UIC Neighborhoods + Health

Validity of secondary data sources to identify bicycle lanes in Chicago, IL, USA

Kelly K. Jones, Shannon N. Zenk
University of Illinois at Chicago
January 25, 2018

Acknowledgements

The authors thank Alana Steffen, PhD for assistance with power calculations and Sandy Slater, PhD for feedback on the street audit tool. The University of Illinois at Chicago Chancellor's Graduate Research Fellowship to the first author provided funding for this research.

Suggested Citation

Jones K, Zenk SN. (2018). Validity of secondary data sources to identify bicycle lanes in Chicago, IL, USA. Retrieved from Neighborhoods and Health website: https://neighborhoodsandhealth.uic.edu/.

Abstract:

High-quality data sources are critical for conducting rigorous research into associations between built environment features and active living. While street audits are recognized as the gold standard for such data, some studies must rely on secondary data sources, which may not have been adequately validated. This case study reports findings from a validity study of local government and OpenStreetMap as sources for data on the location of on-street bicycle lanes in Chicago, IL, USA. We found that local government data was valid, with high accuracy, sensitivity, positive predictive value, specificity, and negative predictive value. We found that OpenStreetMap had adequate accuracy, positive predictive value, specificity, and negative predictive value, but sensitivity was lower in this data source than in the local government source. We recommend using local government sources when secondary data is required.

Validity of secondary data sources to identify bicycle lanes in Chicago, IL, USA

How features of the built environment relate to active living is of interest to public health practitioners [1-4], as well as urban and transportation planners [5, 6]. High-quality data sources are critical to conduct rigorous research that will produce credible results and lead to effective land use and transportation policies. Measuring features of the built environment directly via direct observation is the recognized gold standard, and there are a wide variety of tools available to guide such street audits [1-4]. However, street audits require significant resources in both time and manpower, and as such are infeasible for many studies, particularly those covering large geographic areas. In addition, street audits cannot be performed retrospectively. Therefore, for some studies, secondary data are needed.

Two secondary sources of data for on-street bicycle lanes are local governments and OpenStreetMap (OSM). Local governments may provide data as shapefiles (data files used with geographic information system software), as features in online searchable maps, or as print maps [7]. OSM is a volunteered geographic information database with an Open Data Commons Open Database License to which users contribute and that can be accessed freely through the internet [8]. Data can be accessed directly through the website interface, or downloaded with the use of commercial or open source geographic information software. When choosing to use either data source, researchers must consider reliability (to ensure consistency of identification over time and space) and validity (to ensure that the intended construct is being identified) [9].

This short case study reports findings from a validity study of local government and OSM as sources for data on the location of on-street bicycle lanes in Chicago, IL. Chicago is a Midwestern US city with approximately 2.7 million residents living on 234 square miles. While Chicago is a relatively northern city and is challenged by cold and snowy winters, it has invested heavily in active and public transportation infrastructure to encourage multi-modal streets appropriate for all users. The Chicago Department of Transportation has committed to building a 645-mile bicycle system by 2020 [10]. While Chicago is a leader in ensuring appropriate transportation options for all residents and visitors, the experiences and lessons learned in Chicago are of value to other cities, whether starting out with multi-modal streets, or well on their way.

Methods:

Secondary data on bicycle lanes from the City of Chicago and OSM, both shapefiles, were compared to a street audit performed by the first author [11, 12]. For OSM, features can be identified through tags that differ in specificity. Thus, both a general and a specific measure were created for OSM data. The general measure identified street segments labeled as bicycling-related regardless of specific identification as a bicycle lane, while the specific measure included only those segments labeled as bicycle lanes.

The street audit tool was developed using the Bridging the Gap/Community Obesity Measures Project (BTG-COMP) street segment observation form, the Active Neighborhood Checklist, the WI Active Community Audit Tool, and the Systematic

Pedestrian and Cycling Environmental Scan (SPACES) audits [1-4] and had two-week test-retest reliability of 1.00 on a sample of 10% of all audited segments. A stratified random sample of 250 street segments, composed of 150 local street segments and 100 non-local street segments, was drawn. This provided sufficient power to detect an accuracy of 0.5 with 90% confidence intervals under assumed prevalence of bicycle lanes of 2% for local and 20% for nonlocal street segments. The street audit was conducted in June and July, 2015; the municipal dataset and OSM extractions were retrieved in September, 2015.

Each sampled street segment in each of the three tested secondary datasets (i.e., City of Chicago, general OSM, specific OSM) was classified as true positive (TP), false positive (FP), true negative (TN) or false negative (FN) by comparing the street audit and dataset on presence of a bicycle lane as shown in Table 1, with the street audit findings representing the truth. Table 2 shows the formulas for five validity statistics calculated for each secondary dataset, as well as their interpretation as applied to this study. None of the validity statistics corrected for chance [9].

Table 1: Definitions of street segment classification

Segment Classification	Bicycle lane identified: Audit	Bicycle lane identified: Dataset
True Positive (TP)	Yes	Yes
False Positive (FP)	No	Yes
True Negative (TN)	No	No
False Negative (FN)	Yes	No

Table 2: Formulas of validity statistics, as well as their interpretation as applied to this study

Statistic	Definition	Interpretation
Accuracy	(TP + TN)/ (TP + FP + TN + FN)	Proportion of street segments in the dataset that correctly identified the presence or absence of bicycle lanes
Sensitivity	(TP)/(TP + FN)	Proportion of street segments with bicycle lanes as identified through the audit that had bicycle lanes in the dataset. Low sensitivity indicates undercount of bicycle lanes.
Positive Predictive Value (PPV)	(TP)/(TP + FP)	Proportion of street segments with bicycle lanes in the dataset that had bicycle lanes as identified through the audit. Low PPV indicates overcount of bicycle lanes.
Specificity	(TN)/(FP + TN)	Proportion of street segments without bicycle lanes as identified through the audit that did not have bicycle lanes in the dataset. Low specificity indicates overcount of bicycle lanes.
Negative Predictive Value (NPV)	(TN)/(TN + FN)	Proportion of street segments without bicycle lanes in the dataset that did not have bicycle lanes as identified through the audit. Low NPV indicates undercount of bicycle lanes.

Results:

Out of 250 sampled street segments for auditing, two were parking lots, one was gated private property, and one was a police station, and thus were discarded. Of the remaining 246 segments, 27 (11.0%) had bicycle lanes based on the audit. Twenty-five of the 27 identified bicycle lanes were found on nonlocal streets.

Validity statistics for the three secondary datasets as sources for bicycle lane locations are reported in Table 3. Overall, accuracy, PPV, specificity, and NPV were high, with all reported values greater than 0.90, while there was a range found for sensitivity: 0.96 in the municipal dataset, 0.74 in the general OSM dataset, and 0.41 in the specific OSM dataset.

Table 3: Validity findings for three secondary datasets as sources for bicycle lane locations

Measure	Count Lanes Identified:	Accuracy	Sensitivity	PPV	Specificity	NPV
Audit	27	-	-	-	-	-
Municipal	28	0.99	0.96	0.93	0.99	0.99
OSM general	24	0.96	0.74	0.83	0.98	0.97
OSM specific	12	0.93	0.41	0.92	0.99	0.93

note: PPV is positive predictive value, NPV is negative predictive value, OSM is OpenStreetMap

Discussion:

Through our audit of street segments in Chicago, we found bicycle lanes on 25% of nonlocal streets and 1.3% of local streets. We found that the City of Chicago dataset is a valid source for bicycle lane locations in Chicago, demonstrating high accuracy, sensitivity, PPV, specificity, and NPV. While both OSM measures showed adequate accuracy, PPV, specificity and NPV values, sensitivity was lower, particularly in the specific OSM dataset. Additionally, the specific OSM dataset showed slightly higher PPV than the general OSM dataset.

These findings have important implications for the validity and use of OSM as a source for bicycle lane location data. Street segments with bicycling infrastructure are not all tagged identically in OSM. Lower sensitivity in the specific OSM dataset indicates that there are bicycle lanes in Chicago that lack specific bicycle lane tags, but have more general bicycling tags. However, the lower PPV of the general OSM dataset indicates that inclusion of more general tags to identify bicycle lanes in OSM results in the identification of additional street segments that do not have bicycle lanes as well as segments with bicycle lanes that are not specifically tagged. In other words, use of OSM tags that are too specific may result in undercount while tags that are too general may result in overcount. However our findings suggest that in Chicago the overcount associated with general tags is of a smaller magnitude than the undercount associated with specific tags.

This study was adequately powered to detect accuracy, specificity and NPV, but was underpowered for both sensitivity and PPV because of the low prevalence of bicycle lanes on all city streets, both local and nonlocal. A sampling technique that purposively sampled street segments pre-identified as having bicycle lanes would facilitate adequate power for sensitivity and PPV. Such a sampling technique was not used in this study because it was inappropriate for testing accuracy, specificity, and NPV.

Chicago is a large, dense city with an active bicycling program [10]; therefore, the findings reported here may not be generalizable to all types of other cities. However, our findings suggest that for studies that are unable to audit street segments for bicycle lanes, a dataset provided by the local government may be a very good option. If OSM is considered as a data source, it should be carefully assessed with particular attention paid to tags used to identify bicycle lanes.

Acknowledgments:

The authors thank Alana Steffen, PhD for assistance with power calculations and Sandy Slater, PhD for feedback on the street audit tool. The University of Illinois at Chicago Chancellor's Graduate Research Fellowship to the first author provided funding for this research.

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Pikora T, Bull F, Jamrozik K, Knuiman M, Giles-Corti B, Donovan R: **Developing a reliable audit instrument to measure the physical environment for physical activity.** American Journal of Preventive Medicine 2002, **23**(3):187-194.

2. Slater SJ, Nicholson L, Chriqui J, Barker DC, Chaloupka FJ, Johnston LD: **Walkable communities and adolescent weight.** American Journal of Preventive Medicine 2013, **44**(2):164-168.

3. Hoehner CM, Ivy A, Ramirez LK, Handy S, Brownson RC: **Active neighborhood checklist: a user-friendly and reliable tool for assessing activity friendliness.** American Journal of Health Promotion 2007, **21**(6):534-537.

4. Department of Health Services, Division of Public Health, Nutrition, Physical Activity and Obesity Program, Wisconsin Partnership for Activity and Nutrition: **Wisconsin Worksite Resource Kit to Prevent Obesity and Related Chronic Diseases.** July 2012, P-00399 (9/12).

5. Ewing R: **Active living: A planning subfield comes of age.** Planning 2016, **82**(8):46-47.

6. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J: **An ecological approach to creating active living communities.** Annual Review of Public Health 2006, **27**:297-322.

7. Buehler R, Pucher J: **Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes.** Transportation 2012, **39**(2):409-432.

8. Neis P, Zielstra D: **Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap.** Future Internet 2014, **6**(1):76-106.

9. Gordis L: **Epidemiology.** Philadelphia:Elsevier/Saunders**.** 2009. Print.

10. Chicago Department of Transportation: **Chicago Streets for Cycling Plan 2020.** City of Chicago: Chicago, IL. 2012. Print.

11. Chicago Data Portal. City of Chicago 2016, retrieved from data.cityofchicago.org.

12. OpenStreetMap. OpenStreetMap Contributors 2016, retrieved from www.openstreetmap.org.