HHS Public Access
Author manuscript
Health Place. Author manuscript; available in PMC 2023 June 21.

Published in final edited form as:
Health Place. 2023 May ; 81: 103002. doi:10.1016/j.healthplace.2023.103002.

Association between passively collected walking and bicycling data and purposefully collected active commuting survey data—United States, 2019

Graycie W. Soto, MPH[a,b], **Bryant J. Webber, MD, MPH**[a,c], **Kelly Fletcher, MPH**[d], **Tiffany J. Chen, MSPH**[a,e], **Michael D. Garber, PhD, MPH**[d,f], **Akimi Smith, MPH**[a], **Grete Wilt, MPH**[d], **Michel Conn, MPH**[d], **Geoffrey P. Whitfield, PhD, MEd**[a]

[a]Division of Nutrition, Physical Activity, and Obesity, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention (4770 Buford Hwy, Atlanta, GA, USA 30341)

[b]Oak Ridge Institute for Science and Education (ORISE) Research Participation Program, Oak Ridge, TN, USA 37831

[c]Epidemic Intelligence Service, Centers for Disease Control and Prevention (1600 Clifton Rd, Atlanta, GA, USA 30329)

[d]Geospatial Research, Analysis, and Services Program, Geospatial Research, Analysis, and Services Program, Centers for Disease Control and Prevention (4770 Buford Hwy, Atlanta, GA, USA 30341)

[e]McKing Consulting Corporation (2900 Chamblee Tucker Road, Building 10, Suite 100, Atlanta, GA, USA 30341)

[f]Department of Environmental and Radiological Health Sciences, Colorado State University (Environmental Health Building, 1681 Campus Delivery, Fort Collins, CO, USA 80523)

Abstract

Commercially-available location-based services (LBS) data derived primarily from mobile devices may provide an alternative to surveys for monitoring physically-active transportation. Using Spearman correlation, we compared county-level metrics of walking and bicycling from StreetLight with metrics of physically-active commuting among U.S. workers from the American Community Survey. Our strongest pair of metrics ranked counties (n=298) similarly for walking (rho=0.53 [95% CI: 0.44–0.61]) and bicycling (rho=0.61 [0.53–0.67]). Correlations were higher for denser and more urban counties. LBS data may offer public health and transportation

Corresponding Author: Graycie Soto | qxb1@cdc.gov | 770-488-6809.

Declarations of Interest: none

Disclaimer: Use of data from StreetLight, Inc., does not imply U.S. Department of Health and Human Services or Centers for Disease Control and Prevention endorsement of StreetLight, Inc., or any of its products. U.S. Department of Health and Human Services and Centers for Disease Control and Prevention do not directly or indirectly endorse StreetLight Inc. or any particular organization, product, or service, whether directly or indirectly related to StreetLight, Inc. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the U.S. government.

professionals timely information on walking and bicycling behavior at finer geographic scales than some existing surveys.

Keywords

walking; bicycling; transportation; geographic information systems; population surveillance; mobility data

Introduction

Physical activity is an important health behavior. The *Physical Activity Guidelines for Americans*, 2nd edition, recommends that adults perform at least 150 minutes per week of moderate-intensity equivalent aerobic physical activity to obtain substantial health benefits (U.S. Department of Health and Human Services, 2018). This can occur during active transportation, a domain of physical activity that includes walking, bicycling, and other modes of human-powered movement to get from one place to another. Because many trips in the United States consist of distances amenable to these modes, active transportation is a key target for increasing physical activity.

As active transportation is promoted, its accurate measurement and monitoring will be important. At least four purposefully sampled surveillance systems gather self-reported national data on active transportation. Overall prevalence estimates from these systems differ substantially, underscoring the complexity of measuring this behavior (Whitfield et al., 2015). Although these traditional surveillance systems provide important information, their value is often limited to a certain metric or subpopulation (e.g., walking to school among adolescents) (Fulton et al., 2016; Omura et al., 2021). Moreover, these systems suffer from survey-related biases, coarse geographic resolution, and the intrinsic time lag between data collection and availability (Sallis and Pate, 2021; Whitfield et al., 2015).

Location-based services (LBS) data may complement traditional surveillance by providing more timely access to active transportation data, potentially at smaller geographic scales, than some existing systems. In this paper we use the term *location-based services data* to denote aggregated, passively collected data that are anonymous, derived primarily from mobile devices, and repurposed for public health research and surveillance. This passive collection is unlike traditional active transportation surveillance, which relies on purposeful collection of data from survey respondents (Whitfield et al., 2015). LBS data may be purchased from commercial technology and transportation companies for various applications (Berrigan et al., 2021). Several studies have used LBS data to quantify active transportation (Hunter et al., 2021), conduct research (Lee and Sener, 2019), or to plan and evaluate urban infrastructure (Garber et al., 2022; Musakwa and Selala, 2016; Sanders et al., 2017; Sun et al., 2017). Early evidence suggests that these novel data may most closely resemble purposefully sampled data in areas with high population density (Whitfield et al., 2016), but the validity of LBS data for public health application requires further investigation.

LBS data are available from multiple vendors, one of which is StreetLight Data, Inc. (henceforth, StreetLight), a transportation intelligence company. Unlike vendors who supply data exclusively from activity tracking apps and physical fitness devices, StreetLight also incorporates general LBS data (details of which are presented in the Methods). This may allow StreetLight to capture physical activity data on a larger and more representative sample of the population than the subpopulation that uses these specialized apps and devices (Lee and Sener, 2020). Although this may confer research and surveillance advantages, more evidence is needed. To date, StreetLight data have been used in few scientific investigations, mostly sponsored by transportation authorities or focused on city-level metrics (Kothuri et al., 2022; Lee and Sener, 2020; Cheng et al., 2022). To our knowledge, LBS data from StreetLight have not been compared to data from nationally representative surveys of the U.S. population.

When considering potential enhancements to active transportation surveillance, including those involving mobile technologies like smartphones, an American College of Sports Medicine Consensus Statement indicated that limitations in data collection, compilation, and analysis must be addressed “before these alternative technologies can be used in national surveillance” (Fulton et al., 2016). Some of these limitations may be attributed to biases introduced by using mobile devices and smartphone apps as data sources: the very young, the very old, and those with limited access to technology may be underrepresented or excluded (Lee and Sener, 2020). Addressing the need for validation, our study aimed to provide county-level estimates of convergent validity between measures of walking and bicycling from StreetLight and related but distinct measures of self-reported walking and bicycling to work from the American Community Survey (ACS). Because each data source offers multiple constructs (e.g., different trip types in StreetLight or different populations in ACS), we also sought to determine which pair of constructs exhibited the strongest association. For the construct pair with the strongest association, a secondary objective was to assess whether the association between StreetLight and ACS measures varied by counties’ geographic, demographic, and socioeconomic characteristics.

Methods

Study Design and Sampling Methodology

This study compared anonymized, aggregated LBS data obtained from StreetLight and publicly available active commuting data from ACS in a sample of counties across the United States (n=298 out of N=3142). A purposeful sample of approximately 300 counties was selected to balance cost, efficiency, and representativeness. We aimed to sample counties with both a large share of the national population while achieving good representation across regional geographies and the six levels of the 2013 National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties (Ingram and Franco, 2014). We restricted our sampling frame to the 48 contiguous United States because StreetLight data are only available for these states. We also restricted to those counties above the 10th percentile population value per 2015–2019 ACS 5-year estimates for the most rural category (2,838 people) to ensure that there would be a large enough population in each county to capture a meaningful estimate. From that sampling frame, we first selected all counties (n=68) in the most urban of the NCHS urban-rural categories. To achieve representation across the geographic regions and urban-rural continuum of the contiguous United States, we stratified the remaining counties by Census Bureau-designated regions (n=4) and by the remaining NCHS urban-rural categories (n=5). The number of counties sampled from each of these 20 region/urban-rural strata was proportional to the population in each stratum; within strata, counties were sampled at random. This yielded another 230 counties, for a total of 298 sampled (just shy of the 300-county maximum based on study budget). Additional details on the sampling strategy are available (Supplementary Methods 1).

Pedestrian and Bicycling Data from StreetLight

We purchased county-level LBS data from StreetLight, a transportation intelligence company that sells multimodal transportation data to transportation agencies, commercial industries, and research enterprises. StreetLight uses general LBS data, such as smartphone apps that use location services either in the foreground or background, and active mode-specific LBS data, which originate from sources specific to measuring active modes of transportation (e.g., activity tracking apps or physical fitness devices). StreetLight cleans, filters, and organizes these data, then applies a proprietary algorithm, which was developed using machine learning, to assign likely trip mode. The algorithm was trained using ground truth data from sources in which the travel mode was confirmed, and it was validated with external data sources, including permanent pedestrian and bicycle counters. When the algorithm detects that an individual used different travel modes sequentially (e.g., vehicle then bicycle), it counts each mode as a distinct trip. Detailed methodology on data sources, trip mode identification, validation, and acquisition is provided in Supplementary Methods 2.

Our StreetLight dataset contained the average daily count of pedestrian and bicycle trips for 2019 for each of the 298 sampled counties. StreetLight determined the block group of residence based on where the device spent the overnight hours and was thereby able to differentiate trips made by residents of a given county. Block group of occupation was similarly determined based on daytime position. Given the assumption that work-based trips occur more frequently on weekdays and recreational trips on weekends, StreetLight classified trips by three *day* types (all days of the week, only weekdays, and only weekends) and by two *traveler* types (trips taken by residents of the county and combined resident and non-resident trips beginning in the county). Additionally, for the trips taken by combined residents and non-residents, StreetLight estimated the proportion of trips that were from home to work or vice versa (home-based work). The combinations of these metrics yielded six unique types of trip counts for each county and three additional types of trip counts for home-based work trips:

- Any-purpose trips on all days of the week among residents and non-residents
- Any-purpose trips on weekdays among residents and non-residents
- Any-purpose trips on weekends among residents and non-residents
- Any-purpose trips on all days of the week among only residents

- Any-purpose trips on weekdays among only residents
- Any-purpose trips on weekends among only residents
- Home-based work trips on all days of the week among residents and non-residents
- Home-based work trips on weekdays among residents and non-residents
- Home-based work trips on weekends among residents and non-residents

StreetLight suppressed some bicycling trip count data due to insufficient sample sizes. Across these nine day-traveler-purpose combinations used in our analyses, a maximum of 8 sampled counties had missing bicycling trip count data, for a minimum sample size of 290 counties. For the specific day-traveler-purpose combination used in our subsequent stratified analysis, only one sampled county was missing trip count data, resulting in a bicycling sample size of 297 counties.

We calculated daily rates of pedestrian and bicycling trips from StreetLight as trips divided by a population denominator (per 1,000 people). For combined resident and non-resident trips, the population denominator was derived from the commuter-adjusted daytime population because this best represented the population eligible to make such trips. The daytime population, provided by ACS, adjusts the county population to account for the influx and outflux of workers during the day (U.S. Census Bureau, 2021). For resident-only trips, we examined three different denominators from ACS: all county residents; county workers only; and county non-teleworkers only (details provided below).

Active Commuting Data from American Community Survey

Comparison data on active commuting among workers was obtained from ACS, a nationwide survey administered by the U.S. Census Bureau. Every year 3.5 million households are sampled via the internet, phone, or paper questionnaire, with all household residents asked to complete the survey. Data are pooled each calendar year to create a single-year estimate for geographic areas with at least 65,000 people, and data are pooled from 5 consecutive years to create a 5-year estimate for geographic areas with fewer than 65,000 people. We used the 5-year estimates from 2015–2019 because one-third of the 298 sampled counties were below the 65,000-person threshold. We confirmed agreement between the 1- and 5-year ACS estimates using Bland-Altman plots (Supplementary Methods 3), where the mean difference is close to 0 and few observations are outside two standard deviations (Johnson and Augusta, 2018). Survey response rate (measured at the housing-unit level) ranged from 86.0% in 2019 to 95.8% in 2016.

ACS asks respondents aged ≥16 years about their occupational status, including their primary mode of commuting. Respondents who reported working for pay in the previous week were asked, “How did [you] usually get to work last week?” Twelve response options were provided, including “walked,” “bicycled,” and “worked from home” (U.S. Census Bureau, 2019). Respondents could only choose one method and were instructed to indicate the method of transportation they used for most of the distance. We defined those who selected “walked” as walkers and those who selected “bicycled” as bicyclists. We also used occupational status to define two denominators: all workers (respondents ≥ 16 years who reported working for pay in the previous week) and non-teleworkers (calculated as all workers minus those who reported usually working from home). This resulted in four active commuting values from ACS: the proportion of all workers who walked to work, the proportion of all workers who bicycled to work, the proportion of non-teleworkers who walked to work, and the proportion of non-teleworkers who bicycled to work.

County Characteristics

We used geographic, demographic, and socioeconomic variables from the ACS 5-year estimates (2015–2019) to stratify the counties. These variables included U.S. Census region (Midwest, Northeast, South, West), median age, race/ethnicity (percent of the county population that self-identifies as non-Hispanic White, non-Hispanic Black, or Hispanic), poverty status (percent of the county population below the federal poverty level in the past 12 months), median household income, population density (county population divided by county square mileage, the latter acquired from the U.S. Census Bureau), and education (percent of population with at least some college education) (U.S. Census Bureau, 2020).

We used the 2013 NCHS Urban-Rural Classification Scheme for Counties (Ingram and Franco, 2014) to stratify counties according to five levels of urbanicity (large central metro, large fringe metro, medium metro, small metro, and micropolitan/non-core). We also assessed the overall social vulnerability of counties using the 2018 Social Vulnerability Index (SVI) from the Centers for Disease Control and Prevention (CDC) Agency for Toxic Substances and Disease Registry. The overall SVI metric integrates data on 15 social factors within four thematic areas—socioeconomic status, household composition and disability, minority status and language, and housing type and transportation—to assign each census tract and county a social vulnerability score ranging from 0.00 (least vulnerable) to 1.00 (most vulnerable) (Agency for Toxic Substances and Disease Registry, 2022).

Statistical Analysis

We described the sampled counties (n = 298) and all U.S. counties and county equivalent areas (N = 3142) by the variables described above. For all county descriptors other than census region and urbanicity, we used the sampled counties (n = 298) to divide counties into tertiles. We then applied the tertile values to all U.S. counties and county equivalents to understand how our sample compares to all U.S. counties.

For both walking and bicycling, we developed a correlation matrix to compare StreetLight trip rates (by various combinations of day types, traveler types, trip purposes, and denominators) and ACS active commuting prevalence (among all workers and non-teleworkers). Because most ACS and StreetLight variables were non-normally distributed according to Shapiro-Wilk testing, we used the nonparametric pairwise Spearman's rank correlation coefficient to assess the relationship between StreetLight measures of walking and bicycling trips and ACS active commuting. Because bicycle commuting had null prevalence for 46 counties, we assessed the impact of tied ranks by comparing the bicycling correlations derived from four recommended methods to account for ties (SAS Institute Inc., n.d.). These correlations were similar, suggesting ties did not have an undue effect on the bicycle correlation coefficients; we subsequently used the default method (average ranks) to handle ties. To estimate variability in the measures arising from our sampling strategy, we took a bootstrapping approach, resampling the sampled counties with replacement, stratified by the 20 region/urban-rural strata.

From the correlation matrix we identified the StreetLight–ACS pair with the strongest correlation coefficient, and we used this pair of measures for the stratified analyses. For the stratified analyses, we first calculated the median and interquartile range of the StreetLight and ACS measures for all counties and for each geographic, demographic, and socioeconomic stratum. Second, we calculated Spearman's rho with 95% confidence intervals (CI; based on Fisher's z transformation) between the selected StreetLight and ACS measures by descriptive strata. We applied Cohen's convention (Cohen, 1988) to interpret rho values (low correlation: <0.3; moderate: 0.3 to <0.5; strong: ≥ 0.5). As a sensitivity analysis, we also calculated stratified correlations for the StreetLight–ACS pair with the weakest overall correlation coefficient.

We used R v 4.2 (R Foundation for Statistical Computing, Vienna, Austria) to sample counties (code provided in Supplementary Methods 1), resample counties using a bootstrapping approach, and calculate confidence intervals for Spearman's rho in the correlation matrix. We conducted all other analyses using SAS v 9.4 (SAS Institute Inc., Cary, North Carolina, USA).

Regulatory Information

This activity was reviewed by CDC and was determined to be public health surveillance. It was conducted according to federal law and CDC policy (45 C.F.R. part 46.102(l)(2), 21 C.F.R. part 56, 42 U.S.C. Sect. 241(d), 5 U.S.C. Sect. 552a, and 44 U.S.C. Sect. 3501 et seq.). CDC did not receive any personally identifiable information from ACS or StreetLight.

Results

Sampled County Characteristics

The 298 sampled counties had a combined population of 140.5 million people, or 9.5% of U.S. counties and county equivalents and 43.3% of the U.S. population. The sampled counties differed from all counties and county equivalents by several demographic and socioeconomic characteristics. Because our sampling strategy relatively oversampled large cities, compared to all county or county equivalents, sampled counties tended to have a younger, more affluent population at a slightly lower level of social vulnerability (e.g., median age <37 years: 18.4% of all counties, 33.6% of sampled counties). The prevalence of active commuting in ACS was higher in the sampled counties than in all counties or county equivalents: the median prevalence of walking to work (according to ACS) was 3.15%, and the median prevalence of bicycling to work was 0.73%, vs. 2.67% and 0.55% for all counties or county equivalents, respectively. Restricting to non-teleworkers, the median prevalence of walking and bicycling to work in the sampled counties was 3.33% and 0.78%, respectively, and in all counties and county equivalents was 2.81% and 0.58%, respectively (Table 1).

Pedestrian Metrics

Of the 30 distinct combinations of ACS and StreetLight pedestrian metrics, 6 had a high correlation, 18 had a moderate correlation, and 6 had a low correlation. The strongest correlation was between percent of non-teleworkers walking to work (from ACS) and any-day resident walk trips per 1,000 county residents (from StreetLight): rho = 0.53 (95% CI: 0.44–0.61) (Table 2).

At the county level, 2.11% (IQR: 2.13%) of non-teleworkers walked to work (according to ACS), and residents made 1,149 (IQR: 460) walk trips per 1,000 residents per day (according to StreetLight) (Table 3).

All stratified analyses compared percent of non-teleworkers walking to work with any-day resident walk trips per 1,000 county residents, hereafter referred to as the strongest correlated walking pair of metrics. Of the 36 categories of county characteristics, 20 correlations were strong, 15 were moderate, and one was low (Table 3). The only low correlation was among small metro counties (rho = 0.30; 95% CI: −0.07–0.59). Strong correlations were found for counties in the Northeast (rho = 0.66; 95% CI: 0.48–0.79), in the lowest tertile for Hispanic population percentage (rho = 0.66; 95% CI: 0.53–0.76), in the highest tertile for Black population percentage (rho = 0.62; 95% CI: 0.48–0.73), in the lowest tertile for social vulnerability (rho = 0.62; 95% CI: 0.48–0.73), and others (Table 3). The correlations of counties in the Northeast (rho = 0.66) and in the West (rho = 0.31) are visualized in side-by-side scatterplots (Figure 2).

Bicycling Metrics

Of the 30 distinct combinations between ACS and StreetLight bicycling metrics, 28 had a strong correlation, and two had a moderate correlation (Table 4). The strongest correlation was between percent of non-teleworkers bicycling to work (from ACS) and any-day resident bicycle trips per 1,000 county residents (from StreetLight), with rho = 0.61 (95% CI: 0.53–0.67) (Table 4). When assessing the impact of tied ranks on the correlation coefficients by using the four different methods to account for ties, the rho values were generally within .01 of our original findings.

At the county level, 0.25% (IQR: 0.43%) of non-teleworkers bicycled to work (according to ACS), and residents made 30 (IQR: 32) bicycle trips per 1,000 residents per day (according to StreetLight) (Table 5).

All stratified analyses compared percent of non-teleworkers bicycling to work with any-day resident bicycle trips per 1,000 county residents, hereafter referred to as the strongest correlated bicycling pair of metrics. Of the 36 categories of county characteristics, 24 correlations were strong, 11 were moderate, and one was low (Table 5). The only low correlation was among micropolitan/non-core counties (rho = 0.25; 95% CI: −0.05–0.50). Strong correlations were found for counties in the highest tertile for median income (rho = 0.80; 95% CI: 0.72–0.86), in the highest tertile for Black population percentage (rho = 0.75; 95% CI: 0.65–0.83), in the highest tertile for population density (rho = 0.75; 95% CI: 0.65–0.83), in the middle tertile for White population percentage (rho = 0.74; 95% CI: 0.63–0.82), and others (Table 5). The correlations of large central metro counties (rho = 0.71) and micropolitan/non-core counties (rho = 0.25) are visualized in side-by-side scatterplots (Figure 3).

Sensitivity Analysis

For walking, the weakest overall correlation was between percent of non-teleworkers walking to work (from ACS) and weekend home-based work walk trips per 1,000 daytime population (from StreetLight): rho = 0.11 (95% CI: 0.00–0.22). Stratified correlations for this metric pair were mostly non-significant or weak, with the exception of moderate correlations in large central metro and more densely populated counties, and in the Northeast (Supplementary Results, Table S1). For bicycling, the weakest overall correlation was between percent of all workers bicycling to work (from ACS) and weekend home-based work bicycle trips per 1,000 daytime population (from StreetLight): rho = 0.49 (95% CI: 0.40–0.57). Stratified correlations were weaker with this pair but largely mirrored the pattern observed in the primary analysis (Supplementary Results, Table S2).

Discussion

Passively collected pedestrian and bicycling data from StreetLight ranked counties similarly to purposefully sampled pedestrian and bicycling active commuting data from ACS in this national sample of U.S. counties. Overall, walking correlations were moderate and bicycling correlations were strong. For the stratified analysis using the most strongly correlated pair, correlations were strong overall for both walking (rho = 0.53) and bicycling (rho = 0.61) and tended to be higher in more densely populated, urban counties. These findings suggest that StreetLight data may be sufficiently valid for select public health applications. Most existing surveillance systems for physical activity have a time lag of at least two years and offer poor geographic granularity below the state or county level. Less constrained by these limitations, LBS data may offer public health and transportation professionals an additional tool for assessing walking and bicycling behavior. For example, LBS data may be preferable to traditional surveillance data for conducting a time-sensitive project, such as changes in walking patterns during an epidemic, or for assessing the impact of a community design intervention to promote bicycling within a city or town.

To our knowledge, this is the first peer-reviewed study to compare location-based mobility pedestrian and bicycling data from StreetLight to ACS and examine correlations by county demographic and socioeconomic status characteristics. Our research builds on previous studies that have compared other types of crowdsourced and LBS data to traditional surveillance measures. Whitfield and colleagues found a strong correlation (rho = 0.60) between the number of commuters, as derived from Strava—a GPS-based physical activity tracking platform that allows users to track and share bicycle rides, runs, and other activities—and the number of active commuters, as derived from ACS, in block groups in four large U.S. cities (2016). Similar to our results, they also found correlations were stronger in more densely populated areas, suggesting crowdsourced data more closely approximates purposefully sampled data in densely populated areas. In a study comparing StreetLight data and estimates of pedestrians and bicyclists passing fixed-point, permanent counters, Cheng and colleagues found strong correlations for pedestrian (rho = 0.72) and bicycling measures (rho = 0.69) (2022). Lee and Sener observe that most validation studies are restricted to bicycling metrics, which they attribute to variability and uncertainty of passively collected pedestrian data (2020).

Our study offers new evidence that LBS data may have an important role in future surveillance of walking and bicycling behavior. Among 30 StreetLight-ACS metric combinations, any-day resident trips per 1,000 county residents from StreetLight and percent of non-teleworkers actively commuting to work from ACS demonstrated the strongest Spearman's correlation for both walking (rho = 0.53) and bicycling (rho = 0.61). These strong correlations are especially noteworthy because StreetLight and ACS measure related but distinct physical activity constructs. While both are limited to county residents (i.e., they do not include in-county pedestrian and bicycle trips by non-residents), the ACS measure reflects travel associated with work, while the StreetLight measure reflects trips for any purpose. Therefore, even if each metric perfectly captured their intended construct, the rank correlation between them would likely not be perfect.

Surprisingly, StreetLight metrics restricted to home-based work trips resulted in weaker correlations with ACS active commuting measures (particularly with walking, as observed in the sensitivity analysis). This suggests a disconnect between the intended purposes of the two data sources. Although the exact reasons for this cannot be determined in our study, it could reflect misassigned or unattributed work locations in StreetLight, especially for multi-purpose locations (e.g., a college campus or grocery store) and for traveling occupations (e.g., electricians and plumbers). Regardless, it means that purpose-assigned walking and bicycling data in StreetLight may be less useful for public health applications than its purpose-agnostic walking and bicycling data.

Our study adds to a growing body of evidence that the validity of LBS data may be stronger in areas of higher population density and suggests additional variation by sociodemographic characteristics. For example, the strength of association for the strongest correlated bicycling pair of metrics was higher for counties with a younger median age, a higher proportion of Black residents, and a higher proportion of residents with at least some college education —all characteristics associated with urban counties in the United States (Cromartie and Vilorio, 2019; Day et al., 2016; U.S. Department of Agriculture, 2017). Conversely, LBS data may be less comparable to purposefully sampled data in rural and more sparsely populated areas, as well as in areas that experience high seasonal population changes from tourism. For example, counties with outdoor recreational destinations may see seasonal increases in walking and bicycling that are not related to actively commuting to work by county residents. One outlier county in our sample, Summit County, Colorado—home to the popular outdoor destination Breckenridge and a vast trail network—had a high rate of StreetLight bicycle trips but a low prevalence of ACS active commuting by bicycle. Although these findings are not surprising, they underscore the importance of differentiating between resident and non-resident trips and the need to contextualize and validate these measures with existing survey data. The results also highlight that validity of LBS data in active commuting surveillance may vary geographically.

One concern with these “emerging data sources” is that they may not be fully representative of the walking and bicycling populations due to some degree of selection bias in smartphone app users (Garber et al., 2019; Lee and Sener, 2020). The stronger correlations in our study between StreetLight and ACS metrics in denser and more urban counties may be an example of potential bias: residents of urban areas are more likely to own smartphones and therefore may be more likely to use the various apps that contribute LBS data to StreetLight (Pew Research Center, 2021). Further, more socially vulnerable groups may not be fully represented in these data sources. When using the strongest and weakest correlated pair of metrics, the strength of association between StreetLight and ACS for walking was higher for the counties with the lowest levels of social vulnerability, which may indicate that people in these areas (e.g., areas ranked higher for socioeconomic status, vehicle access, and non-minority households) are more likely to have smartphones contributing LBS data than are people in the counties with the highest levels of social vulnerability. Another less understood but plausible reason may be due to differences in who opts out of sharing their location data with apps. These and other factors may have contributed to the range in correlations of the metrics we tested. This warrants additional research because this could disproportionally affect groups of people that could benefit from public health interventions to increase physical activity (e.g., rural or socially vulnerable populations).

Implications for Physical Activity Surveillance

LBS data sources have the potential to complement traditional physical activity surveillance sources, which rely on purposefully sampled surveys. Although they are unlikely to replace traditional surveillance systems, they may augment these systems by providing novel data beyond what was examined in our study, such as circuity and speed information, and offer more granular geographic details in a timelier manner. Many studies have demonstrated the utility of combining user-generated mobility data with traditional trip counts (Dadashova et al., 2020; Kothuri et al., 2022; Lee and Sener, 2020). Although these data sources are promising surveillance tools, researchers may need to consider not only their validity and representativeness, but also their ethical implications, including privacy concerns and potential lack of understanding by smartphone users about what data they are consenting to share (Breslin et al., 2019; Roy, 2017).

Limitations

There are several limitations to consider when interpreting these findings. First, StreetLight uses a proprietary algorithm to assign trip mode, which limits reproducibility. Second, due to privacy reasons, StreetLight is unable to provide individual-level sociodemographic characteristics, which may have been valuable for data validation. Third, we did not compare single-year estimates from ACS for 2019 to the corresponding year of StreetLight data because ACS only generates single-year estimates for geographic areas with populations with 65,000 people or greater. However, we did confirm agreement between the 2019 and 2015–2019 ACS data, suggesting 2015–2019 ACS data would serve as a good proxy for 2019 ACS data in the comparison with StreetLight. Fourth, we do not have a gold-standard measure to which to compare StreetLight or ACS. Fifth, the StreetLight and ACS metrics have inherent limitations. Incomplete population penetration of mobile devices and smartphone apps, and opt-out of data sharing by users, could introduce selection bias into the StreetLight data, while demographic and socioeconomic disparities in owning and using these devices and apps could result in samples that are less representative of the entire population. Sixth, the ACS question may not capture the usual commute mode over a longer time period because only the primary mode of commuting to work in the past week is captured, and since the ACS measure relies on self-report, it may be subject to recall and social desirability biases. Finally, we performed subgroup analyses on only the strongest and weakest correlated pair of StreetLight and ACS metrics, and the results may be different for other combinations.

Future research

Future research could replicate this analysis across all U.S. counties or at smaller geographic units, such as block groups or census tracts, and could compare StreetLight data to other surveillance system data that measure components of active transportation or physical activity, including the Behavioral Risk Factor Surveillance System, National Health and Nutrition Examination Survey, National Health Interview Survey, American Time Use Survey, and National Household Travel Survey. Follow-up analyses could investigate if inferred sociodemographic attributes of travelers who contribute to the StreetLight data (as assessed by the block groups of the devices' residence) are similar to sociodemographic attributes of all county residents or of active commuters in the county. Future analyses may also explore the various thematic areas or more granular social factors that comprise social vulnerability. Finally, because StreetLight provides trip counts in two-month increments, future studies might use surveillance systems with similar time units to assess for seasonal changes in data validity.

Conclusion

The results of this analysis generally suggest moderate to strong convergent validity between passively collected walking and bicycling data from StreetLight and purposefully sampled active commuting data from ACS. In general, Spearman's correlations were mostly moderate for pedestrian and mostly strong for bicycling data, and, when comparing the strongest and weakest correlated pairs of metrics for walking and bicycling, they were higher for denser and more urban counties in the United States. StreetLight data may have select applications in monitoring walking and bicycling within a community. Because correlations were lower when restricting to commute-based walking and bicycling, caution is advised when using trip purpose data in StreetLight. Should future investigations find LBS data sources to be valid, reliable, and ethically sound, they may complement traditional surveillance data with timelier and more geographically precise estimates of active transportation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This research was supported by the Centers for Disease Control and Prevention. No outside funding was obtained.

References

Agency for Toxic Substances and Disease Registry. (2022). CDC SVI 2018 Documentation https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/pdf/SVI2018Documentation_01192022_1.pdf

Berrigan D, Dannenberg AL, Lee M, Rodgers K, Wojcik JR, Wali B, Tribby CP, Buehler R, Sallis JF, Roberts JD, Steedly A, Peng B, Eisenberg Y, Rodriguez DA (2021). The 2019 Conference on Health and Active Transportation: research needs and opportunities. Int J Environ Res Public Health, 18(22). 10.3390/ijerph182211842

Breslin S, Shareck M, Fuller D (2019). Research ethics for mobile sensing device use by vulnerable populations. Soc Sci Med, 232, 50–57. 10.1016/j.socscimed.2019.04.035 [PubMed: 31059869]

Cheng W, Zhang Y, Clay E (2022). Comprehensive Performance Assessment of Passive Crowdsourcing for Counting Pedestrians and Bikes Mineta Transportation Institute Publications. 10.31979/mti.2022.2025

Cohen J (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Routledge, New York. 10.4324/9780203771587

Cromartie J, Vilorio D (2019). Rural Population Trends Retrieved 7/7/2022 from https://www.ers.usda.gov/amber-waves/2019/february/rural-population-trends/

Dadashova B, Griffin GP, Das S, Turner SM, Sherman B (2020). Estimation of average annual daily bicycle counts using crowdsourced Strava data. Transp Res Rec, 2674, 390–402. 10.1177/0361198120946016

Day JC, Hays D, Smith A (2016). A Glance at the Age Structure and Labor Force Participation of Rural America. U.S. Census Bureau https://www.census.gov/newsroom/blogs/random-samplings/2016/12/a_glance_at_the_age.html

Fulton JE, Carlson SA, Ainsworth BE, Berrigan D, Carlson C, Dorn JM, Heath GW, Kohl HW, 3rd, Lee IM, Lee SM, Másse LC, Morrow JR, Jr., Gabriel KP, Pivarnik JM, Pronk NP, Rodgers AB, Saelens BE, Sallis JF, Troiano RP, . . . Wendel A (2016). Strategic priorities for physical activity surveillance in the United States. Med Sci Sports Exerc, 48(10), 2057–2069. 10.1249/mss.0000000000000989 [PubMed: 27187094]

Garber MD, Flanders WD, Watkins KE, Lobelo F, Kramer MR, McCullough LE (2022). Have paved trails and protected bike lanes led to more bicycling in Atlanta?: a generalized synthetic-control analysis. Epidemiology, 33(4), 493–504. 10.1097/ede.0000000000001483 [PubMed: 35439778]

Garber MD, Watkins KE, Kramer MR (2019). Comparing bicyclists who use smartphone apps to record rides with those who do not: implications for representativeness and selection bias. J Transp Health, 15. 10.1016/j.jth.2019.100661

Hunter RF, Garcia L, de Sa TH, Zapata-Diomedi B, Millett C, Woodcock J, Pentland A, Moro E (2021). Effect of COVID-19 response policies on walking behavior in US cities. Nat Commun, 12(1), 3652. 10.1038/s41467-021-23937-9 [PubMed: 34135325]

Ingram DD, Franco SJ (2014). 2013 NCHS Urban–rural classification scheme for counties. National Center for Health Statistics. Vital Health Stat, 2(166). https://www.cdc.gov/nchs/data/series/sr_02/sr02_166.pdf

Johnson M, Augusta JW (2018). Simple Methods for Repeatability and Comparability: Bland-Altman Plots, Bias, and Measurement Error https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1815-2018.pdf

Kothuri S, Broach J, McNeil N, Hyun K, Mattingly S, Miah MM, Nordback K, Proulx F (2022). Exploring Data Fusion Techniques to Estimate Network-Wide Bicycle Volumes (NITC-RR-1269) https://trec.pdx.edu/research/project/1269/Exploring_Data_Fusion_Techniques_to_Derive_Bicycle_Volumes_on_a_Network

Lee K, Sener IN (2019). Understanding potential exposure of bicyclists on roadways to traffic-related air pollution: findings from El Paso, Texas, using Strava Metro data. Int J Environ Res Public Health, 16(3). 10.3390/ijerph16030371

Lee K, Sener IN (2020). Emerging data for pedestrian and bicycle monitoring: sources and applications. Transp Res Interdiscip Perspect, 4, 100095. 10.1016/j.trip.2020.100095

Musakwa W, Selala KM (2016). Mapping cycling patterns and trends using Strava Metro data in the city of Johannesburg, South Africa. Data Brief, 9, 898–905. 10.1016/j.dib.2016.11.002 [PubMed: 27872887]

Omura JD, Whitfield GP, Chen TJ, Hyde ET, Ussery EN, Watson KB, Carlson SA (2021). Surveillance of physical activity and sedentary behavior among youth and adults in the United States: history and opportunities. J Phys Act Health, 18(S1), s6–s24. 10.1123/jpah.2021-0179 [PubMed: 34465651]

Pew Research Center. (2021). Mobile Fact Sheet https://www.pewresearch.org/internet/fact-sheet/mobile/

Roy AL (2017). Innovation or violation? Leveraging mobile technology to conduct socially responsible community research. Amer J Community Psychol, 60(3–4), 385–390. 10.1002/ajcp.12187 [PubMed: 28944516]

Sallis JF, Pate RR (2021). Creating the future of physical activity surveillance in the United States: better data for better health. J Phys Act Health, 18(S1), s1–s5. 10.1123/jpah.2021-0182 [PubMed: 34465655]

Sanders RL, Frackelton A, Gardner S, Schneider R, Hintze M (2017). Ballpark method for estimating pedestrian and bicyclist exposure in Seattle, Washington: potential option for resource-constrained cities in an age of big data. Transp Res Record, 2605(1), 32–44. 10.3141/2605-03

StreetLight Data. (2020). Bike and Pedestrian Metrics, Methodology, Data Sources, and Validation, Version 4.0 https://learn.streetlightdata.com/bike-and-pedestrian-methodology-validation

Sun Y, Du Y, Wang Y, Zhuang L (2017). Examining associations of environmental characteristics with recreational cycling behaviour by street-level Strava data. Int J Environ Res Public Health, 14(6). 10.3390/ijerph14060644

Census Bureau US. (2019). American Community Survey Questionnaire Retrieved from https://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2019/quest19.pdf

Census Bureau US. (2020). Understanding and Using American Community Survey Data: What All Data Users Need to Know Washington, DC: U.S. Government Publishing Office. Retrieved from https://www.census.gov/programs-surveys/acs/library/handbooks/general.html

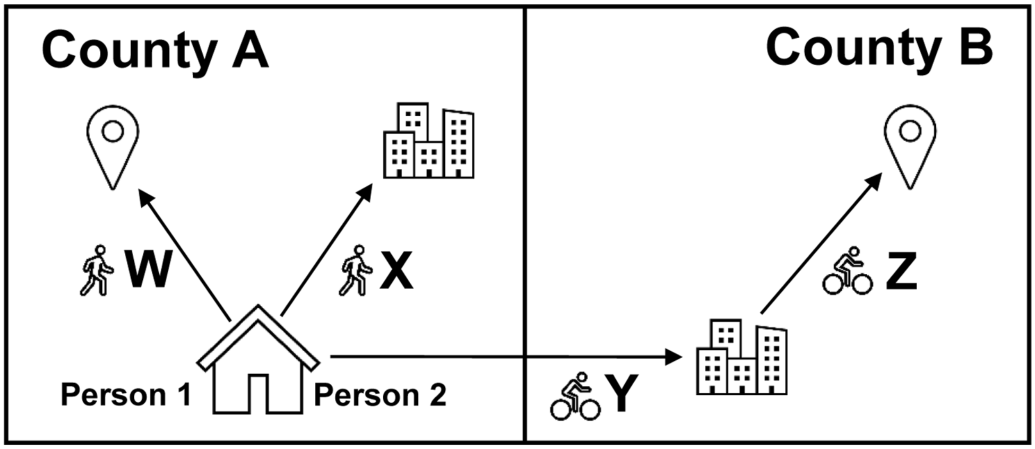
U.S. Census Bureau. (2021). Calculating Commuter-Adjusted Population Estimates Retrieved 7/6/2022 from https://www.census.gov/topics/employment/commuting/guidance/calculations.html

Department of Agriculture U.S.. (2017). Rural Education At A Glance, 2017 Edition https://www.ers.usda.gov/webdocs/publications/83078/eib-171.pdf?v=6913.4

U.S. Department of Health and Human Services. (2018). Physical Activity Guidelines for Americans Washington, DC. Retrieved from https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf

Whitfield GP, Paul P, Wendel AM (2015). Active transportation surveillance - United States, 1999–2012. MMWR Surveill Summ, 64(7), 1–17. https://www.cdc.gov/mmwr/preview/mmwrhtml/ss6407a1.htm

Whitfield GP, Ussery EN, Riordan B, Wendel AM (2016). Association between user-generated commuting data and population-representative active commuting surveillance data - four cities, 2014–2015. MMWR Morb Mortal Wkly Rep, 65(36), 959–962. 10.15585/mmwr.mm6536a4 [PubMed: 27632357]



Trip	Description	Assigned to County A
W	Person 1 walks from home to other destination	By StreetLight only
X	Person 1 walks from home to worksite	By StreetLight and ACS
Y	Person 2 bicycles from home to worksite	By StreetLight and ACS
Z	Person 2 bicycles from work to other destination	By Neither StreetLight nor ACS

Note: Person 1 is a resident of and works in County A. Person 2 is a resident of County A and works in County B.
ACS, American Community Survey.

Home Worksite Other Destination

Fig. 1.
Four active commuting trips and county assignment.

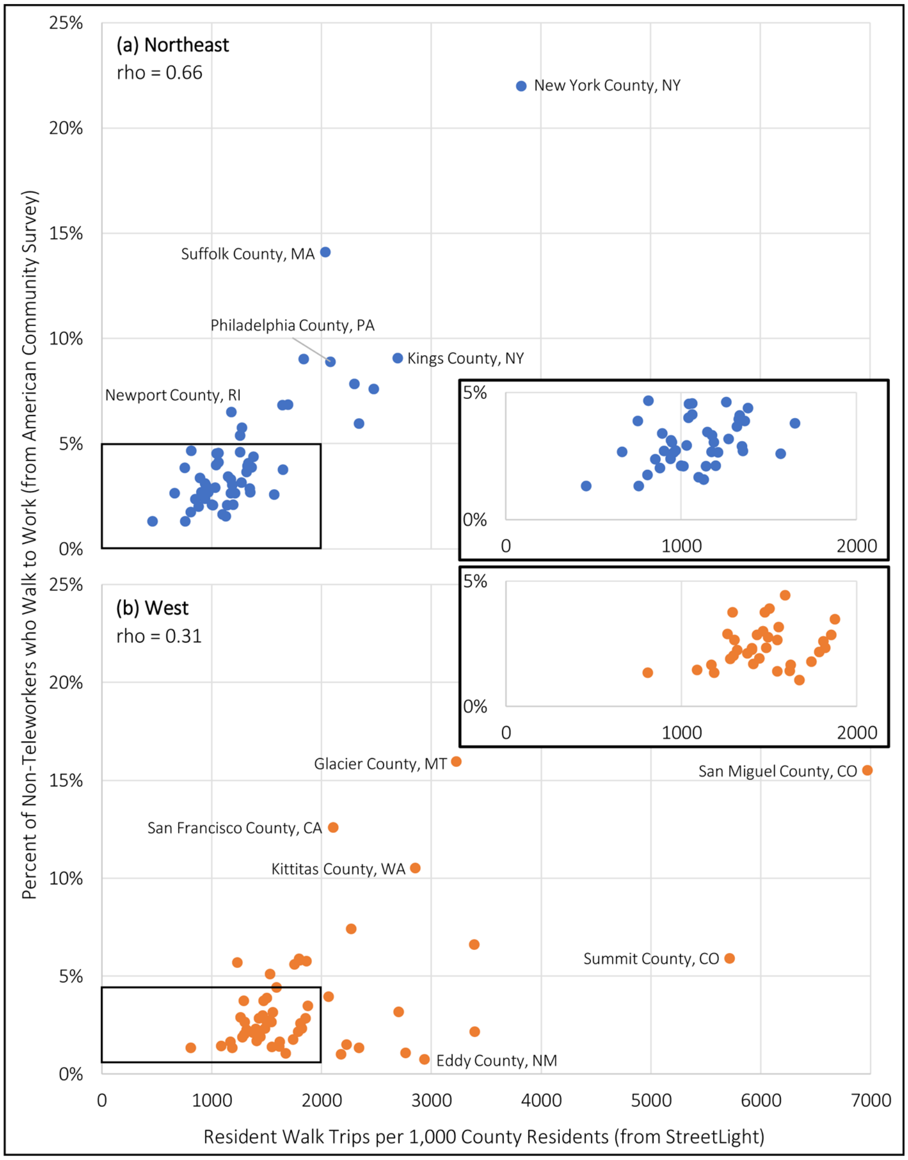


Fig. 2.

Percent walking to work (from American Community Survey) and resident walk trips (from StreetLight)—U.S. counties, (a) Northeast and (b) West, 2019.

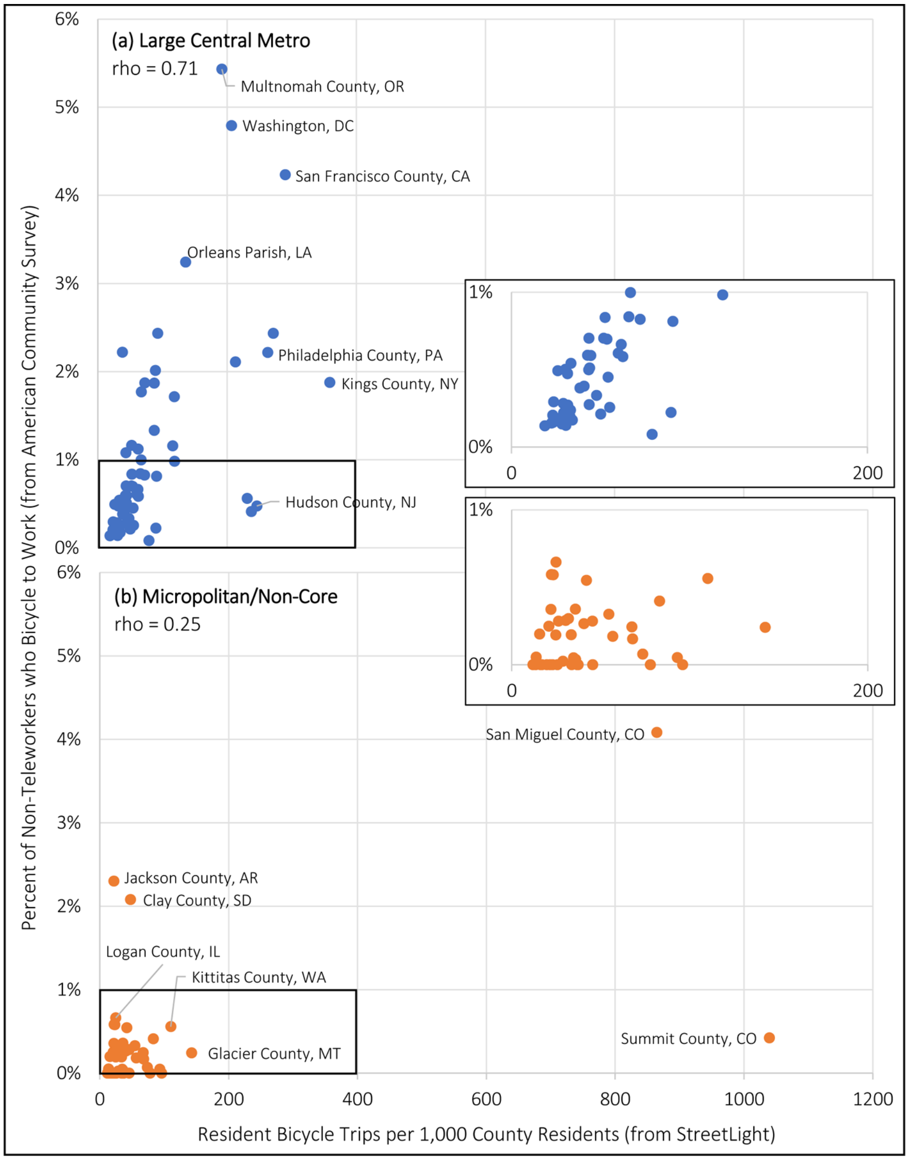


Fig. 3.
Percent bicycling to work (from American Community Survey) and resident bicycle trips (from StreetLight)—U.S. counties, (a) large central metro and (b) micropolitan/non-core, 2019.

Table 1.

Characteristics of all U.S. counties and sampled counties, 2019.

Characteristic	United States	Sampled Counties
No. counties	3142[a]	298
Total population, millions	324.7	140.5
Census region		
Midwest	1055 (33.6%)	66 (22.1%)
Northeast	217 (6.9%)	55 (18.5%)
South	1422 (45.3%)	122 (40.9%)
West	448 (14.3%)	55 (18.5%)
Urbanicity[b]		
Large central metro	68 (2.2%)	68 (22.8%)
Large fringe metro	368 (11.7%)	84 (28.2%)
Medium metro	372 (11.8%)	69 (23.2%)
Small metro	358 (11.4%)	30 (10.1%)
Micropolitan/Non-core	1976 (62.9%)	47 (15.8%)
Population density[c]		
Tertile 1 (< 95.3)	2220 (70.7%)	100 (33.6%)
Tertile 2 (95.3 to <570.5)	689 (21.9%)	99 (33.2%)
Tertile 3 (≥ 570.5)	233 (7.4%)	99 (33.2%)
Median age, years		
Tertile 1 (< 37.0)	579 (18.4%)	100 (33.6%)
Tertile 2 (37.0 to <41.3)	972 (30.9%)	101 (33.9%)
Tertile 3 (≥ 41.3)	1591 (50.6%)	97 (32.6%)
White race		
Tertile 1 (< 61.0%)	663 (21.1%)	100 (33.6%)
Tertile 2 (61.0% to <83.0%)	859 (27.3%)	99 (33.2%)
Tertile 3 (≥ 83.0%)	1620 (51.6%)	99 (33.2%)
Black race		
Tertile 1 (< 2.7%)	1678 (53.4%)	100 (33.6%)
Tertile 2 (2.7% to <12.6%)	779 (24.8%)	99 (33.2%)
Tertile 3 (≥ 12.6%)	685 (21.8%)	99 (33.2%)
Hispanic ethnicity		
Tertile 1 (< 4.3%)	1598 (50.9%)	100 (33.6%)
Tertile 2 (4.3% to <11.0%)	849 (27.0%)	99 (33.2%)
Tertile 3 (≥ 11.0%)	695 (22.1%)	99 (33.2%)
Living below FPL		

Characteristic	United States	Sampled Counties
Tertile 1 (< 10.6%)	786 (25.0%)	100 (33.6%)
Tertile 2 (10.6% to <15.1%)	972 (30.9%)	99 (33.2%)
Tertile 3 (≥ 15.1%)	1384 (44.0%)	99 (33.2%)
Median income		
Tertile 1 (< $53,948)	1843 (58.7%)	100 (33.6%)
Tertile 2 ($53,948 to <$66,641)	885 (28.2%)	99 (33.2%)
Tertile 3 (≥ $66,641)	414 (13.2%)	99 (33.2%)
Some college or more		
Tertile 1 (< 55.2%)	1855 (59.0%)	104 (34.9%)
Tertile 2 (55.2% to <64.3%)	846 (26.9%)	95 (31.9%)
Tertile 3 (≥ 64.3%)	441 (14.0%)	99 (33.2%)
CDC SVI percentile[d]		
Tertile 1 (< 0.32)	999 (31.8%)	100 (33.6%)
Tertile 2 (0.32 to <0.63)	992 (31.6%)	99 (33.2%)
Tertile 3 (≥ 0.63)	1150 (36.6%)	99 (33.2%)
Commute mode, all workers[e]		
Walk	2.67%	3.15%
Bicycle	0.55%	0.73%
Commute mode, non-teleworkers[e]		
Walk	2.81%	3.33%
Bicycle	0.58%	0.78%

Abbreviations: CDC SVI, Centers for Disease Control and Prevention social vulnerability index; FPL, federal poverty level

All data from American Community Survey 2015–2019, except where notated

[a] All counties and county equivalents in the 50 states and District of Columbia in 2010

[b] From the 2013 National Center for Health Statistics Urban-Rural Classification Scheme

[c] Defined as people per square mile

[d] From the Centers for Disease Control and Prevention social vulnerability index 2018; overall vulnerability index, with tertile 3 representing most vulnerable (for all U.S. counties, n = 3141)

[e] Primary mode of transportation to work; restricted to age ≥16 years

Table 2.

Correlations between county ranks for StreetLight measure of walking trips and American Community Survey walking to work—298 U.S. counties, 2019.

From StreetLight			From American Community Survey		rho[a]	95% CI[b]
Day	**Traveler**	**Purpose**	**Denominator**	**Workers**		
All	All	All	Daytime Population	All	0.45	0.37–0.54
All	All	All	Daytime Population	Non-TW	0.46	0.37–0.54
Weekdays	All	All	Daytime Population	All	0.46	0.38–0.55
Weekdays	All	All	Daytime Population	Non-TW	0.47	0.39–0.55
Weekends	All	All	Daytime Population	All	0.39	0.30–0.48
Weekends	All	All	Daytime Population	Non-TW	0.39	0.30–0.48
All	All	HBW	Daytime Population	All	0.26	0.17–0.35
All	All	HBW	Daytime Population	Non-TW	0.26	0.17–0.35
Weekdays	All	HBW	Daytime Population	All	0.30	0.21–0.40
Weekdays	All	HBW	Daytime Population	Non-TW	0.30	0.21–0.40
Weekends	All	HBW	Daytime Population	All	0.11	0.02–0.22
Weekends	All	HBW	Daytime Population	Non-TW	0.11	0.02–0.21
All	Residents	All	Residents	All	**0.52**	0.44–0.60
All	Residents	All	Residents	Non-TW	**0.53** [c]	0.45–0.61
Weekdays	Residents	All	Residents	All	**0.52**	0.44–0.60
Weekdays	Residents	All	Residents	Non-TW	**0.53**	0.45–0.61
Weekends	Residents	All	Residents	All	**0.50**	0.41–0.58
Weekends	Residents	All	Residents	Non-TW	**0.50**	0.42–0.59
All	Residents	All	Workers	All	0.42	0.33–0.51
All	Residents	All	Workers	Non-TW	0.42	0.33–0.51
Weekdays	Residents	All	Workers	All	0.43	0.34–0.52
Weekdays	Residents	All	Workers	Non-TW	0.43	0.34–0.52
Weekends	Residents	All	Workers	All	0.38	0.29–0.48
Weekends	Residents	All	Workers	Non-TW	0.38	0.29–0.48
All	Residents	All	Non-TW	All	0.44	0.35–0.53
All	Residents	All	Non-TW	Non-TW	0.44	0.35–0.53
Weekdays	Residents	All	Non-TW	All	0.44	0.35–0.53
Weekdays	Residents	All	Non-TW	Non-TW	0.44	0.35–0.53
Weekends	Residents	All	Non-TW	All	0.40	0.30–0.49
Weekends	Residents	All	Non-TW	Non-TW	0.40	0.30–0.49

Abbreviations: HBW, home-based work; TW, teleworkers

Bolded results indicate a strong correlation when using Cohen’s convention to interpret rho values (low correlation: <0.3; moderate: 0.3 to <0.5; strong: ≥0.5).

[a]Spearman rho correlations comparing walk trips per 1,000 people (from StreetLight) with percent walking to work (from the American Community Survey).

[b] Bootstrap confidence interval of Spearman rho

[c] Combination with the strongest correlation

Author Manuscript
Author Manuscript
Author Manuscript

Table 3.

County rank correlations between resident-only walk trips on all days for any purpose per 1,000 county residents (from StreetLight) with percent of non-teleworkers walking to work (from the American Community Survey)[a], by county characteristics—298 U.S. counties, 2019.

	N[b]	StreetLight[c] Median (IQR)	ACS[d] Median (IQR)	rho	95% CI[e]
All	298	1149 (460)	2.11 (2.13)	**0.53**	0.44–0.61
Census region					
Midwest	66	1140 (280)	2.19 (1.85)	**0.56**	0.36–0.70
Northeast	55	1177 (405)	3.38 (2.02)	**0.66**	0.48–0.79
South	122	994 (392)	1.48 (1.17)	0.43	0.27–0.56
West	55	1616 (704)	2.65 (2.74)	0.31	0.04–0.53
Urbanicity					
Large central metro	68	1297 (464)	2.77 (3.12)	**0.58**	0.40–0.72
Large fringe metro	84	958 (350)	1.66 (1.34)	0.40	0.21–0.57
Medium metro	69	1131 (454)	2.17 (1.85)	**0.57**	0.39–0.71
Small metro	30	1177 (325)	2.07 (1.08)	0.30	−0.07–0.59
Micropolitan/Non-core	47	1220 (813)	2.60 (3.24)	0.48	0.22–0.67
Population density					
Tertile 1 (< 95.3)	100	1206 (820)	2.09 (2.34)	0.46	0.29–0.60
Tertile 2 (95.3 to <570.5)	99	1063 (380)	1.62 (1.69)	**0.61**	0.47–0.72
Tertile 3 (≥ 570.5)	99	1179 (411)	2.66 (2.65)	**0.53**	0.37–0.66
Median age, years					
Tertile 1 (< 37.0)	100	1212 (555)	2.27 (2.42)	0.49	0.33–0.63
Tertile 2 (37.0 to <41.3)	101	1095 (410)	2.09 (1.63)	0.48	0.31–0.62
Tertile 3 (≥ 41.3)	97	1128 (457)	2.09 (1.96)	**0.59**	0.44–0.70
White race					
Tertile 1 (< 61.0%)	100	1210 (566)	2.13 (2.43)	**0.50**	0.33–0.63
Tertile 2 (61.0% to <83.0%)	99	1068 (430)	2.12 (1.72)	**0.55**	0.39–0.67
Tertile 3 (≥ 83.0%)	99	1148 (508)	2.11 (2.34)	**0.55**	0.39–0.67
Black race					
Tertile 1 (< 2.7%)	100	1311 (696)	2.34 (2.47)	0.43	0.25–0.58
Tertile 2 (2.7% to <12.6%)	99	1065 (422)	2.09 (1.84)	0.48	0.31–0.62
Tertile 3 (≥ 12.6%)	99	1069 (390)	1.90 (1.82)	**0.62**	0.48–0.73
Hispanic ethnicity					
Tertile 1 (< 4.3%)	100	1054 (407)	2.02 (1.92)	**0.66**	0.53–0.76
Tertile 2 (4.3% to <11.0%)	99	1156 (526)	2.15 (1.81)	**0.55**	0.40–0.68
Tertile 3 (≥ 11.0%)	99	1243 (512)	2.30 (2.27)	0.38	0.19–0.53

	N[b]	StreetLight[c] Median (IQR)	ACS[d] Median (IQR)	rho	95% CI[e]
Living below FPL					
Tertile 1 (< 10.6%)	100	1050 (459)	2.02 (2.24)	**0.54**	0.38–0.66
Tertile 2 (10.6% to <15.1%)	99	1213 (503)	2.11 (1.66)	0.47	0.30–0.61
Tertile 3 (≥ 15.1%)	99	1140 (462)	2.16 (2.29)	**0.59**	0.45–0.71
Median income					
Tertile 1 (< $53,948)	100	1089 (420)	1.93 (1.93)	**0.51**	0.35–0.64
Tertile 2 ($53,948 to <$66,641)	99	1260 (513)	2.34 (2.00)	0.41	0.23–0.56
Tertile 3 (≥ $66,641)	99	1064 (567)	2.25 (2.43)	**0.60**	0.46–0.71
Some college or more					
Tertile 1 (< 55.2%)	104	1057 (429)	1.57 (1.58)	0.39	0.21–0.54
Tertile 2 (55.2% to <64.3%)	95	1208 (415)	2.11 (2.18)	**0.56**	0.41–0.69
Tertile 3 (≥ 64.3%)	99	1204 (628)	2.70 (2.62)	**0.55**	0.39–0.67
CDC SVI percentile					
Tertile 1 (< 0.32)	100	1095 (549)	2.35 (2.61)	**0.62**	0.48–0.73
Tertile 2 (0.32 to <0.63)	99	1161 (398)	2.25 (1.60)	**0.53**	0.37–0.66
Tertile 3 (≥ 0.63)	99	1183 (572)	1.88 (1.79)	0.45	0.28–0.60

Abbreviations: ACS, American Community Survey; CDC SVI, Centers for Disease Control and Prevention social vulnerability index; CI, confidence interval; FPL, federal poverty level; IQR, interquartile range

Bolded results indicate a strong correlation when using Cohen's convention to interpret rho values (low correlation: <0.3; moderate: 0.3 to <0.5; strong: ≥0.5).

[a]Combination with the strongest correlation from table 2

[b]Number of counties

[c]Median (IQR) of any-day resident walk trips per 1,000 county residents

[d]Median (IQR) of the percent of non-teleworkers who report walking to work

[e]Confidence interval of Spearman rho based on Fisher's z transformation

Table 4.

Correlations between county ranks for StreetLight measure of bicycling trips and American Community Survey bicycling to work—298 U.S. counties, 2019.

From StreetLight			From American Community Survey		rho[a]	95% CI[b]
Day	Traveler	Purpose	Denominator	Workers		
All	All	All	Daytime Population	All	**0.52**	0.43–0.59
All	All	All	Daytime Population	Non-TW	**0.52**	0.47–0.63
Weekdays	All	All	Daytime Population	All	**0.52**	0.43–0.60
Weekdays	All	All	Daytime Population	Non-TW	**0.52**	0.43–0.60
Weekends	All	All	Daytime Population	All	**0.51**	0.43–0.59
Weekends	All	All	Daytime Population	Non-TW	**0.51**	0.43–0.59
All	All	HBW	Daytime Population	All	**0.52**	0.44–0.60
All	All	HBW	Daytime Population	Non-TW	**0.52**	0.45–0.61
Weekdays	All	HBW	Daytime Population	All	**0.53**	0.45–0.61
Weekdays	All	HBW	Daytime Population	Non-TW	**0.53**	0.45–0.61
Weekends	All	HBW	Daytime Population	All	0.49	0.40–0.57
Weekends	All	HBW	Daytime Population	Non-TW	0.49	0.41–0.58
All[c]	Residents	All	Residents	All	**0.60**	0.53–0.68
All[c]	Residents	All	Residents	Non-TW	**0.61** [f]	0.53–0.68
Weekdays[d]	Residents	All	Residents	All	**0.60**	0.52–0.67
Weekdays[d]	Residents	All	Residents	Non-TW	**0.60**	0.52–0.68
Weekends[e]	Residents	All	Residents	All	**0.59**	0.52–0.67
Weekends[e]	Residents	All	Residents	Non-TW	**0.60**	0.52–0.67
All[c]	Residents	All	Workers	All	**0.58**	0.50–0.65
All[c]	Residents	All	Workers	Non-TW	**0.58**	0.50–0.66
Weekdays[d]	Residents	All	Workers	All	**0.57**	0.49–0.65
Weekdays[d]	Residents	All	Workers	Non-TW	**0.57**	0.49–0.65
Weekends[e]	Residents	All	Workers	All	**0.57**	0.49–0.65
Weekends[e]	Residents	All	Workers	Non-TW	**0.57**	0.49–0.65
All[c]	Residents	All	Non-TW	All	**0.58**	0.50–0.66
All[c]	Residents	All	Non- TW	Non-TW	**0.58**	0.50–0.66
Weekdays[d]	Residents	All	Non- TW	All	**0.57**	0.49–0.65
Weekdays[d]	Residents	All	Non- TW	Non-TW	**0.57**	0.49–0.65
Weekends[e]	Residents	All	Non- TW	All	**0.57**	0.49–0.65
Weekends[e]	Residents	All	Non- TW	Non-TW	**0.57**	0.50–0.65

Abbreviations: HBW, home-based work; TW, teleworkers

Bolded results indicate a strong correlation when using Cohen's convention to interpret rho values (low correlation: <0.3; moderate: 0.3 to <0.5; strong: ≥0.5).

[a]Spearman rho correlations comparing bicycle trips per 1,000 people (from StreetLight) with percent bicycling to work (from the American Community Survey).

[b]Bootstrap confidence interval of Spearman rho

[c]n = 297 counties

[d]n = 295 counties

[e]n = 290 counties

[f]Combination with the strongest correlation

Author Manuscript
Author Manuscript
Author Manuscript

Table 5.

County rank correlations between resident-only bicycle trips on all days for any purpose per 1,000 county residents (from StreetLight) with percent of non-teleworkers bicycling to work (from the American Community Survey)[a], by county characteristics—297 U.S. counties, 2019.

	N[b]	StreetLight[c] Median (IQR)	ACS[d] Median (IQR)	rho	95% CI[e]
All	297	30 (32)	0.25 (0.43)	**0.61**	0.53–0.67
Census region					
Midwest	65	29 (16)	0.25 (0.37)	0.43	0.21–0.61
Northeast	55	32 (31)	0.25 (0.29)	**0.56**	0.35–0.72
South	122	22 (17)	0.14 (0.37)	**0.60**	0.47–0.70
West	55	70 (65)	0.56 (0.74)	0.38	0.13–0.59
Urbanicity					
Large central metro	68	53 (56)	0.57 (0.89)	**0.71**	0.57–0.81
Large fringe metro	84	24 (15)	0.17 (0.25)	**0.60**	0.45–0.72
Medium metro	69	26 (28)	0.23 (0.40)	**0.60**	0.42–0.73
Small metro	30	28 (21)	0.31 (0.59)	0.49	0.15–0.72
Micropolitan/Non-core	46	34 (34)	0.19 (0.36)	0.25	−0.05–0.50
Population density					
Tertile 1 (< 95.3)	99	32 (36)	0.18 (0.36)	0.40	0.22–0.55
Tertile 2 (95.3 to <570.5)	99	25 (22)	0.20 (0.31)	**0.70**	0.58–0.79
Tertile 3 (≥ 570.5)	99	40 (40)	0.45 (0.63)	**0.75**	0.65–0.83
Median age, years					
Tertile 1 (< 37.0)	100	36 (43)	0.40 (0.66)	**0.63**	0.50–0.74
Tertile 2 (37.0 to <41.3)	101	29 (26)	0.25 (0.42)	**0.67**	0.55–0.77
Tertile 3 (≥ 41.3)	96	27 (25)	0.11 (0.33)	0.48	0.31–0.62
White race					
Tertile 1 (< 61.0%)	100	39 (49)	0.28 (0.53)	**0.70**	0.59–0.79
Tertile 2 (61.0% to <83.0%)	99	27 (29)	0.25 (0.46)	**0.74**	0.63–0.82
Tertile 3 (≥ 83.0%)	98	28 (23)	0.19 (0.33)	0.38	0.19–0.54
Black race					
Tertile 1 (< 2.7%)	99	37 (48)	0.25 (0.48)	0.42	0.25–0.57
Tertile 2 (2.7% to <12.6%)	99	28 (26)	0.25 (0.39)	**0.68**	0.55–0.77
Tertile 3 (≥ 12.6%)	99	28 (26)	0.22 (0.47)	**0.75**	0.65–0.83
Hispanic ethnicity					
Tertile 1 (< 4.3%)	100	24 (16)	0.16 (0.34)	0.43	0.26–0.58
Tertile 2 (4.3% to <11.0%)	98	30 (31)	0.27 (0.44)	**0.68**	0.56–0.78
Tertile 3 (≥ 11.0%)	99	44 (49)	0.31 (0.50)	**0.65**	0.52–0.75

	N[b]	StreetLight[c] Median (IQR)	ACS[d] Median (IQR)	rho	95% CI[e]
Living below FPL					
Tertile 1 (< 10.6%)	99	28 (26)	0.21 (0.37)	**0.59**	0.45–0.71
Tertile 2 (10.6% to <15.1%)	99	33 (37)	0.29 (0.43)	**0.61**	0.47–0.72
Tertile 3 (≥ 15.1%)	99	29 (27)	0.25 (0.52)	**0.60**	0.46–0.72
Median income					
Tertile 1 (< $53,948)	100	24 (21)	0.16 (0.38)	0.45	0.27–0.59
Tertile 2 ($53,948 to <$66,641)	98	33 (29)	0.27 (0.41)	0.47	0.30–0.61
Tertile 3 (≥ $66,641)	99	32 (49)	0.29 (0.64)	**0.80**	0.72–0.86
Some college or more					
Tertile 1 (< 55.2%)	104	23 (17)	0.08 (0.27)	0.39	0.21–0.54
Tertile 2 (55.2% to <64.3%)	94	32 (24)	0.28 (0.32)	**0.59**	0.44–0.71
Tertile 3 (≥ 64.3%)	99	43 (49)	0.50 (0.85)	**0.68**	0.55–0.77
CDC SVI percentile					
Tertile 1 (< 0.32)	99	28 (34)	0.21 (0.41)	**0.58**	0.43–0.70
Tertile 2 (0.32 to <0.63)	99	32 (30)	0.27 (0.46)	**0.60**	0.45–0.71
Tertile 3 (≥ 0.63)	99	31 (35)	0.27 (0.47)	**0.63**	0.49–0.73

Abbreviations: ACS, American Community Survey; CDC SVI, Centers for Disease Control and Prevention social vulnerability index; CI, confidence interval; FPL, federal poverty level; IQR, interquartile range

Bolded results indicate a strong correlation when using Cohen's convention to interpret rho values (low correlation: <0.3; moderate: 0.3 to <0.5; strong: ≥0.5).

[a] Combination with the strongest correlation from Table 4

[b] Number of counties

[c] Median (IQR) of any-day resident bicycle trips per 1,000 county residents

[d] Median (IQR) of the percent of non-teleworkers who report bicycling to work

[e] Confidence interval of Spearman rho based on Fisher's z transformation