

RESEARCH

Open Access



# Probabilistic walking models using built environment and sociodemographic predictors

Anne Vernez Moudon<sup>1\*</sup> , Ruizhu Huang<sup>2</sup>, Orion T. Stewart<sup>1,3</sup>, Hannah Cohen-Cline<sup>4</sup>, Carolyn Noonan<sup>5</sup>, Philip M. Hurvitz<sup>6</sup> and Glen E. Duncan<sup>7</sup>

## Abstract

**Background:** Individual sociodemographic and home neighborhood built environment (BE) factors influence the probability of engaging in health-enhancing levels of walking or moderate-to-vigorous physical activity (MVPA). Methods are needed to parsimoniously model the associations.

**Methods:** Participants included 2392 adults drawn from a community-based twin registry living in the Seattle region. Objective BE measures from four domains (regional context, neighborhood composition, destinations, transportation) were taken for neighborhood sizes of 833 and 1666 road network meters from home. Hosmer and Lemeshow's methods served to fit logistic regression models of walking and MVPA outcomes using sociodemographic and BE predictors. Backward elimination identified variables included in final models, and comparison of receiver operating characteristic (ROC) curves determined model fit improvements.

**Results:** Built environment variables associated with physical activity were reduced from 86 to 5 or fewer. Sociodemographic and BE variables from all four BE domains were associated with activity outcomes but differed by activity type and neighborhood size. For the study population, ROC comparisons indicated that adding BE variables to a base model of sociodemographic factors did not improve the ability to predict walking or MVPA.

**Conclusions:** Using sociodemographic and built environment factors, the proposed approach can guide the estimation of activity prediction models for different activity types, neighborhood sizes, and discrete BE characteristics. Variables associated with walking and MVPA are population and neighborhood BE-specific.

**Keywords:** Physical activity, Active travel, Home neighborhood domains

## Background

Physical inactivity is a major public health concern. Two decades after developing programs and incentives to increase activity, the Centers for Disease Control and Prevention's national surveillance data show that only about half of the US population reports engaging in at least 150 min of moderate-intensity physical activity per week [1], which is the level recommended for health [2]. More alarmingly, objective (meaning measured) data from a large representative sample show that the proportion of

people who engage in recommended levels of activity is much smaller, at about 5% of the population [3].

Activity levels are thought to be affected by factors ranging from biology to policy (e.g., zoning laws). Research has investigated how physical activity, and particularly walking, is influenced not only at the individual level by demographic and social factors, as is common in public health, but also at the community and population levels, by the built environment (BE) and policies [4–6]. The BE is of interest to those working to increase physical activity because, unlike individual demographics and social factors, it can be directly modified to support activity, and its modifications can impact the entire populations [7–9].

“Walkability” is a catch-all term, which has been used to characterize activity-supportive BEs. Walkable

\* Correspondence: [moudon@uw.edu](mailto:moudon@uw.edu)

<sup>1</sup>Architecture, Landscape Architecture, and Urban Design and Planning, University of Washington, 1107 NE 45th St, Suite 535, Box 354802, Seattle, WA 98195, USA

Full list of author information is available at the end of the article



environments are known to contain a transportation infrastructure that ensures short, direct, and safe trips to routine or recreational destinations while encouraging modes of travel other than automobiles [10, 11]. Routine destinations are shown to be associated with walking consist of transit stops, grocery stores, banks, coffee shops, and other retail outlets [12–14]. Recreational destinations associated with walking include parks, bodies of water, and facilities such as gyms or recreation centers [15, 16]. However, finding common links between environment and activity remains challenging because the relationship between walkability and actual activity levels, including walking, is complex. Across studies, variations in the conceptualization, measurement, and modeling of activity-friendly environments [17–19] impede the potential for generalization. Only a consistent treatment of environmental predictors and activity outcomes will help untangle their relationship for different populations in different settings [17, 20].

Methodological advantages may accrue from studies using individuals and their environment as units of analysis because they simplify the modeling process, facilitate the interpretation of results, and yield more statistical power [21]. Additionally, the increasing availability of detailed objective data on BE can help avert statistical errors due to the “modifiable areal unit problem,” which occurs when different spatial data are captured at different scales (e.g., counties and ZIP Codes) and/or in spatial units of different shapes (e.g., grids versus “natural” shapes) [22, 23]. As well, disaggregating these data facilitates targeting location-specific BE changes, which may be associated with increased physical activity [24, 25]. Yet with some exceptions [26, 27], few studies have combined both individual-level and disaggregated objective activity and BE data in the context of large populations living within large metropolitan regions.

This study presents an approach to estimate the probability that an individual will engage in physical activity at levels sufficient for health benefits. Using a large population sample in a large metropolitan area, it introduces a clear conceptual model of BE and a robust method to test the predictive ability of both sociodemographic factors and extensive objective measurements of BE characteristics on two outcomes of interest: amount of walking in the neighborhood and total moderate-to-vigorous physical activity (MVPA). The intent is to present a method to identify a concise set of variables that could eventually help construct a walking index combining measured behaviors with walkability indicators.

## Methods

### Participants

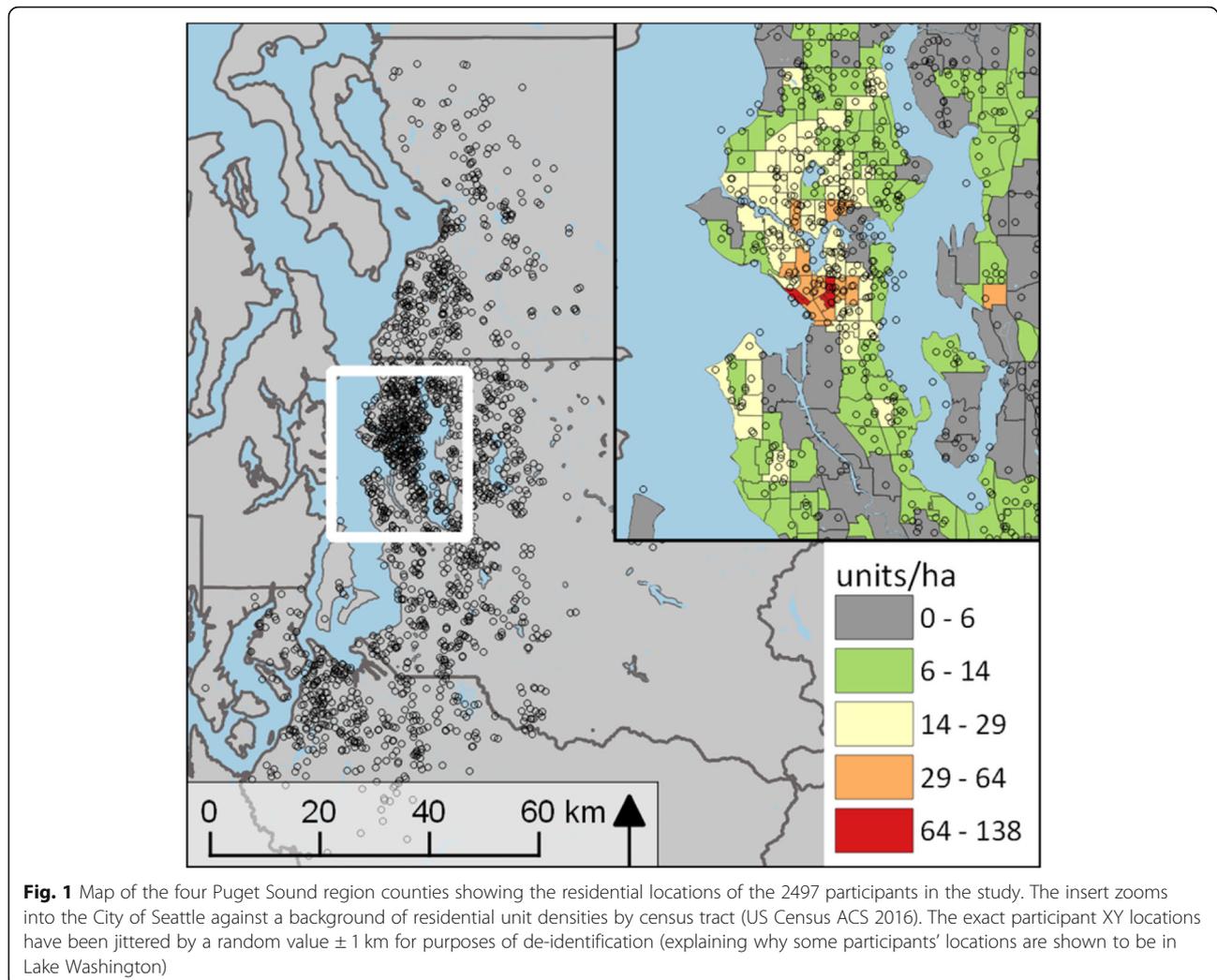
Participants came from a community-based sample of adult twins assembled from the Washington State

Department of Licensing records; a detailed description is available elsewhere [28]. The sample consisted of 2497 individual members of twin pairs living within the four-county Puget Sound region around Seattle, WA (population 3.3 million; 2600 km<sup>2</sup> [1000 mi<sup>2</sup>]) (Fig. 1). Addresses and survey data, including items on sociodemographics, height, weight, general health, common medical conditions, and lifestyle behaviors, came from a Health & Wellbeing questionnaire administered between May 2010 and July 2012. The questionnaire followed an initial Twin Registry enrollment survey and was sent to 11,822 Registry members. The response rate was 74%, with only residents of the Puget Sound region selected for this study. Respondents provided written informed consent as a part of survey completion, and all study procedures were approved by the University of Washington institutional review board.

### Physical activity and sociodemographic measures

We assessed two distinct types of physical activity, MVPA and walking. MVPA came from two survey questions: (1) “Over the past 4 weeks, how many days during a typical week did you exercise vigorously for at least 20 minutes? Vigorous exercise causes heavy sweating or large increases in breathing or heart rate and includes such activities as running, lap swimming, aerobics classes, and fast bicycling” and (2) “Over the past 4 weeks, how many days during a typical week did you exercise moderately for at least 30 minutes? Moderate exercise causes only light sweating or light to moderate increases in breathing or heart rate and includes such activities as brisk walking, bicycling for pleasure, golf, and dancing.” The number of minutes per week of combined vigorous and moderate exercise was calculated as (20 min vigorous exercise × number of days) + (30 min moderate exercise × number of days). For the analyses, MVPA was dichotomized as < 150 and ≥ 150 min per week, the recommended threshold of physical activity [2, 29].

Walking was assessed with two questions: (1) “How many days during a typical week do you walk for recreation, exercise, to get to and from places, or for any other reason in your neighborhood?” and (2) “When you walk in your neighborhood, about how many minutes, on average, do you spend walking each time you walk?” Response options, in minutes, included < 15, 15, 30, 45, 60, 75, and > 90. The question was phrased so that respondents could include all types of walking, brisk or casual, as well as utilitarian or recreational, in their assessment. The number of minutes per week of walking was calculated as (number of days × average number of minutes walking each time). When participants responded “less than 15” or “90 or more” minutes, the calculation used 10 or 90, respectively, for the average



number of minutes. For the analyses, walking was also dichotomized as  $< 150$  and  $\geq 150$  min per week.

Sociodemographic variables were age, sex, race (White and non-White), education level (less than high school graduate, high school graduate or general equivalency diploma, some college, and Bachelor's degree or higher), and total household income in the past year ( $< \$20,000$ ;  $\$20,000$ – $29,999$ ;  $\$30,000$ – $39,999$ ;  $\$40,000$ – $49,999$ ;  $\$50,000$ – $59,999$ ;  $\$60,000$ – $69,999$ ;  $\$70,000$ – $79,999$ ; and  $\geq \$80,000$ ).

#### BE variables

Built environment variables came from four domains that previous studies have associated with walking and physical activity in general: regional context, neighborhood composition, destinations, and transportation [17, 30, 31]. Regional context was defined in terms of terrain (i.e., slope), because the Puget Sound region is notably hilly, and living in Seattle vs. not, because Seattle is its largest, most densely developed city. Variables for neighborhood composition included the density of residential

units and jobs, the presence of neighborhood commercial centers, and residential property values. Destinations were divided into subcategories: food sources (e.g., grocery stores, restaurants), retail and service (e.g., general merchandise, health centers), sports and fitness facilities (e.g., team or solo sports), open space (e.g., parks), and education facilities (e.g., schools). Transportation included both infrastructure (e.g., streets, intersections) and traffic conditions (e.g., vehicular traffic volumes, bus ridership). The list of BE variables and related measures is presented in Additional file 1.

Regional terrain slope came from the USGS National Elevation Dataset (NED 1/3 arc-second raster data, 2012). Neighborhood composition data came from assessor's tax parcel data in the Puget Sound four counties: King (2010), Snohomish (2009), Pierce (2009), and Kitsap (2009). Employment data were developed by the Urban Form Laboratory (UFL) team [32]. Data on food sources and facilities for fitness, services, and retail came from the commercial vendor InfoUSA (2011) and were

classified by the UFL [33]. Parks data were developed by the UFL (2012) based on county and local jurisdictional data [34]. School data came from the National Center for Education Statistics [35]. Streets and intersections were derived from OpenStreetMap data (<https://www.openstreetmap.org/#map=5/51.500/-0.100>; 2015). Traffic volumes and bike facilities came from the Puget Sound Regional Council (<http://www.psrc.org/data/transportation/bikeped-data>; 2006). Bus ridership (2010, 2011) came from the region's transit agencies.

### BE measures

Two commonly used radii defined the home neighborhood [10, 30, 31, 36, 37]. The immediate neighborhood lay 833 m (0.5 mi) from the home parcel, corresponding to a 10-min walking distance, and 1666 m (1 mi) defined the extended neighborhood within a 20-min walk. "Sausage" network buffers captured areas of exposure [36]. The buffers are created by first identifying street segments traversable within 833 m and 1666 m of each respondent's home parcel, then by delineating areas within 100 m of street center lines, and by filling in any gaps inside the buffer.

Measures taken for each BE variable included count (e.g., of residential units, individual destinations, bus riders), length (e.g., of streets), area (e.g., of parks), density (e.g., gross density of residential units in the neighborhood, net density of residential units in the residential parcels of the neighborhood), and shortest route distance from participants' homes. There were 86 initial BE measures. The regional context domain had 2 measures of 2 variables; the neighborhood composition domain had 13 measures of 5 variables; the destinations domain had 56 measures of 41 variables; and the transportation domain had 15 measures of 5 variables (Additional file 1). Based on OpenStreetMap, network buffer and distance measures were calculated in PostgreSQL 9.3.8 and PostGIS 2.1.3 (PostGIS Development Group, 2008) using programmatic wrappers in R 2.15.3.

### Geocoding

Twins' home addresses and food establishments, fitness and sport facilities, general services, and retail businesses from the InfoUSA destinations data (see Additional file 1 for details) were geocoded in ArcGIS 10.2.1 (Redlands, CA) using StreetMapUSA Premium (<http://www.esri.com/data/streetmap>; 2014). For twins' addresses, 60% were matched to a building rooftop point using a match score of 100, 20% were matched with street address interpolation, and 20% were matched manually. For InfoUSA destinations, a 60 minimum geocoding match score required manual geocoding of 36% of the food destinations ( $n = 8293$ ), 33% of the fitness destinations ( $n = 1550$ ), and 37% of the retail destinations ( $n = 7534$ ). Fewer than 3% remained unmatched.

### Analyses

Descriptive statistics were computed as mean, standard deviation, and range or percent. Prediction models used MVPA and walking as two separate outcomes, and BE variables within the immediate neighborhood (833 m) and within the extended neighborhood (1666 m) as predictors. All four models included sociodemographic variables as predictors. All analyses treated twins as singletons, controlling where needed for correlations within twin pairs.

Modeling entailed four steps. First, to reduce the number of BE variables, the initial 86 measures were correlated with the two outcome variables in the two neighborhood sizes using Pearson correlation coefficients. Correlations were small, and the top 25 measures (correlation coefficient  $> 0.09$ ) were retained for each one of the four models. Variables were then tested for multicollinearity, and only one of the highly correlated variables ( $> 0.8$ ) was kept for inclusion in the models. The final number of BE variables entered in the models ranged from 9 to 16 depending on the model.

Second, we assessed the most appropriate form for independent variables by comparing continuous and categorical forms using the Hosmer and Lemeshow goodness-of-fit test [38]. Ultimately, we found that alternative forms did not improve predictive ability and we kept the sociodemographic and BE variables in continuous form for simplicity. The exceptions were the BE variables measuring the distance from participants' homes to nearest BE feature (e.g., youth or solo sports facilities), which needed to be categorized in order to retain the large portion of the sample (25–37%) that had distances larger than the 1666 m extended neighborhood. The categorized distance variables were modeled in continuous form.

The data were divided into training and validation datasets, to develop the prediction models and to determine model fit avoiding artificially high estimates of model fit, respectively. Two thirds of the twin pairs were randomly selected for the training dataset and one third for the validation dataset. The use of twin pairs, not individuals, insured that the training and validation datasets were independent from each other.

Third, a backward elimination approach established the best subset of independent variables to include in each one of the four models and to constitute what we called "index variables." A first logistic regression model included all sociodemographic and BE variables. The variable with the largest  $P$  value was then removed and the model re-fit iteratively, until all remaining predictors had  $P$  values  $\leq 0.20$  (the traditional 0.05 threshold might reduce the power to find appropriate predictors). Age and sex were forced into the models along with education or income or both, as these four factors are widely associated with physical activity and walking [17]. If neither education nor income had  $P \leq 0.20$ , then the

variable with the lowest *P* value was included. Clustered robust standard errors served to properly inflate variance estimates due to the correlation within twin pair.

Fourth, we assessed the best subset model using the goodness-of-fit test and receiver operating characteristic (ROC) curves in the validation dataset. Area under the ROC curve or the *C*-statistic (0.0–1.0) was used, where 0.5 represented no predictive ability beyond random chance and 1.0 indicated strong predictive ability. The ROC curve from the best subset model was compared to a model including only sociodemographic variables to determine if the best subset model differed significantly in predictive ability. All analyses used Stata/SE 12.1 (College Station, TX).

**Results**

Out of 2497 participants, 105 (4%) were excluded due to missing data for sociodemographic, MVPA, or walking variables. Of the 2392 remaining participants, more than 60% were female, almost 90% were White, and the average age was under 40. About 50% had a college degree, and more than 60% had annual household incomes above \$50,000.

Table 1 shows sociodemographic factors and levels of physical activity stratified by the 150 min per week cut point. Only 39% of participants reported at least 150 min of MVPA per week, and only 24% reported at least 150 min of walking per week. For those reporting less than 150 min per week of MVPA, the mean was 61.8 min, while it was 209.8 min for the more active group of

participants (overall mean of 120 min). For walking in the neighborhood, the reported mean minutes per week was 93 min, with 47.4 and 238.9 min reported by the two groups of participants, respectively.

Table 2 shows the sociodemographic and BE predictor variables selected for inclusion in the models. Of the four sociodemographic variables, age, sex, and education remained in all final models, but race was eliminated, as was income in the walking models.

After testing for multicollinearity, BE variables included in the final models came from all BE domains. For MVPA, 3 out of 9 and 5 out of 16 BE variables remained in the models of the immediate and extended neighborhoods, respectively. For the immediate neighborhood, being in Seattle, the number of neighborhood centers (containing at least a retail outlet, a grocery store, and a traditional restaurant) and distance to the nearest youth facility were associated with meeting the 150 min per week MVPA threshold. For the extended neighborhood, gross residential density, the number of supermarkets, solo sport (e.g., boxing, water sports, bicycle rental, golf, tennis, and climbing, as different from team sports, such as football, basketball, baseball, and soccer), and youth facilities remained in the model, as well as the total length of bike lanes.

For walking, 5 each out of 15 and 13 BE variables remained in the immediate and extended neighborhood models, respectively. Both walking models included residential density (net and gross), the number of fitness

**Table 1** Descriptive statistics of the sample by level of physical activity and walking

|   | MVPA <sup>a</sup> |       |          |       | WALKING  |       |          |       |
|---|-------------------|-------|----------|-------|----------|-------|----------|-------|
|   | <150 min          |       | ≥150 min |       | <150 min |       | ≥150 min |       |
|   | N                 | %     | N        | %     | N        | %     | N        | %     |
|   | 1458              | 61    | 934      | 39    | 1819     | 76    | 573      | 24    |
| <b>SOCIODEMOGRAPHICS</b>                            |                   |       |          |       |          |       |          |       |
| Age <sup>b</sup>                                    | 38.5              | 15.9  | 35.6     | 12.2  | 36.6     | 15.2  | 40.0     | 17.0  |
| Male  | 516               | 35.4  | 350      | 37.5  | 670      | 36.8  | 196      | 34.2  |
| White   | 1305              | 89.5  | 830      | 88.9  | 1629     | 89.6  | 506      | 88.3  |
| Education   |                   |       |          |       |          |       |          |       |
| < HS grad   | 26                | 1.8   | 15       | 1.6   | 32       | 1.8   | 9        | 1.6   |
| HS Grad/GED   | 201               | 13.8  | 96       | 10.3  | 240      | 13.2  | 57       | 10.0  |
| Some college  | 553               | 37.9  | 313      | 33.5  | 676      | 37.2  | 190      | 33.2  |
| Bachelors+  | 678               | 46.5  | 510      | 54.6  | 871      | 47.9  | 317      | 55.3  |
| Income  |                   |       |          |       |          |       |          |       |
| < \$50,000  | 570               | 39.1  | 339      | 36.3  | 695      | 38.2  | 214      | 37.4  |
| ≥ \$50,000  | 888               | 60.9  | 595      | 63.7  | 1124     | 61.8  | 359      | 62.7  |
| <b>MVPA/ WALKING</b>                                |                   |       |          |       |          |       |          |       |
| mean minutes/week <sup>b</sup>                      | 61.84             | 46.44 | 209.75   | 56.29 | 47.42    | 41.26 | 238.85   | 91.79 |
| <sup>a</sup> Moderate-to-vigorous physical activity |                   |       |          |       |          |       |          |       |
| <sup>b</sup> Mean + standard deviation              |                   |       |          |       |          |       |          |       |

<sup>a</sup>Moderate-to-vigorous physical activity  
<sup>b</sup>Mean + standard deviation

**Table 2** Sociodemographic and domain-specific built environment variable selection by activity type and neighborhood size

| Neighborhood Buffer   | MVPA  |         | WALKING |         |
|---|-------|---------|---------|---------|
|   | 833 m | 1,666 m | 833 m   | 1,666 m |
| <b>SOCIODEMOGRAPHICS</b>  |       |         |         |         |
| Age   | O     | O       | O       | O       |
| Male  | O     | O       | O       | O       |
| Race  | X     | X       | X       | X       |
| Education   | O     | O       | O       | O       |
| Income  | O     | O       | X       | X       |
| <b>BUILT ENVIRONMENT PREDICTORS BY DOMAIN</b>   |       |         |         |         |
| <b>Regional context</b>   |       |         |         |         |
| land area > 8% slope  | X     | X       | X       | X       |
| indicator for being within Seattle  | O     | X       | O       | X       |
| <b>Neighborhood composition</b>   |       |         |         |         |
| number of residential units per ha of residential area*1000 (NET)   |       | X       | O       |         |
| number of residential units per ha*1000 (GROSS)   |       | O       |         | O       |
| number of residential units   |       |         |         |         |
| number of neighborhood centers  | O     | X       | X       | O       |
| <b>Food destinations</b>  |       |         |         |         |
| number of broad selection food stores   |       | X       |         |         |
| number of supermarkets  |       | O       |         | X       |
| number of limited service restaurants   |       |         |         |         |
| number of fast food/quick service, restaurants  |       |         | X       | X       |
| <b>Retail destinations</b>  |       |         |         |         |
| number of consumables (small, non-chain retail)   |       |         |         |         |
| number of general merchandise used (small)  |       | X       |         | X       |
| number of apparel (small)   |       |         |         |         |
| number of hobby leisure (large)   |       | X       | O       | X       |
| <b>Fitness facilities destinations</b>  |       |         |         |         |
| number of fitness facilities  | X     | X       | O       | O       |
| number of solo sports facilities  |       | O       |         | X       |
| distance to closest solo sports facility  | X     |         |         |         |
| Number of youth sport facilities  | X     | O       |         | X       |
| distance to closest youth sport facility  | O     |         | X       |         |
| <b>Service and recreation destinations</b>  |       |         |         |         |
| number of parks   | X     | X       | X       | X       |
| number of schools   |       | X       |         | X       |
| number of community centers   |       |         | X       |         |
| number of large health services   |       |         |         |         |
| number of small health services   |       |         |         |         |
| <b>Transportation</b>   |       |         |         |         |
| number of 4-way+ intersections  |       | X       |         |         |
| number of 3-way intersections   |       |         | X       | O       |
| total length of bike lanes  |       | O       | O       | O       |
| traffic volume per street length  |       |         | X       | X       |
| average ridership per bus stop  | X     |         |         |         |
| <p><sup>1</sup> O = variable was selected via backward elimination for the specified prediction model (<math>P \leq 0.20</math>) or was forced into the model (only applies to sex, age, education, and income); X = a variable was considered but not selected for the specified prediction model (<math>P &gt; 0.20</math>); blank cell = a variable was removed from consideration for the specified prediction model during the data reduction phase and was never included in the model.</p> |       |         |         |         |

O = variable was selected via backward elimination for the specified prediction model ( $P \leq 0.20$ ) or was forced into the model (only applies to sex, age, education, and income); X = a variable was considered but not selected for the specified prediction model ( $P > 0.20$ ); blank cell = a variable was removed from consideration for the specified prediction model during the data reduction phase and was never included in the model

facilities, and the total length of bike lanes. The 833-m neighborhood walking model also retained being in Seattle and the number of hobby-leisure facilities, while in the 1666-m model, the number of neighborhood centers and the number of 3-way street intersections remained as predictors of walking at least 150 min per week.

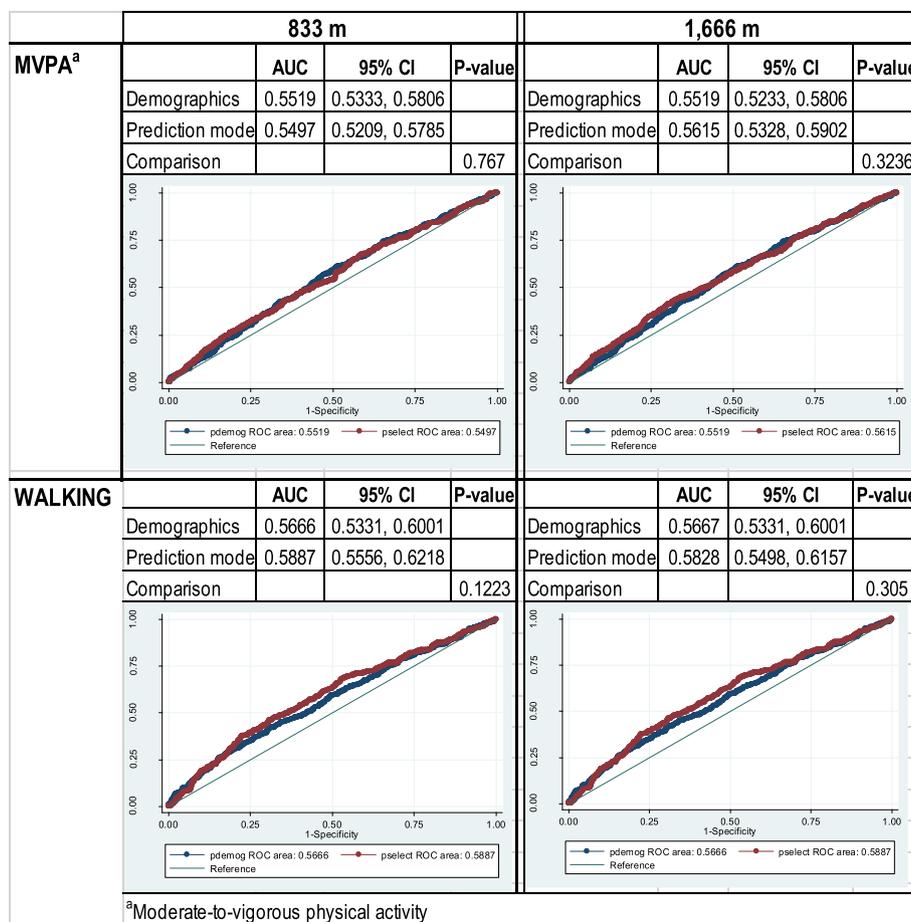
Neither sociodemographic nor BE variables remaining as index variables in the models had strong predictive

ability for MVPA or walking (Table 3). Age and sex variables had different associations with walking and MVPA. Age was negatively related to engaging in sufficient MVPA, but positively associated with walking at least 150 min per week. Males were more likely to engage in MVPA but less likely to walk than their female counterparts. Participants with higher incomes were more likely to reach the recommended MVPA level. In terms of

**Table 3 .**

| Variable                        | MVPA <sup>a</sup> |              |         |        |              |         | WALKING |              |         |        |              |         |
|---------------------------------|-------------------|--------------|---------|--------|--------------|---------|---------|--------------|---------|--------|--------------|---------|
|                                 | 833 m             |              |         | 1666 m |              |         | 833 m   |              |         | 1666 m |              |         |
|                                 | OR                | 95% CI       | P-value | OR     | 95% CI       | P-value | OR      | 95% CI       | P-value | OR     | 95% CI       | P-value |
| Sociodemographic                |                   |              |         |        |              |         |         |              |         |        |              |         |
| Age                             | 0.984             | 0.974, 0.995 | 0.004   | 0.984  | 0.973, 0.994 | 0.003   | 1.017   | 1.006, 1.029 | 0.003   | 1.018  | 1.006, 1.029 | 0.003   |
| Male                            | 1.379             | 0.994, 1.914 | 0.054   | 1.384  | 0.993, 1.927 | 0.055   | 0.845   | 0.593, 1.203 | 0.35    | 0.831  | 0.583, 1.184 | 0.305   |
| Education                       | 1.188             | 0.954, 1.477 | 0.124   | 1.191  | 0.957, 1.484 | 0.117   | 1.202   | 0.931, 1.551 | 0.157   | 1.229  | 0.959, 1.576 | 0.104   |
| Household income                | 1.087             | 1.023, 1.155 | 0.007   | 1.093  | 1.028, 1.162 | 0.004   |         |              |         |        |              |         |
| Regional location               |                   |              |         |        |              |         |         |              |         |        |              |         |
| Seattle                         | 1.534             | 0.961, 2.447 | 0.073   |        |              |         | 1.518   | 0.948, 2.430 | 0.082   |        |              |         |
| Neighborhood composition        |                   |              |         |        |              |         |         |              |         |        |              |         |
| Res units/ha (GROSS)            |                   |              |         | 1.32   | 0.976, 1.786 | 0.071   |         |              |         | 1.477  | 0.942, 2.316 | 0.089   |
| Res units/ha of res area (NET)  |                   |              |         |        |              |         | 1.024   | 0.996, 1.054 | 0.098   |        |              |         |
| Neighborhood centers (count)    | 0.794             | 0.600, 1.512 | 0.107   |        |              |         |         |              |         | 1.181  | 0.979, 1.423 | 0.082   |
| Destinations                    |                   |              |         |        |              |         |         |              |         |        |              |         |
| Supermarkets (count)            |                   |              |         | 0.852  | 0.715, 1.014 | 0.071   |         |              |         |        |              |         |
| Hobby leisure (count)           |                   |              |         |        |              |         | 2.762   | 0.890, 8.568 | 0.079   |        |              |         |
| Fitness facilities (count)      |                   |              |         |        |              |         | 0.916   | 0.839, 0.999 | 0.047   | 0.914  | 0.860, 0.970 | 0.003   |
| Solo sports facilities (count)  |                   |              |         | 1.279  | 1.080, 1.514 | 0.004   |         |              |         |        |              |         |
| Youth sport facilities (count)  |                   |              |         | 1.111  | 0.980, 1.257 | 0.098   |         |              |         |        |              |         |
| Youth sport facility (distance) | 0.941             | 0.862, 1.027 | 0.175   |        |              |         |         |              |         |        |              |         |
| Transportation                  |                   |              |         |        |              |         |         |              |         |        |              |         |
| 3-way intersections (count)     |                   |              |         |        |              |         |         |              |         | 1.001  | 1.000, 1.003 | 0.114   |
| Bike lanes (total length)       |                   |              |         | 1      | 1.000, 1.000 | 0.019   | 1       | 1.000, 1.000 | 0.128   | 1      | 1.000, 1.000 | 0.069   |
| _cons                           | 0.38              | 0.155, 0.931 | 0.034   | 0.269  | 0.119, 0.610 | 0.002   | 0.069   | 0.026, 0.184 | <0.001  | 0.041  | 0.014, 0.120 | <0.001  |

<sup>a</sup>Moderate-to-vigorous physical activity



**Fig. 2** Comparison of receiver operating characteristic (ROC) curves for sociodemographic models (pdemog) versus best models including BE variables (pselect)

neighborhood BE characteristics, participants who reached the 150-min activity threshold were more likely to be located in Seattle, to live in denser neighborhoods with more neighborhood centers and destinations than those who did not reach the health-benefit threshold.

Comparing the ROC curve from the best subset models to that from models including only sociodemographic variables indicated that the addition of BE variables did not significantly improve predictive ability (Fig. 2). The area under the curve (AUC) of the walking models was 0.59 (95% CI 0.56, 0.62) and 0.58 (95% CI 0.55, 0.62) for the immediate and the extended neighborhood, respectively. It was lower, at 0.55 (95% CI 0.52, 0.58) and 0.56 (95% CI 0.53, 0.59), for the 833-m and the 1666-m MVPA models, respectively.

**Discussion**

The proposed method addressed the challenges of assessing the influence of home neighborhood BE characteristics on the probability of being active. It used a clear conceptual construct of the home neighborhood BE and a comprehensive and consistent treatment of BE

characteristics that can be replicated in other studies. The method demonstrated how to systematically reduce the large number of variables and measures necessary to characterize the BE in the modeling process. The final number of BE variables was down to a manageable maximum of 5 per model. Additionally, ROC analyses helped to economically compare the predictive ability of socio-demographic and home neighborhood BE characteristics in assessing the likelihood of an individual meeting the recommended level of physical activity or walking.

In this study, BE characteristics did not improve models predicting walking or MVPA at levels sufficient to benefit the health of adults, implying that for this population, higher neighborhood walkability might not lead to more total MVPA or more neighborhood walking [39, 40]. However, as different BE variables were retained for models predicting MVPA versus walking within the immediate and the extended neighborhood, results suggested that BE effects may vary by activity type and neighborhood size. Furthermore, the final BE index variables came from all of the proposed BE domains,

confirming that these domains identify distinct aspects of BE walkability [17, 31]. Neighborhood composition (measured as density and mixed-use development) remained in all of the models. All models also contained at least one variable in the neighborhood destination domain, which was related to food, general retail, or sport activities, but none retained service and recreation destinations. Transportation domain variables remained in the two walking models and the 1666-m MVPA model. Regional location (measured as being within the City of Seattle) remained in the walking models. Of note, the index variables measured the effect of discrete attributes of the BE, thus potentially guiding interventions targeting changes to the BE that might lead to increased physical activity [21, 24, 25]. In contrast, models using factor analyses or other aggregation techniques to capture the BE cannot provide the information necessary to directly connect analytic results with intervention strategies or policy development.

Regarding the impact of sociodemographic characteristics, age, income, and education had the expected association with activity, in line with other studies [41]. The small proportion of participants reporting at least 150 min of MVPA (39.05%) was more than that of the US adult population as determined using objective measures [3], but less than that the proportion determined using national surveillance instruments [1]. Although this finding lent some support for the utility of our MVPA variable, it was unexpected given the relatively young age of the sample and considering the likelihood that participants over-reported their physical activity via self-report [29]. Also, the proportion of our sample reporting at least 150 min of walking per week was small (less than 24%) [42].

The study limitations included the use of survey-based outcomes, which could lead to over-reporting physical activity. However, in an ongoing study of twins who wore accelerometers and GPS devices over a 2-week period, subjectively measured MVPA of 283 twins correlated significantly with objectively measured MVPA ( $r = 0.47$ ,  $P < 0.001$ ) (Additional file 2). The correlation for walking was  $r = 0.44$ ,  $P < 0.001$ . Further, the survey data covered total MVPA but only neighborhood walking, thus disallowing inferences on the influence of the home BE on either neighborhood MVPA or total walking [43]. Regarding the location of activity, we tested the effect of different neighborhood sizes in the probability of engaging in activity, but could not assess the potential impact of temporally different neighborhood exposures between participants [44]. The cross-sectional study design limits the ability to infer causation between the BE and active behaviors. Finally, lack of data on participant perceptions of their home neighborhood precluded estimating the effects of subjective neighborhood assessment on activity, potentially biasing the findings [45].

## Conclusions

The study introduced a clear approach to estimate the influence of sociodemographic and discrete attributes of the home neighborhood BE on an individual's probability of being active or walking sufficiently to benefit health. It presented a method to identify BE variables of the immediate and extended home neighborhoods likely to be associated with activity.

For the study population, including BE variables in the models did not improve the ability to predict the likelihood of reaching recommended health-sustaining levels of walking or physical activity compared to using sociodemographic variables only. This does not mean that these relationships do not exist. Perhaps, different methods and objective data that would address possible measurement error in the outcome variables would yield different results. Further analyses of possible non-linear associations between BE and behavior outcome variables might also change the results. However, sociodemographic and BE variables associated with neighborhood walking at least 150 min per week differed from those associated with MVPA of the same duration, suggesting that different factors may be involved in the engagement of walking or MVPA. Sociodemographic and BE influences also varied by the assumed size of the home neighborhood. Final BE variables came from all four proposed BE domains, which should be considered in analyses of BE influences on activity in general.

## Additional files

**Additional file 1:** Built environment variables and measures by domain. The file contains a table of the 86 built environment variables and related measures included in the analyses. Variables are classified into the four main domains of the proposed BE conceptual construct, with subdomains identified when needed to further clarify the construct. (PDF 132 kb)

**Additional file 2:** Comparing measured with self-report MVPA and walking data using a sub-sample of 283 twins who wore accelerometers and GPS devices over a 2-week period. The file contains supplemental information on correlations between measured and self-report MVPA and walking. Also provided is a within-person comparison of measured and self-report walking data using a 150- and 100-min per week threshold of recommended activity. (PDF 138 kb)

## Abbreviations

BE: Built environment; MVPA: Moderate and vigorous physical activity; PA: Physical activity; ROC: Receiver operating characteristic; UFL: Urban Form Laboratory at the University of Washington; USGS NED: United States Geological Survey National Elevation Dataset; WA: Washington State

## Acknowledgements

We thank the anonymous reviewers and Ally Avery for helpful comments and suggestions on the revised manuscript.

## Funding

This work was supported by grants from the National Institutes of Health (RC2HL103416 to D. Buchwald and R01AG042176 to G. Duncan). The National Institutes of Health played no role in the study design; the collection, analysis, or interpretation of data; the drafting of this manuscript; or the decision to submit it for publication.

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Authors' contributions**

AVM and GED conceptualized the study and oversaw its development. AVM wrote the first draft. RH and OTS collected and analyzed the BE data. CJN and HCC were responsible for the statistical analyses. PMH oversaw the GIS methods. All authors contributed to the writing of this manuscript and to the interpretation of the results. They have no competing interests to declare in relation to this manuscript. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Participant consent was written, and all study procedures were approved by the institutional review board of the University of Washington.

**Consent for publication**

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Architecture, Landscape Architecture, and Urban Design and Planning, University of Washington, 1107 NE 45th St, Suite 535, Box 354802, Seattle, WA 98195, USA. <sup>2</sup>Texas Advanced Computing Center, University of Texas at Austin, Austin, USA. <sup>3</sup>Present Address: Institute for Population Health Improvement, University of California, Davis, 4800 2nd Avenue, Suite 2600, Sacramento, CA 95817, USA. <sup>4</sup>Center for Outcomes Research and Education, Portland, OR 97213, USA. <sup>5</sup>Initiative for Research and Education to Advance Community Health (IREACH), Washington State University, Seattle, WA 98101, USA. <sup>6</sup>Department of Urban Design and Planning, University of Washington, Seattle, USA. <sup>7</sup>Elson S. Floyd College of Medicine, Department of Nutrition & Exercise Physiology, Washington State Twin Registry, Washington State University Health Sciences Spokane, Spokane, USA.

Received: 19 May 2018 Accepted: 8 May 2019

Published online: 03 June 2019

**References**

- Centers for Disease Control and Prevention (CDC). Adult participation in aerobic and muscle-strengthening physical activities — United States, 2011. *MMWR Morb Mortal Wkly Rep.* 2013;62(17):326–30. PMID: PMC4604926.
- Garber C, Blissmer B, Deschenes M, Franklin B, Lamonte M, Lee I, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc.* 2011;43:1334–59.
- Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40:181–8.
- Giles-Corti B, Donovan RJ. Relative influences of individual, social environmental, and physical environmental correlates of walking. *Am J Public Health.* 2003;93(9):1583–9.
- Ogilvie D, Mitchell R, Mutrie N, Petticrew M, Platt SC. Personal and environmental correlates of active travel and physical activity in a deprived urban population. *Int J Behav Nutr Phys Act.* 2008;5:43.
- Wen M, Kandula NR, Lauderdale DS. Walking for transportation or leisure: what difference does the neighborhood make? *J Gen Intern Med.* 2007;22(12):1674–80.
- Berke EM, Vernez-Moudon A. Built environment change: a framework to support health-enhancing behaviour through environmental policy and health research. *J Epidemiol Community Health.* 2014;68(6):586–90.
- Rundle AG, Heymsfield SB. Can walkable urban design play a role in reducing the incidence of obesity-related conditions? *JAMA.* 2016;315(20):2175–7.
- Giles-Corti B, Vernez-Moudon A, Reis R, Turrell G, Dannenberg AL, Badland H, et al. City planning and population health: a global challenge. *Lancet.* 2016;388(10062):2912–24.
- Sallis JF. Measuring physical activity environments: a brief history. *Am J Prev Med.* 2009;36(4 Suppl):S86–92.
- Frank LD, Sallis JF, Saelens BE, Leary L, Cain K, Conway TL, et al. The development of a walkability index: application to the Neighborhood Quality of Life Study. *Br J Sports Med.* 2010;44(13):924–33.
- Cerin E, Leslie E, du Toit L, Owen N, Frank LD. Destinations that matter: associations with walking for transport. *Health Place.* 2007;13:713–24.
- McCormack GR, Giles-Corti B, Bulsara M. The relationship between destination proximity, destination mix and physical activity behaviors. *Prev Med.* 2008;46(1):33–40.
- Moudon AV, Lee C, Cheadle AD, Garvin CW, Johnson DB, Schmid TL, et al. Attributes of environments supporting walking. *Am J Health Promot.* 2007;21(5):448–59.
- Lee C, Moudon AV. Environmental correlates of walking for transportation or recreation purposes. *J Phys Act Health.* 2006;3(Suppl 1):S77–98.
- Sugiyama T, Neuhaus M, Cole R, Giles-Corti B, Owen N. Destination and route attributes associated with adults' walking: a review. *Med Sci Sports Exerc.* 2012;44:1275–86.
- Saelens BE, Handy SL. Built environment correlates of walking: a review. *Med Sci Sports Exerc.* 2008;40(7 Suppl):S550–66.
- Glazier RH, Creatore MI, Weyman JT, Fazli G, Matheson FI, Gozdyra P, et al. Density, destinations or both? A comparison of measures of walkability in relation to transportation behaviors, obesity and diabetes in Toronto. *Canada PLoS One.* 2014;9(1):e85295.
- Koohsari MJ, Owen N, Cerin E, Giles-Corti B, Sugiyama T. Walkability and walking for transport: characterizing the built environment using space syntax. *Int J Behav Nutr Phys Act.* 2016;13(1):121.
- Troiano RP, Freedson PSC. Promises and pitfalls of emerging measures of physical activity and the environment. *Am J Prev Med.* 2010;38:682–3.
- Matthews SA, Moudon AV, Daniel M. Work group II: Using Geographic Information Systems for enhancing research relevant to policy on diet, physical activity, and weight. *Am J Prev Med.* 2009;36(4, Suppl 1):S171–6.
- Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical analysis. *Environ Plan A.* 1991;23:1025–44.
- Wakefield J. Ecologic studies revisited. *Annu Rev Public Health.* 2008;29:75–90.
- McConville ME, Rodriguez DA, Clifton K, Cho G, Fleischhacker S. Disaggregate land uses and walking. *Am J Prev Med.* 2011;40(1):25–32.
- Rodriguez DA, Aytur S, Forsyth A, Oakes JM, Clifton KJ. Relation of modifiable neighborhood attributes to walking. *Prev Med.* 2008;47(3):260–4.
- Forsyth A, Oakes JM, Schmitz KH. The built environment, walking, and physical activity: is the environment more important to some people than others? *Transport Res D.* 2009;14(1):42–9.
- Stewart OT, Moudon AV, Saelens BE, Lee C, Kang B, Doescher MP. Comparing associations between the built environment and walking in rural small towns and a large metropolitan area. *Env Behav.* 2016;48(1):13–26.
- Strachan E, Hunt C, Afari N, Duncan G, Noonan C, Schur E, et al. University of Washington Twin Registry: poised for the next generation of twin research. *Twin Res Hum Genet.* 2013;16(1):455–62.
- Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S. adults: compliance with the Physical Activity Guidelines for Americans. *Am J Prev Med.* 2011;40:454–61.
- Hurvitz PM, Moudon AV. Home versus nonhome neighborhood: quantifying differences in exposure to the built environment. *Am J Prev Med.* 2012;42(4):411–7.
- Stewart OT, Carlos HA, Lee C, Berke EM, Hurvitz PM, Li L, Moudon AV, Doescher MP. Secondary GIS built environment data for health research: guidance for data development. *J Transport Health.* 2016;3(4):529–39.
- Moudon AV, Sohn DW, Kavage SE, Mabry JE. Transportation-Efficient Land Use Mapping Index (TELUMI), a tool to assess multimodal transportation options in metropolitan regions. *Int J Sustain Transp.* 2011;5(2):111–33.
- Moudon AV, Drewnowski A, Duncan GE, Hurvitz PM, Saelens BE, Scharnhorst E. Characterizing the food environment: pitfalls and future directions. *Public Health Nutr.* 2013;16(7):1238–43.
- Stewart OT, Moudon AV, Fesinmeyer MD, Zhou C, Saelens BE. The association between park visitation and physical activity measured with accelerometer, GPS, and travel diary. *Health Place.* 2016;38:82–8.
- U.S. Department of Education. Common Core of Data for the 2007/2008 school year. Washington, DC: Institute of Education Sciences, National Center for Education Statistics; 2010.

36. Forsyth A, Van Riper D, Larson N, Wall M, Neumark-Sztainer DC. Creating a replicable, valid cross-platform buffering technique: the sausage network buffer for measuring food and physical activity built environments. *Int J Health Geogr.* 2012;11:14.
37. Villanueva K, Knuiman M, Nathan A, Giles-Corti B, Christian H, Foster S, Bull F. The impact of neighborhood walkability on walking: does it differ across adult life stage and does neighborhood buffer size matter? *Health Place.* 2014;25:43–6.
38. Hosmer DW, Lemeshow S. *Applied logistic regression.* New York: John Wiley & Sons; 2000.
39. Lovasi GS, Moudon AV, Pearson AL, Hurvitz PM, Larson EB, Siscovick DS, et al. Using built environment characteristics to predict walking for exercise. *Int J Health Geogr.* 2008;7:10.
40. Oakes JM, Forsyth A, Schmitz KH. The effects of neighborhood density and street connectivity on walking behavior: the Twin Cities walking study. *Epidemiol Perspect Innov.* 2007;4:16.
41. Boone-Heinonen J, Gordon-Larsen P. Life stage and sex specificity in relationships between the built and socioeconomic environments and physical activity. *J Epidemiol Community Health.* 2011;65(10):847–52.
42. Centers for Disease Control and Prevention (CDC). Vital signs: walking among adults—United States, 2005 and 2010. *MMWR Morb Mortal Wkly Rep.* 2012;61:595–601.
43. Chaix B, Kestens Y, Perchoux C, Karusisi N, Merlo J, Labadi K. An interactive mapping tool to assess individual mobility patterns in neighborhood studies. *Am J Prev Med.* 2012;43(4):440–50.
44. Kwan M-P. The uncertain geographic context problem. *Ann Assoc Am Geogr.* 2012;102(5):958–68.
45. Gebel K, Bauman A, Owen N. Correlates of non-concordance between perceived and objective measures of walkability. *Ann Behav Med.* 2009;37(2):228–38.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

