Aggregation and spatial analysis of walking activity in an urban area: results from the Halifax space-time activity survey

K Neatt[1], H Millward[1] and J Spinney[1]
[1]Saint Mary's University, Halifax, Canada

E-mail: kevin.neatt@smu.ca

Abstract. This study examines neighborhood characteristics affecting the incidence of walking trips in urban and suburban areas of Halifax, Canada. We employ data from the Space-Time Activity Research (STAR) survey, conducted in 2007-8. Primary respondents completed a two-day time-diary survey, and their movements were tracked using a GPS data logger. Primary respondents logged a total of 5,005 walking trips, specified by 781,205 individual GPS points. Redundant and erroneous points, such as those with zero or excessive speed, were removed. Data points were then imported into ArcGIS, converted from points to linear features, visually inspected for data quality, and cleaned appropriately. From mapped walking tracks we developed hypotheses regarding variations in walking density. To test these, walking distances were aggregated by census tracts (CTs), and expressed as walking densities (per resident, per metre of road, and per developed area). We employed multivariate regression to examine which neighborhood (CT) variables are most useful as estimators of walking densities. Contrary to much of the planning literature, built-environment measures of road connectivity and dwelling density were found to have little estimating power. Office and institutional land uses are more useful estimators, as are the income and age characteristics of the resident population.

1. Introduction

The promotion of walking is an important goal for both health professionals and urban planners, since walking is an environmentally-friendly and low-cost alternative to motorized travel in urban areas, and an important form of healthy physical activity (Sallis et al., 2004). Walking research is extensive, but has suffered from a lack of objective data on walking behavior, typically relying on recall questionnaires using subjective categories of walking frequencies. In particular, little information is available on the spatial location of walking activity, or the factors which shape the aggregate geography of walking behavior.

This paper addresses this gap in our knowledge. We report and assess the spatial location of walking, employing data from the innovative Halifax STAR (Space-Time Activity Research) time-use and transport survey, which tracked respondent walking behaviour using a wearable GPS-enabled (global positioning system) data device. Our focus in the current paper is on the methods employed to derive the aggregate pattern of walking tracks, for all walking purposes, in the medium-sized Canadian metropolitan area of Halifax.

2. Relevant Literature

Considerable research has focused on relationships between walking activity, health, and the built environment, typically using self-reported quantities of walking. These studies inform and support policies aimed at increasing the mode share of walking, and hypothesize that walking behaviors are significantly affected by the ‘‘walkability’’ of neighborhoods focused on respondents’ homes, and make an implicit assumption that walking largely or exclusively occurs within such neighborhood areas. Early papers on this topic were provided by Atash (1994), Crane (2000), Frank and Engelke (2001), Handy et al. (2002), Moudon and Lee (2003), and Saelens et al. (2003). Many empirical studies have been conducted, and several meta-studies are now available (Ewing and Cervero, 2010; Gebel et al., 2007; Renalds et al., 2010; Sallis, 2009; Saelens and Handy, 2008).

A large proportion of the studies use walkability measures of residential density, street connectivity, and land-use mix derived from an early study by Boarnet and Sarmiento (1998), but incorporated into a single ‘walkability index’ by Frank et al. (2005). Though easy to use and understand, the sub-indexes are inter-correlated, so that their effects are not simply additive, and their separate contributions are not reported. An alternative approach, taken by Lee and Moudon (2006), employs multiple regression to isolate the separate effects of a host of land-use and urban-form measures (and personal characteristics too). Their ‘shotgun’ approach is more statistically rigorous, but far less opaque, and difficult to replicate.

A major problem with walking studies in general has been the quality and reliability of data on walking behaviors. Most studies have relied on subjective recall questionnaires, which gauge walking activity by a small number of categories, rather than by exact number or length of walking episodes. These data are subject to both recall bias and social-desirability bias (Podsakoff et al., 2003; van der Ploeg et al., 2010). A few walking studies have employed time diaries, which are more accurate than recall questionnaires (e.g. Forsyth et al., 2007; Frank et al., 2008b, 2010), and several recent studies have employed a time-diary in combination with accelerometers and/or GPS tracking to measure distances/durations (Forsyth et al., 2007, 2008, Dewulf et al., 2012). Almost all walking studies, however, focus on correlating respondent and neighborhood characteristics with a single measure of respondent walking activity (often a category rather than scale measure).

Following the early work by Boarnet and Sarmiento (1998) and Frank et al. (2005), the most common research approach has been to treat individuals as cases, their subjective recall of walking frequency as the dependent variable, and neighbourhood characteristics as independent variables. Personal characteristics such as age, sex, car ownership, etc. are usually included as control variables. Our study employs more objective, GPS-verified, data on the location of all walking trips, not just home-based ones. We also take a very different approach to analysis, in that our cases are the CT neighbourhoods themselves, and the dependent variables are aggregate measures of walking density.

In a recent study the authors employed GPS-verified time-diary data from the Halifax STAR project to provide a comparison of AT-walking versus leisure walking (Spinney et al., 2012). A second study (Millward et al., 2013) analyzes the frequency and length of walking trips, categorized by origins, purposes, and destinations, and also investigates distance-decay functions for major destinations. The current study builds on this earlier work by focusing on the mapped tracks of walking trips, and how those tracks can be aggregated by neighborhood areas.

3. Study Area and Data Sources

Walking data were collected as part of the Halifax Space-Time Activity Research (STAR) Project, a joint project between Saint Mary’s University and Halifax Regional Municipality in 2007 - 2008. The Halifax STAR project is a unique large-scale survey that collected information from households regarding both travel and time-use activity. Halifax is a medium-sized (c.400,000) metropolitan area in eastern Canada, and the survey area included all urban, suburban, and commuter-belt districts.

GPS and computer-assisted telephone interview (CATI) software were integrated in order to provide the interviewer with a detailed listing and map display of the location and timing of respondents’ out-of-home activities (Spinney and Millward, 2011). The GPS data loggers (HP iPAQ) recorded positions at a resolution of three recordings every two seconds and had an accuracy of sub-ten meters, with many positions having accuracy of within three meters. This enabled a prompted-recall approach for collecting accurate and precise information about the location and timing of respondents’ activity patterns. The survey provided fully geo-referenced 2-day (i.e. 48-h) time diary data from 1,971 randomly selected primary respondents aged 15 years or over. Attempts to stratify for proportional distributions for age, sex, and geography were unsuccessful for younger age groups and rural households, but the diary data approximate a proportional distribution for sex, days of the week, and the four seasons.

Of the 1,971 primary respondents, 1,189 recorded one or more walking episodes (trips). The locations of the 5,005 trips were fully geo-referenced. Although the STAR data includes information on trip purpose, the present study does not distinguish between active transportation and recreation walking purposes. It was felt that neighbourhood design ought to encourage walking for both purposes and that within the confines of this research, it was more appropriate to investigate all walking behaviour rather than focusing on the specific purpose.

Built environment data, including census tract boundaries, waterbodies, and roads, were obtained from Statistics Canada in a shapefile format. Darren Scott of McMaster University provided the following built environment variables: intersection density, areas for each of six land-use categories (retail, commercial, industrial, office, institutional, and parkland) and retail lot coverage ratio. These variables are explained further in the independent variables section.

Three census tract-level socio-demographic variables from the 2006 Statistics Canada census were also used in this study: population, numbers of residents by age cohorts, and household income.

4. Cleaning of Walking Track Data

Data cleaning or "weeding" refers to the removal of unwarranted or inaccurate information in the dataset. The raw walking data file, consisting of 5,005 walking trips and 781,205 individual GPS points (cases) was imported into IBM SPSS Statistics version 21. Weeding was a four step process to remove redundant or inaccurate points, as follows.

1. Delete all points that had reception from fewer than six satellites (2,214 points removed, 778,991 remaining).
2. Eliminate all GPS points that had a horizontal dilution of precision (HDOP) of greater than or equal to eight. HDOP is one quantifiable representation of GPS accuracy: a high HDOP value equals lower GPS accuracy (Wagner and Mueller, 2011). A figure of less than eight was selected as an appropriate cut-off value based on surveying industry standards (47,968 points removed, 731,023 remaining).
3. Select all points with speeds above those normally associated with either walking or running. A threshold of 14 km/h was selected for this, since people are not able to run above that speed (7,613 points removed, 723,410 remaining).
4. Remove all points with a minimum speed of zero km/h. In these cases the respondent was stationary between the recordings, and the readings were redundant (621,506 points removed, 159,699 points remaining).

Additional data cleaning was conducted in ArcGIS version 10. A .csv output file was created in SPSS and imported into ArcGIS, enabling the data points to be mapped by UTM coordinates. The points were then converted into continuous lines based on the unique walking event ID. Walking track lines were then broken at each census tract boundary to facilitate the future aggregation of total walking distance per census tract. A manual, judgement-based editing process was then performed to improve accuracy. Select examples of manual editing procedures are provided here for reference.

Many interpolated walking tracks were not logical and therefore deleted. For example, a walking track might cross a residential housing block on a diagonal, and there may be insufficient data to accurately realign the track along the street network. In such a case the entire walking track would be deleted. Other cases of illogical tracks showed the track crossing lakes or inlets of the ocean. Most such cases were deleted, although a few of them were found (from air-photo inspection) to be possible routes that followed causeways or low-tide trails, and were retained.

Several walking tracks were edited rather than deleted. Figure 1 provides an example in which there were sufficient GPS points to accurately depict the walking track at the beginning and the end of the journey, but insufficient data within the middle section. A manual procedure was performed to create additional nodes and align the walking track with the road system. The result represents a logical walking route that is supported with GPS accuracy at each end of the route.

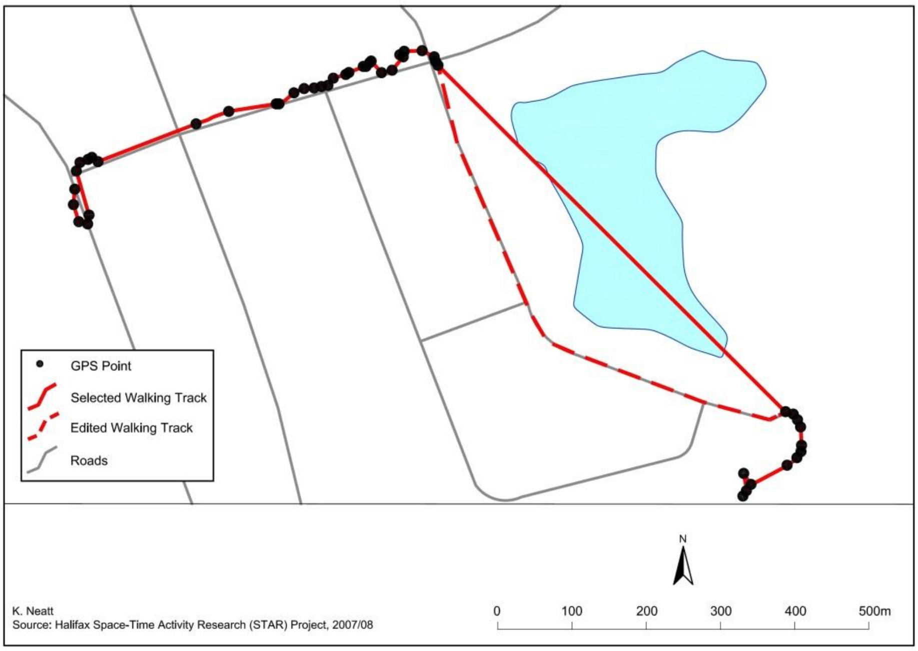

Figure 1. Example of an edited walking track error

5. Neighborhood Walking Densities

A single ArcGIS file was created that contained the walking tracks, roads, water bodies, and census tracts. Figure 2 shows all edited tracks, and allows some generalizations regarding the most walked-in areas. Such areas are mostly in the inner city, surrounding the central business district (CBD) in the south-east portion of the Halifax peninsula. This dense bundle of walking trips is understandable, since the CBD has a high concentration of employment and a mix of offices, retailing, and apartment buildings. In the southern portion of the peninsula, walking is associated with other areas of high employment associated with hospitals and universities. There is also a small cluster of walking in the CBD of Dartmouth, which lies across the harbour from Halifax. In the suburban areas of both Halifax and Dartmouth walking tends to be concentrated along commercial transport corridors, and also within several major parks (e.g. Point Pleasant Park, at the southern tip of the Halifax peninsula).

To more objectively relate the incidence of walking to neighbourhood characteristics, each walking track (or portion thereof) was associated with the particular census tract (CT) within which it was located, and an aggregate walked distance per CT was calculated. Since CT’s vary considerably in area, population, and degree of development, three walking densities were developed to measure walked distance per CT in a comparable manner. These densities were:

- **W/P** = Aggregate walked distance divided by resident population of CT (a measure of people’s propensity to walk)
- **W/R** = Aggregate walked distance divided by CT road length (walking related to available street network)
- **W/DA** = Aggregate walked distance divided by CT developed area (area in residential, commercial, institutional, park/recreation, office, or industrial use) (walking related to the built environment)

IOP Conf. Series: Earth and Environmental Science **34** (2016) 012022 doi:10.1088/1755-1315/34/1/012022

Although a positive correlation among the three density variables is acknowledged, they measure somewhat different aspects of the environment: W/P relates primarily to the demand for walking opportunities, whereas W/R is a proxy for the available supply of walking routes. W/DA relates to aspects of both supply and demand, in that developed areas contain both far more opportunities for walking than do the undeveloped forests and fields, and also far more people. Regarding the W/P measure, one should bear in mind that not all walking within a census tract will be by residents of that neighborhood.

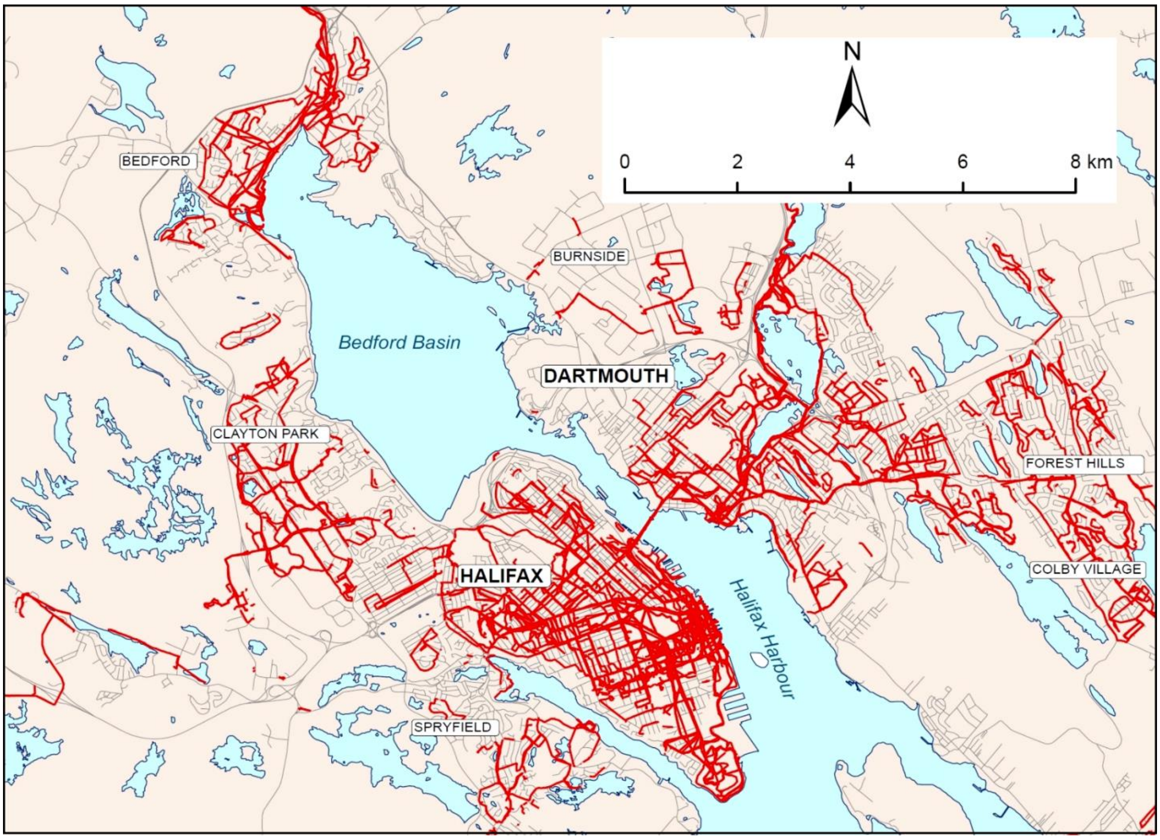

Figure 2. Map of walked routes, related to the road network.

Figure 3 shows the pattern of CT scores for walked distance per road metre (W/R), and clearly indicates that such densities are higher in almost all inner city tracts (the few exceptions being quiet residential areas, with little commercial activity or employment uses). Walking densities are generally low in the suburbs, but moderately higher in tracts containing commercial activity and/or developed parks. Maps for the two other walking densities show fairly similar patterns, although the suburb versus inner-city contrast is more marked for W/DA, and less marked for W/P.

6. Variables Related to Walking Densities

As neighbourhood (CT) built-environment variables we employed the standard components of the Index of Walkability, both separately and as composite indexes, as follows.

- Dwelling density was calculated by dividing the total number of dwellings per census tract by the total amount of developed area within the census tract.
- Intersection density was calculated by dividing the total number of intersections within each census tract by the census tract area. This figure was provided by Darren Scott, McMaster University.

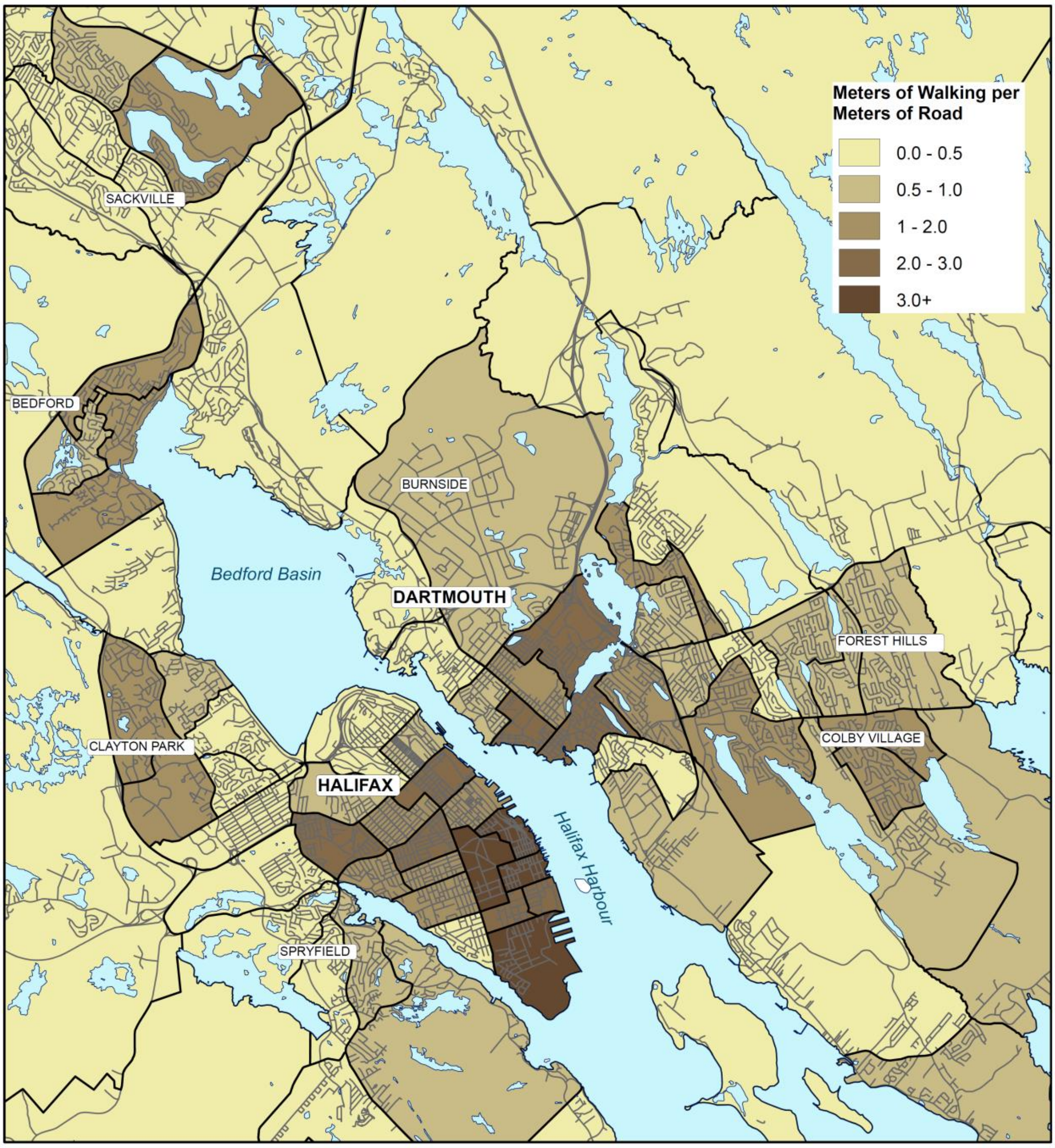

Figure 3. Density of walking per census tract, per road-metre (W/R)

- Entropy is a measure of land-use mixture within a prescribed area and was calculated based on the work of Frank et al. (2006).
- Retail lot coverage ratio was calculated by dividing the total retail building footprint area by total retail parcel area per census tract.
- 3-variable Walkability Index (WI-3) = dwelling density (z-score) + intersection density (z-score) + entropy (z-score).
- 4-variable Walkability Index (WI-4) = dwelling density (z-score) + intersection density (z-score) + entropy (z-score) + Retail Lot Coverage Ratio (z-score).
- In addition, we scored each CT by six separate land use variables, indicating the percentage of developed land in each of residential, commercial, institutional, park/recreation, office, and industrial uses.

- Two broad CT-level socio-demographic control variables were included in the study; age and income. Specifically, age was categorized into three cohorts of Young Adult (age 15 – 39), Middle-aged (age 40-64) and Older Adults (age 65-plus), while income was measured as average household income (in $000's) per CT.

7. Preliminary Results

Stepwise multiple linear regression models were run for each dependent variable (W/P, W/R and W/DA), testing them against built environment and socio-demographic control variables using IBM SPSS Version 21. Some of the models used only the separate components of the walkability indexes, and some used only the composite indices. Some included the control variables, and some didn't.

For all three CT walking densities, consistently better estimates of density were provided by the model that employed the individual walkability index variables (entropy, residential density, intersection density, and retail lot coverage ratio) along with the socio-economic variables (income and age) and the percentage of six developed land uses for each CT. The adjusted R^2 values for these models were 0.562 (W/P), 0.607 (W/R), and 0.683 (W/DA). Socio-demographic and land-use variables were consistently the dominant variables. Specifically, both institutional and office land uses were found to be the two most influential variables for all three walking densities. Average income was also identified as significant in these models for all three walking densities. Retail lot coverage ratio was the only separate walkability component that was significant against W/P, W/R and W/DA. On the other hand, dwelling density was included in model 4 against W/P, but was found not to be significant and was excluded from the equivalent models for W/R and W/DA.

In general, of the four built environment variables normally associated with walkability, retail lot coverage ratio was found to be the only variable significantly related to W/P, W/R, and W/DA. The socio-demographic and land-use variables were found to be better estimators of walking densities than built environment variables. The above analysis suggests that it may be more beneficial to investigate the separate components that influence walking rather than combining walking influences into a pre-determined index. The consistent reoccurrence of both office and industrial land uses in the regression models indicates that walking is associated with areas of employment more than with residential land uses. This notion is consistent with work by Spinney et al. (2013) that suggests much walking for transport originates outside of the home neighbourhood.

References

Atash F (1994) Redesigning suburbia for walking and transit: emerging concepts. *Journal of Urban Planning and Development,* **120**, 48-57.

Boarnet M & Sarmiento S (1998) Can land-use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics. *Urban Studies,* **35**, 1155-1169.

Crane R (2000) The influence of urban form on travel: an interpretive review. *Journal of Planning Literature* **15**, 13-23.

Dewulf B, Neutens T, Van Dyck, de Bourdeaudhuij I & Van de Weghe N (2012), Correspondence between objective and perceived walking times to urban destinations: influence of physical activity, neighbourhood walkability, and socio-demographics. *International Journal of Health Geographics*, **11**(1), 43-52.

Ewing R & Cervero R (2010) Travel and the built environment. *Journal of the American Planning Association,* **76**, 265-294.

Forsyth A, Oakes J, Schmitz K & Hearst M (2007) Does residential density increase walking and other physical activity? *Urban Studies,* **44**(4), 679-697.

Frank L & Engelke P (2001) The built environment and human activity patterns: Exploring the impacts of urban form on public health. *Journal of Planning Literature,* **16**, 202-218.

Frank L, Greenwald M, Winkelman S, Chapman J & Kavage S (2010) Carbonless footprints: Promoting health and climate stabilization through active transportation. *Preventive Medicine,* **50**, S99-S105.

Frank L, Kerr J, Sallis J, Miles R & Chapman J (2008) A hierarchy of sociodemographic and environmental correlates of walking and obesity. *Preventive Medicine,* **47**, 172-178.

Frank L, Sallis J., Saelens B, Leary L, Cain K, Conway T & Hess P (2010) The development of a walkability index: application to the Neighborhood Quality of Life Study. *British Journal of Sports Medicine,* **44**, 924-933.

Frank L, Schmid T, Sallis J, Chapman J & Saelens B (2005) Linking objectively measured physical activity with objectively measured urban form - Findings from SMARTRAQ. *American Journal of Preventive Medicine,* **28**, 117-125.

Gebel K, Bauman A & Petticrew M (2007) The physical environment and physical activity - A critical appraisal of review articles. *American Journal of Preventive Medicine,* **32**, 361-369.

Handy S, Boarnet M, Ewing R & Killingsworth R (2002) How the built environment affects physical activity. *American Journal of Preventive Medicine,* **23**(2S), 64-73.

Lee C & Moudon A (2006) The 3Ds+R: Quantifying land use and urban form correlates of walking. *Transportation Research: Part D,* **11**, 204-215.

Millward,H, Spinney J & Scott D (2013) Active-transport walking behavior: destinations, durations, distances. *Journal of Transport Geography*, **28**, 101-110.

Moudon A & Lee C (2003) Walking and bicycling: an evaluation of environmental audit instruments. *American Journal of Health Promotion,* **18**, 21-37.

Podsakoff P, MacKenzie S, Lee J-Y & Podsakoff N (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*, **88**(5), 879-903.

Renalds A, Smith T & Hale P (2010) A systematic review of built environment and health. *Family & Community Health,* **33**, 68-78.

Saelens B & Handy S (2008) Built environment correlates of walking: a review. *Medicine and Science in Sports and Exercise*, **40**, S550-S566.

Saelens B, Sallis J & Frank L (2003) Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, **25**, 80-91.

Sallis J (2009) Measuring physical activity environments: a brief history. *American Journal of Preventive Medicine,* **36**, S86-S92.

Sallis J, Frank L, Saelens B & Kraft M (2004) Active transportation and physical activity: opportunities for collaboration on transportation and public health research. *Transport. Res. Part A: Policy Pract.,* **38**(4), 249-268.

Spinney J & Millward H (2011) Weather impacts on leisure activities in Halifax, Nova Scotia. *International Journal of Biometeorology,* **55**, 133-145.

Spinney J, Millward H & Scott D (2012) Walking for transport versus recreation: a comparison of participants, timing, and locations. *Journal of Physical Activity and Health,* **9**, 153-162.

Van Der Ploeg H, Tudor-Locke C, Marshall A, Craig C, Hagstromer M, Sjostrom M & Bauman A (2010) Reliability and validity of the International Physical Activity Questionnaire for assessing walking. *Research Quarterly for Exercise and Sport,* **81**, 97-101.

Wagner C & Mueller D (2011) Comparison of bottom-track to global positioning system referenced discharges measured using an acoustic Doppler current profiler*, Journal of Hydrology*, **401**(3), 250-258.