2% perfect matches and 2.6% matches with scores of 80–85 for an overall street segment match rate of 94.8% among the 384 facilities received from the vendor. Unmatched addresses resulted from street numbers outside the address range of the street segments, unmatched street name, mismatched street types, or similar issues. For the remaining 5.2% of facilities, latitude and longitude coordinates provided by the commercial vendor were used, yielding a total of 100% geocoding match rate of the records returned from the commercial vendor.

Following geocoding, a geoprocessing procedure was conducted to “clip” ZIP code polygon boundaries (geographic unit of commercial database) to block group boundaries (geographic unit of field census), resulting in 173 facilities within the study area. Facilities with duplicate identifiers or addresses were eliminated; for each duplicate set, facility type classification was identical and one record was retained, leaving a total of 161 facilities.

Field Census

A full field census of physical activity facilities in the two selected locations was performed from August–October 2005. The field team identified and classified each physical activity facility as one or more pre-specified types (Table 2), obtained curbside Global Positioning System (GPS) readings and photographs, and recorded other pertinent details. GPS measurements were obtained using a Leica GS20 with differential correction and manual corrections for offsets, yielding locations accurate to approximately one to two meters. The field census was treated as the criterion measure.

Count, attribute, and positional errors in the commercial database were quantified by comparing to the field census overall and by facility type.

Data Analysis

Descriptive and statistical analyses were performed in Stata 9.2 (18) and Microsoft Access 2003 (19); spatial analysis was performed in ArcGIS 9.1 (20).

Count error was assessed by classifying each facility according to validity in the field and inclusion in the commercial database. In the field, a facility was classified as an invalid physical activity resource for adolescents if it was not found, not judged to be a recreation facility (e.g., an administrative office), or relevant only for very young children (e.g., playgrounds or other facilities for young children). Concordance was calculated as the number of facilities in agreement divided by the number of facilities. Kappa, a chance-corrected measure of agreement (21), sensitivity, and specificity were calculated using agreement statistic functions in Stata. Because facility types were not mutually exclusive, type-specific statistics may include overlapping resources. Averages weighted by type-specific field census counts provided global measures of agreement. “Insufficient cell size” indicates fewer than two commercial database or field facilities indicated for any given facility type.

Positional error was assessed using Euclidean and network distances calculated based on Universal Transverse Mercator (UTM) coordinates; ArcGIS Network Analyst (22) was used to calculate network route differences, hereafter referred to as “network distances”. To assess the impact of positional error on physical activity facility counts within neighborhoods defined by census geographies or ZIP codes, US Census geography shapefiles (2005), ESRI Streetmap USA files (2004), ZIP code (2005) and physical activity facility GPS and geocoded locations were spatially matched to determine the 5-digit ZIP code, census tract, block group, and nearest street upon which the geocoded and GPS locations fell. Facility coordinates were re-projected from UTM to the geographic coordinate system to match the other GIS layers. The street upon

which geocoded and GPS coordinates were located was considered to match if: (1) the street segment identifier or (2) the street name and type for the nearest street segment matched. Locations were compared for each facility and summarized descriptively.

RESULTS

Count and Attribute Error

All facilities combined—The commercial database and field census showed moderate agreement of presence of any facility, indicated by an overall concordance of 0.42 (Table 3). Kappa and specificity were not valid for overall counts because counts of facilities found in neither the commercial database nor the field census were, by definition, zero. Disagreement resulted primarily from facilities identified in the field but not the commercial database (90 facilities) or facilities in both data sources but not judged as valid recreation facilities (40 facilities), as opposed to facilities in the commercial database not found in the field (16 facilities) (data not shown).

Type-Specific Agreement—Agreement of facility presence by facility type was assessed by comparing the GIS versus field census facility type classifications. Agreement was high, with concordance and specificity 0.82 and 0.90 or greater, respectively. Sensitivity ranged from 0.10 to 0.79 for non-urban and urban facilities combined. In general, kappa statistics indicated slight to substantial agreement, ranging from 0.14 to 0.76. Agreement was generally higher for non-urban than urban facilities. The weighted average (across facility types) of type-specific agreement statistics are listed in Table 3.

Positional Error

Euclidean distances between the commercial database (geocoded address) and field locations (GPS) were substantially larger for non-urban than urban facilities (mean Euclidean distance: non-urban, 757 meters vs. urban, 35 meters), and for large facilities such as golf courses and schools (e.g., mean Euclidean distance: outdoor non-urban, 3,235 meters vs. member non-urban, 302 meters) (Table 4). Network distances were larger but mirrored Euclidean distances (Table 4).

Because many studies use administrative boundaries such as ZIP code or Census tract to define neighborhoods, understanding the impact of positional errors on facility counts within these areas is important. In the non-urban and urban communities, 94 and 100% of geocoded and GPS locations, respectively, fell into the same 5-digit ZIP code and Census tract, dropping to 92 and 98% in the same same block group (Table 5). Further, for 71% of facilities, geocoded and GPS locations fell on the same street (Table 5), suggesting that most positional error resulted from the interpolation of points along the street segment during the geocoding process.

While these results are promising, they include only those facilities identified in both the commercial database and field census. Observed count errors, however, indicate that the commercial database provided an undercount of facilities. Table 6 presents field counts within block groups containing the specified number of physical activity facilities. For example, block groups containing one physical activity facility according to the commercial database actually contained 2.6 and 2.0 facilities in the non-urban and urban community, respectively.

DISCUSSION

Despite rapidly growing use of GIS's in health research and concerns about GIS accuracy, published validation studies of GIS resource data are scant. In this study, we compared a commercial database of physical activity facilities to a field census in two communities. We

found moderate overall agreement, with the main sources of error resulting from facilities not included or misclassified as physical activity facilities in the commercial database.

Classification of physical activity facility type was similar according to the commercial database and field team. While the positional error was large for some facilities, both locations fell into the same Census block group for 95% of facilities.

Completeness and Attribute Accuracy

Discordance in the presence of physical activity facilities, regardless of type, resulted from three types of error: (1) facilities included in the commercial database but not found in the field; (2) facilities included in the commercial database but not identified as valid physical activity facilities in the field; and (3) facilities identified in the field but not in the commercial database. The latter two errors were far more common than the first. Because only records matching pre-specified definitions of physical activity facilities were obtained from the commercial vendor, we were not able to investigate if these facilities were contained in the database but not captured by the *a priori* classification scheme. Administrative offices for physical activity facilities, in the absence of the actual facility itself, were the primary contributors to the second error type. Refinement of the classification scheme for physical activity facilities could potentially reduce both types of error.

Positional Accuracy

Positional error ranged from negligible (5 meters Euclidean distance) to exceptional (almost 20,000 meters, in the case of a golf course) and was larger than distances observed in residential address geocoding validation studies (11–15). However, this study examined physical activity facilities, which are often large in size and could have ambiguous point locations (e.g., golf courses, which cover a relatively large geographic area), rather than residences. Indeed, the facility with the largest positional error was a golf course with a long driveway. Positional error for smaller facilities such as YMCA and member facilities was more consistent with error of residential locations.

For any given facility, GPS and geocoded locations may fall far apart but within the same neighborhood. In studies that use administratively defined neighborhoods, this scenario would neither alter facility counts within neighborhoods nor introduce bias to associations with health-related outcomes. In this study, commercial database and field locations fell into the same block group for 95% of facilities.

The discrepancy between large distance measures and low count error is consistent with the nature of the block group boundaries. Block groups generally contain 300–3000 individuals, so block group size increases with decreasing population density. The largest distances were found in the non-urban community in which block groups averaged 3.7 square miles versus 0.03 square miles in the urban community, yet the larger block group areas increased the likelihood that both locations would fall into the same block group. Large distance measures may have more potential for bias when using buffer-defined neighborhoods with a large radius (e.g., 5-mile radius around each respondent's residence); however, investigation into this issue requires respondent locations and is thus outside the scope of this study.

The assumption that address numbers are uniformly distributed along a street segment is a known source of error in the geocoding process. Further, street segments are shorter in urban areas than less populated areas. Indeed, the commercial database and field locations fell along the same street for 71% of facilities in both communities, while the distances are substantially larger in the non-urban than urban community. That much of the error occurs along the street segment is a promising finding because it reduces the potential network distance between

geocoded versus actual location. It is also likely to be random with respect to resident characteristics, thus biasing built environment-health outcome associations toward the null.

Field counts calculated within a given commercial database count face the issues of completeness as well as positional accuracy. Our commercial database generally underestimated the number of physical activity facilities within a block group, particularly in non-urban communities. However, commercial database and field counts increased together, albeit not in parallel, so rank orders of physical activity counts would be preserved. This error is likely to result in nonsystematic misclassification and thus attenuate associations between physical activity facilities and health outcomes.

Urbanicity Differences

While some differences such as facility size and geocoded street segment length are characteristic of any non-urban versus urban area, other differences were imposed by the study design. Because our field team confined their census area to specified block groups encompassing a larger area in the non-urban community, the maximum distance between the field and geocoded locations was larger and the team was more likely to locate a facility listed in the commercial database in the non-urban relative to the urban community. These considerations are consistent with generally larger positional error but lower count error in non-urban than urban communities. However, few facilities contained in the commercial database were not located in the field, so the latter issue likely has minimal impact.

Strengths & Limitations

While including two dissimilar communities provides validity data for both an urban and non-urban area, these communities are not representative of the range of geographic areas in the U.S and span a small range of urbanicity. The small number of facilities may have resulted in unstable estimates of error, but overall patterns were consistent, suggesting that our estimates are reasonably reliable and valid.

Errors observed in this study can be attributed to several sources. First, it is possible that facilities found in the field but not in the commercial database were new facilities not yet included in the database. We do not currently have access to more recent data to investigate this issue, but this source of error can be expected in any study using commercial data and is therefore important to capture in this validation study. Second, because facilities contained in the commercial database used in this study depended on both the source data as well as the investigator-supplied *a priori* classification strategy, it is difficult to determine the contribution of each to the observed errors. Third, disagreement of the street nearest to GPS versus geocoded locations can be attributed to error in the geocoded coordinates or in the street file. In general, errors observed in this study were reasonable, so these limitations should not reduce confidence in the utility of commercial databases for built environment research.

Geocoding of facility locations yielded a high match rate (94.8%), but for 5.2% of facilities, geocoded coordinates provided by the commercial vendor were used. However, comparison of positional error calculated from the two sources show only small (7.5 meters) difference in mean error, so use of two sources of geocodes is unlikely to impact our results.

Finally, this study assesses one potential GIS layer among many. Evaluation of other GIS data sources is required to more fully describe GIS data validity and potential bias in other GIS-derived measures of the built environment.

Despite these limitations, this study is the first to estimate completeness and attribute and positional accuracy of GIS physical activity facilities data. Criterion measures were provided by a detailed field census and GPS locations accurate to within one to two meters.

Conclusions

Overall, our validation study findings suggest that the commercial database of physical activity facilities may contain appreciable error, but the random nature of the sources of error and small error in neighborhood facilities counts suggest that bias to associations with health-related outcomes is probably small and toward the null. Our findings do not suggest that existing associations between aspects of the built environment and health related outcomes should be called into question.

An important next step is the application of validation findings to individual-level respondent locations and health data, allowing estimation, through simulation or empirical comparisons, of the magnitude of bias to the association between physical activity facilities and health outcomes introduced by this error.

Acknowledgements

The funding of this study comes from the National Institutes of Health NICHD (K01 HD044263-01), NIDDK (DK56350), NIEHS (P30ES10126) and NIA (P30AG024376). The authors would like to thank Allen Serkin, Peter Zambito, Evan Hammer, and Brian Frizzelle of the University of North Carolina, Carolina Population Center, Spatial Analysis Unit for collecting and managing the field census and commercial data, for their assistance and expertise regarding spatial data and analysis, and for their comments on the manuscript. The authors also thank Ms. Frances Dancy for her helpful administrative assistance. There were no potential or real conflicts of financial or personal interest with the financial sponsors of the scientific project.

ABBREVIATIONS

GIS	Geographic Information System
GPS	Global Positioning System
SIC	Standard Industrial Classification codes
UTM	Universal Transverse Mercator
YMCA	Young Men's Christian Association
YWCA	Young Women's Christian Association

References

1. 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;117:417–424. [PubMed: 16452361]
2. Evenson KR, Birnbaum AS, Bedimo-Rung AL, Sallis JF, Voorhees CC, Ring K, et al. Girls' perception of physical environmental factors and transportation: reliability and association with physical activity and active transport to school. *Int J Behav Nutr Phys Act* 2006;3:28. [PubMed: 16972999]
3. Cohen DA, Ashwood JS, Scott MM, Overton A, Evenson KR, Staten LK, et al. Public parks and physical activity among adolescent girls. *Pediatrics* 2006;118:e1381–e1389. [PubMed: 17079539]
4. Li F, Fisher KJ, Brownson RC, Bosworth M. Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *J Epidemiol Community Health* 2005;59:558–64. [PubMed: 15965138]

5. King WC, Belle SH, Brach JS, Simkin-Silverman LR, Soska T, Kriska AM. Objective measures of neighborhood environment and physical activity in older women. *Am J Prev Med* 2005;28:461–9. [PubMed: 15894150]
6. Duncan M, Mummery K. Psychosocial and environmental factors associated with physical activity among city dwellers in regional Queensland. *Prev Med* 2005;40:363–72. [PubMed: 15530589]
7. Handy S, Cao X, Mokhtarian PL. Correlation or causality between the built environment and travel behavior? Evidence from Northern California Transportation Research Part D 2005;10:427–444.
8. Giles-Corti B, Donovan RJ. The relative influence of individual, social and physical environment determinants of physical activity. *Soc Sci Med* 2002;54:1793–812. [PubMed: 12113436]
9. Owen N, Humpel N, Leslie E, Bauman A, Sallis JF. Understanding environmental influences on walking; Review and research agenda. *Am J Prev Med* 2004;27:67–76. [PubMed: 15212778]
10. Huston SL, Evenson KR, Bors P, Gizlice Z. Neighborhood environment, access to places for activity, and leisure-time physical activity in a diverse North Carolina population. *Am J Health Promot* 2003;18:58–69. [PubMed: 13677963]
11. Bonner MR, Han D, Nie J, Rogerson P, Vena JE, Freudenheim JL. Positional accuracy of geocoded addresses in epidemiologic research. *Epidemiology* 2003;14:408–12. [PubMed: 12843763]
12. Zhan FB, Brender JD, De Lima I, Suarez L, Langlois PH. Match rate and positional accuracy of two geocoding methods for epidemiologic research. *Ann Epidemiol* 2006;16:842–9. [PubMed: 17027286]
13. Strickland MJ, Siffel C, Gardner BR, Berzen AK, Correa A. Quantifying geocode location error using GIS methods. *Environ Health* 2007;6:10. [PubMed: 17408484]
14. Zandbergen PA. Influence of geocoding quality on environmental exposure assessment of children living near high traffic roads. *BMC Public Health* 2007;7:37. [PubMed: 17367533]
15. Schootman M, Sterling DA, Struthers J, Yan Y, Laboube T, Emo B, et al. Positional Accuracy and Geographic Bias of Four Methods of Geocoding in Epidemiologic Research. *Ann Epidemiol* 2007;17:464–470. [PubMed: 17448683]
16. Cayo MR, Talbot TO. Positional error in automated geocoding of residential addresses. *Int J Health Geogr* 2003;2:10. [PubMed: 14687425]
17. ESRI. Tele Atlas Dynamap/2000 for Download v15.1. 2005. <http://www.esri.com/data/download/teeatlas/>
18. Stata statistical software: STATA/SE for Unix. 9.1. College Station, TX: Stata Corporation; 2006.
19. Microsoft Office Access [computer program]. 2003. Microsoft Corporation 2003.
20. The Complete Geographic Information System [computer program]. 9.1. Redlands, CA: ESRI Corporation 2005.
21. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, For GG. Tips for learners of evidence-based medicine: 3 Measures of observer variability (kappa statistic). *Cmaj* 2004;171:1369–73. [PubMed: 15557592]
22. Network Analyst [computer program] 9.1. Redlands, CA: ESRI Corporation; 2005.



Figure 1. Illustration of count, positional, and attribute errors

Dark lines represent census block group boundaries; ★ GPS location • Geocoded location

A: Distance between geocoded and GPS locations is moderate, but they fall within the same block group. This error will contribute to positional error but not facility count within this block group.

B: Geocoded and GPS locations are very close, contributing little positional error and no error to facility count. If the commercial database classifies this facility as a member facility, while the field census deemed it a public facility, it would contribute to attribute error.

C: Distance between geocoded and GPS locations is moderate, but they fall into different block groups. This error will contribute to positional and facility count error within this block group. Further, if the upper right block group fell outside of the study area, this error will also contribute to the total count error.

Table 1

Characteristics of two validation communities*

	Non-Urban	Urban
Block group area (mile ²) [mean [†] (SD)]	3.7 (8.6)	0.03 (0.02)
Block group population [mean [†] (SD)]	1,753 (948)	1,213 (782)
Population density (persons/mile ²) [mean [†] (SD)]	1,634 (1,152)	58,581 (35,285)
White non-Hispanic population [mean [†] (SD)]	960 (673)	815 (768)
Median household income [mean [†] (SD)]	\$40,157 (21,794)	\$33,925 (22,059)

SD=Standard Deviation

* Field census of physical activity facilities located in 40 Census block groups in each of two communities conducted in 2005

[†] unweighted average among block groups, using 2000 Census data

Table 2

Physical activity facility types and descriptions

Type	Description
All facilities combined	Aggregate measure across all 9 categories
Instruction	Dance studios, basketball instruction, martial arts
Member	Athletic club and gymnasium, tennis club, basketball club
Outdoor	Sporting and recreation camps, swimming pools
Public facilities	Public beach, pools, tennis courts, recreation centers
Public fee facilities	Physical fitness facilities, bicycle rental, public golf courses
School	Public and private schools and colleges
YMCA	YMCA, YWCA
Youth Organization	Boy/Girl Scouts, youth centers

Table 3

Agreement of facility presence reported by commercial database versus field census

	Census (count)	Agree (count)	Disagree (count)	Concordance*	Kappa (SE)	Specificity (95% CI)	Sensitivity (95% CI)
All facilities combined							
Non-urban & Urban	195	105	146	0.42	--†	--†	0.54 (0.48, 0.60)
Non-urban	124	63	97	0.39	--†	--†	0.51 (0.43, 0.59)
Urban	71	42	49	0.46	--†	--†	0.59 (0.49, 0.69)
Instruction							
Non-urban & Urban	35	228	23	0.91	0.56 (0.06)	0.97 (0.95, 0.99)	0.51 (0.45, 0.58)
Non-urban	24	145	15	0.91	0.60 (0.08)	0.96 (0.93, 0.99)	0.58 (0.51, 0.66)
Urban	11	83	8	0.91	0.46 (0.10)	0.99 (0.96, 1.01)	0.36 (0.26, 0.46)
Member							
Non-urban & Urban	25	228	23	0.91	0.39 (0.06)	0.97 (0.95, 0.99)	0.36 (0.30, 0.42)
Non-urban	19	142	18	0.89	0.34 (0.08)	0.96 (0.94, 0.99)	0.32 (0.24, 0.39)
Urban	6	86	5	0.95	0.52 (0.10)	0.98 (0.95, 1.01)	0.50 (0.40, 0.60)
Outdoor							
Non-urban & Urban	13	240	11	0.96	0.54 (0.06)	0.98 (0.96, 1.00)	0.54 (0.48, 0.60)
Non-urban	12	151	9	0.94	0.58 (0.08)	0.97 (0.95, 1.00)	0.58 (0.51, 0.66)
Urban	1	89	2	--§	--§	--§	--§
Public							
Non-urban & Urban	20	231	20	0.92	0.14 (0.05)	0.99 (0.98, 1.00)	0.10 (0.06, 0.14)
Non-urban	17	142	18	--§	--§	--§	--§
Urban	3	89	2	0.98	0.66 (0.10)	0.99 (0.97, 1.01)	0.67 (0.57, 0.76)
Public Fee							
Non-urban & Urban	19	239	12	0.95	0.64 (0.06)	0.98 (0.96, 1.00)	0.63 (0.57, 0.69)
Non-urban	16	151	9	0.94	0.68 (0.08)	0.97 (0.95, 1.00)	0.69 (0.62, 0.76)
Urban	3	88	3	0.97	0.38 (0.10)	0.99 (0.97, 1.01)	0.33 (0.24, 0.43)
School							
Non-urban & Urban	73	222	29	0.88	0.72 (0.06)	0.92 (0.89, 0.95)	0.79 (0.74, 0.84)
Non-urban	32	147	13	0.92	0.76 (0.08)	0.93 (0.89, 0.97)	0.88 (0.82, 0.93)
Urban	41	75	16	0.82	0.64 (0.10)	0.90 (0.84, 0.96)	0.73 (0.64, 0.82)
YMCA							
Non-urban & Urban	4	246	5	0.98	0.54 (0.06)	0.98 (0.97, 1.00)	0.75 (0.70, 0.80)
Non-urban	2	158	2	0.99	0.66 (0.07)	0.99 (0.97, 1.00)	1.00 (1.00, 1.00)
Urban	2	88	3	0.97	0.38 (0.10)	0.98 (0.95, 1.01)	0.50 (0.40, 0.60)
Youth Organization							
Non-urban & Urban	7	234	17	0.93	0.23 (0.06)	0.95 (0.92, 0.97)	0.43 (0.37, 0.49)
Non-urban	4	152	8	0.95	0.31 (0.07)	0.96 (0.93, 0.99)	0.50 (0.42, 0.58)
Urban	3	82	9	0.90	0.14 (0.09)	0.92 (0.86, 0.98)	0.33 (0.24, 0.43)
Weighted							
Non-urban & Urban	--†	--†	--†	0.91	0.55	0.99	0.59
Non-urban	--†	--†	--†	0.92	0.60	0.99	0.67
Urban	--†	--†	--†	0.87	0.56	0.97	0.64

SE=Standard Error; CI=Confidence Interval

* $\text{Concordance} = \frac{\# \text{Agree}}{\# \text{Agree} + \# \text{Disagree}}$

† Not applicable; by definition, facilities that are not in the commercial database or in the field census were not included in the study data

§ Not calculated due to insufficient cell size

‡ Weighted averages (type-specific field census counts) across facility types. Weighted counts are not applicable.

Table 4
Distance (meters) between geocoded position from commercial database and GPS position from field census, by facility type*

	Count	Euclidean Distance (meters)			Network Distance (meters)		
		Mean (SE)	Min	Max	Mean (SE)	Min	Max
All facilities combined							
Non-urban	63	757 (2,537)	10	19,546	895 (3,260)	2	25,347
Urban	42	35 (37)	5	141	40 (48)	3	185
Instruction							
Non-urban	14	944 (1,545)	12	4,474	1,046 (1,742)	11	4,818
Urban	4	26 (7)	21	37	34 (9)	26	45
Member							
Non-urban	6	302 (222)	124	587	318 (210)	126	607
Urban*	4	33 (30)	5	69	33 (30)	5	69
Outdoor*							
Non-urban	7	3,235 (7,205)	12	19,546	4,151 (9,360)	4	25,347
Public*							
Urban	2	54 (69)	5	103	78 (105)	4	153
Public Fee							
Non-urban	11	2,164 (5,779)	40	19,546	2,749 (7,509)	32	25,347
Urban	1	12 [†]	12	12	8 [†]	8	8
School							
Non-urban	28	303 (314)	27	1,508	331 (358)	27	1,594
Urban	30	37 (40)	5	141	41 (51)	3	185
YMCA							
Non-urban	2	163 (149)	58	268	163 (148)	58	268
Urban	1	5 [†]	5	5	4 [†]	4	4
Youth Organization							
Non-urban	2	163 (149)	58	268	163 (148)	58	268
Urban	2	54 (69)	5	103	78 (105)	4	153

Min=Minimum; Max=Maximum; SD=Standard Deviation

* Urban Outdoor facilities and non-urban Public facilities, and are not reported because none were found in both the GIS and field census. Distance calculated among facilities in both the GIS and field census.

[†] Count was calculated from a single facility, so SD is not applicable.

Table 5
 Agreement of administratively defined neighborhood and nearest street of GPS and geocoded physical activity facility locations.*

Community	# Facilities	5-Digit ZIP Code	Census Tract	Block Group	Street
Non-Urban	63	59 94%	59 94%	58 92%	45 71%
Urban	42	42 100%	42 100%	41 98%	30 71%

* Agreement calculated among facilities in both the GIS and field census.

Table 6
Physical activity facility field census counts within Census block group, by geocoded counts

Facilities count in Block Group(Geocode)	# Block Groups	Non-Urban Mean Count(field)	SD	# Block Groups	Urban Mean Count(field)	SD
0	3	1.0	1.7	14	0.4	0.8
1	10	2.6	1.6	13	2.0	1.2
2	14	3.4	1.5	6	3.5	1.4
3	4	5.5	0.6	2	3.5	0.7
4	3	5.7	0.6	1	5.0	--*
5	1	10.0	--*	1	8.0	--*
6	5	12.6	6.5	3	11.3	0.6

SD=Standard Deviation

* Count was calculated from a single block group, so SD is not applicable.