



HHS Public Access

Author manuscript

Health Place. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

Health Place. 2020 May ; 63: 102332. doi:10.1016/j.healthplace.2020.102332.

Residential neighborhood features associated with objectively measured walking near home: Revisiting walkability using the Automatic Context Measurement Tool (ACMT)

Stephen J. Mooney¹, Philip M. Hurvitz^{2,3}, Anne Vernez Moudon², Chuan Zhou^{4,5}, Ronit Dalmat¹, Brian E. Saelens^{4,5}

¹Department of Epidemiology, University of Washington, Seattle, WA, United States

²Urban Form Lab, University of Washington, Seattle, WA, United States

³Center for Studies in Demography and Ecology, University of Washington, Seattle, WA, United States

⁴Seattle Children's Research Institute, Seattle, WA, United States

⁵Department of Pediatrics, University of Washington, Seattle, WA, United States

Abstract

Many distinct characteristics of the social, natural, and built neighborhood environment have been included in walkability measures, and it is unclear which measures best describe the features of a place that support walking. We developed the Automatic Context Measurement Tool, which measures neighborhood environment characteristics from public data for any point location in the United States. We explored these characteristics in home neighborhood environments in relation to walking identified from integrated GPS, accelerometer, and travel log data from 681 residents of King County, WA. Of 146 neighborhood characteristics, 92 (63%) were associated with walking bout counts after adjustment for individual characteristics and correction for false discovery. The strongest built environment predictor of walking bout count was housing unit count. Models using data-driven and *a priori* defined walkability measures exhibited similar fit statistics. Walkability measures consisting of different neighborhood characteristic measurements may capture the same underlying variation in neighborhood conditions.

Keywords

American Community Survey; EPA Walkability Index; Neighborhood Environment-Wide Association Study; Walking Bouts

Correspondence to: Stephen Mooney, 1959 NE Pacific Street, Health Sciences Bldg, F-262, Box 357236, Seattle, WA 98195, 206.685.1643, sjm2186@u.washington.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 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.

Conflicts of Interest: None Declared

Introduction

Despite considerable evidence that variation in the built environment is associated with variation in walking (Ding and Gebel, 2012; Grasser et al., 2013; Hajna et al., 2015; Hirsch et al., 2013; Owen et al., 2007; Saelens and Handy, 2008), measuring a given place's walkability remains problematic given inconsistently measured built environments (Shashank and Schuurman, 2019). For example, of the six published papers identified in a narrowly scoped 2015 systematic review of the relation between geographic information systems-based walkability measures and objectively measured steps per day, no two studies used the same set of variables to represent walkability (Hajna et al., 2015).

There are several reasons for this inconsistency. First, the word *walkability* does not refer to a single, consistent, construct – the term is sometimes used to characterize an environment where people *choose to* walk but is also sometimes used to characterize an environment that *better supports* walking (Forsyth, 2015). These related but distinct interpretations in the construct of walkability call for different measures; for example, according to the first definition, parking availability might be a component of a walkability measure because low parking availability is a key determinant of active travel mode choices including walking (Rodriguez et al., 2008) but would not be applicable to the second, because parking does not directly support or preclude walking.

Second, many built environment characteristics could plausibly influence an individual's choice to walk, including proximity to destinations, infrastructure such as sidewalks, parks, and transit access (Rundle et al., 2019). These characteristics are often spatially correlated. For example, urban retail destinations often have transit access and sidewalks. However, not all features thought to encourage walking are positively correlated. For example, within cities, park land typically contains little to no retail, such that the presence of parks may be negatively correlated with retail density. Indeed, it may be that park land supports walking, yet does so *less* than higher residential density that might exist in the absence of land being dedicated to parks. In such a circumstance, parkland could be negatively associated with walking in spite of its support for walking. Meanwhile, neighborhood measures that are correlated with residential density but do not have a strong theoretical link to walking, such as count of residents aged 4554, will be positively associated with walking, and thus appear to be a good walkability measure. Figure 1 illustrates these relationships visually.

Third, any given neighborhood characteristic can be measured many different ways. For example, neighborhood commuting behaviors could be measured as the count of neighborhood residents who drive alone, or as the proportion of neighborhood residents who choose active modes, or commuter train ridership counts at a nearby train station (Johnson and Lu, 2011). Moreover, these choices could apply different spatial measurement frameworks; one researcher might define a neighborhood using census tracts and another might use Euclidean or network-based buffers around the home (L. D. Frank et al., 2008).

Because the data to construct neighborhood measures are inconsistently available over space and time (e.g., some jurisdictions have released sidewalk GIS data to the public while others may not have or do not make such data accessible) (Hirsch et al., 2016), individual

researchers frequently choose to construct walkability measures that mutually optimize for face validity and ease of construction given available data, potentially precluding straightforward replication in other contexts (Ioannidis, 2018). Therefore, identifying one or several variables using nationally and longitudinally available data that predict walking behavior with as much precision as more complex or location specific indices may remove a key barrier to better understanding the relationship of urban form to walking behavior and subsequent health outcomes across a variety of locations.

The notion that decisions to walk incorporate both individual preferences and socio-ecological environments (Sallis et al., 2015) raises several questions with biomedical analogs: first, which of the many features that are correlated with walking are the best leverage points for policy intervention to increase walking (e.g., changing zoning or adding transit service)? This question parallels the biomedical goal to identify the leverage points in cellular processes for pharmaceutical intervention, as when scanning differences in RNA transcription markers between diseased and disease-free study participants (Rundle et al., 2012). Second, can we use data-driven approaches to identify individual or composite measures that capture walkability more accurately than indices constructed from *a priori* selected variables? This question parallels the biomedical goal of accounting for within-metabolism systemic feedback in toxicology (Maresca et al., 2015) and microbiome research (Turnbaugh et al., 2006) using mixture models or principal component analysis.

These analogies suggest we may use tools previously used to study biological systems to assess the social/spatial systems in which neighborhood walking occurs. First, to explore this notion, we developed the Automatic Context Measurement Tool (ACMT), an R package that uses nationally and freely available spatial data in the U.S. to assemble a panel of 146 neighborhood measures comprising built and natural environment, residential demographics, and behaviors for any location in the United States. We applied the ACMT to measure the neighborhoods surrounding the residences of a well-defined cohort with detailed walking measures based in King County, WA. Then, to identify “candidate walkability measures”, we conducted a Neighborhood Environment-Wide Association Study (NE-WAS) (Lynch, 2019; Lynch et al., 2017; Mooney et al., 2017), analyzing each neighborhood measure independently. Second, to explore whether a dimensionality reduced composite measure might better explain walking than any particular neighborhood measure on its own, we conducted a principal component analysis of all the neighborhood measures that we selected and explored the first component’s association with walking near home.

Methods

Setting

We used data from the Travel Assessment and Community (TRAC) study. TRAC is a three-wave cohort study of adult residents of King County, WA, designed to assess the travel and physical activity impacts of a light rail system that opened in 2009 (Author, 2014c). King County is a large and geographically diverse county including the City of Seattle.

Participants

Recruitment for TRAC began in July 2008 and continued through July 2009. The sampling and recruitment process has been described in more detail elsewhere (Moudon et al., 2009). About half of the participants were selected from among people living within 1.6 km of at least one future light rail stop and half were selected from among people living other neighborhoods within King County that matched rail stop-adjacent neighborhoods on household income, racial/ethnic composition, residential property values, residential density, land use mix, bus ridership, and housing type. Households in eligible areas were contacted using address and phone information obtained from MSG Marketing Systems Group, a consumer marketing company. Enrollment processes ensured participants were adults 18 years old or older who gave informed consent, were able to walk unassisted for at least 10 minutes, and completed the travel log and survey in English. Follow-up assessments occurred 1–2 and 3–4 years after the July 2009 opening of the light rail line. From 801 TRAC participants who consented, we excluded 120 for whom personal monitoring data for at least 10 hours of one day in one of the three waves were not available. The Seattle Children’s Research Institute Institutional Review Board approved the study.

Survey Measures

Participants reported their age, gender, race/ethnicity, height, and educational attainment at baseline. At baseline and each follow-up wave, they reported annual household income, weight, count of other household members, and count of motor vehicles available to the household. For analysis to best reflect the sample, we categorized age as: < 40, 40–65, or > 65; race/ethnicity as: non-Hispanic White or other; annual household income as: less than \$50,000, \$50,000 - \$100,000, or over \$100,000. All categories were chosen for direct comparison with prior analyses of these data using *a priori* spatial measures to assess determinants of walking near home (Huang et al., 2019). Very little survey data were missing (10% of respondents did not report household income; no other measure was missing more than 5% of responses); nonetheless, we used chained equations to impute missing survey responses (Buuren and Groothuis-Oudshoorn, 2010) and combined estimates from five imputed datasets using Rubin’s rules (Rubin, 2004).

Travel Behavior Measures

At each wave, participants were asked to wear an Actigraph GT1M or GT3X accelerometer and a GlobalSat DG-100 portable or QStarz BG-1000XT global positioning system (GPS) data logger for seven days while also recording their travel in a paper diary over the same period. Accelerometers were set to aggregate counts in 30-second epochs. GPS devices were set to collect location data at 30-second intervals, and these GPS, accelerometer, and travel log data were merged in “LifeLogs” that synchronized these data by timestamp (Author, 2014b). The study team timed contact with participants to maximize monitoring at the same time of the year as prior waves (70% of those monitored in waves 1 and 2 started wave 2 within 4 weeks of their wave 1 month/day start date, as did 39% of those monitored in waves 2 and 3).

Walking Bouts

Using an algorithm described in detail elsewhere (Author et al., 2013), we identified “walking bouts” -- periods of time in which the accelerometer indicated the participant was physically active and the GPS and/or travel diary data were consistent with walking behavior. Specifically, walking bouts were defined as sustained intervals as at least 5 total minutes of accelerometer readings over 500 counts per 30-second epoch over a period, or at least 7 minutes with no more than 2 minutes below that threshold and either GPS information (e.g., speed range of 2–6 km/h) or travel diary entries indicating walking as the travel mode. We considered a bout to be within the home neighborhood but not at home if any of its GPS points fell within the 833 m buffer around the home location but at least one point fell more than 125 m from home defined as above as the centroid of GPS locations within a parcel (Author, 2019b; Author, 2014a).

Residential Neighborhood Measures

We used King County’s E-911 address point GIS data to identify a land parcel corresponding to each participant reported home address and considered the centroid of that parcel to represent the residential location. For subjects whose home address geocoded to a parcel larger than 2 acres (e.g., large apartment complexes), we considered the centroid of all that individuals’ GPS points within that parcel to be the home location for that individual participant (Author, 2014a). We considered the 833 meter Euclidean buffer (approximately ½ mile, corresponding approximately to a 10-minute walk) around that home address to constitute the home neighborhood. We chose Euclidean buffers rather than network buffers (Oliver et al., 2007) or sausage buffers (Forsyth et al., 2012) to minimize computational complexity and to ensure comparability with prior work with these data (Author, 2019b).

Because our goal was to conduct a NE-WAS (Mooney et al., 2017) to identify a measure or measures of neighborhood context that would function in King County but could easily be used across other spatial contexts, we selected candidate measures that were available in a standardized way across the United States. Using an informal literature review, we identified three sources of nationally available standardized data that included variables to capture neighborhood context that could plausibly affect walking. Specifically, we used the American Community Survey (United States Census Bureau, n.d.), the National Land Cover Database (Multi-Resolution Land Characteristics Consortium, n.d.), and the Environmental Protection Agency’s National Walkability Index database (Environmental Protection Agency, n.d.). We identified 141 neighborhood measures (76 count and 65 proportion) from the 2010 American Community Survey, four from the 2011 National Land Cover Database, and one from the National Walkability Index, which itself draws on other data sources released from 2010-2012.

Some of these residential population measures, such as number of adults aged 45-54 living in the neighborhood, have no clear theoretical link to walking behavior. These measures act to illustrate the noise that can be found in observational associations – that is, finding a measure with no theoretical relevance to walking is the best predictor of walking behavior raises concerns that our results arose by chance or that the measure acts a proxy for a theoretically important construct we are failing to capture.

Neighborhood measures are described in more detail in the web material. Briefly, our measure set comprises two types: (A) measures of a place's natural and built environment (e.g., proportion of undeveloped land, residential density) and (B) measures of neighborhood residents -- social composition and behavior measures that may be influenced by or associated with built environment (e.g., commute mode, proportion of never-married adults [who, for example, are more likely to live in apartment buildings]). Any of these measures may be able to account for differences in neighborhood support for walking in statistical models. However, whereas natural and built environment measures capture place-specific amenities which policy makers might intervene on to increase 'walkability', social composition is not typically a walkability target. For example, a zoning intervention affects allowable residential density directly; and increased densities may increase traffic congestion that may in turn indirectly affect commute mode choice. In contrast, a policy seeking to modify commute mode behavior directly, such as a subsidized transit pass, would not typically be considered a 'walkability' intervention.

ACMT

Next, we developed the ACMT to compile standardized measures of neighborhood context for any location in the United States. While the National Land Cover Database is available as a 30 m grid surface that can easily be aggregated to larger spatial scales, the American Community Survey and National Walkability Index data are available by census tract. In order to estimate the value of each variable within each 833 m neighborhood Euclidean buffer of participants' residence, the ACMT uses GIS overlay functions that allocate portion of the total tract value to each "clipped" spatial subset of a tract within a buffer based on the ratio of the subset's area to total tract area. Each subset's values for each variable were then summed to estimate the value for the Euclidean buffer as a whole. This process allows us to estimate measures in participant-specific buffers that often overlap with multiple census tracts under the assumption that all variables are evenly distributed across each tract. Web Figure 1 is a dendrogram displaying a hierarchical clustering of these measures. ACMT code is available at <https://github.com/smooney27/ACMT>.

Statistical Analysis

First, we regressed the total number of walk bouts over all three waves on each residential neighborhood measure separately, adjusting for individual age, gender, race/ethnicity, and household income. We selected walking bout count rather than total duration walking in order to focus on the decision to walk; however, we assessed duration of time walking in a sensitivity analysis, as detailed below. We fit negative binomial models, using number of monitored days over the three waves as an offset. We selected the negative binomial distribution to account for the over-dispersed distribution of walk bouts in the population. We developed Manhattan plots to visually assess the strength of association between neighborhood variables and walk bouts per monitored day.

Next, we compared five mixed effects negative binomial models using specific factors to predict number of home neighborhood walk bouts: one using no variable to characterize the neighborhood, one using the EPA's walkability index, one using local individual tax parcel-based residential density and employment density (which were previously found to predict

home neighborhood walking in this sample (Author, 2019b)), one using the first principal component from a principal component analysis of all 146 residential neighborhood measures, and one using the built environment measure from the NE-WAS analysis with the strongest statistical association. We tested the principal component analysis approach to explore how in order to explore the hypothesis that a measure compositing variation from multiple measurable neighborhood factors could better predict walking than any individual measure alone. We used 10-fold cross-validation to measure model predictive accuracy, considering the median value from across the five imputed datasets to best estimate root mean log squared error (RMLSE). All models adjusted for individual age, gender, race/ethnicity, and household income.

We also performed several opportunistic analyses. First, principal component analysis has previously been used to select candidate measures in a prior data-driven analysis of neighborhood predictors of health (Lynch et al., 2017). Building on the principal component analysis described above, we replicated that work's analytic plan to see which measures might represent the best individual candidate measures. This work is detailed in Web Appendix 1. Second, we repeated the analysis using total duration of walk bouts as the outcome in order to ensure our findings were not an artifact of shorter walk bouts in higher walkability neighborhoods. Walk duration analyses used linear rather than negative binomial models.

All analyses used R version 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria).

Results

People and walk bouts

Data were available to study walking behavior from a total of 681 TRAC participants over three waves of data collection (Table 1), though not all participants were monitored in each wave. Compared with the King County adult population, a greater proportion of study participants were female (63%), non-Hispanic white (80%) and highly educated (70% college graduates). During monitored weeks, participants averaged 7 walk bouts per week in the defined home residential neighborhood and 6 walk bouts per week elsewhere in the county. The number of walk bouts in the home neighborhood was right-skewed: 113 participants never walked near home, whereas one participant engaged in 78 walk bouts during the course of the study. Number of walk bouts near home was positively correlated within people across waves ($r = 0.50$ between waves 1 and 2, and waves 2 and 3).

Home neighborhoods of residents

Residential neighborhood characteristics varied among study participants. For example, the estimated proportion of neighborhood residents who worked outside the home and drove alone to work ranged from 21% to 79%, suggesting participants experienced considerable variation in neighborhood transportation behaviors. There was strong positive inter-correlation between some neighborhood measures (Web Figure 3), with 16% of measure pairs having correlation coefficients above 0.8 or below -0.8 . Figure 2 shows scatterplots of selected pairs of neighborhood characteristics.

Residential neighborhood predictors of walking near home

In participant characteristic-adjusted negative binomial models considering each neighborhood measure independently, 92 (63%) were significantly associated at the $p < 0.05$ level after a Benjamini-Hochberg false discovery rate correction, and 80 of the 146 (55%) measures remained significantly associated with number of walk bouts after a more conservative Bonferroni correction (Figure 3). The 5 most strongly associated neighborhood measures, all of which were positively associated with walking bouts were (1) count of residents who self-reported White race, (2) count of male residents, (3) count of adults aged 15 and over, (4) count of residents who are US citizens and were born in the United States, and (5) count of residents who were born in the United States. These five measures were all correlated with each other with correlation coefficients > 0.9 .

Housing unit count (ranked tenth) was the highest ranked built/natural environment measure, and the next most strongly predictive built/natural environment measure was the EPA walkability index (ranked thirty-third). The list of neighborhood measures ordered by strength of association with walk bout count is given in Web Table 1.

Principal component analysis revealed that 44% of variation in the neighborhood measures was explained by the first principal component, which generally described features of denser housing conditions. The second principal component, which explained a further 14% of the variation, generally described the pervasiveness of traditional transportation and household patterns. Details regarding the identified components are summarized in Web Figure 2

Measure predictive accuracy comparison

In all models, neighborhood variables were significantly associated with walking bouts and the individual-level variables were not. The model including number of housing units had the best predictive accuracy and the EPA walkability index had the worst, but the differences were small (RMLSE: 4.30 for housing units, 4.36 for EPA index, Table 2). The model using principal components did not perform better than residential density and employment density alone. The increase in odds of an additional walk bout near home was associated with a one standard deviation increase in each of the measures was similar across all measures (Table 3).

Sensitivity Analyses

Associations between total walking duration and neighborhood environment characteristics were generally weaker than associations between bout count and neighborhood characteristics (only 2 measures – count of widowed residents and count of separated residents – were significantly associated after Bonferroni correction). However, the overall pattern remained that walking was more strongly associated with count measures than proportion measures and that many neighborhood measures were associated with walking duration (Web Figure 4).

Discussion

In this analysis, we explored the notion that neighborhood features that encourage walking are strongly inter-correlated such that measures of each can act interchangeably to predict counts of measured walk bouts. Consistent with previous work (Author, 2019b), we found that higher scores of walkability measures near the home were associated with more walking near the home. The strengths of association were similar between data-driven processes explored in the present analyses and home environment measures defined *a priori* (Author, 2019b), suggesting that most or all walkability measures, regardless of contributing components, may be picking up on the same variability in the neighborhoods. This suggests that, at least given the current nationally available measures, walkability researchers who are considering developing a new walkability index might be better served selecting a single reliable and longitudinally available measure available freely everywhere, perhaps residential density, for future studies.

We caution, however, that our measure set comprised not only measures of natural (e.g., land devoted to open space) and built environment (e.g., residential density) but also measures of social context and behaviors that are shaped by neighborhood environments but not directly modifiable by urban design (e.g., proportion of residents over age 85). While either type of measure may be appropriate for analysts hoping to control statistically for differences between neighborhood support for walking, only the former type of measure are of interest for planners looking to encourage walking. In King County, neither measure type was clearly superior at predicting walking. However, this similarity may not hold in other contexts; one of the benefits of our approach of using only nationally available data for local measurement is the opportunity to compare identically assessed measures in different geographic contexts.

Our finding that neighborhood measures were highly inter-correlated is consistent with prior work exploring multiple measures of neighborhood environments (Lynch et al., 2017). We used principal component analysis not only to compare with this prior work but also because it is a standard and widely used dimensionality reduction technique. An alternate dimensionality reduction approach to identify a neighborhood typology is latent class analysis. Latent class analysis has been used previously in some neighborhood studies (Adams et al., 2015, 2013, 2012; Weden et al., 2011), and poses an important direction for future research on the tradeoffs between nationally available, standardized measures and more local measures. However, dimensionality reduction approaches cannot distinguish between variation due to underlying features and variation due to data source; a substantial concern for ACMT-based analyses pulling from separate but disproportionately represented datasets. In this light, complex systems approaches exploring how neighborhood features affect each other (e.g., do increases in transit commute mode share lead to increases in population density) could shed important light on sources of intercorrelation.

Of all 146 neighborhood measures we compiled, the best individual predictor was the number of housing units within the ¼ mile buffer around home. While the notion that residential density is associated with walking has been established on theoretical and empirical grounds (Carlson et al., 2015; L. D. Frank et al., 2008), and the urban design

literature supports mechanistic explanations for this association (e.g., high density housing correlates with more non-residential destinations nearby and lower parking availability (Manville et al., 2013)), some prior work has also suggested this association arises due to structural confounding (Forsyth et al., 2007; Oakes et al., 2007). We caution that the atheoretical approach to this analysis, wherein we measured and analyzed 146 neighborhood measures, increases the likelihood that any given characteristic was most associated with walking by chance. Much as genetic studies require validation in another cohort before results can be interpreted as substantively meaningful, NE-WAS results should be presumed to have arisen by chance until demonstrated otherwise (Ioannidis et al., 2001). We will pursue this replication ourselves while also making the ACMT publicly available so other researchers can as well.

In addition to replication, our study highlights several other areas for important future work. First, for expedience, we assumed a standard neighborhood definition of 833 m Euclidean buffers and a linear relationship between neighborhood characteristics and walking bouts. Future work should explore different neighborhood definitions (James et al., 2014) and different analytic approaches, potentially including distributed lag models (Baek et al., 2016) and flexible functional forms (Savitz et al., 2014). Poisson-gamma ('hurdle') and zero-inflated models that distinguish between contextual correlates of walking at all and of duration walked among walkers may shed further light on different aspects of walkability (Huang et al., 2019). Such analyses may not only help understand scale of neighborhood walkability impacts but also the risk that candidate neighborhood measures are selected due to artifacts arising from modeling choices.

A key strength of this study is our use of nationally available data to create neighborhood measures, allowing replication in other geographic contexts. Our findings also benefit from objective and spatially precise measures of walking. That our walk bouts occurred in the same spatial context where our measures were taken partially mitigates the 'residential effect fallacy', wherein spatial autocorrelation in neighborhood features leads to overestimates of the impact of neighborhoods (Chaix et al., 2017). Other strengths include a fairly large sample size for objectively measured data and a geographic context, King County, Washington, which has substantial variability in urban form.

However, our results should be understood in light of several key limitations. First, we constructed environmental measures only around participants' residences. We defined the residential neighborhood as an 833 meter circular buffer around the home, a substantial oversimplification of the area comprising complete environmental exposures. As noted above, future work should explore the impacts of buffer size on measures of walkability. Second, our study participants were selected to be in neighborhoods adjacent to a new light rail line or in neighborhoods identified as similar to those neighborhoods in terms of demographics and built environment; thus the sampled neighborhoods did not represent a random geographic or representative sample of King County. This selection process may have limited the environmental heterogeneity of the sample. Moreover, while King County has racial and ethnic diversity comparable to the broader United States (in the 2010 US Census, 65% of King County residents report non-Hispanic ethnicity and white race, as compared to 63% of the United States as a whole), our sample is less ethnically diverse

(79% non-Hispanic White). King County residents also have substantially higher income (2010 median household income of \$68,000 as compared with \$49,000 nationally) than the nation as a whole, as did our sample. We anticipate replicating this approach in other populations in other geographic contexts in order to confirm findings, and plan to make the ACMT available to other researchers as well.

Third, one tension in using data-driven techniques such as principal component analysis with standardized measures is that the patterns underlying the principal components are by definition confined to the observed data and therefore the variability they incorporate is limited by the range of observed variability. The first principal component identified as explaining 43.6% of the variation in these measures would likely explain less variation if applied to similar measures compiled in a different geographic context or with a different sample of participants. If dimensionality reduction techniques prove useful for defining walkability from numerous measures in the future, identifying the appropriate sampling frame for compiling measures will be a key component of the research (Author et al., 2019a). Finally, our results represent an exploratory analysis of candidate measures. Further work, perhaps focusing on people moving between neighborhoods (Hirsch et al., 2014), are needed to explore causal relationships between neighborhood variables and walking.

In conclusion, we conducted a NE-WAS study to identify neighborhood environment correlates of walking near home in a sample of adults in King County Washington. Our study used the newly developed Automatic Context Measurement Tool, which is available to other research teams who might conduct similar neighborhood studies, to compile these environment measures. We found that measures were highly inter-correlated and that differences in predictive accuracy between measures were modest and may have arisen due to chance. Neither metrics that capture environmental use (e.g., commute times) nor metrics that capture directly modifiable features of the built environment (e.g., housing units) were consistently better predictors of walking, though the latter may be more useful to planners focused on environmental change. While our findings are preliminary, they suggest that most environmental measures of walkability, including ones created using nationally available data sources, can substitute for each other with little loss in information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this article was supported by the National Heart, Lung, and Blood Institute (award R01HL091881) and the National Library of Medicine (K99LM012868).

References

Adams MA, Ding D, Sallis JF, Bowles HR, Ainsworth BE, Bergman P, Bull FC, Carr H, Craig CL, De Bourdeaudhuij I, others, 2013 Patterns of neighborhood environment attributes related to physical activity across 11 countries: a latent class analysis. *International journal of behavioral nutrition and physical activity* 10, 34. [PubMed: 23497187]

- Adams MA, Sallis JF, Conway TL, Frank LD, Saelens BE, Kerr J, Cain KL, King AC, 2012 Neighborhood environment profiles for physical activity among older adults. *American Journal of Health Behavior* 36, 757–769. [PubMed: 23026035]
- Adams MA, Todd M, Kurka J, Conway TL, Cain KL, Frank LD, Sallis JF, 2015 Patterns of walkability, transit, and recreation environment for physical activity. *American journal of preventive medicine* 49, 878–887. [PubMed: 26232902]
- Baek J, Sánchez BN, Berrocal VJ, Sanchez-Vaznaugh EV, 2016 Distributed lag models: examining associations between the built environment and health. *Epidemiology (Cambridge, Mass.)* 27, 116.
- Buuren S van, Groothuis-Oudshoorn K, 2010 mice: Multivariate imputation by chained equations in R. *Journal of statistical software* 1–68.
- Carlson JA, Saelens BE, Kerr J, Schipperijn J, Conway TL, Frank LD, Chapman JE, Glanz K, Cain KL, Sallis JF, 2015 Association between neighborhood walkability and GPS-measured walking, bicycling and vehicle time in adolescents. *Health & place* 32, 1–7. [PubMed: 25588788]
- Chaix B, Duncan D, Vallée J, Vernez-Moudon A, Benmarhnia T, Kestens Y, 2017 The “Residential” Effect Fallacy in Neighborhood and Health Studies. *Epidemiology* 28, 789–797. [PubMed: 28767516]
- Ding D, Gebel K, 2012 Built environment, physical activity, and obesity: what have we learned from reviewing the literature? *Health & place* 18, 100–105. [PubMed: 21983062]
- Environmental Protection Agency, n.d. Smart Location Database. URL <https://www.epa.gov/smartgrowth/smart-location-mapping> (accessed 7.29.19).
- Forsyth A, 2015 What is a walkable place? The walkability debate in urban design. *Urban design international* 20, 274–292.
- Forsyth A, Oakes JM, Schmitz KH, Hearst M, 2007 Does residential density increase walking and other physical activity? *Urban studies* 44, 679–697.
- Forsyth A, Van Riper D, Larson N, Wall M, Neumark-Sztainer D, 2012 Creating a replicable, valid cross-platform buffering technique: the sausage network buffer for measuring food and physical activity built environments. *International journal of health geographics* 11, 14. [PubMed: 22554353]
- Frank LD, Bradley M, Kavage S, Chapman J, Lawton TK, 2008 Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation* 35, 37–54.
- Frank LD, Kerr J, Sallis JF, Miles R, Chapman J, 2008 A hierarchy of sociodemographic and environmental correlates of walking and obesity. *Preventive medicine* 47, 172–178. [PubMed: 18565576]
- Grasser G, Van Dyck D, Titze S, Stronegger W, 2013 Objectively measured walkability and active transport and weight-related outcomes in adults: a systematic review. *International journal of public health* 58, 615–625. [PubMed: 23224518]
- Hajna S, Ross NA, Brazeau A-S, Bélisle P, Joseph L, Dasgupta K, 2015 Associations between neighbourhood walkability and daily steps in adults: a systematic review and meta-analysis. *BMC public health* 15, 768. [PubMed: 26260474]
- Hirsch JA, Diez Roux AV, Moore KA, Evenson KR, Rodriguez DA, 2014 Change in walking and body mass index following residential relocation: the multi-ethnic study of atherosclerosis. *American journal of public health* 104, e49–e56.
- Hirsch JA, Meyer KA, Peterson M, Rodriguez DA, Song Y, Peng K, Huh J, Gordon-Larsen P, 2016 Obtaining longitudinal built environment data retrospectively across 25 years in four US cities. *Frontiers in public health* 4, 65. [PubMed: 27148512]
- Hirsch JA, Moore KA, Evenson KR, Rodriguez DA, Roux AVD, 2013 Walk Score® and Transit Score® and walking in the multi-ethnic study of atherosclerosis. *American journal of preventive medicine* 45, 158–166. [PubMed: 23867022]
- Huang R, Moudon AV, Zhou C, Saelens BE, 2019 Higher residential and employment densities are associated with more objectively measured walking in the home neighborhood. *Journal of Transport & Health* 12, 142–151. [PubMed: 31598466]
- Ioannidis JP, 2018 Why replication has more scientific value than original discovery. *Behavioral and Brain Sciences* 41.

- Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG, 2001 Replication validity of genetic association studies. *Nature genetics* 29, 306. [PubMed: 11600885]
- James P, Berrigan D, Hart JE, Hipp JA, Hoehner CM, Kerr J, Major JM, Oka M, Laden F, 2014 Effects of buffer size and shape on associations between the built environment and energy balance. *Health & place* 27, 162–170. [PubMed: 24607875]
- Johnson GD, Lu X, 2011 Neighborhood-level built environment and social characteristics associated with serious childhood motor vehicle occupant injuries. *Health & place* 17, 902–910. [PubMed: 21571572]
- Lynch SM, 2019 Towards Systematic Methods in an Era of Big Data: Neighborhood Wide Association Studies, in: *Geospatial Approaches to Energy Balance and Breast Cancer*. Springer, pp. 99–117.
- Lynch SM, Mitra N, Ross M, Newcomb C, Dailey K, Jackson T, Zeigler-Johnson CM, Riethman H, Branas CC, Rebbeck TR, 2017 A neighborhood-wide association study (NWAS): example of prostate cancer aggressiveness. *PloS one* 12, e0174548. [PubMed: 28346484]
- Manville M, Beata A, Shoup D, 2013 Turning housing into driving: Parking requirements and density in Los Angeles and New York. *Housing Policy Debate* 23, 350–375.
- Maresca MM, Hoepner LA, Hassoun A, Oberfield SE, Mooney SJ, Calafat AM, Ramirez J, Freyer G, Perera FP, Whyatt RM, 2015 Prenatal exposure to phthalates and childhood body size in an urban cohort. *Environmental health perspectives* 124, 514–520. [PubMed: 26069025]
- Mooney SJ, Joshi S, Cerda M, Kennedy GJ, Beard JR, Rundle AG, 2017 Contextual Correlates of Physical Activity among Older Adults: A Neighborhood Environment-Wide Association Study (NE-WAS). *AACR*.
- Moudon AV, Rutherford S, Saelens B, Hallenbeck ME, Turkiyyah G, 2009 A report on participant sampling and recruitment for travel and physical activity data collection.
- Multi-Resolution Land Characteristics Consortium, n.d. National Land Cover Database (NLCD). URL <https://www.mrlc.gov/> (accessed 7.29.19).
- Oakes JM, Forsyth A, Schmitz KH, 2007 The effects of neighborhood density and street connectivity on walking behavior: the Twin Cities walking study. *Epidemiologic Perspectives & Innovations* 4, 16. [PubMed: 18078510]
- Oliver LN, Schuurman N, Hall AW, 2007 Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International journal of health geographics* 6, 41. [PubMed: 17883870]
- Owen N, Cerin E, Leslie E, Coffee N, Frank LD, Bauman AE, Hugo G, Saelens BE, Sallis JF, 2007 Neighborhood walkability and the walking behavior of Australian adults. *American journal of preventive medicine* 33, 387–395. [PubMed: 17950404]
- Rodriguez DA, Aytur S, Forsyth A, Oakes JM, Clifton KJ, 2008 Relation of modifiable neighborhood attributes to walking. *Preventive medicine* 47, 260–264. [PubMed: 18436297]
- Rubin DB, 2004 Multiple imputation for nonresponse in surveys. John Wiley & Sons.
- Rundle A, Ahsan H, Vineis P, 2012 Better cancer biomarker discovery through better study design. *European journal of clinical investigation* 42, 1350–1359. [PubMed: 22998109]
- Rundle AG, Chen Y, Quinn JW, Rahai N, Bartley K, Mooney SJ, Bader MD, Zeleniuch-Jacquotte A, Lovasi GS, Neckerman KM, 2019 Development of a neighborhood walkability index for studying neighborhood physical activity contexts in communities across the US over the past three decades. *Journal of urban health* 96, 1–8.
- Saelens BE, Handy SL, 2008 Built environment correlates of walking: a review. *Medicine and science in sports and exercise* 40, S550. [PubMed: 18562973]
- Sallis JF, Owen N, Fisher E, 2015 Ecological models of health behavior. *Health behavior: Theory, research, and practice* 5, 43–64.
- Savitz DA, Bobb JF, Carr JL, Clougherty JE, Dominici F, Elston B, Ito K, Ross Z, Yee M, Matte TD, 2014 Ambient fine particulate matter, nitrogen dioxide, and term birth weight in New York, New York. *American journal of epidemiology* 179, 457–466. [PubMed: 24218031]
- Shashank A, Schuurman N, 2019 Unpacking walkability indices and their inherent assumptions. *Health & place* 55, 145–154. [PubMed: 30580962]

- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI, 2006 An obesity-associated gut microbiome with increased capacity for energy harvest. *nature* 444, 1027. [PubMed: 17183312]
- United States Census Bureau, n.d. American Community Survey. URL <https://www.census.gov/programs-surveys/acs> (accessed 7.29.19).
- Weden MM, Bird CE, Escarce JJ, Lurie N, 2011 Neighborhood archetypes for population health research: Is there no place like home? *Health & place* 17, 289–299. [PubMed: 21168356]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- We assessed neighborhood predictors of walking bouts in King County, WA
- 92 of 146 neighborhood measures predicted frequency of walking bouts
- Housing unit count was the strongest predictor of walking
- The Automatic Context Measurement Tool promotes replicable research
- ACMT code is available at <https://github.com/smooney27/ACMT>

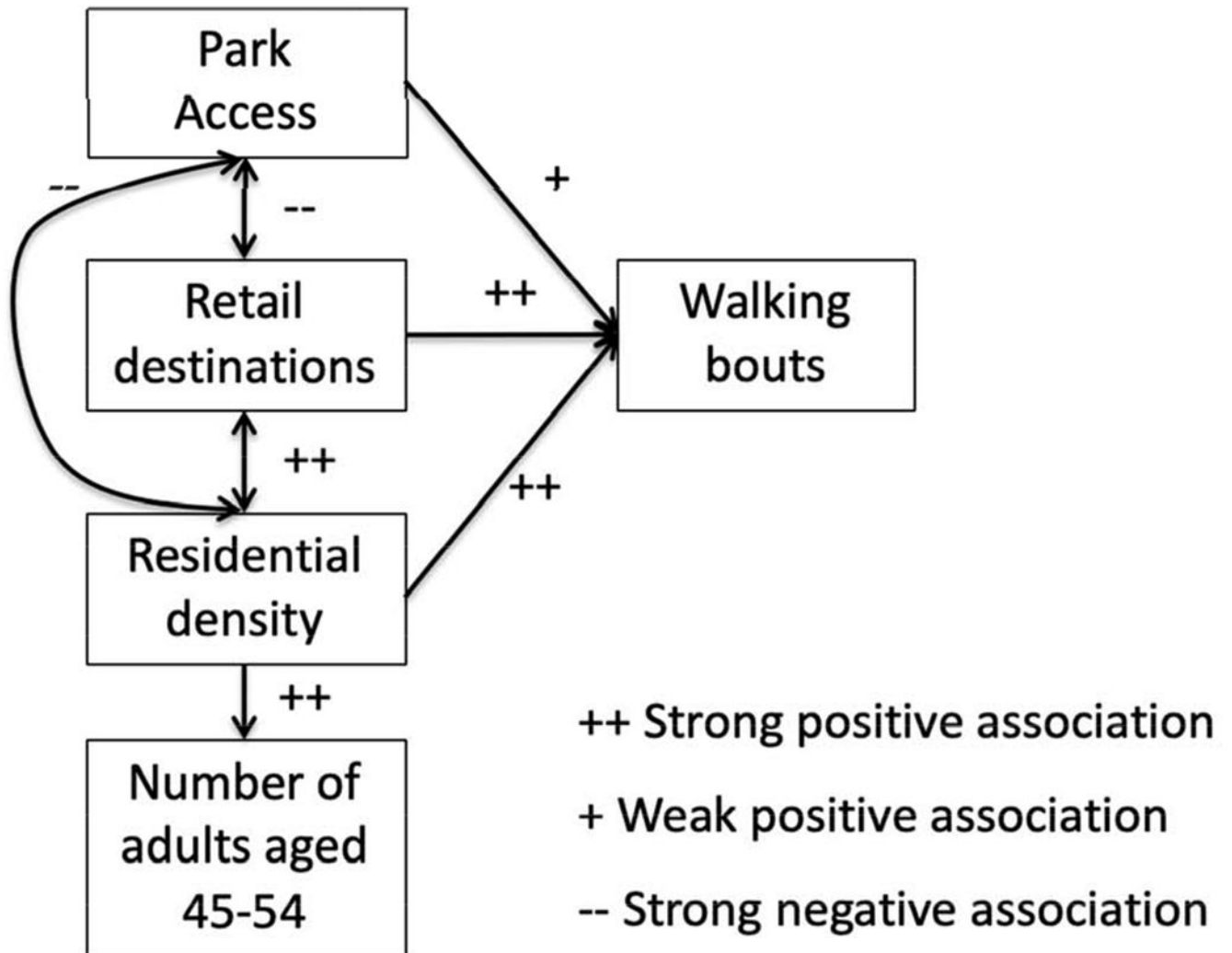


Figure 1. Schematic diagram of potential set of associations wherein retail density and residential density encourage walking bouts strongly and park access encourages walking bouts weakly. Number of adults aged 45-54 is strongly tied to residential density and thus to walking. In such a situation, number of adults aged 45-54 might be a better predictor of walking than park access.

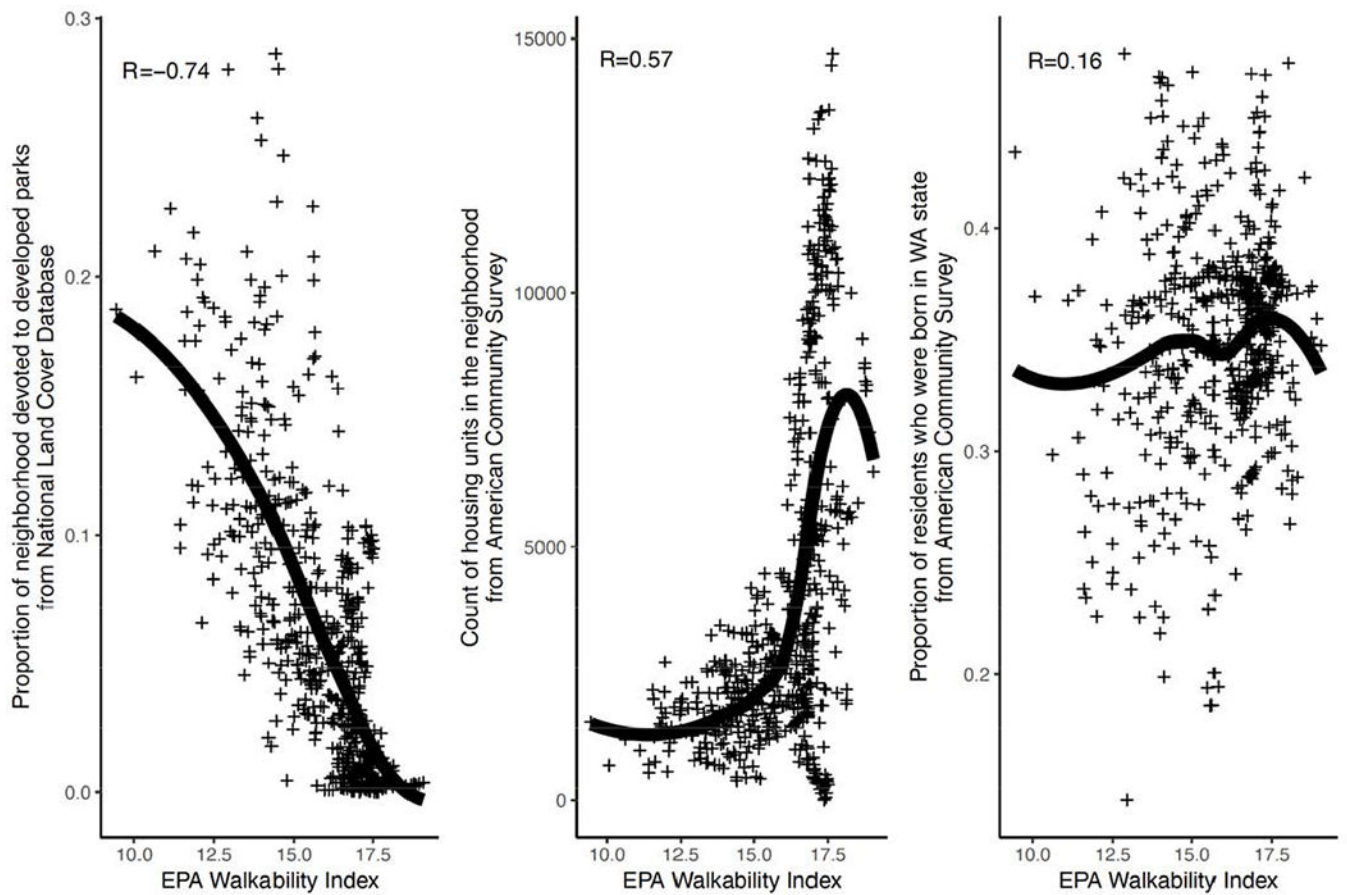


Figure 2.

Scatter plots showing correlations between EPA walkability index and selected neighborhood measures the 2011 National Land Cover Database and the 2010 American Community Survey 5-year estimates. Lines show locally smoothed regression trends. Measures were selected to include two measures thought to encourage walking (park land, housing units) and one not expected to (proportion of residents born in Washington State).

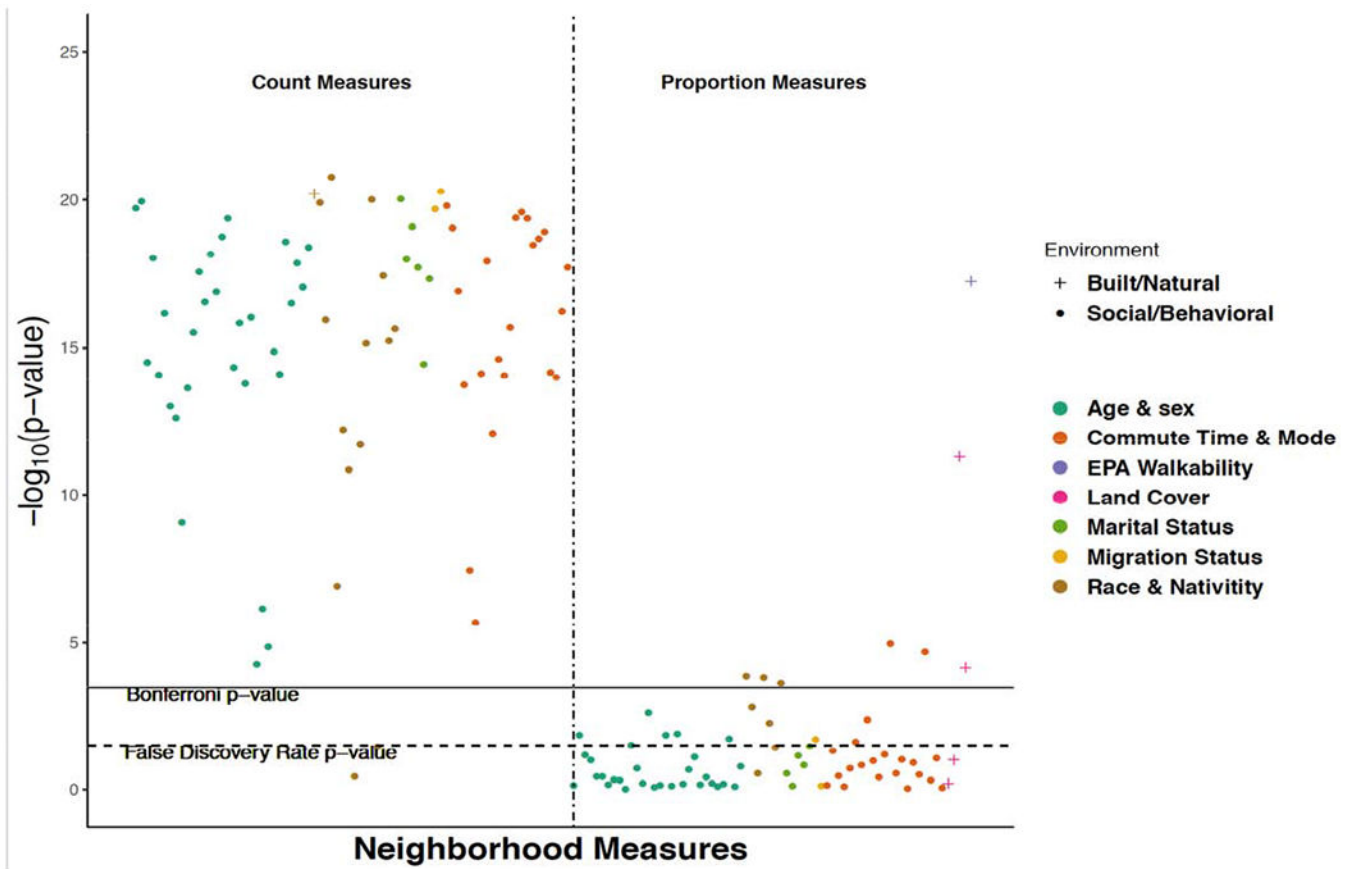


Figure 3. Manhattan Plot showing strength of association (represented as negative log p-value for coefficient from a covariate-adjusted negative binomial model) between the neighborhood measures and number of walk bouts within 833m of home among TRAC study participants, 2009-2013

Table 1.

Selected demographic and neighborhood characteristics of 681 Travel Assessment and Community study participants

Characteristic	Overall
Gender, N (%)	
	<i>Female</i> 428 (63)
	<i>Male</i> 253 (37)
Age, Mean (SD)	51 (13)
Age Category	
	<i>18-39</i> 161 (24)
	<i>40-65</i> 452 (66)
	<i>66+</i> 106 (16)
Race/Ethnicity, N (%)	
	<i>Non-Hispanic white</i> 537 (79)
	<i>Hispanic or non-white</i> 146 (21)
Household Income, N (%)	
	<i>Less than 50K</i> 212 (31)
	<i>50K-100K</i> 209 (31)
	<i>More than 100K</i> 191 (28)
Educational Attainment, N (%)	
	<i>Less than College Grad</i> 185 (27)
	<i>College Graduate</i> 474 (70)
Walk Bouts per Week, Mean (SD)	
	<i>Total</i> 12 (6)
	<i>Near Home</i> 6 (5)
Examples of Neighborhood Characteristics Within 833m of Home Parcel Centroid, Mean (SD)	
	<i>Population</i> ¹ 9,610 (6,929)
	<i>Housing Units</i> ¹ 3,560 (2,865)
	<i>Proportion of Employed Residents who Drive to Work</i> ¹ (Range: 0-100) 64 (10)
	<i>EPA Walkability Index Score</i> ² (Range: 0-20) 16 (2)
	<i>Undeveloped Land Proportion</i> ³ (Range: 0-100) 3 (5)

¹ American Community Survey

² Environmental Protection Agency data

³ National Land Cover Database

Table 2.

Selected fit statistics from four models of number of residential neighborhood walk bouts using different walkability measures

	Pseudo-R ²	AIC	Cross-Validated RMSE*	p-value of LRT test vs. No Walkability
Individual-level Measures Only	0.646	3669	90.6	1
EPA Walkability	0.736	3592	80.5	<0.01
Residential Density & Employment Density	0.754	3580	76.8	<0.01
Principal Component #1	0.740	3592	76.2	<0.01
Principal Component #1 & #2	0.748	3587	76.0	<0.01
Housing Units	0.749	3583	75.1	<0.01

*RMSE = root mean squared error

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3. Odds ratios for each additional walk bout from adjusted models predicting walking

Model	No Walkability Measure	EPA Walkability Measure	A priori Res Density & Emp Density	Principal Component #1	Principal Components #1 and #2	Housing Unit Count
Age						
Less than 40	--	--	--	--	--	--
40-65	0.94 (0.79,1.12)	0.99 (0.84,1.17)	1.03 (0.87,1.21)	1.03 (0.87,1.22)	1.04 (0.88,1.23)	1.02 (0.86,1.20)
66 and above	1.33 (1.04,1.70)	1.38 (1.10,1.74)	1.27 (1.01,1.60)	1.30 (1.03,1.63)	1.31 (1.05,1.66)	1.25 (1.00,1.58)
Gender						
Female	--	--	--	--	--	--
Male	1.13 (0.97,1.30)	1.05 (0.92,1.21)	0.98 (0.85,1.12)	0.99 (0.86,1.13)	0.98 (0.85,1.13)	0.97 (0.85,1.12)
Household Income (dollars)						
Less than 50,000	--	--	--	--	--	--
50,000-100,000	0.80 (0.67,0.95)	0.86 (0.73,1.02)	0.85 (0.72,1.00)	0.86 (0.72,1.01)	0.85 (0.72,1.00)	0.86 (0.73,1.02)
Over 100,000	0.85 (0.70,1.02)	0.93 (0.78,1.11)	0.97 (0.81,1.15)	1.00 (0.84,1.20)	0.99 (0.83,1.18)	1.00 (0.84,1.19)
Educational Attainment						
Less than College Graduate	--	--	--	--	--	--
College Graduate	1.10 (0.92,1.30)	0.99 (0.84,1.17)	1.04 (0.88,1.22)	1.01 (0.85,1.19)	1.00 (0.85,1.18)	1.03 (0.87,1.21)
Race/Ethnicity						
Non-Hispanic White	0.93 (0.77,1.12)	0.93 (0.78,1.11)	0.95 (0.80,1.13)	0.93 (0.78,1.11)	0.92 (0.78,1.10)	0.93 (0.78,1.11)
Neighborhood Measures (per standard deviation)						
	--	1.40 (1.30,1.50)	1.34 (1.24,1.46) ^a	1.37 (1.28,1.47)	1.37 (1.28,1.47) ^c	1.39 (1.30,1.49)
	--	--	1.07 (0.99,1.16) ^b	--	0.92 (0.86,0.98) ^d	--

^aResidential Density

^bEmployment Density

^cPrincipal Component #1

^dPrincipal Component #2