The Association Between Residential Built Environment and Walking Habits Among Older Women in the Portland Metro-Area

A Thesis
By

Vivian Wai-Ying Siu

Presented to the Department of Public Health and Preventive Medicine
Oregon Health & Science University
In partial fulfillment of the requirements for the degree of
Master of Public Health
June 2011

Department of Public Health and Preventive Medicine

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's thesis of

Vivian Wai-Ying Siu

has been approved

Dr. Yvonne L. Michael, Mentor/Advisor

Dr. Rochelle Fu, Member

Dr. William E. Lambert, Member

Table of Contents

List of Tables and Figures

List of Abbreviations

BRFSS	Behavioral Risk Factor Surveillance System
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CI	95% Confidence Interval
GIS	Geographic Information System
IDW	Inverse Distance Weighting
MAX	Metropolitan Area Express
OR	Odds Ratio
PCA	Principal Component Analysis
PPEF	Portland Pedestrian Environmental Factor
RLIS	Regional Land Information System
SOF	Study of Osteoporotic Fractures
TAZ	Traffic Analysis Zone
UGB	Urban Growth Boundary

Acknowledgements

I am grateful for the time and effort my thesis committee members have devoted to provide guidances and advices in helping me to complete my thesis project. Thank you to my mentor/advisor Dr. Michael, for working with me closely and patiently throughout my project. I have learned a lot about study design and research methodology from her. Thank you Dr. Fu for her help and advice on the statistical analysis. I really appreciate her quick responses to my questions and help with reviewing and understanding the findings. Thank you Dr. Lambert, for providing an objective perspective, and specific suggestions and comments to improve my work.

I would also like to extend my thank to Mark Bosworth for his help on identifying and locating the available spatial data and providing his GIS expertise on methodologies to quantify the built environment attributes. I am also thankful for my friend, Heidi Guenin, who offered her help to review and edit an earlier draft.

Finally, I would like to thank my family and friends for their love, support, and encouragement. I am grateful that I can count on them whenever I need someone to talk with, their supports have kept me motivated along the way.

Abstract

Objective: The built environment may shape walking habits among older women. This study was designed to address the gaps in literature related to quantifying the built environment in order to investigate the association between the built environment and walking. Standardized and high resolution geographic units were established to objectively measure built environment attributes, and subsequent use of cluster analysis to assess the mixed effect among built environment attributes facilitated the examination of the association between urban forms and walking among women 65 years or older residing in the Portland metro area.

Methods: Data from the Metro Regional Land Information System were adopted to establish high resolution objective built environment measures related to: accessibility of bus stops, light rail stations, commercial areas, and park areas; route connectivity; and population density. Cluster analysis was performed to define areas/urban forms that shared similar built environment attributes. A sample of 2005 baseline individual level data from the Portland center of the Study of Osteoporotic Fracture was linked to the derived clusters and Census block group information. A logistic regression model was constructed to investigate the associations between urban form and walking outcome (dichotomized as <5 blocks/day or ≥5 blocks/day), adjusting for age, education, self-reported health, lifetime smoking quantity, other exercise, BMI, history of stroke, and block group percent of live alone population in poverty.

Results: Six urban form clusters were established: central city, city periphery, suburb, urban fringe with poor commercial area access, urban fringe with poor park access, and satellite city. The adjusted model showed a lower odds of walking ≥5 blocks/day for older women residing in the city periphery (OR=0.69, CI: 0.55-0.85), suburb (OR=0.69, CI: 0.50-0.95), and urban fringe with poor commercial area access (OR=0.40, CI: 0.18-0.88) as compared to those residing in the central city. There was no difference in walking outcome observed for urban fringe with poor park access (OR=0.98, CI: 0.52-1.84) or satellite city (OR=0.53, CI: 0.16-1.71) as compared to the central city.

Conclusions: The uses of standardized small geographic units for establishing built environment measurements and cluster analysis to identify urban forms were useful and effective in redefining the built environment without the using administrative boundaries. This approach also accounted for multiple attributes when assessing the association between built environment and walking habits. Among the six urban forms explored, central city, with a combination of high population density, high street connectivity, and convenient access to amenities (especially to transit and commercial areas), may effectively support and encourage older women to engage in a modest amount of daily walking. The use of standardized small geographic units and clusters to define built environment can facilitate other studies investigating the impact of built environment and health. The findings may also inform planning decisions and help to shape pedestrian-friendly communities that provide public health benefits in the long run.

Introduction

The impact of built environments on walking behavior has been a popular research topic in the past two decades. Studies have shown a positive association between built environments and utilitarian walking, while the association between built environments and recreation walking were mixed.[1,2] The relationship between built environments and walking for both purposes combined is unclear. A better understanding of the relationship between built environments and walking behaviors could lead to urban design and planning policies that promote walking, thereby generating public health benefits. However, the challenge of quantifying built environment measurements (for example, defining attributes, data scale, data quality) and the challenge created by diverse methodologies adopted among studies make it difficult to compare and ascertain the associations. Few previous studies have quantified the built environment attributes objectively and at high resolution; the use of cluster analysis to identify different urban forms and to consider the combined effect of different attributes together have also been limited.[2-5] This research gap provided an opportunity for the current study to develop refined built environment measures, identify distinct urban forms, and expand on the existing literature investigating the association between built environments and walking.

Background

Built environment influences physical activity

Urban form is believed to influence the physical activities people engage in daily. For example, increasing urban sprawl is connected to the trend of increasing vehicle miles traveled per person over the past half a century.[6] Increasing urban sprawl also correlates with high levels of sedentary activity, such as increased television watching, and with decreased physical exertion at work and home.[6] Physical inactivity is a significant public health problem, because it is a known risk factor associated with many chronic diseases, including coronary heart disease, development of type II diabetes, colon cancer, hypertension, obesity, osteoporosis, and breast cancer.[7-9] Physical inactivity is also the fourth leading risk factor for global mortality, with 6% of deaths worldwide attributed to this risk factor.[10] The costs of physical inactivity extend beyond illness and mortality, but also lost in productivity; the conservative estimate of economic cost due to absence of leisure-time physical activity was 2.4% of the 1995 US health care expenditure and was 9.4% when also considering indirect impact from obesity.[11] Regular physical activity is positively associated with psychological well being, joint health, increased bone mass, pulmonary rehabilitation, and lower risk of obesity and certain types of cancers related to immune deficiencies.[7]

Current trend and research needs

An extensive body of literature exists investigating the relationship between built environments and physical activity. One growing focus of this body of research is to understand the social and environmental determinants that promote physical activities and healthy aging.[12] As planning efforts to promote New Urbanism and Smart Growth grow,[13] the number and proportion of the elderly population in the United States are also increasing.[14] The demographic composition is shifting rapidly towards an aging population as the baby boomers enter retirement age. Therefore, understanding urban form typologies and how their effects shape older adults' health behavior can inform decision makers and the public as they plan for an aging society. In the long run, livable environments that promote physical activities among older adults support chronic disease management and reduce the risks of diseases among a majority of the population. Arguably, a place that works for older adults would promote physical activities for people of other age groups, because the older adult population is one of the most vulnerable groups for poor health outcomes tied to physical inactivity. The benefit of creating a healthy society for older adults extends beyond an improvement in health and quality of life; it also minimizes future healthcare burdens related to physical inactivity.

Older adults, physical activities, and evidence of impact from built environment

Older adults in general engage in relatively low levels of physical activity[7,10] and in low intensity activities such as walking and gardening for exercise,[12] compared to younger adults. According to the Center of Disease Control and Prevention's (CDC) 2008 Behavioral Risk Factor Surveillance System (BRFSS) data, the no leisure-time physical activity prevalence for population 65 or above was 32.7% for the US and 23.2% for Oregon.[15] These data did not include the share of population who were inactive or had insufficient physical activity to meet the CDC-recommended guideline of 30 minutes per day, 5 or more days per week of moderate-intensity activities or 20 minutes per day, 3 or more days per week of vigorous-intensity activities (150 minutes total per week).[15] Older women are of particular concern, since women tend to have lower levels of physical activity than men in any age group.[7,16] Women also tend to live longer than men, so they constitute a large share of the vulnerable population. Since walking is the primary physical activity many older women participate in, increasing this health behavior through changes to the built environment may have tremendous public health significance.

Several studies focused on older populations in various neighborhoods and cities found positive associations between the level of walking and the accessibility of facilities and the density of housing and population[17-21] as well as among older women specifically.[22,23] Based on the findings from this literature review, this study aimed to verify and examine the impact of neighborhood built environment on walking among older women.

Measurement of the built environment

Various methods have been used to quantify built environment and to study the association between specific neighborhood characteristics and walking. A few neighborhood features have been identified as important determinants of walking. Studies with built environments measured at the individual level typically gathered information by surveying individual perceptions of the environment[20,21,24,25] by aggregating neighborhood measures from secondary data such as Census,[3,19,22,24-29] or by measuring these characteristics within a certain distance of the subjects' residences.[20,22,25,30-32] Previously identified built environment characteristics and existing data organization approaches guided the selection and construction of the neighborhood measures in the current study.

Neighborhood characteristics that are commonly measured include: residential density, street connectivity, accessibility to transit services, land use mix, and retail floor area ratio. Residential density measures the compactness of the living environment; it is defined as the number of housing units divided by residential land area, with a greater amount of walking associated with higher residential density.[27,30,33,34]

Street connectivity measures the street network design and is defined as the number of intersections within an area[30,33,34] or the percent of four-way intersections within an area.[27] Greater street connectivity represents street networks that are more compact and that tend to be in grid structures, with relatively more four-way intersections than the three-way intersections or cul-de-sacs that are commonly observed in suburban areas. A higher street connectivity is believed to promote walking by providing more direct pathways and more route options for pedestrians.[1]

The other common variables typically attempt to quantify the mix of services and amenities. The accessibility to transit variable measures the number of transit stops in the area[32] or the distance to transit stops.[26] A higher number of transit stops near residences and close residential proximity to transit services increase the desirability of travel by public transportation, increasing the frequency and amount of walking through time spent walking to and from the service stops.[1]

Land use mix is measured using: entropy scores;[30,33,34] destination count/densities of commercial, cultural, and recreational facilities within an area;[20,22,27,32] presence of services within a distance buffer from the subject of interest;[32] and distance to services[25] or a proxy, such as year built for housing stock.[22] In general, a place with a diverse land use mix would have a high entropy score, a high destination count, short distances to services, and older houses (pre-1950). More pedestrian activities are expected in areas with a diverse land use mix.

The retail floor area ratio measures the area of retail floor area divided by the retail land area.[30,33] A high value for this measure indicates a high retail density area with less setback of the structure and less parking area, while a low value indicates low intensity of commercial activities. Areas with high retail floor area ratio are associated with greater pedestrian activities.

While some studies evaluated the effect of individual neighborhood attributes on walking,[26,28,32,35,36] many studies have examined built environment factors in combination, using approaches such as summation indices,[37,38] propensity scores,[27] the Frank index for regression analysis,[25,30,34,35] and factors derived from principal component analysis (PCA).[39]

Although these methods vary in strength and weakness, cluster analysis was used in the current study for several reasons. First, cluster analysis has the ability to account for the interaction effects among all variables or the varying effect each individual built environment variable has on walking.

Although evaluating the effect of individual neighborhood attributes could help identify important attributes, there was limited ability to account for the mixed effects between different attributes, which might magnify or reduce the impact on walking as compared to the expected summation effect from the same attributes. Moreover, regression models that were often used to examine individual factors could exclude neighborhood attributes with multi-linearity problem, in which variables that were highly correlated with each other were dropped from the model. However, some of these variables might have interaction effects that were important in the model that researchers would want to retain. Built environment index, PCA, and cluster analysis could account for the interactive effects and multi-linearity problem different built environment attributes exerted on walking that could only be partially considered in regression models. Second, cluster analysis might be relevant and have greater validity than a built environment index. Objective measures developed by PCA and cluster analysis were found to have stronger agreement with the experts-assessed built environment conditions measurement system, Portland Pedestrian Environmental Factor (PPEF), than the agreement between the built environment index, a simple composite index with the summation of scores from all items measured, and the PPEF.[40] Third, cluster analysis has the advantage of a relatively straightforward interpretation of the resulting clusters or areas with similar urban forms. Although PCA addresses the interaction effects, the results from a PCA are abstract and difficult for audiences to interpret.

The use of cluster analysis to examine the association between neighborhood environment and walking among older women has the potential to add value to the existing body of literature, since few studies have used a cluster analysis to study the topic, especially with the use of localized neighborhood environment. Riva et al.[19,41] completed a cluster analysis of active living potential, which measured the population density (total population/km^2), land use mix (entropy index), and accessibility variables (count of supermarket, banks, pharmacies and libraries within 1km network buffer) by Census disseminated area. This study concluded that active living potential was positively associated with utilitarian walking among adults of 45 years old or above. Previous reports have also applied the cluster analysis method in the following ways: identification of neighborhood types to investigate the relationship with travel purpose;[3,42] Census tract clustering of the area population and housing characteristics to study the relationship with travel behavior;[3] Traffic Analysis Zone clustering of transportation and land use characteristics to evaluate neighborhood and residential location choices;[43] and walking and biking prevalence clustering to identify built environment and individual factors that are related to extremely high or low levels of walking and biking.[4] Additional studies using similar or improved cluster analysis methods could strengthen the findings on the relationship between built environments and walking behaviors among individuals.

Specific aims of current research

The goal of this research is to answer two questions: 1) Is there an association between different urban forms and a modest amount of daily walking among older women? 2) If so, what is the magnitude of the effects of various urban forms on walking among older women? I hypothesized that urban form shapes health behavior. Specifically, I hypothesized that a compact urban form promotes walking among older women. Compact neighborhoods with high population density, high street connectivity, and close proximity to transit services, commercial areas, and parks were posited to be favorable to pedestrian activity. High connectivity and convenient access to amenities were hypothesized to shape transportation choices and encourage utilitarian and leisure walking, and high population density was anticipated to support investment of infrastructure and amenities. Older women residing in these compact areas are expected to walk more than those residing in remote areas with low population density, poor street connectivity, and poor accessibility to transit services, commercial areas, and park areas.

To achieve the specific aims, a set of objective and localized built environment measures linkable to individual level data was developed to assess the association between the built environment and walking. Using objective data allowed for a systematic approach to measuring built environment that was consistently reported for the subjects and replicable elsewhere or at different time points. Quantifying built environment attributes at high resolution minimized the geographic unit of the measure and lessened the potential bias associated with the use of coarsely grouped measures, in which neighborhood characteristics might be wrongly assumed as the experience for the individuals (i.e., the ecological fallacy).

Materials and Methods

This study was a cross-sectional design using existing data from a larger longitudinal cohort study, the Neighborhood and Obesity Study, which investigated relationships between neighborhood design and obesity. The sample of older women was a subset of participants from the Study of Osteoporotic Fractures (SOF), a multicenter observational study that gathered a series of anthropometric and demographic data, fracture and medical history, and information on functioning, quality of life, and lifestyle through self-reported questionnaires and clinical exams since its baseline of 1986.[44] The SOF recruited healthy, community-dwelling women, with a total of 9704 subjects included as the initial cohort between the four centers in U.S. cities: Baltimore, Pittsburgh, Minneapolis, and Portland.[44] Portland, Oregon was one of the centers with a matching set of historically archived spatial data, providing information sufficient to make the current analysis feasible.

The baseline data from the Portland SOF cohort was adapted for this cross-sectional analysis. The inclusion criteria for subjects were limited by a combination of the SOF design and spatial data available. The study sample from SOF included 2419 Caucasian women ages 65 and older from the Portland cohort assembled in 1986. While the sample recruited for the SOF study also included subjects residing in the Vancouver, Washington region, detailed spatial information required for establishing the neighborhood measures for the current analysis were only available for the Portland metro area urban growth boundary (UGB). Since it was impossible to quantify the built environment characteristics needed for the 347 subjects (14.3%) who resided outside of the Portland metro area UGB, they were excluded from the study. Additionally, 67 subjects (2.8%) were excluded because they provided postal boxes or addresses that were impossible to geocode; thus, the residences for these subjects were not linkable to the built environment information. Ultimately, the analytic sample included 2005 subjects residing within the Portland UGB, representing 82.9% of the study sample.

Built-Environment Measures

As part of the overall Neighborhood and Obesity Study, attributes that characterize the built environments were identified and quantified. These attributes allowed for the categorization of different types of neighborhood built environments that was linkable to individual records on walking and other health information and socioeconomic characteristics. Metro, a locally elected regional government, coordinates with local jurisdictions to maintain spatial data within the Regional Land Information System (RLIS) for planning purposes. RLIS was used to develop the objective exposure measures for use in this project.

Replicating the Frank index[35] was not possible, because the information available in the Portland metro region did not match the needed inputs for measuring net residential density and land use mix. The regional aerial imagery and parcel level land use information that were needed for deriving the net residential density and the land use mix measures were not available until the mid-90s; alternative methodology was developed to quantify the density of urban settlement and land use. A set of six neighborhood measures were constructed using the existing Metro RLIS administrative data and ArcGIS 9.3 (ESRI, Redlands, CA). These six variables were: 1) distance to bus stop; 2) distance to light rail – these two variables both gauged the accessibility to transit services by measuring the distance to the closest bus stop and MAX station respectively; 3) distance to commercial area; 4) distance to park areas – the distances to commercial area and park area measures were proxies for the land use mix, as they measured the access to potential destinations; 5) intersection density – which was a street connectivity measure expressed as the number of intersections divided by the area; and 6) population density – which described the residential distribution. Since the spatial information was unavailable for 1986, data available for the earliest possible year were used as surrogates. The 1988 archival transit data from Trimet, the local transit agency, was digitized to develop the distance to bus stop and light rail station variables. Metro-maintained zoning and park data from 1990 were used to create the distance to commercial area and park area variables. A 1988 streets file was established for measuring the intersection density variable, and 1990 Census block population data was adopted for developing the population density variable. The use of spatial information at a slightly later time point assumed that change in urban form was a gradual process and that information would still be a valid representation of built environment for baseline.

The six built environment variables were developed in a raster or grid environment for the entire Portland metro area, rather than using other geographies with aggregated information, to minimize the bias associated with grouping in a larger geographic unit. Each grid cell was set at 264 feet by 264 feet (or 80 meters by 80 meters). Each side would take roughly one minute of brisk walk to complete; the cell size approximated a Portland city block. The high resolution cell size also ensured linkage of localized resident neighborhood measures to the subjects based on their mapped resident address location.

There were four distance variables, measuring the accessibility to the closest bus stop, light rail station, commercial area, and park area. The variables were measured in Euclidean distance and expressed in feet. The shorter the distance, the better the accessibility to the corresponding amenity. The two density variables, population density and intersection density, were created similarly using the kernel density function, accounting for the street intersections within a quarter mile buffer from each grid cell and the block group population over a one mile buffer, respectively. They were measured as the number of intersections or persons per square mile, and a higher number signified a greater density. Complete details for the exposure measures development process and criteria used to define and standardize the data were documented separately in Appendix A.

The reliability and validity of the built environment measures were also evaluated using a combination of GIS analysis processes and statistical tests. First, the six built environment raster measures were replicated using the GIS Model Builder function to automate the data development process. The pair-wise differences between the duplicate and original set of raster measurements were calculated to detect potential procedural errors; identical measurements were observed for all grid cells with the step by step and automated data development methods. Second, vector/direct distance and density measurements for the six attributes were established for the study sample. The differences between the raster and vector built environment measures were compared using a combination of descriptive statistics and statistical tests. The results indicated that the differences between raster and vector methods for all built environment measures were good or reasonable for use in the current study.[45] The combined effect of the six built environment attributes was operationalized to identify different types of urban forms using cluster analysis.

Walking

The impact of built environments on walking habits was of interest, regardless of whether the walk was for utilitarian or leisure purposes. The total daily walking measure was a derived variable combining responses from two questions in the original self-administered SOF questionnaire. These two questions prompted for the average number of city blocks or equivalent distance (12 blocks = 1 mile) walked per day for exercise and for normal routine, such as shopping. The results from these two questions were added to generate a continuous total number of blocks walked per day measure.

The health outcome examined in the current study was the total number of blocks walked, which combined the blocks walked for utilitarian and leisure purposes. The daily number of blocks walked variable was dichotomized into walking less than five blocks a day or walking five or more blocks a day for exercise and other reasons combined. The walking of less than five blocks a day category was treated as the reference group. The cutoff of five blocks was equivalent to about 400 meters or a quarter of a mile. This amount of walking has been found to provide health benefit to sedentary older adults.[46] The walking outcome is denoted as "walk", "walking", or "walked" for the rest of the discussions in this paper.

Covariates

A number of variables from the SOF were considered as potential confounder and interaction terms. These variables measured individual characteristics, medical history, and health behaviors. Variables such as age, years of education, weekly calories burned from a combination of low, medium, and high intensity exercise, BMI, and amount of smoking (in pack years) were expressed as continuous variables. A few variables were categorical, including: health rating compared to others (excellent, good, fair, poor, very poor) and compared to self a year ago (much better, somewhat better, about the same, somewhat worst, much worst), and marital status (married, widowed, separated, divorced, never married). Medical history variables were reported as binary variables (yes/no), including, whether the subject had a stroke, or had arthritis/rheumatism.

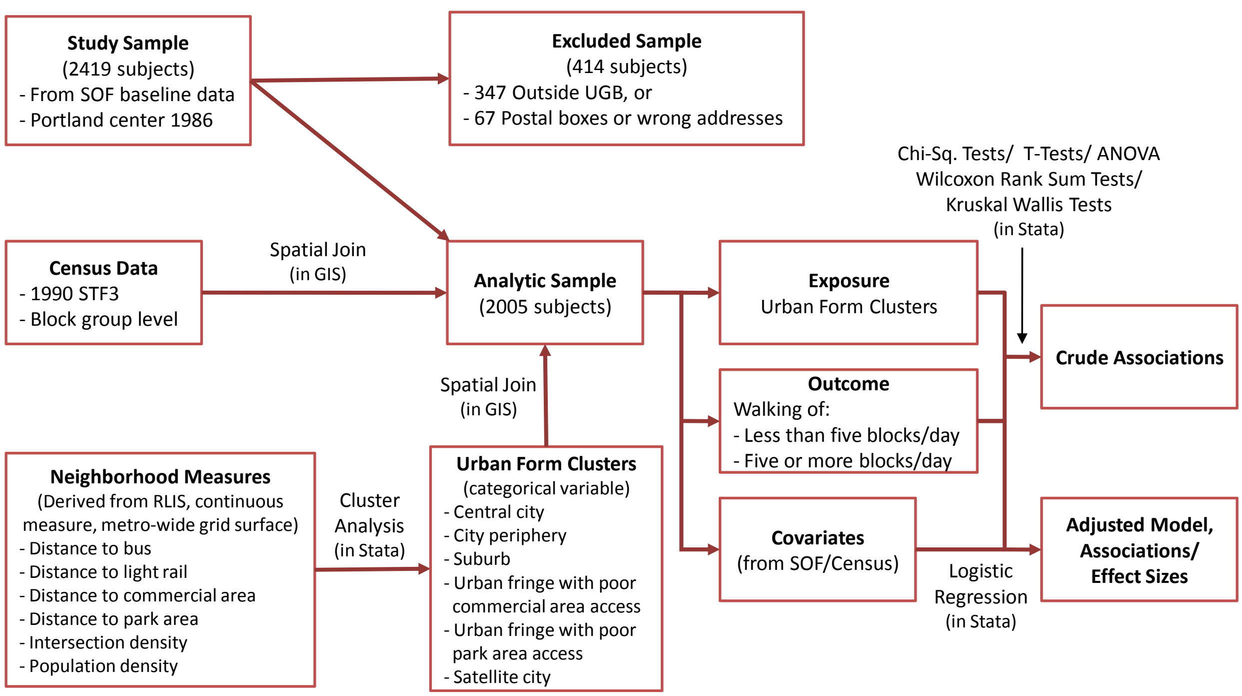
Additionally, Census block group data were included to compensate for the lack of socioeconomic characteristics collected through the SOF questionnaire. Measures of subjects' socioeconomic status were developed using population and household data from the Census. These data were aggregated and summarized in relation to subject residences, providing surrogate socioeconomic status measures. Some of the variables examined include: percent Hispanic population; percent population 65 or older; percent female-headed household with children; percent population with high school or above education; percent population with college or above education; percent labor force in managerial, professional, or specialty services occupation; median household income; percent population in household with interest, dividends, and rental income; percent population living alone and in poverty; and percent 65 plus population living alone and in poverty.

Several of the confounders were classified or reclassified to allow for more meaningful interpretation in the final model. The education variable was dichotomized into 12 years or less education or more than 12 years of education, indicating a cutoff at high school level education. Self-rated health was reclassified into 2 categories, “excellent/good” and “fair/poor/very poor”. The original five categories would result in over 20% of the cells with a count of less than 5 when the variable was tabulated by urban form clusters. Smoking, measured in pack years, was classified into three categories: 0 pack years, 1-40 pack years, and over 40 pack years. The classification would identify non-smokers, light smokers, and heavy smokers. Exercise was measured as the calories consumed per week including the sum of low, medium, and high intensity activities. This variable was dichotomized into 2500kcal or less and more than 2500kcal. Last, BMI was derived from the height and weight of the subjects. A standard classification of underweight (<18.5), normal (18.5-24.9), overweight (25.0-29.9), and obese (≥30.0) was adopted according to the World Health Organization international classification. The underweight and normal categories were later combined due to low number cell counts for the underweight group when the variable was tabulated by urban form clusters. The first category for each variable was set as the reference group in the logistic regression model.

Statistical Analysis

The central goal of this study was to assess the association between built environment and walking habits among older women in the Portland Metro area. All statistical analyses were conducted using Stata 11 (StataCorp, College Station, TX). The structural diagram for the research analysis processes is shown in Figure 1. A cluster analysis was conducted to operationalize the built environment attributes into similar urban form environments, which provide predictive power on the association between built environment and walking. The neighborhood measures were standardized to ensure comparability among measures with different scales. The use of cluster analysis classified each grid cell in the Portland metro area into a cluster based on the K-median partitioning method. Calinski-Harabasz stopping rule was adopted to determine the optimal number of clusters that would maximize the between group differences while minimize the within group differences between the attributes considered, clustering with the highest F-statistics value yielded the optimal results.[47] Grid cells with similar composition of neighborhood attributes would be aggregated into the same group, converting the six continuous neighborhood measures into urban forms that shared similar built environment attributes. The resulting clusters were then mapped to evaluate face validity.

Figure 1. Data Organization and Research Processes



The first approach to understand the sample data was to examine the descriptive statistics. Descriptive statistics such as histograms, means, standard deviations, medians, ranges, and skewness tests were generated to provide graphical and numerical summaries of the distribution pattern of the data and to check the normality of the exposure, outcome, and the covariate variables. These variables were also summarized by residence in the urban growth boundary (yes/no). To evaluate selection bias associated with exclusion criteria, differences by residence in the UGB were assessed differently by data type for each variable: student t-tests for normally distributed continuous variable, Wilcoxon rank-sum tests for skewed continuous variables, Pearson's chi-square tests for categorical variables, and proportion tests or Fisher's exact tests (applicable to small sample only) for binary variables of each subcategory for the categorical variables. Cross tabulations of population characteristics by clusters and walking behavior were also conducted to test differences between groups using Wilcoxon rank-sum tests, Kruskal-Wallis tests, chi-square tests, and ANOVA tests.

Chi-square tests were used to test whether the proportion of older women walking five or more blocks a day differed between different clusters/urban forms. Additionally, the crude odds ratios were estimated using an unadjusted logistic regression. The binary walking variable was regressed on those residing in the cluster with greatest accessibility to amenities and highest density, who had twelve or less years of education, who rated their health as excellent/ good compared to others, who never smoked, who burnt 2500k or less calories per week through exercise, who were underweight or normal, and who had no history of stroke.

To estimate the independent effect of the urban forms on a modest amount of walking among older women, a logistic regression model was built. All potential confounders were examined by testing their associations with both the urban form clusters and the walking variables using chi-square tests, t-tests, ANOVA tests, Wilcoxon rank-sum tests, Kruskal-Wallis tests. Only variables with a univariate p-value less than 0.25 were considered in the preliminary model. Scaling of the continuous variables was then conducted by categorizing continuous variables included in the preliminary model. The categorization decisions were informed by a combination of locally weighted scatter plot smoothing (LOWESS) curves, data distribution, and cut off standards that were of significance or commonly adopted according to literature. Each categorized variable was returned to the preliminary model to ensure the univariate p-value was still less than 0.25. Interaction terms were also evaluated. In particular, interaction terms between urban form clusters and socioeconomic indicators were examined because findings from other studies suggested a significant interaction between neighborhood walkability and income in predicting utilitarian walking;[48] neighborhood characteristics was also hypothesized to have different magnitude of impact on health by socioeconomic status.[49]

Some of the variables potentially interacting with the urban forms clusters examined included: education at individual level, education at block group level, percent Hispanic population, percent population 65 years or older, percent population in poverty. Additionally, education and percent poverty among the live alone population were also tested. Interaction terms with a p-value less than 0.25 were considered for the multivariate model. All variables that fulfilled these criteria were entered into the model and a manual process was used to establish the model to ensure no subgroup of a categorical variable would be removed if the variable remained overall significant. The variable with the highest p-value was removed from the model one by one until all variables considered in the model were significant at $\alpha=0.05$. Prepackaged backward and forward selection processes were performed to check the results from the manual process. Interaction terms were tested after the main effects model had been established to ensure the effects from interaction terms were not overlooked. Diagnostic tests, such as goodness of fit, measures of fit and potential outliers were examined to ensure the model was robust.

The final logistic regression model was established to assess the association between urban forms and whether the sample of residents walk five or more blocks a day. The analysis also estimated magnitude of association between different urban forms and walking among older women after controlling for confounders. Graphical representation of the odds ratios and 95% confidence intervals of the effect sizes for various urban form clusters, smoking groups and BMI groups were examined. Although not part of the goal for this study, the pair-wise comparison of the impact on walking by different urban forms were also explored to identify urban forms that might have impacted walking differently but were not hypothesized initially. Separate models were fitted to assess the trends/gradient effect observed in the probability of older women committing a modest amount of walking per day with changes in smoking amount and BMI.

Results

The individual characteristics for the analytic sample were generally similar to the excluded set of potential subjects residing outside of the UGB, except for level of exercise, stroke history, and marital status (see Table 1). The excluded subjects were less likely to have a history of stroke, and more likely to be married. Census information suggested that the population and environment outside of the study area might be very different from the included area: the area outside of the study area tended to have a lower share of older adults 65 and over, lower education attainment, lower share of the population with managerial, professional, or specialty services occupations. In terms of household characteristics, compared to the study area, areas outside the UGB had: a lower share of population receiving interests, dividends, and rental income; a higher share of live alone households in poverty; and a higher share of live alone households with householders over 65 years old that in poverty. Table 1 presents the individual and Census block characteristics comparing the SOF sample residing in the Portland metro area to those outside of the UGB.

Table 1. Descriptive Statistics for the Analytic Sample versus Excluded Sample

Median (Inter-Quartile Range) or N (%)	Inside UGB (n=2005)		Outside UGB (n=414)		P-Value
Variable	**Median or N**	**IQR or %**	**Median or N**	**IQR or %**	
Age	72	68 - 76	71	68 - 75	0.13
Education (yrs)	12	12 - 14	12	12 - 14	0.80
Exercise (kcal/week)	929	300 - 1976	1099	457 - 2080	0.02*
BMI	26	23 - 29	26	23 - 30	0.09
Daily block walked	9	3 - 20	9	3 - 20	0.94
History of Stroke	97	4.87%	10	2.44%	0.03*
History of Arthritis	1209	61.25%	262	64.53%	0.22
Smoking (packyears)	0	0 - 11	0	0 - 11	0.46
Self rated health					0.61
Excellent	617	30.77%	119	28.74%	
Good	1031	51.42%	230	55.56%	
Fair	325	16.21%	60	14.49%	
Poor	31	1.55%	5	1.21%	
Very Poor	1	0.05%	0	0.00%	
Compare to self 12 months ago					0.72
Much better	132	6.58%	33	7.97%	
Somewhat better	185	9.23%	40	9.66%	
About the Same	1427	71.17%	295	71.26%	
Somewhat worst	252	12.57%	44	10.63%	
Much worst	9	0.45%	2	0.48%	
Marital Status					0.03*
Married	1005	50.12%	236	57.00%	
Widowed	759	37.86%	148	35.75%	
Separated	10	0.50%	2	0.48%	
Divorced	170	8.48%	21	5.07%	
Never married	61	3.04%	7	1.69%	
Census Block Group Variables					
Percent Hispanic population	1.88%	6.10% - 3.77%	1.96%	0.73% - 4.13%	0.16
Percent population 65 or above	13.76%	9.43% - 17.80%	12.60%	10.16% - 17.51%	0.002*
Percent children household headed by female householder	17.89%	10.37% - 26.11%	16.56%	5.86% - 29.69%	0.12
Percent with high school or above education	84.50%	78.17% - 90.59%	82.37%	77.67% - 86.92%	<0.0001*
Percent with college degree or above education	20.84%	12.78% - 32.31%	14.62%	10.63% - 19.14%	<0.0001*
Percent labor force in managerial, professional, or specialty services occupations	25.36%	18.32% - 35.86%	23.93%	18.01% - 30.46%	<0.0001*
Median household income (1989 dollar)	28226	23289 - 35633	30893	24861 - 37140	0.82
Percent population in household with interest, dividends, and rental income	46.53%	38.21% - 57.25%	43.73%	34.10% - 50.98%	<0.0001*
Percent population in living alone household are in poverty	11.76%	6.34% - 19.10%	13.64%	6.58% - 21.76%	0.03*
Percent 65+ population in living alone household are in poverty	12.04%	0.00% - 25.00%	17.24%	0.00% - 37.84%	0.0001*

** significant at 0.05 level*

The k-median cluster analysis resulted in six distinct urban form clusters across the metro region. The cluster analysis results were mapped in Figure 2 to provide a spatial representation of how grid cells that were grouped into different clusters were distributed. The mapping of urban form clusters in the Portland Metro-area served as a visual validation tool for the cluster analysis results that were performed statistically. The map demonstrates reasonable clustering with grid cells in the same cluster locating in a fairly continuous manner spatially; the distribution of the six clusters was also sensible and consistent with local knowledge of the environmental characteristics of the Metro region.

Figure 2. Spatial Representation of Urban Form Clusters , Portland Metro Area

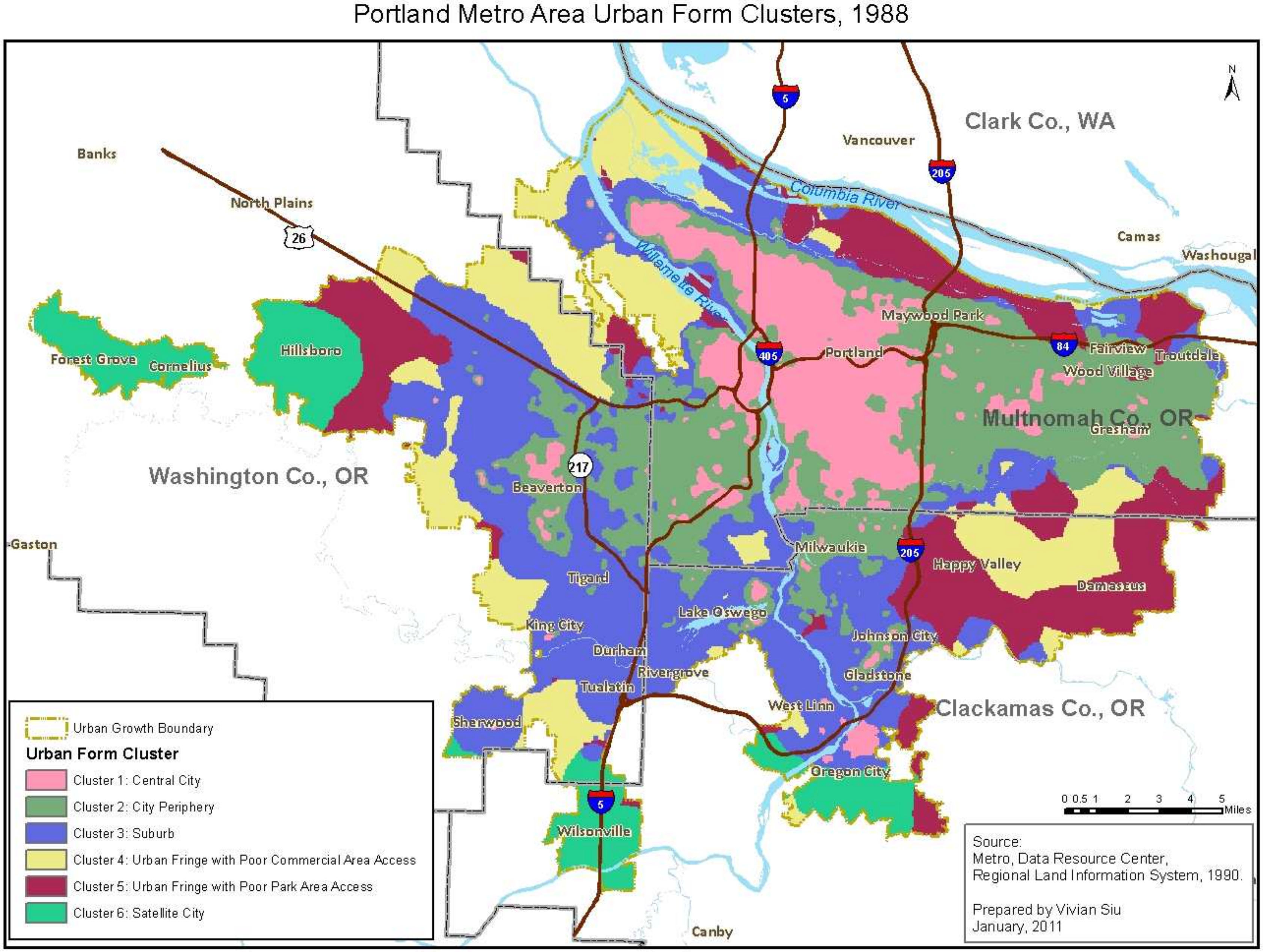


Table 2 provides description of the urban form clusters based on the summary statistics of the built environment attributes. Clusters 1, 2, and 3 showed a gradual decrease in accessibility and density for all built environment attributes: with cluster 1 being the "best" urban form with close proximity to transit, commercial areas, and park areas, while the intersection and population density were also the highest of the six clusters. The accessibility and density decreased from the central core to cluster 2 and dropped even further for cluster 3. Although clusters 4, 5, and 6 generally had poorer accessibility to amenities and lower density than cluster 3, there were certain attributes in each of these clusters that were similar to another cluster and that provide some insight as to how specific attributes influence walking habits. For instance, cluster 4 and 5 shared low intersection and population density, and they had similar measures for access to bus or light rail services; they differed in that cluster 4 had the greatest distance to commercial areas while cluster 5 had the greatest distance to park areas. Cluster 6 was similar to cluster 3 in terms of access to commercial areas and density, but it had the greatest distance to transit services among the six clusters.

Table 2. Summary of Built Environment Attributes by Cluster, Portland Metro Area

	Distance to (in meter)				Density (number per km^2)	
Median (IQR)	**Bus Stop**	**Light Rail Station**	**Commercial Area**	**Park**	**Intersection**	**Population**
Cluster 1: Central city	161 (80 - 241)	2968 (1451 - 5465)	228 (80 - 341)	402 (241 - 569)	111 (92 - 138)	2591 (2062 - 3027)
Cluster 2: City periphery	241 (114 - 402)	3817 (2037 - 7701)	332 (161 - 580)	433 (241 - 688)	53 (38 - 69)	1473 (1125 - 1793)
Cluster 3: Suburb	433 (228 - 809)	12876 (9927 - 15253)	628 (322 - 995)	410 (180 - 688)	24 (10 - 42)	777 (407 - 1199)
Cluster 4: Urban fringe with poor commercial area access	1778 (1048 - 2680)	11273 (7816 - 15331)	2255 (1835 - 2747)	724 (402 - 1074)	5 (0 - 13)	62 (0 - 334)
Cluster 5: Urban fringe with poor park area access	1821 (1018 - 3030)	8459 (5413 - 12436)	1049 (613 - 1527)	1884 (1479 - 2421)	7 (2 - 17)	166 (10 - 367)
Cluster 6: Satellite city	6112 (4474 - 8697)	23356 (21622 - 26551)	515 (241 - 900)	764 (433 - 1249)	26 (8 - 50)	641 (280 - 1093)

Based on the observed medians, the inter-quartile ranges for the built environment attributes, and the cluster map, descriptive terms were assigned to the resulting urban form clusters in Table 3 for simplified reference and interpretation. Cluster 1 or central city, which had the greatest accessibility to amenities including: bus stop, light rail station, commercial area, and park area, highest street connectivity, and highest population density was considered as the most compact urban form and was set as the reference group because this environmental setting was hypothesized to encourage walking.

Table 3. Cluster Descriptions

Urban Form Cluster	Description
Cluster 1: Central city	areas with highest intersection and population density; greatest access to bus stop, light rail station, commercial areas and park areas; most of the city of Portland and the cores of surrounding cities such as Beaverton.
Cluster 2: City periphery	areas immediately surrounding the central city cluster with slightly poorer access to amenities and about half the density of central city. This cluster included areas such as outer east Portland, the remaining part of City of Portland on the west side, part of Beaverton, Gresham, and Milwaukie.
Cluster 3: Suburb	areas with about 500m (0.4 mile) distance to amenities except for light rail service, which was much more inaccessible; moderate intersection and population density. Areas such as the remainder of Beaverton, Tigard, Lake Oswego, Tualatin, West Linn, and Gladstone displayed this type of urban form.
Cluster 4: Urban fringe with poor commercial area access	areas with low population density, low street connectivity, poor transit services, poor commercial area access (more than 2km), but a relatively reasonable access to park areas (roughly 800m/0.5 mile). This cluster encompassed areas such as Forest Park and the east side of Happy Valley.
Cluster 5: Urban fringe with poor park area access	areas with low population density, low street connectivity, poor transit services, poor park area access (more than 2km), but a slightly more reasonable access to commercial areas (about 1km/0.6 mile). This cluster included areas such as the west side of Happy Valley, Damascus, areas along the Columbia River, and areas between Beaverton and Hillsboro.
Cluster 6: Satellite city	areas with attributes similar to the suburb, but with extremely poor access to transit services. Places such as Hillsboro, Cornelius, Forest Grove, Wilsonville, and Oregon City were in this category.

Table 4 shows the cross tabulation and associations of population characteristics by clusters, and Table 5 shows the cross tabulation and associations of population characteristics by walking outcome. The test of association between the urban form clusters and walking indicated that there was not a significant difference in proportion of older women walking (p-value = 0.13). According to the results from a series of tests of associations between the covariates and the exposure/outcome variables, potential confounders identified and considered (see Tables 4 and 5) in the preliminary model included: age; education; self-reported health; lifetime smoking quantity; other exercises; BMI; history of stroke; history of arthritis; percent block group population with Hispanic origin; percent 25 plus population with high school or above education; percent 25 plus population with college degree or above education; percent worker with managerial, professional, or specialty services occupation; median household income; percent of population with interest, dividends, or rental income; and percent of live alone population in poverty.

The estimated odds ratios for the crude model are also presented in Table 5. The crude odds ratios suggested that compared to central city (cluster 1), subjects residing in all other clusters had a lower estimated odds for walking. However, only the difference between the city periphery (cluster 2) and the central city (cluster 1) was statistically significant. Specifically, women living in the city periphery were less likely to walk each day compared to women residing in the central city (OR: 0.22, CI: 0.04- 0.36) (see Table 5).

Table 4. Characteristics of Sample by Urban Form Cluster

N (%) or Median (IQR)	Portland Metro Area	Cluster 1 (n=881)	Cluster 2 (n=787)	Cluster 3 (n=245)	Cluster 4 (n=28)	Cluster 5 (n=51)	Cluster 6 (n=13)	P-Value
Walking								0.13
less than 5 blocks a day	700 (35.00%)	282 (32.16%)	297 (37.74%)	84 (34.43%)	14 (50.00%)	18 (35.29%)	5 (38.46%)	
5 or more blocks a day	1300 (65.00%)	595 (67.84%)	490 (62.26%)	160 (65.57%)	14 (50.00%)	33 (64.71%)	8 (61.54%)	
Percent live alone population in poverty, 1989	11.76% (6.34% - 19.10%)	15.13% (7.80% - 23.33%)	11.04% (6.07% - 17.54%)	8.27% (4.83% - 13.18%)	6.33% (0.00% - 11.11%)	14.39% (11.11% - 23.39%)	12.73% (11.02% - 23.61%)	0.0001
Age	72 (68 - 76)	72 (68 - 76)	71 (68 - 76)	71 (68 - 75)	71 (69 - 75.5)	72 (69 - 75)	68 (68 - 72)	0.22
Education level								0.02
12 years or less	1195 (59.66%)	529 (60.11%)	469 (59.67%)	132 (53.88%)	15 (53.57%)	40 (78.43%)	10 (76.92%)	
more than 12 years	808 (40.34%)	351 (39.89%)	317 (40.33%)	113 (46.12%)	13 (46.43%)	11 (21.57%)	3 (23.08%)	
Self-rated health								0.19
Excellent/Good	1648 (82.19%)	714 (81.04%)	644 (81.83%)	216 (88.16%)	23 (82.14%)	40 (78.43%)	11 (84.62%)	
Fair/ Poor/ Very Poor	357 (17.81%)	167 (18.96%)	143 (18.17%)	29 (11.84%)	5 (17.86%)	11 (21.57%)	2 (15.38%)	
Smoking								0.04
0 packyears	1277 (63.79%)	556 (63.25%)	504 (64.12%)	158 (64.49%)	19 (67.86%)	35 (68.63%)	5 (38.46%)	
1 - 40 packyears	531 (26.52%)	232 (26.39%)	210 (26.72%)	64 (26.12%)	9 (32.14%)	10 (19.61%)	6 (46.15%)	
more than 40 packyears	194 (9.69%)	91 (10.35%)	72 (9.16%)	23 (9.39%)	0 (0%)	6 (11.76%)	2 (15.38%)	
Exercise (kcal per week)								0.05
2500kcal or less	1669 (83.24%)	751 (85.24%)	652 (82.85%)	192 (78.37%)	24 (85.71%)	42 (82.35%)	8 (61.54%)	
more than 2500kcal	336 (16.76%)	130 (14.76%)	135 (17.15%)	53 (21.63%)	4 (14.29%)	9 (17.65%)	5 (38.46%)	
BMI								0.13
Underweight/ Normal (less than 25.0)	726 (36.21%)	313 (35.53%)	283 (35.96%)	104 (42.45%)	8 (28.57%)	14 (27.45%)	4 (30.77%)	
Overweight (25-29.9)	824 (41.10%)	351 (39.84%)	338 (42.95%)	98 (40.00%)	12 (42.86%)	19 (37.25%)	6 (46.15%)	
Obese (30.0 or above)	455 (22.69%)	217 (24.63%)	166 (21.09%)	43 (17.55%)	8 (28.57%)	18 (35.29%)	3 (23.08%)	
Stroke								0.17
No history of stroke	1896 (95.13%)	836 (96.09%)	748 (95.04%)	227 (93.03%)	26 (92.86%)	46 (90.20%)	13 (100.00%)	
With history of stroke	97 (4.87%)	34 (3.91%)	39 (4.96%)	17 (6.97%)	2 (7.14%)	5 (9.80%)	0 (0%)	

Table 5. Characteristics of Sample by Walking Habit Outcome and Estimated Crude Odds Ratios

N (%) or Median (IQR)	Daily Walking		Crude OR (95% CI)	P-Value
	less than 5 blocks	5 blocks or more		
Built environment				0.1311
Cluster 1: Central city	282 (40.29%)	595 (45.77%)	reference	
Cluster 2: City periphery	297 (42.43%)	490 (37.69%)	0.78 (0.64-0.96)	
Cluster 3: Suburb	84 (12.00%)	160 (12.31%)	0.90 (0.67- 1.22)	
Cluster 4: Urban fringe w/ poor commercial area access	14 (2.00%)	14 (1.08%)	0.47 (0.22- 1.01)	
Cluster 5: Urban fringe w/ poor park area access	18 (2.57%)	33 (2.54%)	0.87 (0.48- 1.57)	
Cluster 6: Satellite city	5 (0.71%)	8 (0.62%)	0.76 (0.25- 2.34)	
Percent live alone population in poverty, 1989	12.24% (6.46% - 20.56%)	11.40% (6.17% - 18.58%)	0.35 (0.15- 0.80)	0.013
Age	72 (69 - 77)	71 (68 - 75)	0.95 (0.93- 0.97)	<0.001
Education level				<0.001
12 years or less	460 (65.81%)	731 (56.27%)	reference	
more than 12 years	239 (34.19%)	568 (43.73%)	1.50 (1.24- 1.81)	
Self-rated health				<0.001
Excellent/Good	521 (74.43%)	1123 (86.38%)	reference	
Fair/ Poor/ Very Poor	179 (25.57%)	177 (13.62%)	0.46 (0.36- 0.58)	
Smoking				0.0413
0 packyears	516 (61.36%)	1029 (65.63%)	reference	
1 - 40 packyears	227 (26.99%)	409 (26.08%)	0.96 (0.78- 1.19)	
more than 40 packyears	98 (11.65%)	130 (8.29)	0.67 (0.49- 0.91)	
Exercise (kcal per week)				<0.001
2500kcal or less	638 (91.14%)	1026 (78.92%)	reference	
more than 2500kcal	62 (8.86%)	274 (21.08%)	2.75 (2.05- 3.68)	
BMI				<0.001
Underweight/ Normal (less than 25.0)	216 (30.86%)	507 (39.00%)	reference	
Overweight (25-29.9)	288 (41.14%)	535 (41.15%)	0.79 (0.64- 0.98)	
Obese (30.0 or above)	196 (28.00%)	258 (19.85%)	0.56 (0.44- 0.72)	
Stroke				<0.001
No history of stroke	641 (92.50%)	1250 (96.53%)	reference	
With history of stroke	52 (7.50%)	45 (3.47%)	0.44 (0.29- 0.67)	

Table 6 presents the final model after controlling for age, education, self-reported health, lifetime smoking quantity, other exercises, BMI, history of stroke, and block group percent of live alone population in poverty. No interactions between the urban form clusters and the Census variable or with the education variable met criteria for inclusion in the final model. The adjusted model indicated that there were independent associations between types of urban form clusters and daily walking. Women residing in city periphery (cluster 2), suburb (cluster 3), and urban fringe with poor commercial access (cluster 4) were significantly less likely to walk daily. There were no significant differences in walking outcome among residents from all other urban form combinations.

Consistent with the observation from the crude model, residence in city periphery (cluster 2) was statistically significantly associated with less walking compared to residence in the central city (cluster 1). Older women residing in city periphery were less likely than women residing in the central city to walk each day (OR: 0.69, CI: 0.55-0.85). In addition, women in suburb (cluster 3) and in urban fringe with poor commercial access (cluster 4) were also significantly less likely to walk daily compared to those living in the central city. Older woman in the suburb had a lower odds of walking than those residing in the central city (OR: 0.69, CI: 0.50-0.95), while older woman residing in urban fringe with poor commercial areas access was less likely to walk compared to a woman in the central city (OR: 0.40, CI: 0.18-0.88). There was not a significant difference in walking among older women residing in the urban fringe with poor park areas access (cluster 5) compared to those in the central city (OR: 0.98, CI: 0.52-1.84). Older women residing in satellite city (cluster 6) were also less likely to walk compared to those in the central city (OR: 0.53, CI: 0.16-1.71); however, the difference was also not statistically significant. Due to the small samples of older women residing in urban fringe with poor commercial area access (cluster 4) and satellite city (cluster 6), the findings for the comparison between cluster 4 versus cluster 1 and cluster 6 versus cluster 1 had greater uncertainties and thus were not conclusive. Pair-wise comparisons of effect sizes between the non-central city clusters were also explored. Appendix B summarizes the relative effect sizes in terms of odds ratios and 95% confidence intervals between different urban form cluster combinations and walking. Central city was found to have greater effect in promoting walking among older women than city periphery, suburb, and urban fringe with poor access to commercial area. However, there were no significant differences in walking between older women residing in the central city and urban fringe with poor park area access or between central city and satellite city. Additionally, no significant differences were noted in the proportion of older women walking comparing all the non central city clusters.

From observing the estimated odds ratios in Table 6, the directions of effect other adjusted variables had on walking were also reasonable. The likelihood of women walking decreased with increasing age or increasing neighborhood block group share of live alone population in poverty. Likelihood of walking increased for older women with more than 12 years of education compared to those with 12 years or less education (OR: 1.32) or who burned more than 2500kcal per week through exercises compared to those who exercise less (OR: 2.49). On the other hand, the likelihood for walking as compared to the reference group decreased for women who: rate their health fair, poor, or very poor (OR: 0.52); smoked 1-40 pack-years (OR: 0.88) or over 40 pack-years (OR: 0.66); were overweight (OR: 0.83), or obese (OR: 0.59); and had a stroke (OR: 0.57) (see Table 6).

Table 6. Estimated Odds Ratios and 95% Confidence Intervals, Adjusted Logistic Regression Model

	Est. OR	95% CI	P-Value
Built environment			0.0057
Cluster 1: Central city		Reference	
Cluster 2: City periphery	0.69	(0.55, 0.85)	
Cluster 3: Suburb	0.69	(0.50, 0.95)	
Cluster 4: Urban fringe w/ poor commercial area access	0.40	(0.18, 0.88)	
Cluster 5: Urban fringe w/ poor park area access	0.98	(0.52, 1.84)	
Cluster 6: Satellite city	0.53	(0.16, 1.71)	
Percent live alone population in poverty, 1989	0.28	(0.11, 0.70)	0.007
Age	0.95	(0.93, 0.96)	<0.001
Education level			0.007
12 years or less		Reference	
more than 12 years	1.32	(1.08, 1.62)	
Self-rated health			<0.001
Excellent/Good		Reference	
Fair/ Poor/ Very Poor	0.52	(0.41, 0.67)	
Smoking			0.0427
0 packyears		Reference	
1 - 40 packyears	0.88	(0.70, 1.11)	
more than 40 packyears	0.66	(0.48, 0.92)	
Exercise (kcal per week)			<0.001
2500kcal or less		Reference	
more than 2500kcal	2.49	(1.84, 3.37)	
BMI			0.0004
Underweight/ Normal (less than 25.0)		Reference	
Overweight (25-29.9)	0.83	(0.66, 1.03)	
Obese (30.0 or above)	0.59	(0.46, 0.77)	
Stroke			0.01
No history of stroke		Reference	
With history of stroke	0.57	(0.37, 0.88)	

For the urban form cluster, smoking quantity, and BMI variables, since there were three or more categories, each of these variables were stratified to examine the relative effect sizes. The adjusted odds ratios and their corresponding 95% confidence intervals are displayed graphically in Figure 3a-c to show the effect sizes for the subgroups. Separate models fitting the smoking and BMI variables as continuous forms indicated a significant linear relationship for smoke amount (p=0.003) and BMI (p<0.001), adjusting for the same variables. The observed results suggest that there is a general decline in probability that an older woman will walk daily if she smokes more or has a higher BMI.

Figure 3a. Estimated Effect Sizes and Confidence Intervals of Built Environment Clusters on Walking, Adjusted Multivariate Model

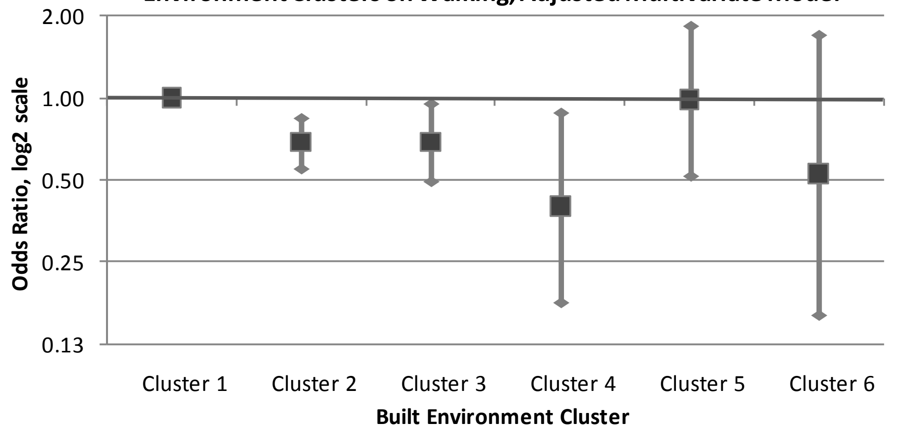


Figure 3b. Estimated Effect Sizes and Confidence Intervals of Smoking Status on Walking, Adjusted Multivariate Model

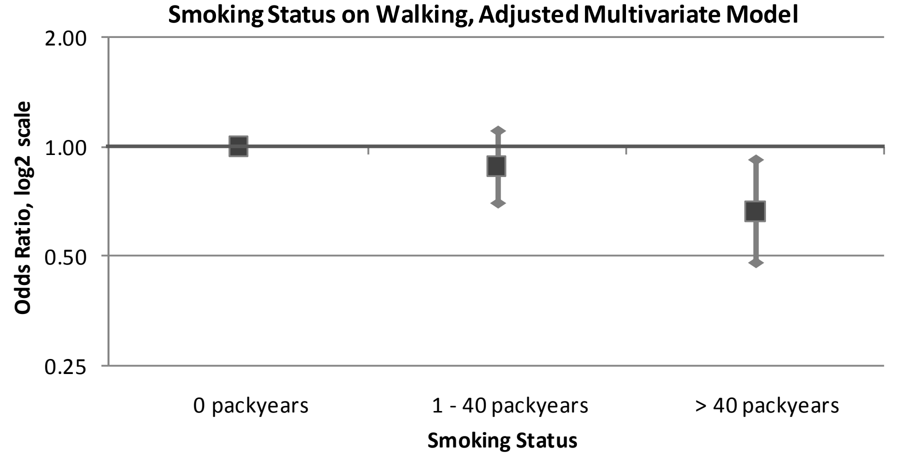


Figure 3c. Estimated Effect Sizes and Confidence Intervals of BMI on Walking, Adjusted Multivariate Model

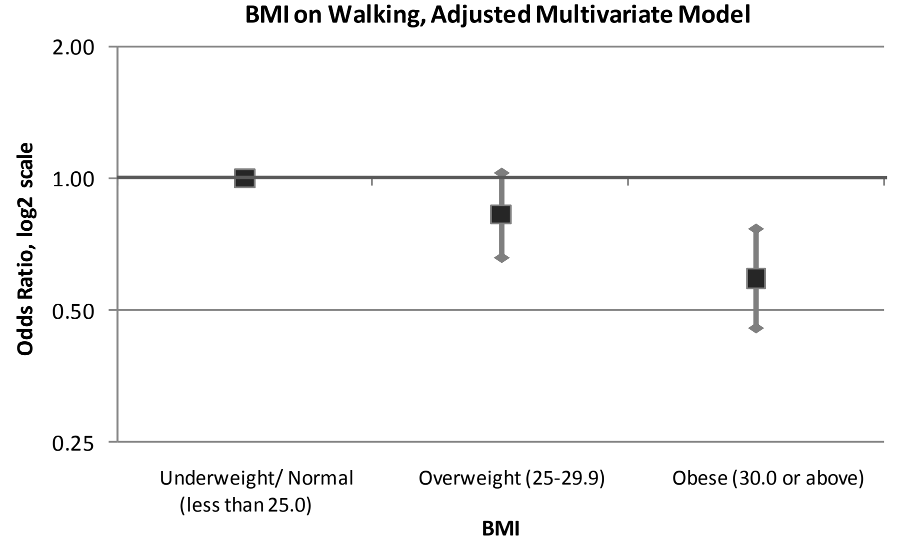
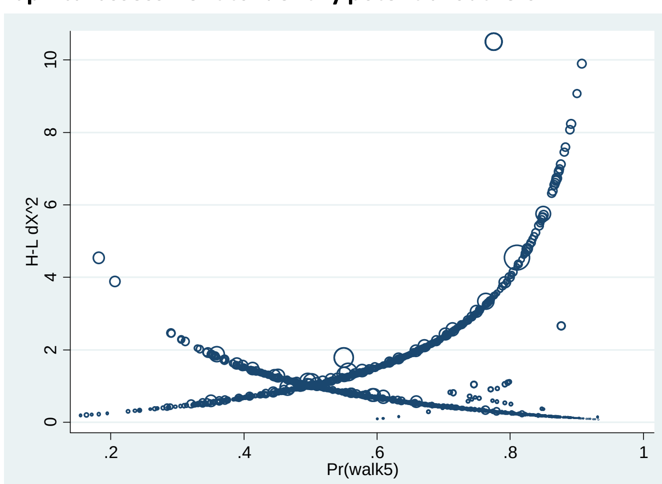


Figure 4 shows the results from the model diagnostic. The Hosmer-Lemeshow test assessed the deviation of expected values from the observed values for each decile group. This test indicated the fitted model was good (p=0.84). The graph in Figure 4 was used to help identify potential outliers and observations with high leverage. Four potential outliers were identified with a change in Pearson chi-square statistic of over 10, a change in deviation of more than 6, or a Cook's distance of more than 0.2. The graph shows the change in Pearson's chi-square statistic against the predicted probability for walking, weighted with Cook's distance. Simultaneous removal of the four observations that were potential outliers or with high leverage did not change the value of the estimated coefficients or the overall model fit substantially; therefore, all observations were retained in the analytic data set used as the basis of the final model.

Figure 4. Model Diagnostic: Goodness of Fit and Assessing Potential Outliers

Goodness of Fit	Test Statistics	P-Value
Hosmer-Lemeshow Chi-sq (8)	4.20	0.8387

Graphical assessment to identify potential outliers



Discussion

This study utilized spatial data, gathered for administrative purposes, to create objective neighborhood measures that were linkable to individual level data. This approach enabled evaluation of the association between neighborhood design and walking behavior. Cluster analysis provided a way to consider different neighborhood attributes jointly and to classify the Portland metro area into six unique urban forms/ clusters: central city, city periphery, suburb, urban fringe with poor commercial area access, urban fringe with poor park area access, and satellite city. As hypothesized, different urban forms/clusters were associated with a modest amount of walking (walking of five city blocks or more a day) among older women after adjusting for age, education, self-rated health, smoking, exercise, BMI, history of stroke, and block group poverty among people living alone.

Among the six urban forms identified and compared, central city appeared to be an environment where older women were more likely to engage in a modest amount of daily walking compared to environments like the city periphery, suburb, and urban fringe with poor commercial area access. The overall high population density, high street connectivity, supported by convenient access to amenities, especially to transit and commercial area, might effectively support and encourage older women to engage in a modest amount of daily walking. The built environments seemed to influence walking habits; improvement or worsening of all six built environment attributes occurred together. Central city displayed a relatively “ideal” setting in which overall accessibility and density were greater than any other urban forms identified while older women residing in this area were more likely to walk. A possible explanation for this observation was that built environment attributes work together to provide a pedestrian friendly place. Good infrastructure and design, such as availability of transit services and high street connectivity, encourage people to navigate around the area.[5,28,50] The existence of commercial businesses and parks provides destinations that attract people to visit.[2,13,20] The high population density provides the capacity to support businesses and transportation services. [2,13,20,28] In combination, these components are associated with increased walking among older women.

Although the city periphery cluster, suburb cluster, and urban fringe cluster with poor commercial area access all displayed significantly lower odds for older women to walk compared to the central city, there were no significant differences observed between these urban forms. This might suggest the changes in magnitude for the built environment measures were not substantial between these urban forms, or there might be attribute thresholds that a combination of measures had to exceed to indicate an urban form effective in encouraging older women to walk.

The two urban forms that did not show lower odds of walking compared to the central city were urban fringe with poor park area access, and satellite city. Areas classified as urban fringe with poor park area access had similar population and street intersection density, as well as similar distances to transit services as the urban fringe with poor commercial access. Although the small sample sizes from the two urban fringe clusters resulted in greater uncertainty in the findings, the results provided some indications that there were varying strengths among built environment attributes in influencing walking habits. Access to destinations, were possible differences that could account for the significant difference in walking for urban fringe cluster with poor commercial area access but not for urban fringe cluster with poor park area access when each cluster was compared with the central city. The findings suggested that in low density areas, proximity to a commercial area might play a stronger role in promoting walking than proximity to a park.

The effect of satellite city on older women's walking behavior was not significantly different from that of central city. Nevertheless, the estimated effect indicated residents of the satellite city had a lower odds of walking (OR=0.53, CI: 0.16-1.71) compared to those residing in the central city. However, the small sample size of 13 residents in the satellite city cluster created a large confidence interval, making it more difficult to ascertain the association. Further studies are needed to ascertain whether areas that are somewhat self-sufficient but physically disjoint from the metro area have an effect on walking that is similar to the effect of the central city.

Few studies have quantified built environment characteristics using high resolution geographic unit or used clusters to describe different urban forms. The studies that have used cluster analysis in investigating the association between the built environment and walking generally support the key components identified in the current study, particularly the study by Riva et al. 2009,[19] in which homogeneous spatial units were established and optimized into zones with similar characteristics. However, in Riva et al.'s study, these units were refined from geographic data from the Canadian Census and were standardized roughly by population. In the current study, the spatial units (grid cells) were self-established and standardized by area. Both Riva's investigation and the current study involved built environment predictors/ active living potential indicators such as access, land use, and density. The specific methods to quantify these variables were slightly different, potentially due to data limitations for the study areas. For accessibility, distance to closest feature was documented in the current study while in Riva et al.'s study the average number of target destinations within a 1km buffer was measured. For land use attribute, land use mix entropy index (Frank index) adopted by Riva was not replicable in the current study; thus, accessibility to commercial areas and park areas were used as indicators for land use patterns. Population densities were calculated similarly for both studies, except interpolation was involved in the current study to provide measures for in-between areas instead of assuming the block group density. Additionally, the current study measured street intersection density, an indicator for street connectivity, but this was not included in the analysis by Riva and colleagues.

The cluster analysis results from both studies showed similar patterns that were capable of identifying areas with different density and characteristics such as transportation corridor; however, the use of more refined standardized spatial units in the current study yielded a cluster map that appeared to be smoother and spatially more contiguous than the version by Riva and colleagues. Findings from

Riva et al. suggested that variations in total walking were significant only at Census tract level but not for the cluster zones; number of 10-minute episodes of walking was significantly higher in the central urban zone compared to the low density suburban zone. The current study found no differences in walking between different clusters but a significant association between built environment and total walking comparing the central city to the city periphery, suburb, or urban fringe with poor commercial access.

Other studies that utilized cluster analysis also performed the grouping using administrative geographic units such as TAZ or Census tract.[3,43] One study[4] performed cluster analysis of the individual walking/biking prevalence; it was found that higher street connectivity and shorter block length were associated with active transportation. In other studies, transit availability near residence has been associated with greater transit mode share regardless of car ownership;[3] additionally, cities with great diversity in housing types, such as those with transit-oriented developments and alternative developments, would allow for more sensitive reflection of true transportation-land use preferences than places that lack such variety.[43]

According to the stated similarity and difference between the current study and existing literature, the key distinction of the current study was the use of more refined self-established spatial unit for developing neighborhood attribute measures. This methodology allowed for slightly different approaches to quantify the built environment that detached the notion of a defined neighborhood. Although there were differences in methodology and study focuses, findings from these studies involving cluster analysis revealed the importance of attributes such as transit accessibility, street connectivity, and a combination of access, land use, and density associating with walking/ transportation choices. The use of cluster method to identify urban forms was also a reasonable approach for assessing the built environment and walking.

Several data limitations in the current study must be recognized. First, 17% of the original study sample was excluded from the current analysis as a result of the inclusion/exclusion criteria. This may lead to selection bias if the selection of sample is dependent on the built environment (exposure) and the walking outcome. Since the sample for this study was originally recruited for the SOF, detailed built environment and walking were not determining factors for the subject recruitment and selection bias from this source was unlikely. However, the comparison of included women to excluded women on multiple individual and neighborhood characteristics suggests limited bias associated with the selection. Significant differences were noted for population and neighborhood characteristics such as level of exercise, history of stroke, marital status, neighborhood (Census block group) share of: older population; population with high education attainment; population participating in managerial, professional specialty services occupations; households with interest, dividends, and rental income; and live alone households in poverty among householder of all ages, as well as among those who were 65 or older. Potentially, the inclusion criteria might have created some selection bias, as residents within and outside of the UGB experience different built environment characteristics and socioeconomic status. However, the walking outcome between the included and excluded sample did not show any differences; therefore, if any selection bias existed, it was unlikely to affect the direction of the association. The geographic specificity of the inclusion/exclusion criteria used in the current study limit the external validity of the findings, meaning that a valid inference can be made for the study area population based on the study findings, but the same inference would be less generalizable to people residing in other places.

Second, there is a potential for self-selection bias since older women who are active may prefer to live in areas like the central city, where there are reasonable transportation supports and convenient access to commercial and park areas that enable them to live without being auto-dependent. However, neighborhood selection is only one of the many factors affecting mobility of the older population. Generally, there are three sets of factors determining the mobility of the elderly population: life cycle events, environmental factors, and economic status.[51] The older adult population is also a very diverse group with varying housing preferences and mobility rates by age, gender, and resources.[14,51-53] The complexity of migration decisions and patterns among older adults makes assessing potential self-selection bias challenging. If self-selection bias existed in our sample, the share of older women who live in the central city and walk five or more blocks/day would be elevated, while the share of those who met the walking cutoff in the more suburban/rural clusters would be lowered. This would tend to overestimate the effect size (lower the odds ratios) for walking when comparing the more suburban or rural environment to the central city. The findings from existing literature[54,55] on the topic suggest that individual's preference may not be consistent with the actual living environment or affect walking habits. In other studies,[56] intra-urban migration pattern among elderly suggests that older adults who are wealthy and socially active tend to move away from central city, while more diverse, individuals of lower income relocate within similar neighborhoods. Neighborhood preferences also do not play a strong role in relocation decision among those who moved.[14] This pattern is opposite to the suspected mobility towards the central city, and point to evidences that older adults' mobility is affected by factors such as housing cost and factors other than neighborhood preference. Therefore, self-selection may have limited impact on the observed association in the current study, but the role of self-selection bias cannot be ruled out entirely. The observed association of between built environments and the walk habits may be prone to the problem of reverse causality, making it more difficult to assess the actual causal relationship between the built environment and walking if self-selection bias is present.

Third, measurement error in the walking variable might have lead to information bias if the error was differential by neighborhood cluster. Walking was self-reported as the number of city blocks or equivalent (12 blocks= 1 mile) walked a day. Although provided with various query terminologies such as "city blocks" and a distance conversion in the question to ensure responses were consistent, the definition of "block" could still vary substantially between urban and suburban cities. Blocks in the suburban cities are typically larger than the average city block, so residents in the suburban area may report a smaller number of blocks than those living in the urban area, even though they may have walked the same distance. Under-reporting among suburban or rural residents could cause more subjects in the suburban area to have been misclassified as not walking five or more blocks, even if the walk distance was sufficient. If present, the misclassification would lead to a bias (lower odds ratios) of the effect size on walking outcome among older women when comparing the suburban or rural built environments to the central city. Additionally, the measure of walking could be double-counted for those who work for leisure purposes since the questionnaire included walking as a type of low-intensity exercise. However, the correlation between number of blocks walked per day and calories burned for low-intensity exercise show a low correlation (r=0.27), suggesting the potential problem of double-counting should not be introducing an error of large magnitude.

Fourth, although the use of metro-wide raster measures created a continuous grid surface at a refined scale that freed the measures from an arbitrary neighborhood definition and minimized potential bias due to the use of coarse group level data, a few assumptions were inherent in the built environment measures used in the current study. The accessibility attributes were measured using Euclidean (straight line) distance; the adoption of this methodology did not necessarily represent actual walking distance. Walking paths can be influenced by the street network and physical barriers such as freeways, rivers, and hills. Thus, Euclidean distance tends to underestimate the actual walking distance. However, Euclidean distance is correlated to travel time on roads,[57] and more complex distance measures such as Manhattan distance also tend to overestimate walk distance. The use of Euclidean distance measurement is the simplest method, and it generates reasonably accurate measures for walking experience in urban areas and small region.[58]

Another limitation is that the measured proximity to the closest amenities of interest, including bus stop, light rail station, commercial area, and park, do not account for other aspects that may influence personal choices. Some of these factors such as frequency of transit service, the number of bus lines serving a transit stop, the type of commercial services, and size of park likely played a role in influencing how far subjects were willing to travel for these amenities. Also, some of the spatial data, such as street network (for calculating street connectivity) and transit stops data, used to create the six built environment attributes were retrofitted by digitizing paper maps or historical information. Definitions for the commercial and park variables also involved decisions that might affect the neighborhood attributes measurement; thus leading to a loss of the true associations between built environment and walking. In the case of the commercial variable, aggregating all types of commercial areas, including both retail and office zones but not commercial parks, as target destinations over-counted the number of commercial areas and likely increased the commercial access measured for some subjects. Although areas such as commercial parks were excluded, some destinations that were purely office spaces where people do not visit for any services or areas zoned as commercial for future development but were inconsistent with actual land use on site might still have been included. Categorization of the park variable also presents a similar measurement problem; the number of park, green spaces, and recreation facilities were under-counted and likely captured a less park accessible environment than ground truth for some subjects. Private open spaces such as those managed by private entity or school grounds were not accessible to everyone at all times. Even though some people might utilize these facilities, they were not included as target park destinations. The population density variable was created using the block group level data. The smoothing of population density at block group level using the Inverse Distance Weighted (IDW) method has some inherent error because a large geographic area was used. Alternative spatial interpolation methods were explored and compared, and IDW was determined to be most appropriate strategy because it could avoid extrapolation or generation of negative measurements.[59] The observed change in population density in the Metro-area for all the interpolation methods were also similar, providing some assurance of validity and reliability. Since no alternative data sources were available, means for verifying the validity of the derived data was limited. The errors in measurement of the built environments might have slight effect on how the clusters were classified, but the impact was judged to be minimal overall. The potential measurement error related to over-counting of commercial areas or under-counting of park areas would be applicable to the entire study area; therefore, the change in magnitude would be relative and should not affect the cluster grouping. The use of Euclidean distance measures might reduce the walk distance measured compared to the ground truth, especially for less compact areas because of the street network differences; this could under-estimate the association between built environment and walking. However, other measurement errors, such as destination preferences, use of block group population measurement, or incomplete information in certain jurisdiction would likely exaggerate the urban-rural differences and over-estimate the true associations between the built environment and walking. It is difficult to determine whether these measurement errors will balance out or not, but the overall effect is assumed to be modest.

Fifth, residual confounding due to inaccuracy of measurements, variables that are poorly measured or modeled, may potentially explain the associations observed. For instance, the socioeconomic status variables were aggregated at Census block group level rather than at individual level; thus, these measurements might not reflect the true status for individual subjects residing in the block. Although interactions were not found between the urban forms/clusters and the block group socioeconomic status, measurements at individual level might provide better control of confounding to the associations between the built environment and walking. Lack of control for unmeasured confounders, due to data availability, may also influence the observed associations. Potential confounders that could not be adjusted in the current analysis include neighborhood safety and crime, sidewalk quality, and topography. The effect of residual confounding on the association between built environment and walking is uncertain.

Sixth, the research design was cross-sectional, and the measurement of exposure and outcome variables were documented at a specific time point; therefore, the findings from this project might not reflect the long-term cumulative effects between the resident neighborhood and walking outcome. In addition, the findings from this project might have limited generalizability; the study was geographically specific to the Portland metro area in 1986, the subjects were all Caucasian women who were 65 years or older, and the study excluded subjects who were not geocoded or resided outside of the UGB at baseline. Thus, the findings might not be applicable to older women living in the Portland Metro-area in more contemporary periods, elsewhere in the U.S., or other countries. The findings might also not be generalizable to younger women, men, or people of other races. Despite the apparent limitations on generalizability in the current study, the findings contribute to a growing literature on the association between the built environment and walking, and future research will eventually establish patterns, by age, gender, race, and location.

Last, the data used for this study were over twenty years old. The individual health data at baseline was most complete and provided the best representation of the population of older women; thus the dataset was adopted for the current study. The findings from the study might not be relevant to what would have been observed in the Portland metro area in contemporary sense because urban processes would have altered the composition of neighborhood factors and the specific clusters that were identified might not reflect the current status. The urban process and development experienced in Portland might also have altered the association between built environment and walking. However, the findings and relative associations may still be applicable to areas in the Portland metro-area that are unchanged or transformed to different urban forms identified. The findings may also provide rich information and value that are applicable to other places with similar urban form settings. This information provide general guidance and direction to changes in built environment that can encourage walking among older women.

Although the current study has methodological limitations, the strength of the association is moderately strong, and is not likely explained by chance. From Figure 3a, there is also an indication of gradient effect of decreasing likelihood for women to walk if they lived in clusters further away from the central city. Additionally, existing literature has shown consistent results across studies, different demographic groups, and different locations. Based on review of the literature, there are few longitudinal studies on the association between built environment and walking; however, findings from the available longitudinal studies suggest there is an association between housing density, availability of transit, and utilitarian walking.[28] This finding has been consistent with the conclusions from several meta-analyses and cross-sectional studies that there is an association between built environment and utilitarian walking while the findings has been less conclusive for leisure/recreational walking purposes.[1,2] It is unlikely that chance variation would explain the consistent demonstration of an association between measures of the built environment and walking in studies using differing methods and taking place in a wide variety of urban settings.

Research on the association between the built environment and walking behavior generates additional questions and hypotheses for further investigation. Findings from this research also provide insights and directions for building a healthy community and improving quality of lives.

Review of findings from the literature pointed to the effects of the built environment factors and the role of built environment in influencing utilitarian walking.[2,3,60] The current study also concludes that accessibility to amenities, density, and other built environments attributes combine to make central city setting more pedestrian friendly compared to other areas. However, the questions that remain provide opportunities for future study.

One of the ideas is to explore current urban forms in the same region or in other places to find out whether there were thresholds for the built environment factors that would encourage older women to walk. Comparing a wider range of built environments may yield additional clusters with differing attribute combinations that might allow for a more in depth comparison of relative impact on older women's walking habits. A replicated study using newly recruited sample and current spatial data could provide results that not only explore the impact of contemporary built environments, but it would also yield more up-to-date information for the public and decision makers to create a pedestrian friendly environment. Furthermore, it is unclear whether older women in satellite cities are more likely to conduct a modest level of walking, given better access to commercial areas. An intervention study may advance the understanding on the role of planning policies or built environment modifications have on walking.

Conclusions

The demonstration of an association between built environments and a combination of utilitarian and recreational walking has been mixed in the scientific literature. Although a variety of methods have been used to define and quantify characteristics of the built environment, challenges remain with the limited availability of spatial information, especially at refined geographic units. In addition, recognition of the complex interaction between neighborhood attributes has prompted many studies to examine the association by using the combined built environment measures. However, few studies have used cluster analysis to classify attributes of the built environment. This study pulls together various objective, high-resolution, and standardized measures to identify neighborhood clusters that portray the walking potential for individuals.

The analysis of the built environment as a predictor for older adults walking behaviors demonstrated that the built environment characteristics of the central city are associated, with increased utilitarian and leisure walking among older women relative to metropolitan areas classified as the city periphery, the suburb, or the urban fringe with poor commercial area access. However, areas classified as satellite city, or urban fringe with poor park area access were not associated with reduced levels of walking for older women. Although additional studies are needed to verify this finding, there is some indication that suburban and rural communities have the potential to create built environments to support active living in older women.

This project builds upon theory and scientific literature related to the urban built environment and physical activity via the use of methods that quantify built environment characteristics without committing to a predefined neighborhood boundary, and empirically use of built environment measures in combination to create a summary characterization of neighborhood type. The use of standardized small geographic units to measure built environment characteristics, and cluster analysis to classify neighborhoods has been used in a limited number of published investigations, and the successful application of these methods suggests potential utility for future studies. The findings also have implications for the improvement of public health, potentially informing and guiding urban planning, land use policies, and implementation of programs intended to promote active living. The findings describe broad classes of built environment types, whose features appear to work together to promote walking among elderly women. The regionally specific information from this analysis describes walking behavior for varying neighborhood physical settings. These findings can act as an informative tool that encourages discussions among decision makers and citizens during planning processes, and aid the Portland Metro-area communities in identifying opportunities for changes in the built environment that improve physical activity and livability. The long term benefits of shifting urban form to be more pedestrian-friendly are critically important with the aging population; changes in built environment would promote higher levels of physical activity among older adults and likely among people of other age groups. This change, in turn, may lower the risk of diseases related to the lack of physical activity in the population.

References

1. Owen N, Cerin E, Leslie E, et al. Neighborhood walkability and the walking behavior of Australian adults. *Amercian Journal of Preventive Medicine*. 2007; 33 (5): 387-395.

2. Saelens BE, Handy SL. Built environment correlates of walking: a review. *Medicine & Science in Sports Exercise*. 2008; 40 (7 Suppl): 550-566.

3. Lin J, Long L. What neighborhood are you in? Empirical findings of relationships between household travel and neighborhood characteristics. *Transportation.* 2008; 35: 739-758.

4. Huang L, Stinchcomb DG, Pickle LW, Dill J, Berrigan D. Identifying clusters of active transportation using spatial scan statistics*. American Journal of Preventive Medicine*. 2009; 37(2): 157-166.

5. Ewing R, Cervero R. Travel and the built environment: a meta-analysis. *Journal of the American Planning Association.* 2010; 76(3): 265-294.

6. Brownson RC, Boehmer TK, Luke DA. Declining rate of physical activity in the United States: What are the contributors? *Annual Review of Public Health*. 2005; 26: 421-443.

7. Brownson RC, Remington PL, Davis JR. *Chronic Disease Epidemiology and Control*. 2nd Ed. American Public Health Association: Washington D.C.; 1998: 191-213.

8. Saelens BE, Papadopoulos C. The importance of the built environment in older adults' physical activity: a review of the literature. *Washington State Journal of Public Health Practice*. 2008; 1**:** 13–21.

9. World Health Organization. *Global strategy on diet, physical activity and health*. 2004. Available at: http://www.who.int/dietphysicalactivity/strategy/eb11344/strategy_english_web.pdf

10. Mathers C, Stevens G, Mascarenhas M. Global health risks: mortality and burden of disease attributable to selected major risks. *Department of Health Statistics and Informatics, World Health Organization*. 2009. Available at: http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_full.pdf

11. Colditz GA. Economic costs of obesity and inactivity*. Medicine & Science in Sports Exercise.* 2009; 31 (11 Suppl): 663-667.

12. Prohaska T, Belansky E, Belza B, et al. Physical Activity, Public Health, and Aging: Critical Issues and Research Priorities. *Journal of Gerontology*. 2006; 61B(5): 267-273.

13. Handy SL, Boarnet MG, Ewing R, Killingsworth RE. How the built environment affects physical activity: views from urban planning. *American Journal of Preventive Medicine*. 2002; 23 (2S): 64-73.

14. Neal MB., Chapman N, Dill J, et al. Age-related shifts in housing and transportation demand. *College of Urban and Public Affairs, Portland State University.* 2006.

15. Centers for Disease Control and Prevention. *U.S. physical activity statistics: Behavioral Risk Factor Surveillance System.*2010. Available at: http://www.cdc.gov/nccdphp/dnpa/physical/stats/index.htm

16. Bungum TJ, Thompson-Robinson M. Modes and Locales of Physical Activity: Older Adults. *Californian Journal of Health Promotion*. 2008; 1: 40-48.

17. Forsyth A, Michael OJ, Lee B, Schmitz KH. The built environment, walking, and physical activity: Is the environment more important to some people than others? *Transportation Research: Part D.* 2009; 14(1): 42-49.

18. Nagel CL, Carlson NE, Bosworth M, Michael YL. The relation between neighborhood built enivornment and walking activity among older adults. *American Journal of Epidemiology*. 2008; 168: 461-468.

21. Riva M, Gauvin L, Apparicio P, Brodeur JM. Disentangling the relative influence of built and socioeconomic environments on walking: the contribution of areas homogenous along exposures of interest. *Social Science & Medicine*. 2009; 69: 1296-1305.

20. Rodríguez DA, Evenson KR, Diez Roux AV, Brines SJ. Land Use, Residential Density, and Walking: The Multi-Ethnic Study of Atherosclerosis. *American Journal of Preventive Medicine*. 2009; 37(5): 397-404.

21. Shigematsu R, Sallis JF, Conway TL, et al. Age differences in the relation of perceived neighborhood environment to walking. *Medicine & Science in Sports & Exercise*. 2009; 41(2): 314-321.

22. King WC, Belle SH, Brach JS, Simkin-Silverman LR, Soska T, Kriska AM. Objective Measures of Neighborhood Environment and Physical Activity in Older Women. *American Journal of Preventive Medicine*. 2005; 28(5): 461-469.

23. Mujahid MS, Diez Roux AV, Shen M, et al. Relation between Neighborhood Environments and Obesity in the Multi-Ethnic Study of Atherosclerosis. *American Journal of Epidemiology*. 2008; 167(11): 1349-1357.

24. Badland HM, Schofield GM, Witten K, et al. Understanding the Relationship between Activity and Neighbourhoods (URBAN) Study: research design and methodology. *BioMed Central Public Health*. 2009; 9: 224-234. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2716337/

25. Cerin E, Leslie E, du Toit L, Owen N, Frank LD. Destinations that matter: associations with walking for transport. *Health & Place*. 2007; 13: 713-724.

26. Boarnet MG, Greenwald M, McMillan TE. Walking, urban design, and health: towards a cost-benefit analysis framework. *Journal of Planning Education and Research*. 2008; 27: 341-358.

27. Boer R, Zheng Y, Overton A, Ridgeway GK, Cohen DA. Neighborhood design and walking trips in ten US metropolitan areas. *American Journal of Preventive Medicine*. 2007; 32(4): 298-304.

28. Coogan PF, White LF, Adler TJ, Hathaway KM, Palmer JR, Rosenberg L. Prospective study of urban form and physical activity in the black women's health study. *American Journal of Epidemiology*. 2009; 170 (9): 1105-1117.

29. Gauvin L, Riva M, Barnett T, et al. Association between neighborhood active living potential and walking. *American Journal of Epidemiology*. 2008; 167(8): 944-953.

30. Frank LD, Sallis JF, Conway TL, Chapman JE, Saelens BE, Bachman W. Many pathways from land use to health. *Journal of the American Planning Association*. 2006; 72 (1): 75-87.

31. Li F, Fisher KJ, Brownson RC, Bosworth M. Multilevel modelling of built environment characteristics related to neighbourhood walking activity in older adults. *Journal of Epidemiology Community Health*. 2005; 59: 558-564.

32. McCormack GR, Giles-Corti B, Bulsara M. The relationship between destination proximity, destination mix and physical activity behaviors. *Preventive Medicine*. 2008; 46: 33-40.

33. Frank LD, Sallis JF, Saelens BE, et al. The development of a walkability index: application to the neighborhood quality of life study. *British Journal of Sports Medicine*. 2009. Doi: 10.1136/bjsm.2009.058701.

34. Leslie E, Coffee N, Frank L, Owen N, Bauman A, Hugo G. Walkability of local communities: using geographic information systems to objectively assess relevant environmental attributes. *Health & Place*. 2007; 13: 111-122.

35. Frank LD, Schmid TL, Sallis JF, Chapman J, Saelens BE. Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *American Journal of Preventive Medicine*. 2005; 28(2S2): 117-125.

36. Lovasi GS, Moudon AV, Pearson AL, et al. Using built environment characteristics to predict walking for exercise. *International Journal of Health Geographics*. 2008; 7: 1-13.

37. Berke EM, Koepsell TD, Moudon AV, Hoskins RE, Larson EB. Association of the Built Environment With Physical Activity and Obesity in Older Persons. *American Journal of Public Health*. 2007; 97(3): 486-492.

38. Li F, Harmer P, Cardinal BJ, Vongjaturapat N. Built environment and changes in blood pressure in middle aged and older adults. *Preventive Medicine*. 2009; 48: 237-241.

39. Diez-Roux AV, Nieto J, Muntaner C, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. *American Journal of Epidemiology*. 1997; 146(1): 48-63.

40. Rodríguez DA, Young HM, Schneider R. An easy to compute index for identifying built environment that support walking. Paper presented at: Transportation Research Board meeting; 2006.

41. Riva M, Apparicio P, Gauvin L, Brodeur JM. Establishing the soundness of administrative spatial units for operationalising the active living potential of residential environments: an exemplar for

designing optimal zones. *International Journal of Health Geographics*. 2008; 7(43): doi:10.1186/1476-072X-7-43.

42. McNally MG, Kulkarni A. An assessment of the interaction of the land-use transportation system and travel behavior. *Institute of Transportation Studies, University of California, Irvine*. 1996.

43. Levine J, Inam A, Tomg GW. A choice-based rationale for land use and transportation alternatives: evidence from Boston and Atlanta. *Journal of Planning Education and Research*. 2005; 24: 317-330.

44. Cummings SR, Black DM, Nevitt MC, et al. Appendicular bone density and age predict hit fracture in women. *Journal of the American Medical Association*. 1990; 263(5): 665-668.

45. Siu V. Reliability and validity of spatially derived neighborhood design measures for the Portland metro area, 1988-2006. Internship Report for Oregon Health & Science University Master of Public Health degree. 2009.

46. Diehr P, Hirsch C. Health benefits of increased walking for sedentary, generally healthy older adults: using longitudinal data to approximate an intervention trial. *Journal of Gerontology: Biological Sciences Medical Sciences*. 2010; 65A(9): 982-989.

47. Calinski T, Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics- Theory and Methods*. 1974; 3(1): 1-27.

48. Sallis JF, Saelens BE, Frank LD, et al. Neighborhood built environment and income: examining multiple health outcomes. *Social Science & Medicine*. 2009; 68: 1285-1293.

49. Yen IH, Michael YL, Perdue L. Neighborhood environment in studies of health of older adults: a systematic review. *American Journal of Preventive Medicine*. 2009; 37(5): 455-463.

50. Schlossberg M, Agrawal AW, Irvin K, Bekkouche VL. How far, by which route, and why? A spatial analysis of pedestrian preference. *Mineta Transportation Institute, San Jose State University*. 2007; Report 06-06.

51. Wiseman RF, Roseman CC. A typology of elderly migration based on the decision making process. *Economic Geography*. 1979; 55(4): 324-337.

52. Sergeant JF, Ekerdt DJ, Chapin R. Measurement of late-life residential relocation: why are rates for such a manifest event so varied? *Journal of Gerontology: Social Sciences*. 2008; 63B(2): 92-98.

53. Lin G. Elderly migration: household versus individual approaches. *The Journal of the Regional Science Association International*. 1997; 76(3): 285-300.

54. Frank LD, Saelens BE, Powell KE, Chapman JE. Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? *Social Science & Medicine*. 2007; 65: 1898-1914.

55. Handy S, Cao X, Mokhtarian PL. Self-Selection in the Relationship between the Built Environment and Walking. *Journal of the American Planning Association*. 2006; 72(1): 55-74.

56. Wiseman RF, Virden M. Spatial and social dimensions of intraurban elderly migration*. Economic Geography*. 1977; 53(1): 1-13.

57. Phibbs CS, Luft HS. Correlation of Travel Time on Roads versus Straight Line Distance. *Medical Care Research and Review*. 1995; 52: 532-42.

58. O'Sullivan D, Unwin DJ. *Geographic Information Analysis*. Hoboken, New Jersey: John Wiley & Sons, Inc; 2003.

59. Childs C. Interpolating Surfaces in ArcGIS Spatial Analyst. *ArcUser*. 2004; July-September: 32-35.

60. Lee C, Moudon AV. Neighborhood design and physical activity*. Building Research & Information.* 2008; 36(5): 395-411.

Appendix A.
Methodology for Developing the Portland Metro-wide Neighborhood Measures

In the Neighborhood and Obesity study, six objective neighborhood design measures were created to characterize the built-environment across six time points. These measures were developed using ArcGIS 9.3 (ESRI, Redland, CA) and existing administrative Geographic Information System (GIS) data from the Metro Regional Land Information System (RLIS).

These neighborhood design measures were created in a raster or grid environment with grid size set at 264 feet by 264 feet (approximately one city block distance or one minute walk time), yielding a total of 159839 grid cells in the Portland metro area. The advantage of creating the neighborhood measure in the raster environment is its ability to generate a large amount of data for the Metro area quickly. However, compared to the direct subject measurement, the raster measurement causes a reduction in measurement accuracy, as the usage of grid cells generalized the spatial environment and the associated information. Therefore, a small cell size was adopted to minimize this problem. Developing Portland Metro-wide measure allows for easy linkage of the localized individual level built-environment exposure data to the subjects who have moved, to newly recruited subjects, or to other research projects' cohort; thus this approach to create neighborhood measures can facilitate research efforts.

Neighborhood Design Measures

Objective measures were created for the six neighborhood design aspects: accessibility to bus, accessibility to light rail, accessibility to commercial area, accessibility to park, intersection density and population density. For each measure, data were generated for six time points, 1988, 1990, 1994, 1998, 2002, and 2006, resulting in a total of thirty-six variables.

Additional criteria were applied to the raw data from Metro to help better define and standardize the indicators. The available data available contained additional information that were either not of interest to the researchers or experienced refinement in information categorization at recent time points; hence, regrouping or exclusion of certain spatial information would make the data consistent for analysis. The detailed neighborhood measure development processes for each variable are documented as follow.

Accessibility Measures

Accessibility to Bus Service

Accessibility to bus service was measured as the straight-line (Euclidean) distance from each grid cell centroid to the nearest bus stop. A shorter distance represents greater accessibility. Since the transit data was not available from the RLIS for time points 1988, 1990, and 1994, archive data obtained from Trimet for 1988 were digitized and used to represent bus services available for 1988, 1990, and 1994, assuming changes in bus stops were minimal.

The bus stop point data did not require further data cleaning; therefore, distance analysis was conducted to create Metro-wide raster data for each time point. Spatial Analyst, a GIS extension tool, was used to calculate the Euclidean distance from each grid cell centroid to its closest bus stop. The measure was expressed in feet and rounded to the nearest integer using raster calculator.

A few limitations remain with this measure. Despite the ease of implementation, the Euclidean measure does not necessarily represent actual walking distance, as walking paths can be influenced by the streets network and physical barriers such as freeways, rivers, and hills. Thus, Euclidean distance tends to underestimate the actual walking distance. However, there are literatures suggesting Euclidean distance is correlated to travel time on roads.[57] Another limitation with this measure is that it only measures the proximity to the closest bus stop but does not account for the frequency of transit service or the number of bus lines serving a bus stop. These factors likely affect choices of transportation mode and bus service stop.

Accessibility to Light Rail Station

Accessibility to light rail station was measured as the straight-line (Euclidean) distance from each grid cell centroid to the nearest light rail station. A shorter distance represents greater accessibility to light rail station. Since transit data was not available from RLIS for time points 1988, 1990, and 1994, archive data obtained from Trimet for 1988 were digitized and used to represent light rail services available for 1988, 1990, and 1994, assuming changes in light rail stop were minimal.

The raw spatial data were examined, and additional information such as future stops and proposed stops were excluded. Such non-existing light rail stations for 2002 and 2006 have been removed because the station status information was not available, and no spatial points for future or proposed stops existed at previous time points. This process ensured the spatial data used for analysis only include existing light rail stops for each time point. Spatial Analyst, a GIS extension tool, was used to calculate the Euclidean distances from each grid cell centroid to its closest light rail stop. The measure was expressed in feet and rounded to the nearest integer using raster calculator.

Similar to the accessibility to bus service measurements, the accessibility to light rail station measurements were prone to limitations related to Euclidean distance measure and frequency of light rail services.

Accessibility to Commercial Area

Accessibility to commercial area was measured as the straight-line (Euclidean) distance from each grid cell centroid to the closest commercial zone. The shorter the distance represents greater accessibility to the commercial zone. Zoning data available from comprehensive plans (1990 and 1993) and RLIS (1998, 2002, and 2006) provide extractable information on areas designated for commercial purposes. Zoning data for 1990 also represented the zoning patterns for both 1988 and 1990, while zoning data from 1993 was used as the proxy for 1994 zoning pattern.

Data for all available years were examined for consistency, and only commercially zoned and mixed use areas were included. Since the 1990 and 1993 comprehensive plan data are inconsistent with the zoning data for later years, a zoning code equivalency list was established based on selected areas from the 1998 zoning data to determine areas to be included as commercial zone from the 1990 and 1993 data. For each time point, all spatial data with a general zoning class of “COM-Commercial” were included as commercial area except for those with a zone class of “CI- Campus Industrial” or “INST-Institutional”. With new mixed use categories included in the RLIS data for recent years, the commercial area selection criteria expanded to include data with general zoning class of “MUC- Mixed use commercial” in 2002 and “MUE- Mixed use employment” and “MUR- Mixed use residential” in 2006.

In preparing the spatial data for analysis, the selected commercial spatial data were rasterized and converted into points. Centroids were created to represent commercially zoned areas that had tiny spatial areas or had elongated shapes that could not be rasterized. These small areas centroids were appended to the rasterized points to prepare the data set for the distance analysis. Spatial Analyst was used to calculate the Euclidean distances between each grid cell centroid and the closest commercial zone. The measure was expressed in feet and rounded to the nearest integer using raster calculator.

The accessibility to commercial area measurements were prone to limitations related to Euclidean distance measure.

Accessibility to Park

Accessibility to park was measured as the straight-line (Euclidean) distance from each grid cell centroid to the closest park area. A shorter the distance represents greater accessibility to park. Park data for 1990 also represented park distribution in 1988, and the 1995-96 data were used as proxy for 1994. Since park data from RLIS experienced significant improvement over time and expanded on the type of park categories, a definition of park is necessary to refine and standardize the data. Generally, the number of park categories increased over time, some of the categories available included: park, trail, open space, natural area, private common area, country club, recreational center, pool, tennis court, golf course, school grounds, fairground, community gardens, and cemetery.

Only shapes that were categorized as "park", "open space", "greenway", "trail", or "natural area" were considered for the analysis, because these are areas accessible to public. In addition, filters on park size were adopted to exclude park spaces of unreasonable sizes. Park areas smaller than 660 feet, the size of the smallest park (Mills End Park) in the City of Portland, were excluded from the analysis. Conversely, forest park covers an extensive area and has few access points; shapes of access points were created to represent the access instead of using the entire park area for measurement.

Selected park spatial data were rasterized and converted into points. Parks that were too small or with elongated shapes that could not be rasterized were identified; centroids were then created to represent these areas. These small areas centroids were appended to the rasterized points so the data set became ready for the distance analysis. Spatial Analyst was used to calculate the Euclidean distances between each grid cell centroid and the closest park area. The measure was expressed in feet and rounded to the nearest integer using raster calculator.

The accessibility to park measurements were prone to limitations related to Euclidean distance measure. Some people may utilize some of the excluded private and recreational areas rather than the closest park space.

Density Measures

Intersection Density

Intersection density was measured as the number of street intersections in a quarter mile buffer divided by area. A greater number represents higher intersection density. This variable was derived from RLIS streets files. The 1988 data also represented the 1990 intersection density pattern.

For the purpose of measuring intersection density relevant to walking, two exclusion criteria have been applied to the raw data. First, streets from Clark County, Washington available from Metro's RLIS streets files were removed, because these streets are located beyond the study area and may affect the analysis. This is achieved by using "selection by location" tool and exporting streets within the Clackamas, Multnomah, and Washington County. Second, information on freeway and freeway ramps from RLIS streets files were taken out, since freeway intersections were not accessible to pedestrians. These records were denoted by street type codes beginning with "11*" and were "less than 1200", the "select by attribute" query was used to identify and exclude these spatial data.

"Intersect" tool in GIS was used to create street intersection points from the freeways-removed tri-county area street reference; this yielded an intersection point for each node (e.g. when two lines intersect, there will be 4 points created instead of 1). The intersection points were assigned X, Y coordinates and then grouped together as single intersection using the "collect event" tool in the spatial statistics toolbox. Finally, density tool in Spatial Analyst was used with a search radius of a quarter mile (1320 feet) to yield the intersection kernel density (# of intersections/ Area in sq. ft.) for each cell's quarter mile buffer. The measure was expressed in feet and rounded to the nearest integer using raster calculator.

The Kernel density method was used instead of a point density method, because Kernel density can provide a smooth data surface. The Kernel density method also puts the greatest weight on the value for the cell of interest with a dissipating weight as the distance increases. A limitation to the density variable was that cell measurements near the edge of the study area may be underestimated if the cells are also located near the edge of Clackamas, Multnomah, or Washington counties. The edge effect is due to the lack of spatial information outside of the tri-county area, and the kernel density search was unable to account for streets intersection outside of the county boundary. This should not be a serious problem for the cells in the study area, because streets data of the entire Clackamas, Multnomah, or Washington counties were used in the spatial analysis, and the urban growth boundary is well within a quarter mile from any of the three counties boundary. There should be sufficient buffer area for the kernel density analysis to calculate valid measures.

Population Density

Population density was measured as the number of persons in a square mile buffer divided by area. A greater number represents higher population density. This variable was derived from the Clackamas, Multnomah, and Washington County population data by Census block group. The data were available from the ESRI Business Analyst for 1990, 2000, and 2008. There are some discrepancies in the tri-county total population between block group data from ESRI's Business Analyst and US Census Bureau for 1990 and 2000. However, since 2008 block group level population data was not available from the US Census Bureau, the ESRI Business Analyst dataset was adopted for this analysis.

Block group level population data were expressed as block group centroids through running the "feature to point" tool. Then, density tool in Spatial Analyst was used with a search radius of a square mile (5280 feet) to yield the kernel population density (# of persons/ Area in sq. mile) for each cell within the tri-county area. The spatial analysis was conducted for year 1990, 2000, and 2008. The 1990 derived raster data represented the population density for 1988, and population data for 1994, 1998, 2002, and 2006 were interpolated from the available years using the following formula and raster calculator in GIS.

Population for the Target Year
= initial population + (average annual change in population * years elapsed between the initial year and the target year)

$$P_1 + [\frac{(P_2 - P_1)}{(Y_2 - Y_1)} * (Y_T - Y_1)]$$

where
P_1 = Population at first time point with available data,
P_2 = Population at second time point with available data,
Y_1 = Year for the first time point
Y_2 = Year for the second time point
Y_T = Year for the target time point

For example, Population in 1994 would be calculated as follow:

$$P_{1994} = P_{1990} + [\frac{(P_{2000} - P_{1990})}{10} * 4]$$

A one mile search radius was used for creating the population density variable, because the block centroids are more sparsely distributed than the intersection points. A larger search radius allowed for the inclusion of greater amount of data in the analysis to create more generalized measures. Since values from cells located farther from the core carry less weight, the extra information should not affect the measures substantially. Similar to the intersection density variable, population density cell measurements used the kernel density method and were potentially affected by the edge effect, underestimating the density near the study area boundary.

Organization of Derived Built Environment Measures

The derived neighborhood design measures for each of the six variables and the six time points were appended in a table associated with the grid centroids located within the Portland Metro area. The grid lines and grid centroids were created using the “create fishnet” function in the GIS-Data Management toolbox. The grid centroids were then clipped to the study area (urban growth boundary). Finally, the raster data from each of the 36 variables were added to the table using a combination of GIS tools such as “extract values to points” tool from the spatial analyst toolbox, “add field”, “calculate field”, “delete field” tools from the data management toolbox, and model builder. The data for each neighborhood design measure within the Portland Metro area were summarized in Table A.

Table A. Summary of Built Environments Attributes within the Portland Metro Area, 1988-2006

Built Environment Attributes	1988		1990		1994		1998		2002		2006	
	Median	Inter-Quartile Range	Median	Inter-Quartile Range	Median	Inter-Quartile Range	Median	Inter-Quartile Range	Median	Inter-Quartile Range	Median	Inter-Quartile Range
Distance to bus stop (m)	509	180 - 1765	509	4513 - 14587	509	255 - 1249	360	290 - 1031	332	8 - 60	341	247 - 1509
Distance to light rail station (m)	9337	180 - 1765	9337	4513 - 14587	9337	255 - 1249	5363	290 - 1031	4469	8 - 60	4432	247 - 1509
Distance to commercial area (m)	586	180 - 1765	586	4513 - 14587	586	255 - 1226	515	255 - 885	469	7 - 58	402	331 - 1625
Distance to park (m)	569	180 - 867	569	2591 - 9847	515	241 - 1024	402	180 - 649	360	10 - 63	332	381 - 1773
Intersection density (# intersections/ km^2)	29	161 - 796	29	2052 - 9048	29	180 - 938	35	180 - 628	41	13 - 68	44	423 - 1902
Population density (persons/ km^2)	834	161 - 805	834	2037 - 8950	983	161 - 840	1120	180 - 580	1217	14 - 70	1291	456 - 2015

Appendix B

Table B. Comparison of the Relative Impact of Walk Habit Outcome between Urban Form Clusters

OR (95% CI)	Reference Group					
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
Cluster 1: Central city	1.00	1.45 (1.17, 1.81)*	1.45 (1.05, 2.01)*	2.50 (1.14, 5.48)*	1.02 (0.54, 1.91)	1.90 (0.58, 6.19)
Cluster 2: City periphery	0.69 (0.55, 0.85)*	1.00	1.00 (0.73, 1.37)	1.72 (0.78, 3.76)	0.70 (0.37, 1.32)	1.31 (0.40, 4.26)
Cluster 3: Suburb	0.69 (0.50, 0.95)*	1.00 (0.73, 1.38)	1.00	1.72 (0.76, 3.89)	0.70 (0.36, 1.38)	1.31 (0.39, 4.37)
Cluster 4: Urban fringe with poor commercial area access	0.40 (0.18, 0.88)*	0.58 (0.27, 1.28)	0.58 (0.26, 1.32)	1.00	0.41 (0.15, 1.10)	0.76 (0.19, 3.10)
Cluster 5: Urban fringe with poor park area access	0.98 (0.52, 1.84)	1.43 (0.76, 2.69)	1.43 (0.72, 2.81)	2.45 (0.91, 6.58)	1.00	1.87 (0.50, 6.99)
Cluster 6: Satellite city	0.53 (0.16, 1.71)	0.76 (0.23, 2.49)	0.76 (0.23, 2.55)	1.31 (0.32, 5.34)	0.54 (0.14, 2.01)	1.00

** significant at 0.05 level*