Title
Active Travel, Built Environment and Transit Access: A Micro-Analysis of Pedestrian Travel Behavior

Author
Abdel-Salam, Gaby Hamdi

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Active Travel, Built Environment and Transit Access:
A Micro-Analysis of Pedestrian Travel Behavior

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Transportation Science

by

Gaby Hamdi Abdel-Salam

Dissertation Committee:
Professor Douglas Houston, Chair
Professor Michael G. McNally, Co-Chair
Professor Jean-Daniel Saphores

2014

DEDICATION

To

my parents, my second mom Samiha, my husband Hisham, my daughters Rana, Salma, Janna, my sisters Hala, Sahar, and my friends.

Their love and support for me over the years has enabled me to reach this far in my education and life. I am forever indebted to them all.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

ACKNOWLEDGMENTS

My deepest appreciation and gratitude to my committee chair, Professor Douglas Houston, who has guided me throughout the years of my Doctoral degree. His insights and persistent help paved the way for this research.

I would like to thank my co-chair, Professor Michael G. McNally for his help, for opening the doors to transportation modeling and proving that it is a boundless science.

I would also like to thank my committee member, Professor Jean-Daniel Saphores for his guidance in the Transportation Science program and for introducing me to travel demand analysis, a science I will forever treasure.

In addition, I would like to express my thanks to Professor Marlon Boarnet who has sparked my interest in transportation planning during my Masters degree and who has guided my initial work of my Doctoral degree.

The Expo Study was supported by the Active Living Accelerometer Loan Program, the Haynes Foundation, the Lincoln Institute of Land Policy, the San Jose State Mineta Transportation Institute, the Southern California Association of Governments, the University of California Transportation Center, the University of California Multi-Campus Research Program on Sustainable Transportation, and the University of Southern California Lusk Center for Real Estate.

Finally, I thank my colleagues for all their support, encouragement and help. A special thanks to Andy Hong for processing the accelerometer data, Steve Spears, Dongwoo Yang and Wei Li for all their contributions to project management, data collection and post-processing.

CURRICULUM VITAE

EDUCATION

University of California, Irvine
Ph.D. in Transportation Science 2014
Dissertation: "Active Travel, Built Environment and Transit Access: A Micro-Analysis of Pedestrian Travel Behavior"

University of California, Irvine
Masters in Urban and Regional Planning 2009

Thesis: "Making Ends Meet: A Program Evaluation of the Self-Sufficiency Calculator"

American University in Cairo, Egypt
B.A. in Economics 1994
Minor: Computer Science Programming

AWARDS

Graduate Student Researcher Scholarship
(University of California, Irvine) June 2014 – September 2014

Dissertation Fellowship
(University of California, Irvine) September 2013 – June 2014

Graduate Student Researcher Scholarship
(University of California, Irvine) September 2009 – June 2011

Best Professional Report of 2009
(University of California, Irvine) Aug. 2008 – June 2009

Best Community Scholar of 2009
(University of California, Irvine) August 2008 – June 2009

Undergraduate Scholarship
(American University in Cairo, Egypt) September 1991 – February 1994

TEACHING EXPERIENCE

Graduate Reader 2012
Graded assignments, midterms and final examinations

RELATED EXPERIENCE

Orange County United Way

Intern Aug. 2008 – May 2009

Analyzed data, completed a program evaluation of the ChildNet Assessment Manager, produced a statistical report for the Self-Sufficiency Calculator and trained social workers on a pilot database program

Transystems

Transportation Intern June 2014 – Present

Performed bus shelter measurements and collected GPS readings for field site visits of light rail and bus stations. Conducted aerial image and physical bus shelter condition analyses for ADA compliance status reports.

POSTER PRESENTATIONS

"Walking to Transit: Transit Proximity and Physical Activity in Low-income Communities"
Active Living Research Conference and
University of California Transportation Center Conference 2013

"Segment-level Analysis of the Environmental Context of Moderate-to-Vigorous Physical Activity during Active Travel"
Active Living Research Conference and
University of California Transportation Center Conference 2014

LANGUAGES

Arabic – native language
English – speak fluently and read/write with high proficiency

ABSTRACT OF THE DISSERTATION

Active Travel, Built Environment and Transit Access:

A Micro-Analysis of Pedestrian Travel Behavior

By

Gaby Hamdi Abdel-Salam

Doctor of Philosophy in Transportation Science

University of California, Irvine, 2014

Professor Douglas Houston, Chair

Professor Michael G. McNally, Co-Chair

The introduction of Senate Bill (SB 375) in 2008 stimulated more research linking travel behavior to the built environment. Smart growth tools mandated by this bill aim to reduce vehicle miles traveled (VMT), greenhouse gas (GhG) emissions and promote alternative modes to motorized travel. These tools encompass an array of land use improvements that are expected to influence active travel. Potential changes in the built environment may impact the frequency, amount and even the selection of routes for walking.

Data used in this dissertation was obtained from Phase I of the Expo Study, a three-phase travel survey of residents living near the Expo Light Rail Line in Los Angeles, CA. Respondents carried GPS devices and accelerometers to track locations and activity levels; and completed seven-day trip logs. Phase I of the survey was administered in Fall 2011, prior to the introduction of the Expo Line in April 2012.

This dissertation is comprised of three research topics. The first topic uses a "place-oriented" approach to examine where active travel occurs in neighborhoods adjacent to the Expo Light Rail Line. This chapter is based on the Behavioral Model of Environments, which emphasizes the influence of the physical environment on individuals' travel behavior and route choices. Results indicate that the routes selected by pedestrians have higher densities of commercial and retail centers and better access to more transit stations.

The second research topic uses an ecological modeling approach. Multilevel analysis of the effects of the built environment on active transport was performed in three geographic levels of aggregation near respondents' homes. Examination of land uses at the half-mile extent yield the least number of significant results. In contrast, land uses examined at the segment-level and quarter-mile distance from homes emphasize the importance of street connectivity and green space on increasing transport-related physical activity (TPA). This suggests the importance of analyzing the data at finer geographic levels.

The third research topic proposes a practical methodology of pedestrian route analysis in which observed GPS-tracked routes were examined and compared to GIS-simulated shortest paths. The two route types were compared over deviations in trip-level travel indices, respondents' socio-demographic traits, time of day variations and differences in objectively measured built environment features along both sets of routes. Results suggest that observed routes diverged more from shortest paths with increasing distance and were more circuitous beyond the 2.4-mile threshold. Most walks were completed after the AM Off Peak time. With the exception of the Evening time, observed routes were found to be much longer in all time periods especially in the AM Peak time. Moreover, higher densities of commercial centers, local businesses and green spaces were observed more for GPS-tracked routes than for shortest paths. These routes also had more street intersections and transit stops. Overall, results imply that pedestrians selected routes that were longer than the respective shortest paths and that may have been due to greater access to amenities and activity centers.

INTRODUCTION

This dissertation is comprised of three research topics. The first topic utilizes the Behavioral Model of Environments in a road segment-level examination of active travel. This model states the pertinence of the built environment on travelers' travel choices. Pedestrians sampled in this study preferred routes with commercial and retail centers and those that had better accessibility to more transit stations. The same routes however, lacked sufficient green spaces that were found to be equally effective in attracting greater levels of transport-related physical activity.

The second research topic uses the ecological modeling approach in a multilevel examination of the impacts of the built environment on active transport. The natural nesting of the data set where individuals reside within households and households within neighborhoods permits this analysis and allows the inclusion of variables at the various three levels in the models as controls. The final models show that the odds of MVPA occurring increases with more connective streets and planting more acreage of trees at all three microenvironment geographic extents when individual-level and household-level characteristics are accounted for. Further, increasing the number of transit stops, streets with higher traffic volumes, higher densities of: residential, industrial and commercial uses within a quarter- and half-mile radius from participant homes lower the odds of MVPA in the models.

In the third research topic, I discuss an alternative method to traditional route choice modeling in which actual GPS-tracked pedestrian routes are compared to GIS-simulated shortest paths. This methodology is less cumbersome and involves less computational steps than traditional route choice modeling techniques and accounts for heterogeneity among travelers. The routes are compared by socio-demographic variations among the sample, deviations in several trip-level travel indices, time-of-day benchmarks and differences in the built environment features along each set of paths.

CHAPTER ONE. BACKGROUND

1.1 Planning and Policy Context

The onset of the New Urbanist movement in the 1980's aimed at reversing some effects of post WWII planning such as low density and sprawled developments. The term urban sprawl was coined to depict dispersed parcels that are created in a "leapfrog" pattern that often require reliance on motorized vehicles (Frumkin, 2002). This type of development pattern has resulted in a chain reaction of increasing car pollution and lowering physical activity levels that led to increased sedentary lifestyle rates and therefore increasing health risks of cardiovascular diseases and strokes (Frumkin, 2002).

The New Urbanist movement merges disciplines from urban design and planning that together could help in reducing vehicular use while creating communities seeking to support and encourage pedestrian activity and transit usage (Handy, Boarnet, Ewing, & Killingsworth, 2002). Continuing this approach, planners implemented smart growth tools for developments typified with diversity in uses, well-connected street networks, higher densities and the promotion of alternative travel modes such as public transportation (Werner, Brown, & Gallimore, 2010).

A key motivation for studies examining the link between the built environment and travel behavior has been to identify aspects of the urban form which are associated with lower vehicle miles traveled (VMT) and associated reductions in greenhouse gas (GhG) emissions. This research focus became increasingly important in California after the passage of Senate Bill 375 (SB 375) in 2008. The bill's intent is to reconcile all future demand for new housing and transportation needs. It mandates that Regional Transportation Plans (RTP) of the local

Metropolitan Planning Organizations (MPOs) direct future developments towards denser, mixed-use, transit-oriented communities to collectively reduce VMT and GhG (Institute for Local Government, 2014).

One strategy for reaching GhG reduction goals is through the adoption of smart growth policies that support or encourage alternative travel modes to pollution-producing vehicles. Smart growth is a planning concept that promotes compact, walkable, mixed-use neighborhoods that are well-connected and allow easy access to public transit (U.S. Environmental Protection Agency, 2013). Thus, promoting public transportation and transport-related physical activity will facilitate these goals.

As directed by SB 375, MPOs in California are required to adopt and incorporate Sustainable Communities Strategy (SCS) into their Regional Transportation Plan (RTP) to reduce GhG emissions. As a result, the Southern California Association of Governments (SCAG) has advanced the reduction of VMT and consequent GhG emissions by implementing a regional 2012-2035 RTP and SCS to transportation, land use, housing and environmental planning. Specifically, the plan promotes intensive public transportation investments, in-fill densification and transit-oriented-developments (TODs) along transportation corridors. It also identifies High-Quality Transit Areas (HQTA's) as areas "within one-half mile of a well-serviced transit stop" along transit corridors with high frequency transit service particularly in peak demand time (SCAG, 2014). The goal is to have 53% of new employment growth locations and 51% of housing developments within HQTAs between the years 2008 and 2035 (SCAG, 2014).

The potential reduction in VMT and associated GhG emissions from implementing smart growth elements and infill developments to promote vehicle travel alternatives are greater in older suburban communities. This is because these developments typically have higher densities of residential and diverse commercial uses and are likely to have urban designs that promote transit access and non-motorized travel (Boarnet, Joh, Siembab, & Fulton, 2010). In addition, infill developments will likely increase densities further which in turn could lead to higher traffic congestion levels, decreased vehicle mobility, and increased usage of alternates to motorized travel (Boarnet et al., 2010). The question is: which types of land uses are optimum for retrofitting older suburbs and neighborhoods along transit corridors? Although previous studies have begun to address this question (Boarnet et al., 2010; Sallis et al., 2006; Schlossberg & Brown, 2004), our knowledge remains limited regarding which factors at the block- and street-level encourage reduced VMT and encourage greater transit ridership and increased physical activity.

1.2 The Benefits Of Transport-Related Physical Activity (TPA)

Studies of the relationship between the built environment (BE) and travel behavior have accelerated in the past few decades, but transport-related physical activity (TPA) remains an understudied area particularly as it relates to public transit ridership (Horowitz, 1982).

There has been a growing consensus regarding the positive effects of smart growth developments on increasing transport-related physical activity (TPA) as an alternative to vehicle use (Adams et al., 2011; Badland, Duncan, Oliver, Duncan, & Mavoa, 2010a; de Nazelle et al., 2011). Typically, TPA is discussed in the literature as only two modes: walking or cycling.

Since 1995, national walking trip shares have been on the rise in the U.S. (Figure 1-1). These findings are very promising especially for the transportation and public health fields since physical activity has many positive health implications.

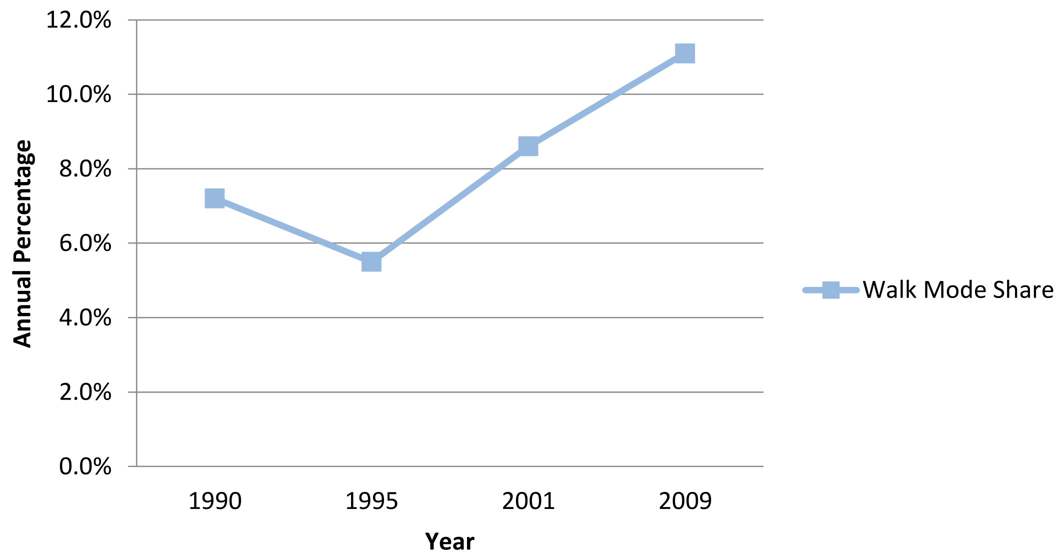


Figure 1-1: Annual U.S. Shares of Walk Trips
(Source: Adapted from Table 1 (Litman, 2012) obtained from NHTS Data)

TPA reduces the health risks from chronic diseases, colon and breast cancer, diabetes, ischemic heart disease, obesity as well as mortality (Frank, Kerr, Sallis, Miles, & Chapman, 2008; Leitzmann et al., 2007; Oliver, Badland, Mavoa, Duncan, & Duncan, 2010; Saarloos, Kim, & Timmermans, 2009). Relative to being sedentary, achieving at least the recommendations of moderate physical activity (30 minutes, five days/week) is correlated with a 32% reduction in mortality risk (Leitzmann et al., 2007).

In addition, all forms of active transport have environmental and traffic impacts. Switching to TPA modes of travel instead of vehicular travel, could mitigate and reduce pollution, Ghg emissions and traffic congestions (de Nazelle et al., 2011; World Health Organization (WHO), 2011), although some concerns have been raised that pedestrians traveling along high-traffic roadways could experience heightened exposure to vehicle-related pollution (de Nazelle et al., 2011; Houston, Wu, Yang, & Jaimes, 2013).

1.3 Contributions of this Research

Most of the studies in the literature were performed on a neighborhood geographic scale but the objectives of this dissertation require a fine-grain block- and street-level analysis. Many authors agree that TPA measurement accuracy obtained from individual or household level analyses, surpasses the estimates resulting from analyses on the block-group level (Cerin, Conway, Saelens, Frank, & Sallis, 2009; Handy et al., 2002). In addition, large parcel sizes were observed to be deterrent to pedestrians whereas smaller lots of retail and food establishments were found to encourage more walking (Lee & Moudon, 2006b). With this in mind, the spatial scale of this research provides a more detailed analysis of distance gradient effects to and from transit stations and the block- and street-level characteristics associated with greater TPA.

Previous studies have also largely neglected the importance of the built environment that augments the positive synergistic effects between public transit and TPA. They have focused mainly on how the built environment enhances neighborhood walkability in order to account for facets which they believed to have a confounding effect if neglected (Badland et al., 2010a; Duncan, Badland, & Mummery, 2009; Lachapelle, Frank, Saelens, Sallis, & Conway, 2011; Owen et al., 2007). One study found that 10 % of the variations in walking trips were explained by built environment elements, the 3 Ds + R (destinations, distance, density & routes) after controlling for socio-demographic characteristics, but neglected to investigate the role of neighborhood design and only relied on stated preference not objectively-measured data (Chanam Lee & Moudon, 2006b). This proposal fills this gap in the literature by including proxies for all the five Ds of the built environment (density, diversity, design, destination accessibility & distance to transit) and the link to physical activity especially as the mediating effect of transit is explored.

In the following chapters, I propose to examine street routes from a "place-based" perspective that will add valuable insights into our understanding of the environmental context of TPA. I hypothesize that aesthetically pleasing routes bordered by high concentrations of green spaces and those leading to transit stations with commercial and retail activity centers are more conducive to active travel.

By examining differences among GPS routes to transit stations and the underlying built environment features, I hypothesize that the average daily physical activity accumulated by transit riders surpasses that of non-transit riders. Here transit access is modeled as an effect mediator that promotes greater objectively measured non-motorized travel. In addition, comparisons of 'observed' GPS routes to 'objectively' measured shortest-distance path estimates will yield insights into the optimum preferred paths that pedestrians choose in their travel patterns. Shortest-distance routes are not always chosen by pedestrians who may prefer more circuitous paths (Papinski & Scott, 2011), which warrants further investigation of path-specific attributes promoting more TPA.

This research will discuss a fine level of analysis that I believe can be more precise in quantifying impacts of major public transportation investments on travel behavior. Moreover, implementing smart growth tools along transit corridors could potentially lead to an increase in non-motorized travel over passenger vehicle use that helps accomplish SB 375 goals.

CHAPTER TWO. OVERVIEW OF THE EXPO STUDY

2.1 L.A. Metro's Exposition Line

The proposed research will analyze data from a study of the impacts of a new light rail service in south Los Angeles, the Exposition (Expo) Line, connecting Downtown, Los Angeles and Culver City. The length of this light rail segment between both points is about 8.6 miles with an average travel time of about 30 minutes (Metro, 2014). Phase II of the Expo Line (not a focus of the current research) is under construction from Culver City to Santa Monica. Figure 2-1 below shows Phase I construction and the route of the Expo Line that officially began service on April 28, 2012. Running on electric overhead catenary wires makes the Expo Light Rail Line a sustainable system designed to connect to the current 70 stations of the Metro transit network (Metro, 2014).

Figure 2-1: Exposition Light Rail Map in Los Angeles, CA
(Source: http://www.metro.net/projects_studies/exposition/images/expo_ph1_fact_sheet.pdf)

2.2 The Expo Line Study Research Design

The Expo Line Study is a quasi-experimental before-after study by design where the study area is divided into experimental and control neighborhoods. The project area is about 12 square miles between the Exposition and Crenshaw corridors in Los Angeles, CA. The purpose is to examine changes in travel behavior of the residents in close proximity to the new Expo Line including the effects on changes in physical activity behavior. The main objective of the study was to examine the before/after impacts on travel behavior from the introduction of the new light rail line along the Exposition Boulevard that opened in April, 2012. Neighborhoods in the study area (Expo study) were divided into two groups: an experimental and a control group. By design, the two study groups are comprised of similar socio-demographic and built environment attributes so that the only resulting effects on travel behavior should be due to the introduction of the new Expo Line.

Experimental or treatment neighborhoods were selected from the half-mile radius closest to the six western new transit stations as opposed to those new stations in the east to avoid bias from the overlapping service of the existing Blue Line light rail and the Silver Line rapid bus. Further, the neighborhoods surrounding the stations bordering the University of Southern California were also excluded because of differences in the demographic spectrum of the residents and because the larger population is comprised of transient students whose travel behavior may not reflect the travel patterns of the existing permanent residents of the corridor.

Neighborhoods selected as the control group are located more than half-mile to two miles away from the new Expo Line stations and along corridors designated to receive future light rail line extensions. The half-mile radius catchment area for studying treatment effects on travel behavior has been widely used in transit-oriented development studies. Recently, a study by Guerra and Cervero of 1,500 transit stations across the United States to forecast ridership relative to varied catchment area radii; found that the half-mile radius was within reasonable bounds for treatment effects on transit use nationwide for residential access while a quarter-mile radius was more suitable for employment access (Guerra, E. & Cervero, 2013). Therefore, the choice of quarter-mile to a half-mile for the experimental group was selected to capture these treatment effects. Effects of the new Expo service are hypothesized to have diminished if not disappeared completely beyond the half-mile radius from transit stations which justify the location choices of the control group.

The map in Figure 2-2 shows the spatial distribution of the households selected from the Expo study divided into either an experimental or control group. Households in both groups agreed to participate in this three-wave study by completing the survey materials over three different periods: Fall 2011 (before the Expo Line was constructed), Fall 2012 (after the Expo Line began service) and Fall 2013 (one-year after service began).

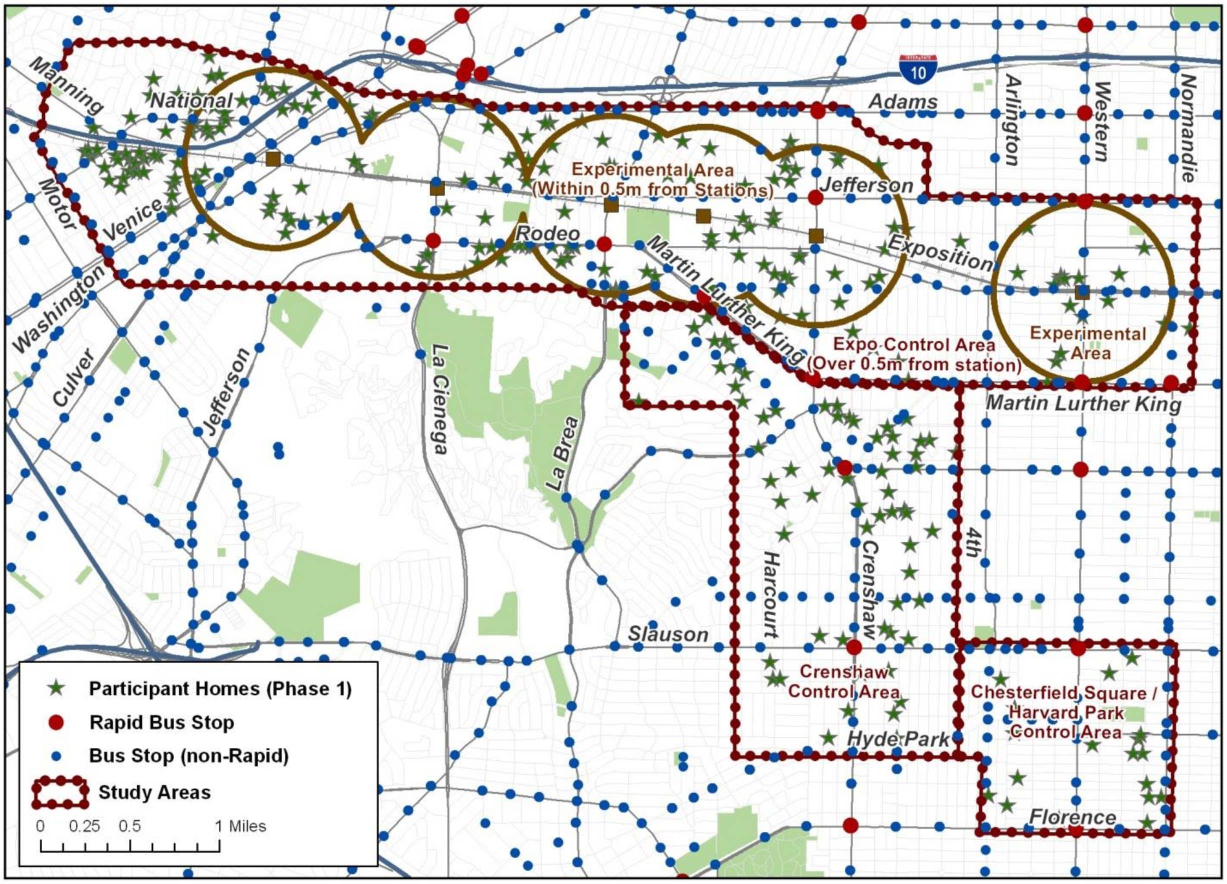


Figure 2-2: EXPO Study Area Experimental & Control Neighborhoods
Source: (Boarnet, Houston, & Spears, 2013)

Figure 2-3 below is a map of the selected households within the ¼-mile and ½-mile buffer from transit stations as well as those located beyond the ½-mile buffer and therefore considered as control in Phase 1 of the study. Expo study researchers contacted the same households to participate in each of the three waves of the study: Fall 2011, Fall 2012 and Fall 2013. There were 284 households in the full sample from Phase 1 (Fall 2011) for both groups (138 experimental & 146 control) and in Phase 2 (Fall 2012) there were a total of 204 households (103 experimental & 101 control). In addition, 143 households in Phase 1 were given GPS units (to track location) and accelerometer devices (to track physical activity intensity) and were classified as the mobile tracking group. Out of the 143 households, 117 had valid data and were used in the final mobile tracking group. In Phase 2, only 105 (out of 143 households) participated in the mobile tracking assignment. The main focus of this dissertation however, pertains to data from Phase 1 only before the introduction of the Expo Light Rail Line. Data from Phase 2 and Phase 3 were not analyzed here since they were not relevant.

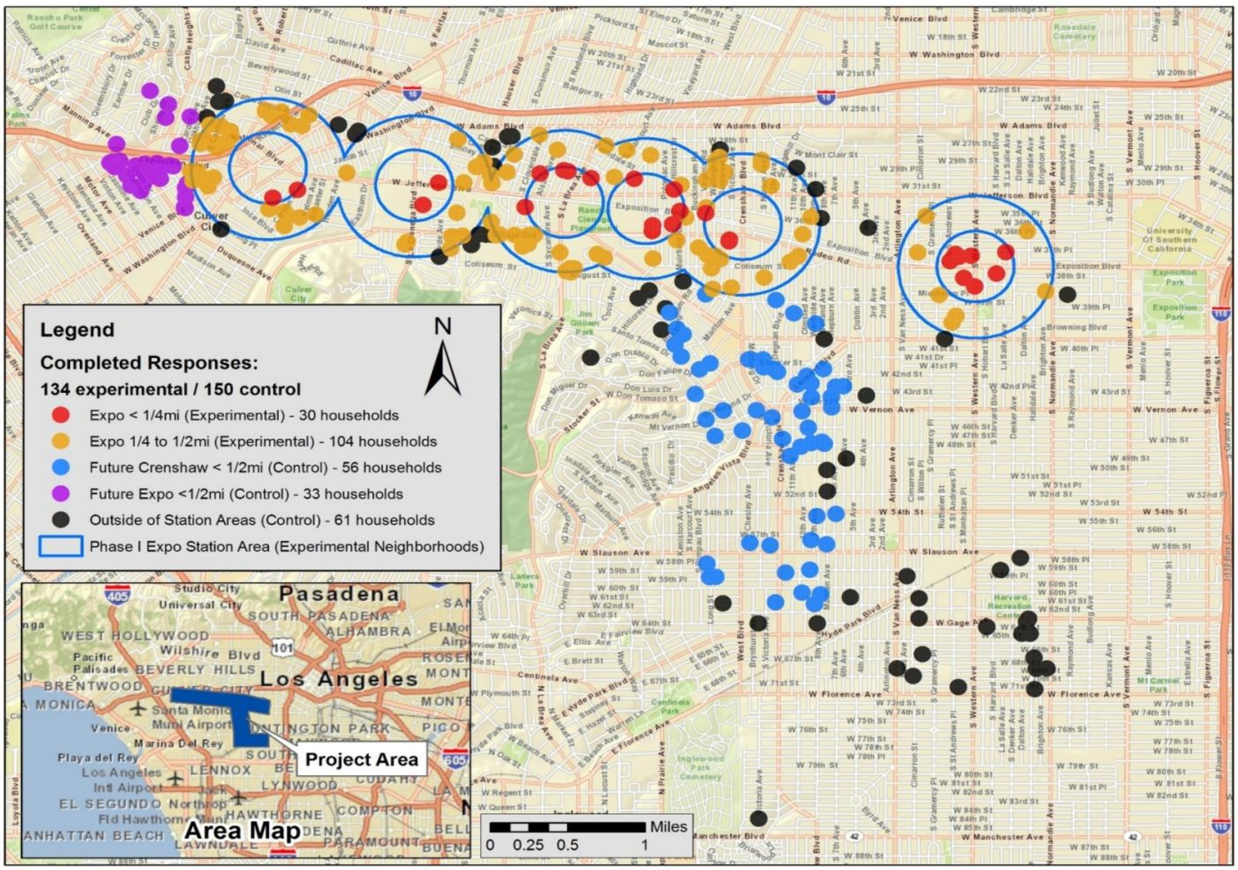


Figure 2-3: EXPO Study Area Phase 1
Source: (Boarnet et al., 2013)

2.3 Participant Recruitment

All households were sent an invitation letter (in English and Spanish) to participate in the study. As shown in Table 2-1, households from Phase I that would participate in the 'standard' baseline survey were promised a $15 gift card to their local grocery store and if they agreed to participate in the 'mobile-tracking' survey they would be given a $30 gift card. A total of 27,275 survey invitation letters were mailed out and 284 households completed the survey (a response rate of about 1%).

The potential participants were requested to complete an initial survey either by accessing the project website online or contacting a project team member directly. All survey materials were previously submitted and approved by the University of California, Irvine Institutional Review Board (IRB). The initial survey consisted of questions regarding their household characteristics and general travel behavior. The participants were also prompted for their willingness to carry a GPS unit. As a result of the questionnaire, the potential respondents were separated as: the web-based group (completed remaining survey materials online), the paper-based group (completed survey materials in hard copy) and the mobile-tracking group (completed survey materials as hard copy and indicated their willingness to carry an accelerometer and GPS units). Figure 2-4 shows a sample of the devices used in the Expo Survey: the Actigraph Gt1M accelerometer to track vertical and lateral movements every one-minute (on the left-hand-side) and the GPS device, QT-1000x (QSTAR) to track locations every 15 seconds (on the right-hand-side).

TABLE 2-1: EXPO SURVEY RESPONSES FOR PHASE I

Sample	**Expo Core Phase I**
Study Area	**Expo Study**
Time Period	9/11-2/12
Participation Rate	1%
Total Addresses Mailed	27,275
Final Household Surveys Completed	284
Final Household Sample (Usable)	276
Final Households of Mobile Tracking	117
Incentive	
Standard	$15
Mobile-Tracking	$30

Source: (Boarnet et al., 2013)

Figure 2-4: Accelerometer and GPS devices used in the Expo Survey
Source: (Boarnet et al., 2013)

A follow up package with the relevant materials and prepaid postage envelopes were then mailed to the participating households except for those defined as the mobile-tracking group. The package included instructions on how to proceed, seven-day trip logs (one for each member at least 12 years old) and vehicle mileage logs (one for each vehicle owned). The mobile-tracking group met directly with a researcher who then trained the individuals on how to use and charge the devices and was also asked to complete the seven-day trip logs and vehicle mileage logs[1]. After the seven-day surveys were completed, the participants met once more with a researcher who gathered the survey materials[2], the GPS and accelerometer units.

Participants in the web-based group were given a unique password and username and were instructed to complete a “baseline survey” developed on SurveyGizmo (web application) and the seven-day trip and vehicle mileage logs. Their responses were recorded by the application and later downloaded into secure project computers at UCI for confidentiality. Table 2-2 below shows the number of households with completed household surveys by survey material and by group type.

TABLE 2-2: PHASE 1 SURVEY DETAILS

Survey Material	**Households**	**Total Persons/ Vehicles**	**Days/ Vehicle Days**
Baseline	304		
Vehicle Mileage Logs	238	337	2308
7-day Travel Logs	288	494	3239
Response Type	**Households**		
Mobile-Tracking	143		
Paper-based	59		
Web-based	82		
Total	284		
Group			
Control	150		
Experimental	134		

Source: (M. Boarnet et al., 2013)

[1] The material was identical to those provided to the web-based and paper-based groups.
[2] Researchers checked the documents for completion.

CHAPTER THREE. DATA AND METHODS

3.1 Data Description

The data used in this dissertation was extracted from the original data set of the 117 households from Phase 1 (before the introduction of the Expo Light Rail) of the Expo project. One participant from each household was asked to complete a baseline survey and a seven-day trip log. The seven-day trip logs recorded daily trips by mode. The baseline survey prompted the user on household-related data as well as information on the primary respondent. The household data included: annual income, vehicle and bicycle ownership, and household size. Primary respondents' data ranged from socio-demographic questions on his/her race, age, height, weight, gender, and employment status to attitudinal inquiries towards travel behavior and safety perceptions. Further, the respondents' weight and height were also collected and used to calculate BMI, a proxy for obesity.

In addition, some respondents also agreed to carry GPS devices and accelerometers to track their location and physical activity levels and were designated as the mobile-tracking group. After cleaning and validating the data from these devices, the resulting data points ranged from four to seven valid days.

Data obtained from the accelerometer and GPS devices needed to be matched to provide a momentary activity spectrum with the relevant location information. Prior aggregation of the data points was required since the accelerometers recorded readings of the different physical activity levels in one-minute increments; while the GPS device traced locations for the individuals in 15-second epochs. Therefore, the GPS readings needed to be aggregated to the one-minute level to obtain a one-to-one minute match with the accelerometer readings.

Further, built environment (BE) attributes were gathered and augmented to the Expo data set. This data set includes street-level characteristics within 0.5 mile buffers around Expo respondents' homes and land use catchment buffer areas of 40-meters to test the influence of the built environment on active travel. The BE matrix was accumulated from several data sources: 2012 transit line and stop locations (Los Angeles Metropolitan Authority, Metro), 2010 TIGER street network, 2005 existing land use data from SCAG (Southern California Association of Governments), WALKSCORE.COM (for a walkability index at the street address level) and 2005 Annual Average Daily Traffic (AADT) volumes from CALTRANS, and high-resolution (2 feet) land cover data for 2002-2005 which combined QuickBird remote sensing data with aerial photographs for the city of Los Angeles, CA (McPherson, Simpson, Xiao, & Wu, 2011).

3.2 Accelerometer Data Quality Control and Validation

The data collected from GPS and accelerometers were processed and validated for quality. The processing for the accelerometer data included two steps: data reduction and data calibration. Only valid hours and days are defined in the initial data reduction phase. During this phase, different criteria measures were defined such as periods of device non-wear time, which entailed 20-60 consecutive minutes of zero activity counts that were identified and excluded. Data outliers such as counts > 16,000 per day were also excluded during the data reduction phase. In addition, valid hours were also determined as 8-10 valid hours or a total of 600 minutes per day. Next, the valid days were identified which were on average four days per week per participant. Finally, participant activity bouts were extracted and were generally in 10-minute sessions separated by two-minute breaks. Table 3-1 below shows the different criteria identified in the data reduction phase. Criterion ‘C’ was deemed optimum and was generally used as the benchmark measure.

Table 3-1: Data Reduction Criteria

Criteria	A	B	C	D	E
Non-valid minutes (exclude continuous 0s)	20 min	20 min	60 min	60 min	60 min
Outliers	>16,000	> 16,000	>16,000	> 16,000	> 16,000
Valid Hours	8 h	10 h	8 h	10 h	9 h
Valid Days	4 days	4days	4 days	4 days	4 days
Samples with 4 valid days	93	80	99	87	96
Counts/valid day/person	218,466	221,629	210,295	217,270	213,992

Source: (M. Boarnet et al., 2013)

The resulting processed data were then classified into physical activity variables in the data calibration phase. MeterPlus software was utilized in this phase to process the accelerometer data and to produce the resulting dependent variables. The variables generally fall into one of the following categories:

- Motion-based (e.g. counts/day)
- Temporal (e.g. average daily MVPA in minutes)
- Energy expenditure
- Total physical expenditure (e.g. Kcal/day[3])

[3] Kilocalorie, a unit of heat needed to raise the temperature of 1kg water by 1 degree at 1 atmospheric pressure

- Physical activity energy expenditure (e.g. METs in minutes/day or hours/day[4])
- Activity-based variable (daily/weekly time spent walking/running…)

3.3 GPS Travel/Location Classification

GPS data points under went several automated and manual reviews to identify stationary or location episodes and vehicular and active travel periods. Raw GPS data record positional information (latitude and longitude) and a time stamp. From these two variables, speed and distances were obtained and then classified into location and travel periods by mode using two methods: an automated process and a manual review.

3.3.1 Automated GPS Review

The automated method produced batches of processed GPS points using R statistical software) then project researchers systematically reviewed the results manually for each participant. Each GPS point went through a series of decision queries to define whether it is a location type or a travel mode based on the following algorithm:

1. Is the speed > 6 mph?
2. Yes ➡ then code as a vehicle trip[5].
3. No ➡ Is the GPS point within 20 meters from a specific location?
4. Yes ➡ Code as a location type (residential or non-residential).

---

[4] Metabolic Equivalent of Task, a measure for the rate of energy consumption during a physical activity

[5] Private vehicle or transit trip, defining exact travel mode type was done in the manual review stage.

5. No ➡ Is the previous or subsequent point a vehicle trip?

6. Yes ➡ Code as a vehicle stop.

(for private vehicle or transit – used to ensure brief periods of vehicle stops under three minutes such as time at traffic lights or stop signs, etc. and is considered an in-vehicle period).

7. No ➡ Code as an outdoor stay location or outdoor walking.

3.3.2 Manual GPS Data Review

The resulting software reviewed GPS points were then loaded into ArcGIS for further extensive review. The data points were overlay on the existing Expo road network, aerial photography images, transit network and transit station data to correct any misclassifications. The following assumptions were used during this manual review process of the GPS points:

1. Walking periods (under 6 mph and not classified as a given stationary location) generally required substantial manual review since they often entail a good amount of stopping and starting. They generally were classified as such due to the closeness of sequential GPS points in a given direction that aligned with a typical location pattern (along a sidewalk, etc.). In contrast, vehicular periods typically had more dispersed GPS points and were generally easily distinguishable from walking periods.

2. Cycling, another form of non-motorized mode was excluded although cyclists can typically reach or exceed the 6 mph benchmark. This is because a quick analysis of the baseline survey data showed no cycling trips made. In addition, previous findings proved that some accelerometers cannot distinguish differences in physical activity

intensity because the full range of vertical and horizontal movements are not captured for cycling mode (Oliver et al., 2010).

3. Train and bus trips were differentiated and coded in the manual review stage by overlaying the GPS points in ArcGIS on aerial images of the Expo study area. Transit trips were generally classified in the automated classification stage as “vehicle” trips and during the manual review stage GIS staff compared “vehicle” trip routes with the location of active bus and train routes and transit stations. Vehicle trips were re-classified as transit trips if their origin and destination locations corresponded with transit station locations and if the route taken corresponded with transit routes.

4. Vehicle stops that were longer than three minutes (i.e., a brief stop such as a long traffic light that was not a given location) were manually reviewed and re-classified as vehicle trip periods as needed.

5. GPS points overlapping or within 20 meters from a large parcel (e.g. University of Southern California, Costco, etc.) were coded as one location point even though participants may have traveled within this larger parcel (between building, etc.).

6. Sometimes due to GPS signal interferences, sequential 15-second GPS points had an erratic pattern and briefly (usually under 1 minute) were located far from the actual GPS device location. Such periods were manually reviewed and as needed re-coded to correspond with a given location or part of a vehicle trip.

Figure 3-1 shows the results after completing the automated and manual GPS data reviews. The map shows the difference between two trips, one was completed by car and the other one was a walking trip.

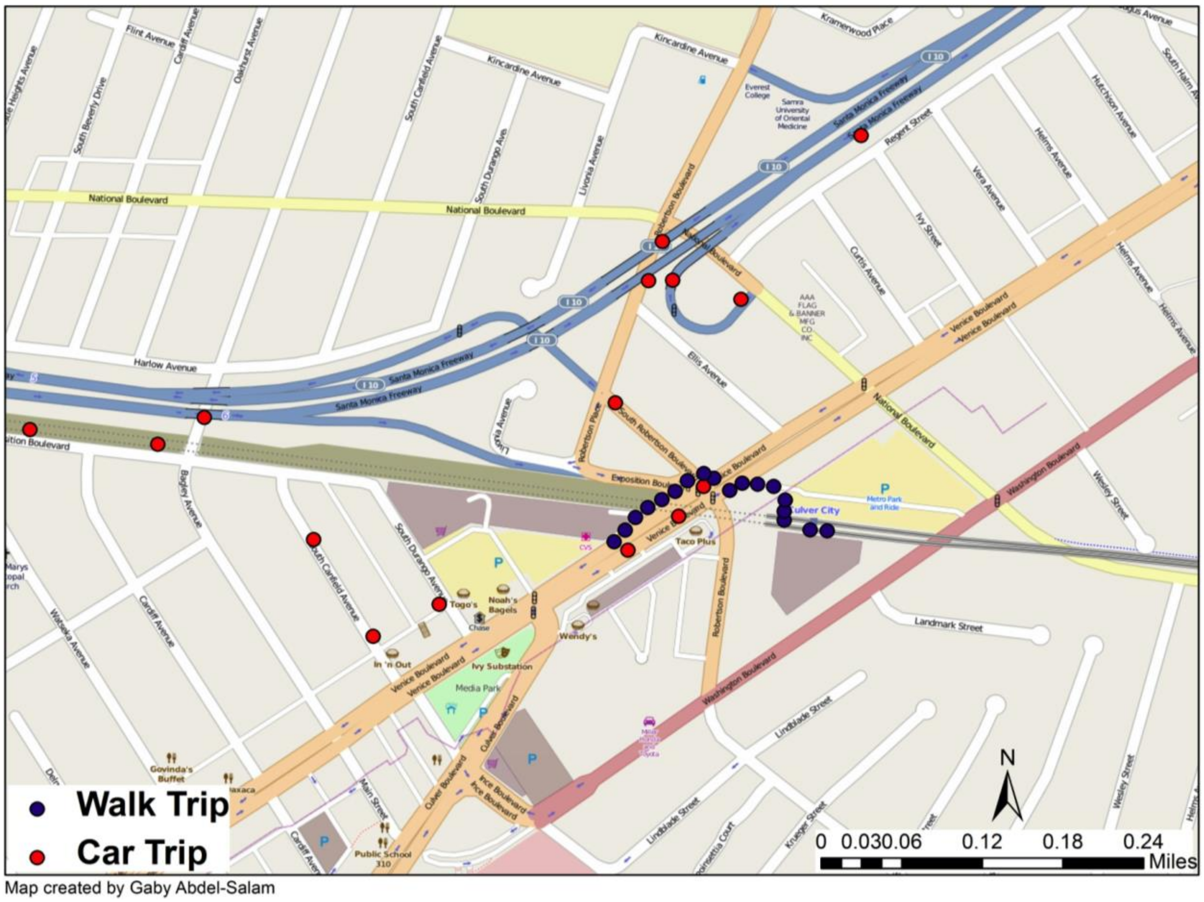


Figure 3-1: Distinguishing Between Different Trip Modes

3.4 Description of Existing Transit Network

The Expo study area is primarily served by the Los Angeles Metropolitan Transit Authority (Metro) for public transportation and secondarily by two transit agencies: Culver City Bus and the Big Blue Bus. Metro currently manages a total of 25 bus lines in the Expo study area that includes: local, rapid, express and shuttles with a total of 473 transit stops that connects the downtown Los Angeles CBD area to Santa Monica in the west and to the LAX/South Bay area to the south-west. In addition, Metro also operates the Expo Line that services our study area with the six new light rail stations.

CHAPTER FOUR

SEGMENT-LEVEL ANALYSIS OF THE ENVIRONMENTAL CONTEXT OF WALKING

INTRODUCTION

Active transportation, transport-related physical activity (TPA) and walking-for-transport are interchangeable terms that define walking/biking as a travel mode. This form of transport has been growing in popularity in the public health and transportation planning fields because of its obvious positive benefits on health and its contribution in decreasing vehicular travel that ultimately reduces traffic congestion and GhG emissions.

The analysis performed in this chapter uses the *Behavioral Model of Environments (BME)* that emphasizes the importance of the physical environment on an individual's choice for active travel. This is achieved by examining the associations between active travel and the built environment features: at origins and destinations and along routes taken and the physical attributes and traffic conditions of the road segments.

Objectives

The objectives of this research is to present a "place-oriented" approach for understanding "where TPA occurs" within walking distance of participant homes in a travel study of the Expo Light Rail Line. Initially I explain the various geographic extents that have been utilized in current research outlining the superiority of the road segment unit of analysis in producing finer and detailed regression results. Thus, I use a segment-level analysis of the built environment, land use, and transit access factors that have been previously hypothesized in the literature to influence walking and moderate-to-vigorous-physical-activity (MVPA).

Research Questions

The research questions I try to answer here are:

1. What roadway segment-level built environment, land use, and transit access factors are associated with greater walking and moderate-to-vigorous-physical-activity (MVPA)?

2. How can we translate these findings to policy measures to promote more active transport?

Results indicate that the routes frequented by pedestrians have higher densities of commercial and retail uses, neighborhood employment and more public transit stations. However, the same routes are lacking in green spaces although the analyses performed at the road segment-level have shown a positive association between green space and active transport especially during episodes of elevated physical activity levels.

This noticeable absence in green spaces along the 'walked' routes point out to the need for more progressive smart growth policies that aim to improve landscape designs to promote alternative modes to vehicular travel. Moreover, these policy reforms may be particularly pertinent in low-income communities whose residents may be more likely to use public transit and active transport than motorized vehicles. Thus, by increasing green spaces and improving pedestrian pathways; we would likely encourage more individuals to walk to destinations and to transit, which reduces their reliance on vehicular modes and contributes to the overall goals of SB 375.

4.1 Theoretical Context

Studies in the active transportation literature are multidisciplinary in the sense that they rely on concepts from urban and transportation planning as well as from the public health disciplines. From the public health perspective, physical activity is exerted during active transportation [walking and bicycling] and therefore healthy living is promoted. This has been a major concern in the public health field due to the rising dependence on motorized forms of travel that is aggravated by an inactive life style (Lee, 2004).

On the urban planning side, social and built environment characteristics are hypothesized to directly affect active transportation. Residential infill developments, densification, higher street connectivity, better job access and promoting public transit are just a few components to counter post World War II sprawled developments. These developments in the past catered to automobile users and neglected street design concepts for bikers and pedestrians (Lee, 2004). In addition, walking and bicycling relate to transportation planning, as they are forms of non-motorized travel which if promoted could alleviate some of the pollution and congestion concerns on the road network.

The main conceptual framework that combines concepts from the three above disciplines is the *Behavioral Model of Environments.* In contrast to the *Ecological Modeling Approach* that focuses on the individual's interactions to his/her physical, social and cultural environment (Giles-Corti, Timperio, Bull, & Pikora, 2005; Sallis et al., 2006) the *Behavioral Model of Environments (BME)* magnifies the importance of the physical environment on an individual's choice for active travel. The main elements of *BME* are: the trip origin and destination (OD), the built environment (BE) features of the OD's and the road/route features utilized during active travel episodes (Moudon & Lee, 2003). Origins and destinations identify the travel purpose and therefore answer the *Where* question for the TPA activity that revolves in the spatio-physical and spatio-behavioral realms (Moudon & Lee, 2003). The BE features along the routes taken could be viewed through the spatio-physical realm as it pertains to the types, mix and number of land uses such as activity centers. Route BE attributes answer the *Why* question for pedestrian and cyclist traffic (Moudon & Lee, 2003). Finally, road/route features encompass the spatio-physical and spatio-behavioral realms and aspects such as route quality and safety concerns of pedestrians or bikers that are often the focal points (Moudon & Lee, 2003).

In addition, all three BME components affect the policy-based and spatio-psychosocial realms. The former affects urban and transportation planning laws by shaping the physical environment and the latter is more relevant to attitudinal and perception-based concepts. A graphical display of the four realms of the *Behavioral Model of Environments* and its three components is shown in Figure 4-1 below.

This chapter is an evaluation of the environmental context of roadway segments and therefore employs concepts from all four realms of the BME. The five D's of the built environment are included in the spatio-physical realm and these encompass characteristics of the routes and activity centers of the OD's. Travel mode choice (motorized or non-motorized) and incidence of collisions constitute the spatio-behavioral realm. Elements pertaining to street/route quality such as aesthetics, pleasantness of route and other TPA perceptions comprise the spatio-psychosocial realm. Finally, local and regional urban and transportation plans affect the policy-based realm.

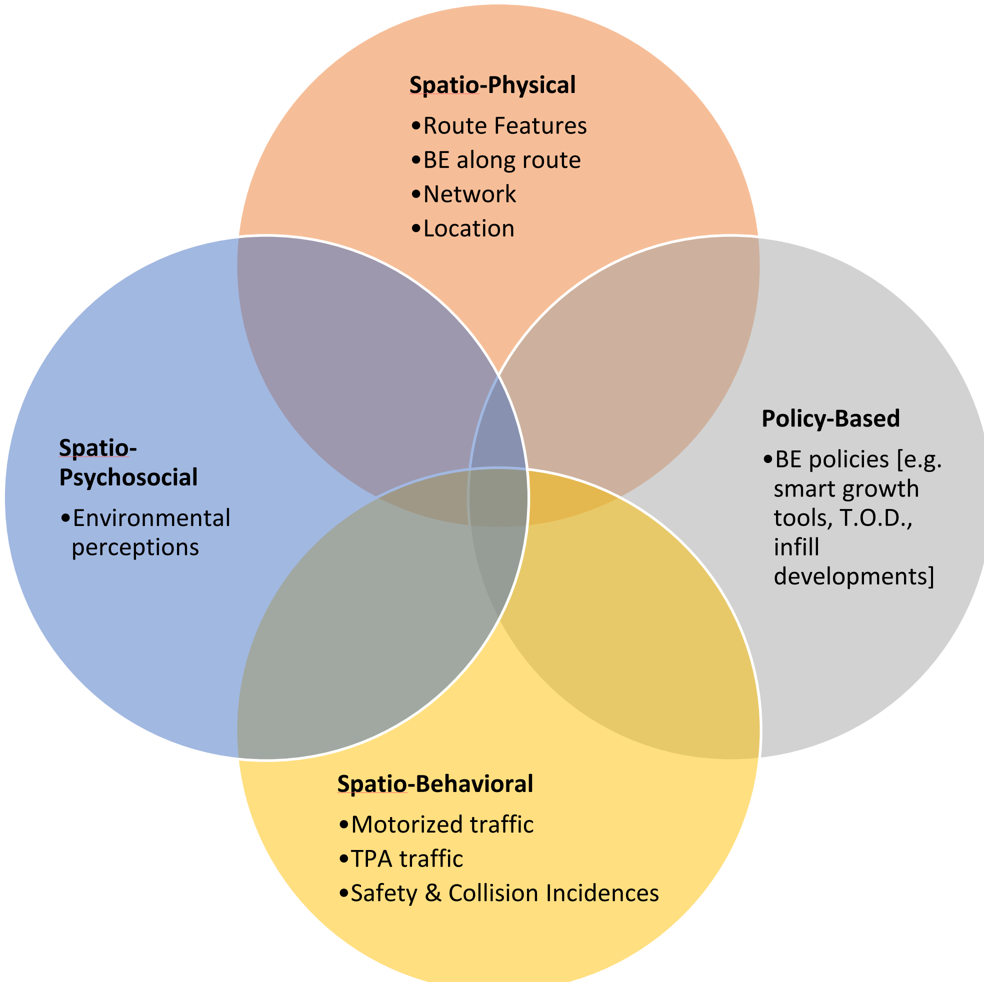


Figure 4-1 Behavioral Model of Environments
(Source: Adapted from Table 2 (Moudon & Lee, 2003))

4.2 Literature Review

In this section, common themes in the planning and public health literature are discussed as they relate to the relationship between the built environment, physical activity and transit use. As mentioned earlier, one objective of this chapter is to unravel the features of the built environment that encourage physical activity pertaining to transport. Thus, I discuss next the environmental context where active transport occurs in general and as it relates to transit use. I also introduce the main contribution of this chapter, active transport that is objectively measured at the street segment-level via matched GPS-accelerometer data points.

4.2.1 Where Physical Activity Occurs and Associated Built Environment Factors

The environmental context of walking and physical activity has received considerable attention in recent years, but many questions remain unanswered. Recent physical activity studies have examined where physical activity occurs over the course of the day specifically in near-home sites. One study found that 90% of all research linking physical activity and health risks considered the built environmental factors only in immediate residential surroundings (Leal & Chaix, 2011). Another concluded that only 46% of all activity bouts occurred in the participants' neighborhoods and that residents with higher moderate-to-vigorous-physical-activity (MVPA) levels usually lived in areas with denser population and housing that have improved street connectivity and better access to more parks (Rodriguez, Brown, & Troped, 2005).

Several studies provide insights into the relationship of TPA and exposure to "green" space and have observed a positive influence of aesthetically pleasant green spaces (vegetation areas) on incidence of physical activity (Giles-Corti, Broomhall, et al., 2005; Rainham et al., 2012;

Rodriguez et al., 2005). A recent systematic review of the literature found that the majority of TPA studies (66%) found a positive correlation of physical activity and green space (Lachowycz & Jones, 2011). Other studies considering accessibility, found that the odds of walking increases by 50% for adults with greater access to large, attractive public open spaces (Giles-Corti, Broomhall, et al., 2005). Similarly, by classifying neighborhoods into green space percentiles, one study concluded that the odds of MVPA increased by 39% for residents in high green space locations (90^{th} percentile) over those in lower levels (10^{th} percentile) (Almanza et al., 2012). Further, another study concluded that the odds were generally higher for a period of MVPA occurring in green spaces compared to non-green spaces among school children (O.R. 1.37 for boys and O.R. 1.08 for girls) (Wheeler, Cooper, Page, & Jago, 2010).

Studies of the built environment and MVPA have focused more on near-home locations for children. One study observed that 63% of all activity bouts in children over the day, were observed in the vicinity of their residence (Jones, Coombes, Griffin, & van Sluijs, 2009). Another examined the "urbanicity" of an area and found that urban students expended three times the amount of MVPA of suburban and rural students mostly near their homes/schools (30% of MVPA) or in other residential areas (10% of MVPA) (Rainham et al., 2012).

National travel surveys suggest that the majority of trips (61%) within ½ mile or less from home were walking trips (Figure 4-2). Increasing distance away from home reduces the mode share for walking (51% for ≤ one mile and 27% for ≤ three miles) while simultaneously increasing trips made by other modes like transit and private vehicles. Several of these short-distance trips are tours in more complex traveling such as trip chains by automobiles to several destination stops, walking to access transit trip chains and walking to access parked vehicles (Litman, 2012).

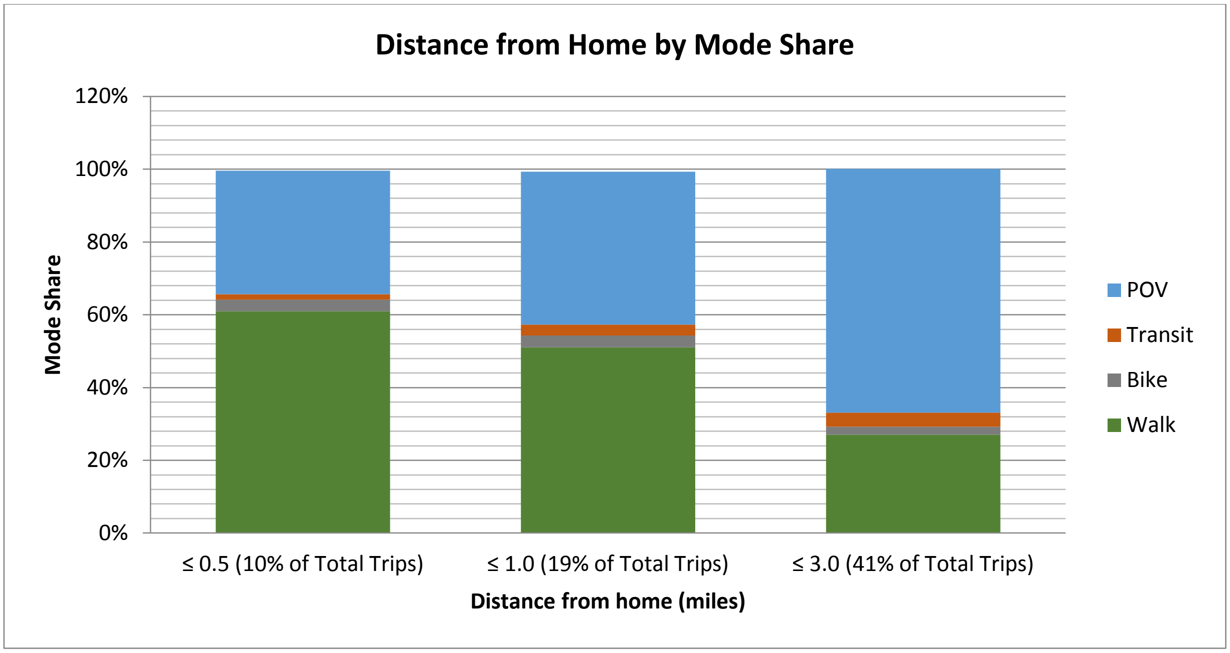


Figure 4-2: National Average Distance from Home by Mode Share
*POV: Privately Owned Vehicle
(Source: Adapted from Table 3 (Litman, 2012) obtained from NHTS Data)

Furthermore, there have been numerous studies linking place 'walkability' to non-motorized travel. Researchers often subdivide environments into tiers for promoting active transport (high walkability) and those that deter it (lower walkability). The consensus has been that more walkable locations generally promote more non-motorized travel (Cerin et al., 2009; L. D. Frank et al., 2008; Frank, Schmid, Sallis, Chapman, & Saelens, 2005; Owen et al., 2007). Moreover, twice as many people (37% vs. 17%) who live in highly walkable environments compared to those residing in low walkable ones meet the daily recommended moderate physical activity threshold of ≥ 30 minutes/day (Frank et al., 2005). Denser neighborhoods featuring more amenities attract on average an additional 60 minutes per week of transport-related walking and 75 minutes per week for leisure walks (Adams et al., 2011). Area walkability is comprised of the smart growth characteristics including density, diversity, design, destinations and distance to transit. A synergy may occur among all these built environment variables which results in greater non-motorized travel after controlling for socio-demographic factors (L. D. Frank et al., 2008).

A few studies focused on built environment correlates to TPA away from home. One research performed factorial analysis of neighborhood factors and concluded that smaller parcels (including grocery type stores and amenities serving basic needs) attract higher walking levels (Lee & Moudon, 2006b). Another study found that pedestrian friendly locations featuring a diversity of recreational amenities differ by an additional 13 minutes of daily MVPA than their counterparts (Adams et al., 2011).

Further, studies in the public health and active living literature have used matched GPS[6]-accelerometer[7] data to examine MVPA levels during "active" commuting (walking/cycling) (Cooper et al., 2010; Duncan et al., 2009; Oliver et al., 2010) and to identify the land use, green space, and walkability of areas associated with greater MVPA (Krenn, Titze, Oja, Jones, & Ogilvie, 2011; Quigg, Gray, Reeder, Holt, & Waters, 2010; Rainham et al., 2012; Rissel, Curac, Greenaway, & Bauman, 2012).

Overall, a review of the current literature shows some important trends in examining the context or 'where' physical activity and TPA occurs in relation to the physical environment and public health. About 90% of current studies that examine the link between cardio-metabolic risk factors and the built environment only considered the immediate surroundings to the respondents' residence (Leal & Chaix, 2011); although about 60% of MVPA has been found to

---

[6] For location tracking
[7] To measure physical activity intensity

be expended farther than one kilometer away from both home and work locations of respondents (Troped, Wilson, Matthews, Cromley, & Melly, 2010). Table 4-1 provides a sample list of current studies and locations where TPA was analyzed. This indicates a gap in the literature where the trends have been to focus only on immediate surroundings to the residential settings. In contrast, the street segments evaluated in this current chapter are in the vicinity of study participants' residence (within 0.5 mile) and encompass the entire Expo study area which includes commercial activity centers, public transit stations and green spaces. Thus, a multitude of locations are examined to better assess the environmental context of TPA. Moreover, the available studies have also assessed the influence of daily environmental exposures on individual MVPA, but the proposed analysis is the first "place-based" study to identify segment-level factors associated with MVPA. Results of this chapter are intended to inform efforts and public policy officials to transform street environments to promote walking and to achieve the goals of SB 375.

TABLE 4-1: CURRENT STUDIES OF 'WHERE' PHYSICAL ACTIVITY OCCURS

Residential Locations	Green Spaces	Urbanicity (urban/suburban/rural)	Area Walkability	Away From Home
Boarnet et al., 2011	Almanza et al., 2012	Rainham et al., 2012	Adams et al., 2011	Lee & Moudon, 2006
Frank et al., 2005	Giles-Corti et al., 2005	Rodriguez et al., 2007	Cerin et al., 2009	Troped et al., 2010
Hoehner et al., 2005	Kaczynski & Henderson, 2007		Frank et al., 2005	
Jones et al., 2009	Lachowycz & Jones, 2011		Frank et al., 2008	
Leal & Chaix, 2011	Rodriguez et al., 2005		Owen et al., 2007	
Rodriguez et al., 2005	Wheeler et al., 2010		Werner et al., 2010	
Saarloos at al., 2009				
Saelens et al., 2003				
Sallis et al., 2006				

4.2.2 TPA, the "Active" Commute, and Public Transit Access

Non-motorized travel along commute routes has been associated with higher MVPA levels. For example, a cross-sectional study of school children observed a 43% increase in the mean accelerometer counts per minute along their school commutes in comparison to counts measured at playgrounds; which amounted to twice the activity levels (Cooper et al., 2010).

TPA includes intermediary trips that lead to public transit hubs, and the general consensus in the literature is that there is a positive correlation between overall TPA, MVPA, and transit access after controlling for socio-demographic factors. A cross-sectional study using multivariate analysis for the 2001 NHTS dataset found that on average transit users walked 19 minutes daily to/from their stations (Besser & Dannenberg, 2005). This effect was more pronounced among low-income minority groups although the sample underestimated lower income households because NHTS surveys are generally administered by phone and lower income groups may not have phone access. Another study found that train commuters walked on average 30% more steps per day compared to car users, an effect amounting to four times the likelihood of physical activity for train users versus private vehicle users (Wener & Evans, 2007).

Similarly, some longitudinal studies have found a noticeable increase in TPA associated with transit use as opposed to vehicular travel. Activity bouts were noticeably higher for transit riders and the obesity rates among new and continuing riders were significantly lower (26% and 15% respectively) compared to non-transit riders (65%) (Brown & Werner, 2007, 2008). These authors noted, however, that the TPA observed was not enough to meet the national physical activity guidelines[8] and that the effects of the new light rail station may have been underestimated because of another pre-existing rail station farther away and the lack of neighborhood parks and leisure spaces (Brown & Werner, 2007). In addition, a recent review of the literature pertaining to transit-related physical activity found that public transportation

[8] Adults are recommended to complete 150 minutes of weekly moderate intensity activity or 75 minutes per week of vigorous intensity exercise (U.S. Department of Health and Human Services, 2014).

commuters walk on average between 8-33 additional minutes per day than their non-commuting counterparts (Rissel et al., 2012).

Increasing investments in transit-oriented developments (TODs) may potentially have a substantial impact on TPA. A key assumption for these developments is the walkability of the surrounding environment. Many have argued that pedestrians are willing to walk more than the standard half-mile radius from their residence in order to reach the desired destinations (Canepa, 2007) provided there are well-connected streets and sidewalks and a diversity of uses along the way (Guerra, E. & Cervero, 2013).

Again, their findings were limited to the neighborhood scale and none have inspected actual routes taken by pedestrians. This proposal aims to fill this gap in the literature by understanding the environmental context of the first/last leg of transit-oriented trips and will help identify for policymakers urban design features that could be enhanced in smart growth communities to encourage walking and transit ridership.

4.2.3 The Environmental Context of Walking at the Segment-Level

Most studies of the environmental context of walking in the public literature have examined the factors associated with individual-level TPA behavior or MVPA outcomes (Adams et al., 2011; Almanza et al., 2012; Badland et al., 2010a; Boone-Heinonen, Gordon-Larsen, Guilkey, Jacobs, & Popkin, 2011; Brown & Werner, 2007; Cerin et al., 2009; Cerin, Leslie, du Toit, Owen, & Frank, 2007; Chaix et al., 2013; Giles-Corti, Timperio, et al., 2005; Oliver et al., 2010; Rissel et al., 2012) or neighborhood-level factors associated with TPA behavior (Boarnet, Forsyth, Day, & Oakes, 2011; Boarnet et al., 2010; Canepa, 2007; de Nazelle et al., 2011; Frank et al., 2005;

Greenwald & Boarnet, 2001; Handy et al., 2002; Schlossberg & Brown, 2004); but an alternative approach is to conduct place- or roadway segment-level analysis to identify the context-specific factors which make particular blocks and roadway segments more conducive for walking. Since built environment improvements are often implemented at the street level, this segment-level of aggregation will help more directly identify the characteristics of the urban streetscape which is associated with greater TPA and could help inform and focus efforts to tailor street improvements to promote more active communities.

Recent studies have demonstrated that matched GPS-accelerometer data provides highly spatially resolved information on daily locations and travel patterns/routes. These studies have examined MVPA levels during "active" commuting (walking/cycling) (Cooper et al., 2010; M. J. Duncan et al., 2009; M. Oliver et al., 2010; Troped et al., 2008) and to identify the land use, green space and walkability of areas associated with greater MVPA (Cooper et al., 2010; Krenn et al., 2011; Quigg et al., 2010; Rainham et al., 2012; Rissel et al., 2012; Rodriguez et al., 2005).

These studies assess the influence of daily environmental exposures on individual-level MVPA outlines, but the proposed analysis is the first "place-based" study to identify roadway segment-level factors associated with TPA and MVPA. It extends recent studies which have used GPS-based locational data matched with accelerometer-based physical activity (PA) monitoring to examine when and where PA occurs through moment-by-moment analysis of the behavioral context of PA (Chaix et al., 2013). They provide valuable insights into the behavioral contexts of physical activity, and generally suggest that higher daily rates of moderate-to-vigorous physical activity (MVPA) for children and youth are associated with greater daily exposure to green space in near-home environments and spaces occupied during daily activities (Almanza et al., 2012; Lachowycz & Jones, 2011; Rodríguez et al., 2012).

The proposed segment-level analyses will contribute to this growing “place-oriented” literature which uses matched GPS-accelerometer data to understand “where TPA occurs”. Whereas other studies have examined the amount of MVPA which occurs in parks and green space, this is the first study which (1) examines the roadway segment-level factors associated with TPA and MVPA and (2) assesses these relationships for an adult sample, and (3) investigates the role of transit access on the level of TPA and MVPA on roadway segments. Previous studies may have examined these relationships in different geographic settings for an adult sample (Badland, Duncan, Oliver, Duncan, & Mavoa, 2010b; Houston, 2014; Mitra & Buliung, 2012) but they did not unravel these links at the roadway segment-level. Resulting “place-based” findings will inform efforts to transform street environments to promote walking and to achieve the goals of SB 375.

4.3 Contributions of the Chapter

Most of the studies reviewed so far were performed on a neighborhood geographic scale but the objectives of this chapter require a fine-grain street-level analysis.

Many authors agree that the accuracy of TPA estimates obtained from individual or household level analyses, surpasses those resulting from analyses on the block-group level (Cerin et al., 2009; Handy et al., 2002). In addition, large parcel sizes were observed to be deterrent to pedestrians, whereas smaller lots of retail and food establishments were found to encourage more walking (Lee & Moudon, 2006b). With this in mind, the spatial scale of this dissertation will provide a more detailed analysis of distance gradient effects to and from transit stations and the street-level characteristics associated with greater TPA.

Some studies have also largely neglected the importance of the built environment which augments the positive synergistic effect between public transit and TPA. They have focused mainly on how the built environment enhances neighborhood walkability in order to account for facets which they believed to have a confounding effect if neglected (Badland et al., 2010a; Duncan et al., 2009; Lachapelle et al., 2011; Owen et al., 2007). One study found that 10 % of the variations in walking trips were explained by built environment elements, the 3 Ds + R (destinations, distance, density & routes) after controlling for socio-demographic characteristics, but neglected to investigate the role of neighborhood design and only relied on stated preference not objectively-measured data (Lee & Moudon, 2006b). This research on the other hand, fills this gap in the literature by including proxies for all the five Ds of the built environment (density, diversity, design, destination accessibility & distance to transit) and the link to physical activity especially as the mediating effect of transit is explored.

In this chapter, I aim to examine street routes from a "place-based" perspective which will add valuable insights into our understanding of the environmental context of TPA travel. I hypothesize that aesthetically pleasing routes bordered by high concentrations of green spaces and routes leading to transit stations with commercial and retail activity centers are more conducive to pedestrian travel.

In addition, the analysis in this chapter was performed at a fine level of analysis which I believe can be more precise in quantifying impacts of major public transportation investments on travel behavior. Moreover, implementing smart growth tools along transit corridors could potentially lead to an increase in non-motorized travel over passenger vehicle use which helps accomplish SB 375 goals.

4.4 Methodology

In this section, I compare methods of calculating the various land use variables in four types of catchment areas outlining differences among them. The fourth is the preferred method used in this chapter, the street segment buffers; which will be discussed in detail as well as the steps taken to construct it. In addition, I discuss how the unit of analysis was created, the street segment-level which was based on the 40-meter segment buffers and the main variables used in the chapter. Finally, I include the model specification for the binary logit followed by an overview of the methodology (Figure 4-7) utilized.

4.4.1 Comparison of Different Land Use Calculation Methods

There are generally four spatial assessment techniques to measure the built environment using a geographic information system such as ArcGIS. Because of the variation in their geographic scale, these techniques consequently differ in their impacts on travel behavior. Differences and limitations of each method are outlined next.

4.4.1.1 Census Tract or Neighborhood-level Measurement

This method implements a large geographic scale such as census tracts, or neighborhood census block groups or even census blocks (Choi, Wang, Delgado, & Ryu, 2007; Handy, 1993) to measure the built environment. This method may be more suitable for regional analyses than for localized travel impact assessments because it is less sensitive to a more focused area.

Therefore, this geographic scale may not fully capture the effects of the built environment on active travel as it is too large to be an appropriate 'walkable' environment. In addition, this method of land use calculation is also less accurate for locations in the boundaries or outskirts as opposed to the centers thus injecting bias into the land use measurement estimates (Oliver, Schuurman, & Hall, 2007).

4.4.1.2 Circular Buffers

This technique utilizes a more detailed spatial analysis approach whereby circular buffers of various sizes are created. Usually, an individual's place of residence or work location is first geocoded and used as the center of these buffers (Chaix et al., 2013; Duncan, Aldstadt, Whalen, & Melly, 2012; Owen et al., 2007) depending on the outcome variable measured (e.g. non-work vs. work travel behavior) and for which the immediate microenvironment is assessed. Even though this approach is tailored to the traveler's immediate environment; it is still inaccurate for areas that have natural obstructions (e.g. canyons, lakes, etc.) or infrastructural barriers (e.g. 'big box' retail type uses, airports, etc.) (Oliver et al., 2007). The reason for this is, these land use types are still accounted for in the measurements but the traveler may not be able to traverse these areas easily or even not at all.

4.4.1.3 Land Use Polygon-based Measurement

This measurement technique uses the pre-ordained land use polygon or parcel as the basis for the land use calculation. This technique may still not be suitable to assess active travel behavior since a pedestrian may only be exposed partially to these uses such as store fronts on the sidewalks or walkways in a residential area, but he/she may not be affected by the full size of the land use. Therefore, a single predominant land use type in a microenvironment may "skew" the results where it would end up overestimating its effect on the traveler's behavior (Oliver et al., 2007). This is especially true for large land use parcels. Figure 4-3 shows an example from the Expo study area where large land use polygons of an industrial use and another for a high-density single family residential area may overestimate the effects that these types of developments may have in the analyses. Their large polygon structure could exert more weight in contrast to a smaller polygon or parcel.

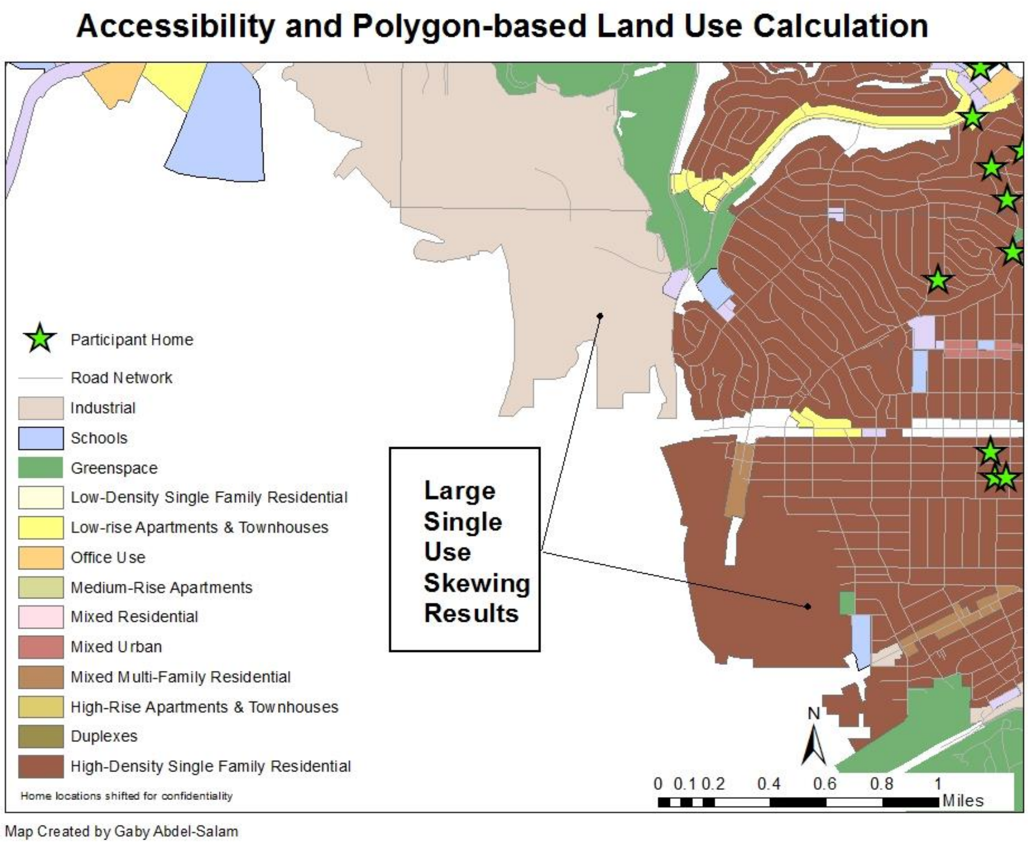


FIGURE 4-3: LARGE LAND USE PARCELS INFLUENCING PEDESTRIAN BEHAVIOR

4.4.1.4 Road Segment-based Buffers

This approach addresses the shortcomings of the above methods and assesses the built environment on the road segment-level. In addition, because segment-based measures involve path choice analysis, unlike other methods; the likelihood of self-selection bias is diminished since these paths are usually not correlated with home and work location preferences as they are in zone-based measures (Guo, 2009).

Since the main interest here is on active travel, the road segment-level buffer method was utilized in this chapter because it was deemed the most sensitive to a pedestrian's experience since it is implemented at a very fine spatial scale. I followed the approach outlined in Oliver et al. (2007) in creating the road segment-based buffers with a few alterations. Similar to the Oliver et al. (2007) study which used road segments within 950 meters (0.5903 miles) from the respondent's postal code, I selected the road segments within 0.5 miles (904.672 meters) from respondents' homes in the Expo study. This decision was based on previous planning and public health literature on the suitable extent of a 'walkable' environment (see Guerra, E. & Cervero, 2013). Further, in contrast to Oliver et al. (2007) who used 50 meter buffer sizes around the road segments; I utilized 40 meter buffers in this chapter to capture the proportion of land use parcels that fall within this buffer extent and which represents the immediate contextual walking environment of a pedestrian. This spatial scale therefore better characterizes the experience of an active traveler during his/her walk trip ignoring inaccessible uses along the route and is less sensitive to larger parcels whose weight may be overestimated in the analyses. As explained by Oliver et al. (2007), a 100-meter buffer would have been too large as it may have included larger parcels that would skew the land use proportions or even inaccessible pedestrian uses. In contrast, a 25 meter buffer would have been too small whereby essential built environment features may have been overlooked especially along wider streets e.g. streets with more than three lanes (Oliver et al., 2007).

Figure 4-4 below depicts three of the above most commonly used land use calculation methods. The blue half-mile circular buffer is the most commonly used method to estimate land use proportions but as shown in the example, there are many areas that do not have a road network and therefore may not have sidewalks for pedestrians inside the circle. This is probably because of the underlying natural feature that obstructs the road connectivity. In comparison, the pink dashed polygon shape in the figure represents a high-density residential use which extends outside of the half-mile radius or 'walkable' range for this respondent and which is a large parcel (almost two-thirds of the half-mile buffer) that would positively skew the estimate on residential use. Finally, the green 40 meter road segment buffers shown below was the preferred method utilized in this chapter and it provides the finest level of spatial analysis that more accurately depicts the pedestrian's contextual environment and the respective influence of the built environment along these segments.

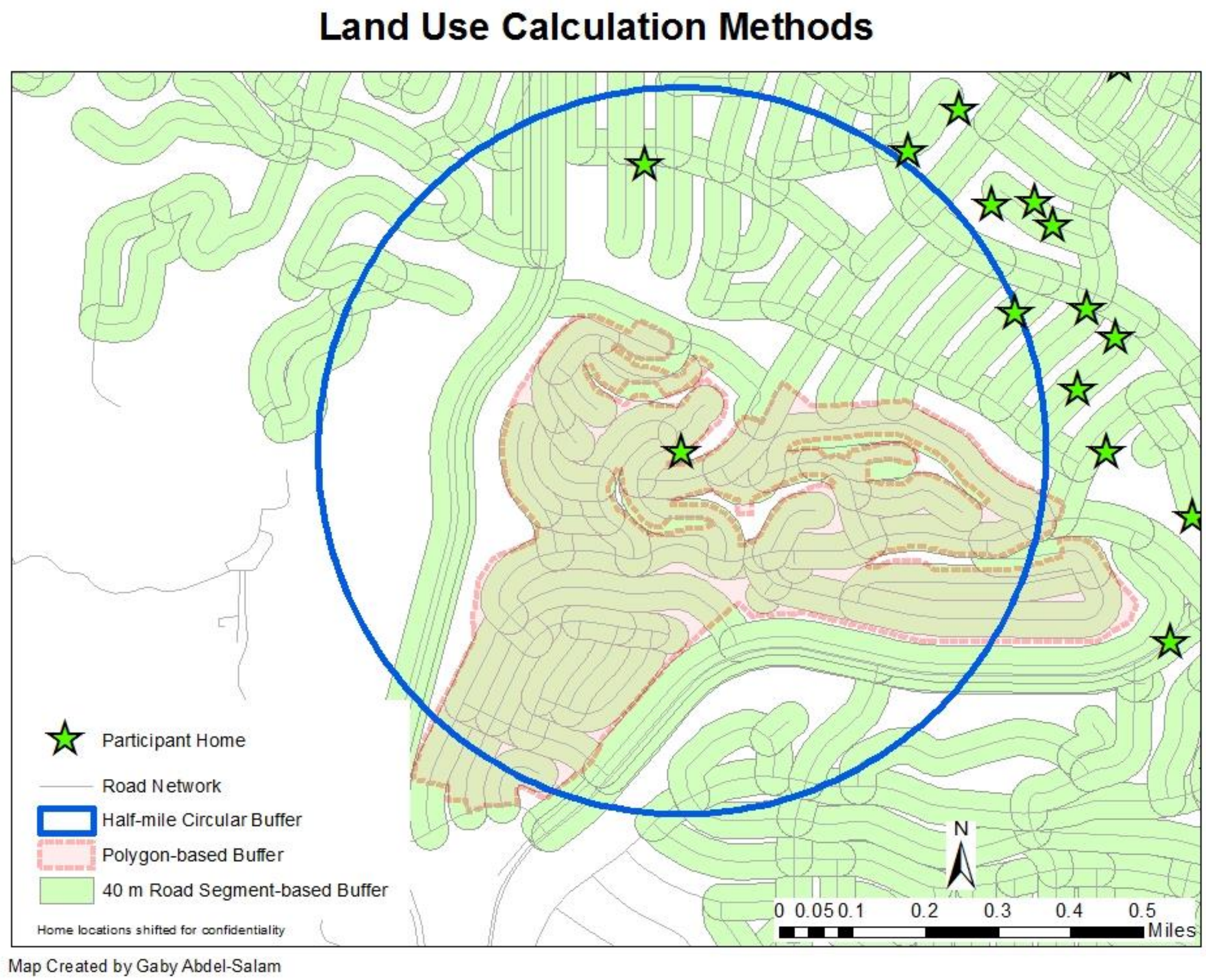


FIGURE 4-4: CIRCULAR-BASED VS. POLYGON-BASED VS. SEGMENT-BASED BUFFERS

4.4.2 Creating the Unit of Analysis (Road Segments)

All analyses in this chapter are performed on the road segment level for Phase One of the Expo study. As explained in the earlier section, segment-based buffers were created in ArcGIS to depict the pedestrian's microenvironment. First, the survey respondents' home locations were geocoded to the existing 2010 TIGER street network file for the Expo study area. Using the Expo respondents' home locations as a point of origin; I created circular buffers of two different sizes (one-mile and a half-mile radius) around each address point. Road segments falling completely within the buffers were extracted from the existing California TIGER road network for the Expo study area.

In addition, it was important to only include in the analysis the road segments in Los Angeles County to match the addresses of the survey respondents. This was a challenge since the street network is continuous across counties and do not stop at county boundaries. Therefore, I could not use partial road segments or subdivide them to extract only those in LA County. Instead, a flag variable was created and used later in the regressions taking on a value (= 1) if the road segment was in LA County and equal to zero otherwise and was based on the 'CountyFP' code from the TIGER street file.

After preliminary analyses and reviewing existing literature, the road segments in the one-mile radius buffers (within one-mile from Expo respondents) were eliminated in favor of the half-mile radius set of road segments. This is consistent with the current literature that recommends that at least 60% of the catchment areas for walking to be within a one-quarter mile from home (five-minute walking distance) to a half-mile (ten-minute walking distance) in the case of transit station access (Canepa, 2007). Therefore, following previous research, all further analyses in this chapter will focus on road segments within a half-mile radius from Expo respondents' homes.

The next step was to create 40 meter buffers around each of the extracted road segments. One reason for this spatial size choice is that it was previously found to have an influence on the incidence of physical activity along the paths. In addition, the 40 meter buffers and slightly larger buffer radii are hypothesized to better capture the immediate built environment features along a pedestrian's route (Oliver et al., 2007).

The 40 meter segment buffers were used to calculate the various land use types inside this catchment area. A comprehensive land use database shapefile was obtained via SCAG for the year 2005 and loaded into GIS. The resulting shapefile is now a segment-level file for Expo road segments within half-mile from participants. The database file (.dbf) for this new shapefile was then exported and merged by household I.D. with the original segment-level dataset.

The maps in Figure 4-5 and 4-6 show an outline of the road segments extracted and the segment-based buffers in the one-mile and half-mile radii respectively. Figure 4-5, shows that a total of 8,374 road segments were extracted from the underlying road network but were not used in further analyses. All analyses in this chapter were based on the 5,649 road segments shown in Figure 4-6 which were also extracted from the Expo road network that met the half-mile radius criteria.

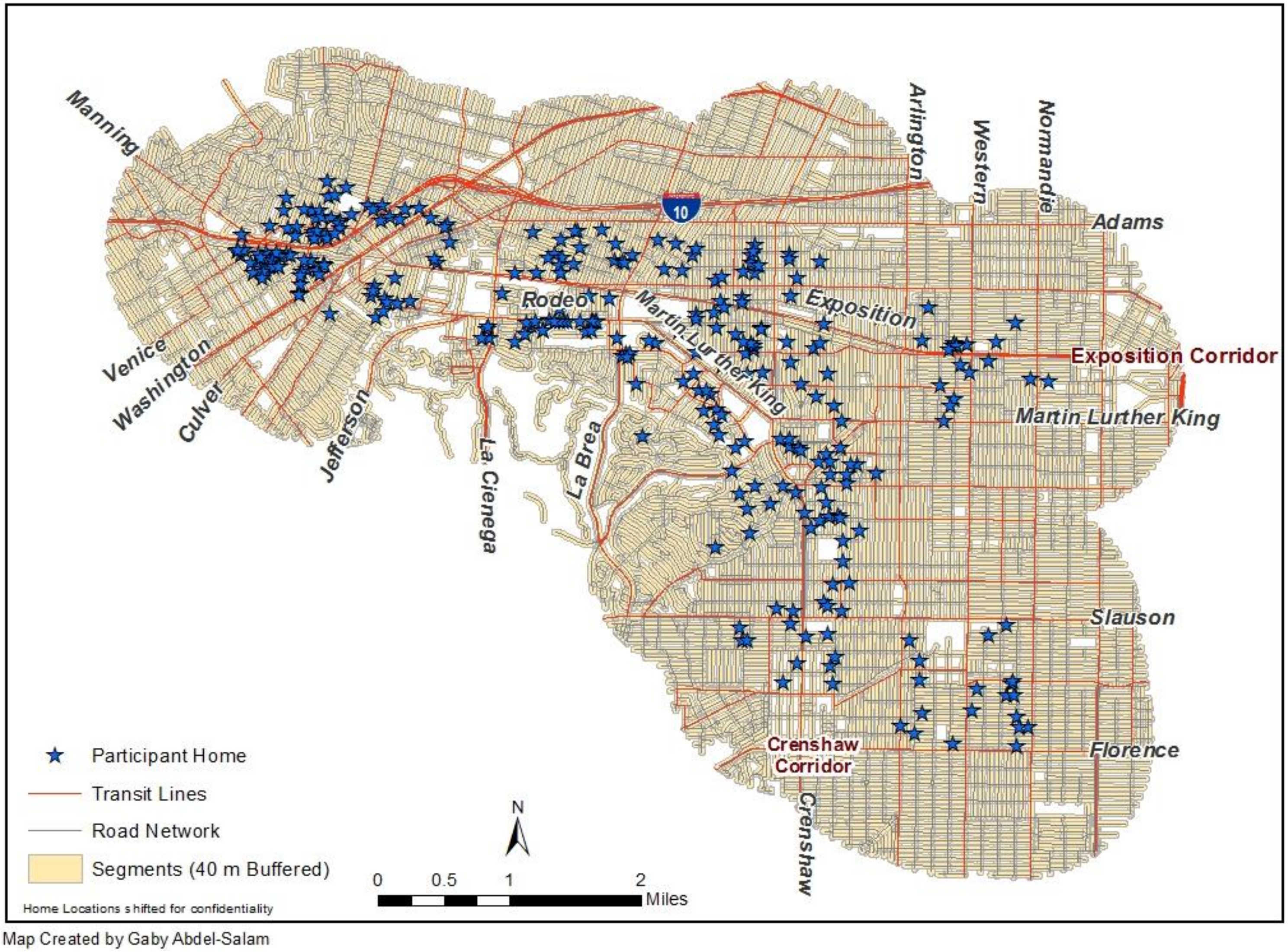
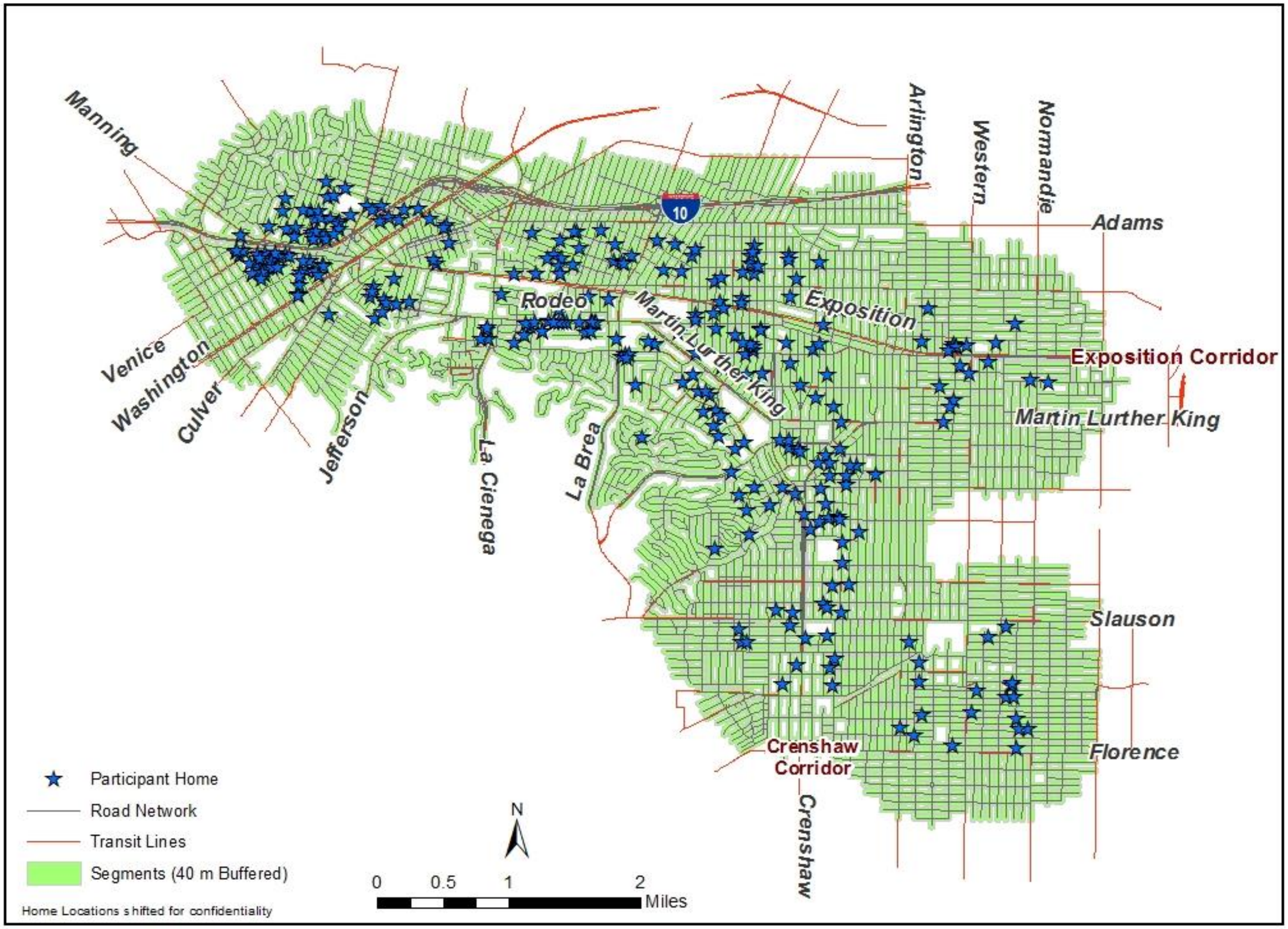


Figure 4-5: Initial Buffered Road Segments Extracted One-mile Radius (N = 8,374)



Map Created by Gaby Abdel-Salam

Figure 4-6: Buffered Road Segments Used In The Analyses One-half mile Radius (N = 5,649)

4.4.3 Key Variables

4.4.3.1 The Outcome Variables

As mentioned earlier, the dataset used in the analyses is comprised of road segment-level information for the land uses, traffic volume, transit stop information and road attributes. To create the outcome variables, this segment-level dataset was merged with the matched accelerometer-GPS momentary level dataset to extract the travel mode status.

Two dependent variables were generated to represent non-motorized travel and TPA. The first is a dichotomous variable that was generated to represent incidences of walking (including walking to transit). This variable is a proxy for active travel or TPA and is binary (= 1 if walking was observed on any road segment; = 0 otherwise). This outcome variable was extracted from the travel mode status field coded during the manual GPS review process explained earlier.

The second outcome variable analyzed was average daily minutes of MVPA measured at the segment-level. This variable was computed using the accelerometer data measured by the Actigraph device worn by the survey participants. Raw accelerometer data were processed via the statistical software R and Meterplus (software that cleans and rates vertical and lateral movements into a physical activity spectrum ranging from sedentary to vigorous activity). Resulting values were saved in a database .dbf format and exported to a GIS platform and joined to cluster shapefiles holding information of location, status (mode of travel), and a date and time stamp (in 15-second intervals). To create daily averages of MVPA, this variable was summed over the valid 24 hour period as explained in section 3.2 (excluding non-wear time and invalid data points).

The shapefile now holding a count of daily MVPA minutes at the participant level (for participants in the mobile tracking group) needed to be transformed to the street segment-level. In GIS, the street segment shapefile based on the TIGER data was created and joined to the shapefile with participant physical activity information. This resulted in a one-to-many array whereby for each street segment, more than one participant may have walked on the same segment per day due to the proximity of the households in the Expo study to one another. Therefore, the MVPA values were aggregated once more over each street segment using the segment I.D. as an identifier.

4.4.3.2 Creating the Land Use Variables

As mentioned earlier, land use data were based on SCAG parcel database for 2005. This includes data on type of use, year of aggregation, geometric area, and the standard land use code utilized by SCAG. This database also includes information for many counties however, only land use data for the Expo study area were extracted and used in the GIS platform. The "geoprocessing" feature in GIS was used to combine each of the 40-meter buffered road segments with SCAG's land use database using the "intersect" tool. This resulted in the creation of new buffers which are now populated with the relevant buffer size information, coordinates (with Expo home as center) and the respective land use polygon information.

These shapefiles still retained the old values for polygon area and perimeter measurements and therefore these values needed to be recalculated because of the new intersection with the 40 meter buffers. The new calculations were completed via the "calculate geometry" tool and the geometric unit for "area" was selected as acres. Another field was also added to compute the proportion of each land use type. This step was necessary to obtain the densities of each land use type utilized in later regressions.

4.4.3.3 Green Space Variables

Green space was defined for census block-groups in the study area using GIS technology and aerial photographs. These variables were calculated from two-feet resolution maps based on QuickBird remote sensing and aerial photography orthographic images (McPherson et al., 2011). After the creation of these variables, a distribution was calculated for the two extreme types of land covers and based on it, percentiles were calculated. From the resulting categories, two opposing dichotomous variables were created: the first = 1 for the 90th percentile of green space (= 0 otherwise), and therefore representing the greenest land cover and the second = 1 for the 90th percentile of impervious land or extreme bare land cover that may include street medians and sidewalks (= 0 otherwise).

4.5 Model Specification

The regression models include two dependent variables: a dichotomous variable for incidence of walking on any segment and a count variable denoting segment-level average daily MVPA minutes. A binary logit regression was utilized for the former outcome variable and a negative binomial was deemed appropriate for the latter variable. The decision for the regression type for the latter variable was chosen after reviewing the descriptive statistics and distribution of this discrete dependent variable. The distribution of the MVPA (Y) variable was not a normal distribution. It was skewed to the right and therefore this does not justify the use of an OLS regression. Although an OLS regression would provide similar trends, direction of associations and significance values; the extreme positive skew of the MVPA variable could result in biased estimates and values may not converge to Y = 0. Therefore, the assumption of normal distribution under the OLS regression may be relaxed under the negative binomial or Poisson regressions. It is also important to note that a Poisson regression was initially performed on this outcome variable but since its variance exceeds the mean, an indication of overdispersion; the negative binomial specification was utilized instead.

4.5.1 Binary Logit for Incidence of Walking

The method used for the first outcome variable is a simple logistic regression estimable by maximum likelihood and it is interchangeably called a binary logit because the dependent variable, *WalkFlag*, is dichotomous. The dependent variable (= 1) if incidences of walking (in general or to transit stops) were observed on any road segment in the Expo study area (Phase 1 data) and (= 0) if no incidences of walking was detected.

Using an ordinary least squares (OLS) method would probably yield the correct signs and significance of the coefficients however, I did not utilize it here because of three reasons:

- The errors of this regression are heteroskedastic, i.e. the variance of the errors of the dependent variable is different at varying values of the independent variables, which is a violation of one of the OLS assumptions.
- The distribution of the error terms is not normal.
- The resulting predicted probabilities of the outcome variable can exceed 1 or can be lower than 0 which violates the dichotomous definition of the variable.

Since the outcome variable Y is dichotomous, we can therefore use a binary logit regression to estimate the probability of *WalkFlag* (Y) occurring on any road segment. *Y* is an (n x 1) response vector; therefore, we model $p = \Pr(Y = 1)$ as in equation (1) below in reduced form:

$$log\left[\frac{p}{(1-p)}\right] = \beta_0 + BE\beta_1 + G\beta_2 + T\beta_3 + e \quad \text{--------------------- (1)}$$

β_0 is an (n x 1) vector of the intercept values, β_1, β_2 and β_3 are (m x 1) vectors of the regression coefficients, e is an (n x 1) vector of random errors.

The independent variables are: *BE* [an (n x m) matrix of the built environment characteristics], *G* [an (n x m) matrix for a "green space spectrum" contrasting the greenest level at the 90th percentile and non-green impervious parcels]; *T* [an (n x m) matrix representing the traffic volume and a flag for medium and high traffic].

The normal assumptions of equation (2) are: expected values of the errors are zero $E[e] = 0$ and their variance and covariance is as follows: $\text{var}(e) = \sigma_e^2$, and $\text{cov}(e_j, e_{j'}) = 0$ for $j \neq j'$. The last term means that we assume no covariance for the error terms.

Equation (1) is estimable by full maximum likelihood (ML). All models were estimated by the PROC LOGISTIC command for non-linear response variables using SAS 9.2 software.

4.5.2 Negative Binomial for Segment-level MVPA Minutes

The second outcome variable *MVPA* or (Y) is a discrete count variable for daily average segment-level minutes of MVPA. This variable is an aggregate of all moderate-to-vigorous physical activity performed during walking periods per day on each segment.

MVPA is an (n x 1) variable with mean (μ); we can use a negative binomial regression to estimate $log(\mu)$ as in equation (2) below in reduced form:

$$log(\mu) = \beta_0 + SD\beta_1 + BE\beta_2 + G\beta_3 + T\beta_4 + e \text{ -------------------- } (2)$$

β_0 is an (n x 1) vector of the intercept values, β_1, β_2, β_3 and β_4 are (m x 1) vectors of the regression coefficients, e is an (n x 1) vector of random errors.

The independent variables are: *SD* [an (n x m) matrix of the main participants' socio-demographic traits], *BE* [an (n x m) matrix of the built environment characteristics], *G* [an (n x m) matrix for a "green space spectrum" contrasting the greenest level at the 90th percentile and non-green impervious parcels]; *T* [an (n x m) matrix representing the traffic volume and a flag for medium and high traffic].

The normal assumptions of equation (2) are: the expected values of the errors are zero $E[e]=0$ and their variance and covariance are as follows: $\mathrm{var}(e) = \sigma_e^2$, and $\mathrm{cov}(e_j, e_{j'}) = 0$ for $j \neq j'$. The last term indicates that we assume no covariance for the error terms.

Equation (2) is estimable by full maximum likelihood (ML). All models were estimated by the PROC GENMOD command, the generalized linear model for discrete variables using SAS 9.2 software.

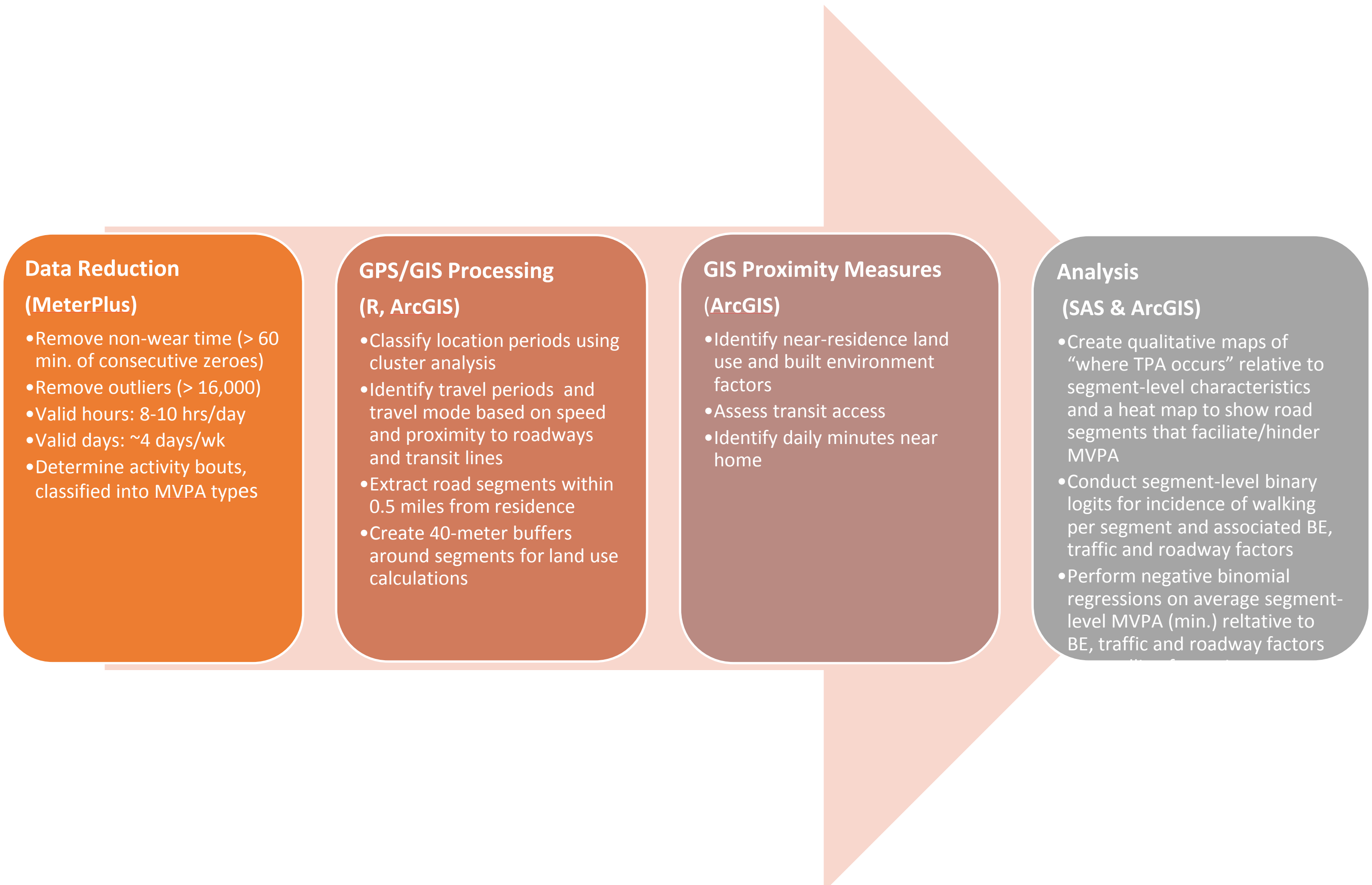


FIGURE 4-7: METHODOLOGY OVERVIEW

4.6 Model Results

In this section, the relationship between the incidence of walking and MVPA were explored using descriptive statistics and qualitative maps of where MVPA occurs. Multivariate analyses were also performed as binary logit and negative binomial regressions to find the probability of walking or the incidence of walking (including to transit) per segment and to predict the impacts of the independent variables on segment-level MVPA. In addition, odds ratios were also calculated to define the magnitudes of each of the built environment features, road segment characteristics and the traffic-related variables on MVPA and the incidence of walking.

4.6.1 Descriptive Statistics

Descriptive statistics were initially performed for key variables as a preliminary analysis. This included a combination of box plots, distribution plots and pie charts. All the variables were transformed to the segment-level of analysis which included aggregations from household level (for socio-demographics), parcel level (some land use variables) and the one-minute date and time stamps of the matched accelerometer-GPS data points (physical activity and active travel variables).

Socio-demographic characteristics for the segment-level sample participants were explored and the results are shown in the pie charts of Figure 4-8. Almost half of the sample (43%) is low-income earning a household income of $35,000 annually or less. Over half of the participants (56%) are of African-American decent and the second largest race (31%) is the White group. Unfortunately, 59% of the sample chose not to state their education level and only 32% having a high school diploma or higher. Moreover, 57% of the sample is female, 66% are unemployed and the average age is 59 years.

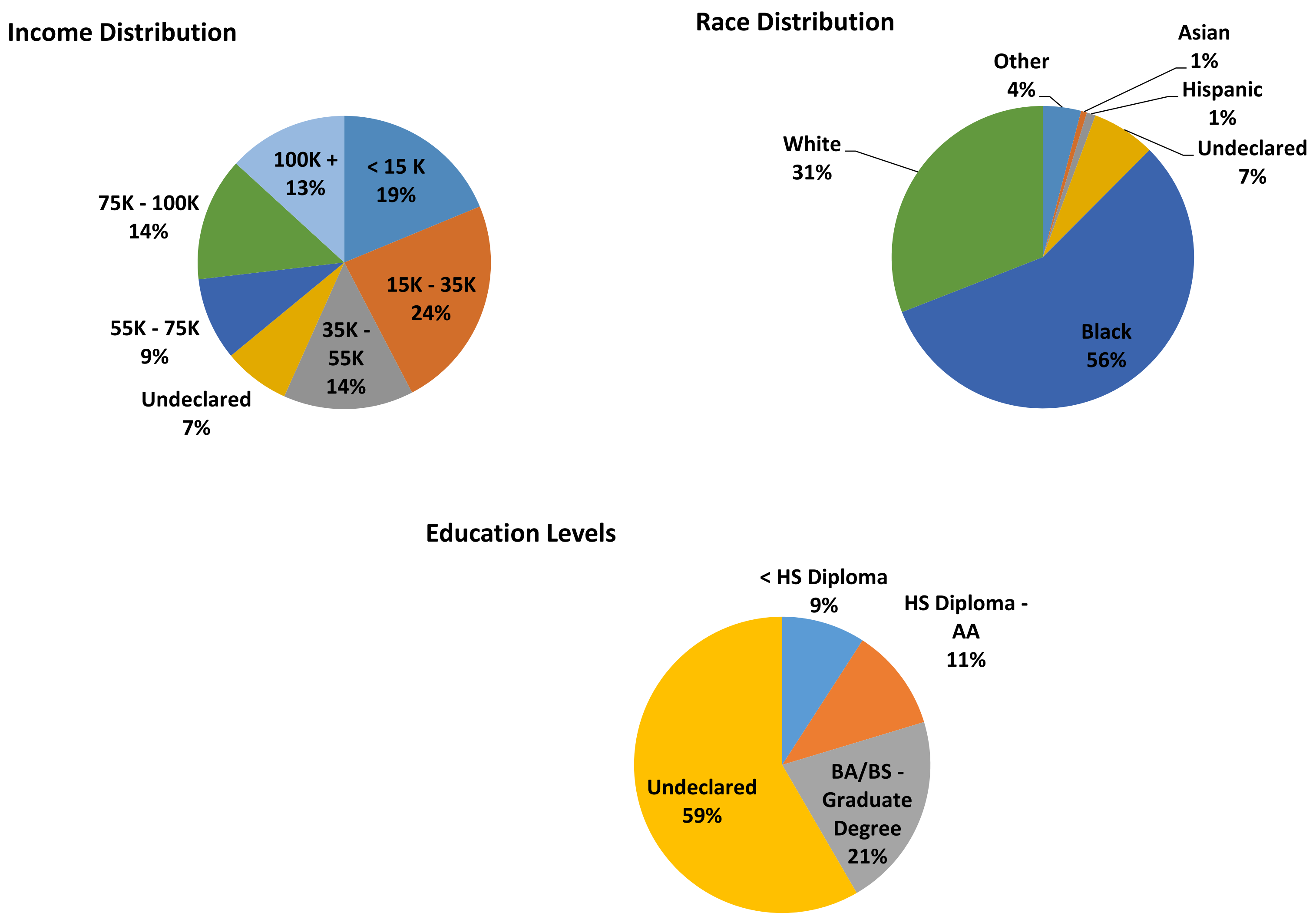
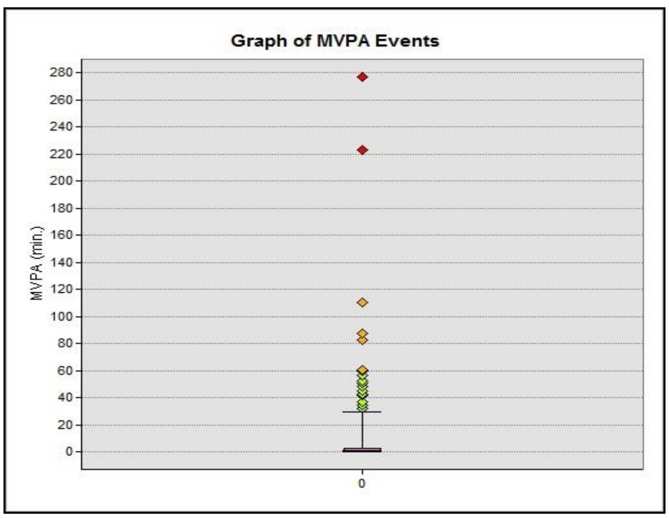


Figure 4-8: Socio-demographic Characteristics of Participants (N = 954)

Figure 4-9 is a box plot for the outcome variable MVPA (min.) measured at the segment-level. The mean of this variable (should be in the center of the box) is 4.61 minutes per segment. The light green and yellow diamonds at the end of the whiskers are mild outliers in the sample and they start from about 30 to about 110 minutes of MVPA per segment. The red diamonds are the two extreme outliers: one above the 220-minute mark and the other just below 280 minutes.



Mean	4.61
Variance	297.13
Min.	0
Max.	276.80

Figure 4-9: Box Plot of Segment-level MVPA Minutes (N = 954)

Figure 4-10 shows the distribution of the same outcome variable showing that it has a discrete positively skewed distribution. This variable is also over-dispersed since the variance is much greater than its mean. This fact was also confirmed in the negative binomial regression results in Table 4-5(a) and Table 4-5 (b) and under the *Dispersion parameter* heading which will be discussed further under section 4.6.5 Goodness of Fit Measures.

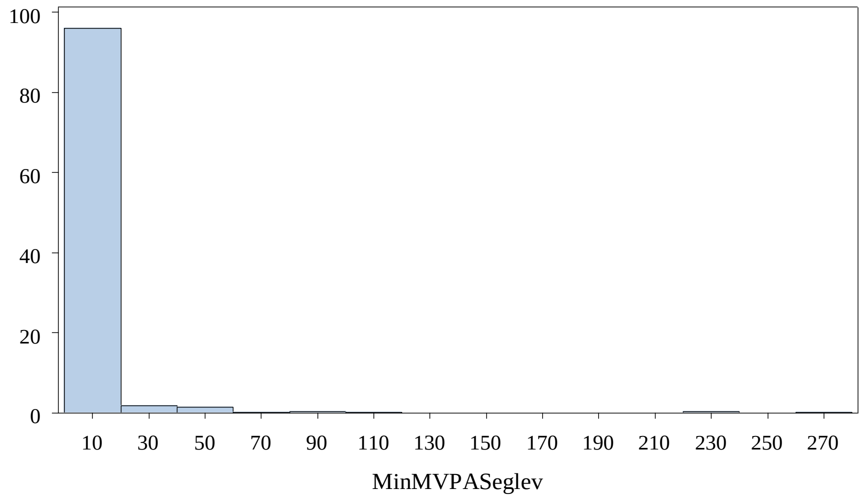


Figure 4-10: Distribution of Segment-level MVPA (min.)

Distribution plots were also created for different land use variables as shown in Figure 4-11. The plots show variables that reflect the commercial density, a variable combining commercial and retail uses, residential and neighborhood employment densities. The combined variable includes retail stores and restaurants. It is important to note that the first two land use variables were never used together in the regression analyses because they overlap in many entries and are therefore highly correlated; instead their use was alternated in the models.

All the land use variables are skewed to the right. The majority (60%) of the road segments has zero or very small percentages of residential and neighborhood employment densities and 35% or the segments have almost no commercial or retail uses.

The flag variable for traffic volume is dichotomous (=1) if medium-to-high traffic is observed per street segment and (= 0) otherwise. The majority (64%) of the street segments of the Expo study area have medium-to-high traffic.

The graphs of Figure 4-12 display the road segment features: the road segment length, density of street intersections, density of unclassified parcels and the distribution of the number of transit stops. These road features are all skewed to the right except for the unclassified parcel variable which has a fairly normal distribution.

Figure 4-13 shows the percentile distribution of the green space and impervious land cover variables. There are three green space variables: tree density, irrigated and non-irrigated lawn densities. In contrast, the impervious land cover density is only one variable which is at the opposite end of the green space spectrum. The distribution of the green space variables is positively skewed to the right while the distribution of the impervious land cover variable is the opposite, skewed to the left.

The distributions of Figure 4-13 provide some important insights into the aesthetic appeal of the Expo study road segments. For example, the majority of the road segments (92%) are mostly impervious or concrete leaving very little space for green areas. In contrast, only 3% of the road segments have: a tree density between 25% and 51%, irrigated lawn between 35% and 43.5% and a density of 10% non-irrigated lawn. These observations provide an incentive for policymakers in the urban planning and public health fields to increase the green space densities especially in light of recent studies that concluded a positive relationship between active travel (and/or physical activity) and green spaces (Coombes, Jones, & Hillsdon, 2010; de Nazelle et al., 2011; de Nazelle, Rodríguez, & Crawford-Brown, 2009; Quigg et al., 2010; Rainham et al., 2012).

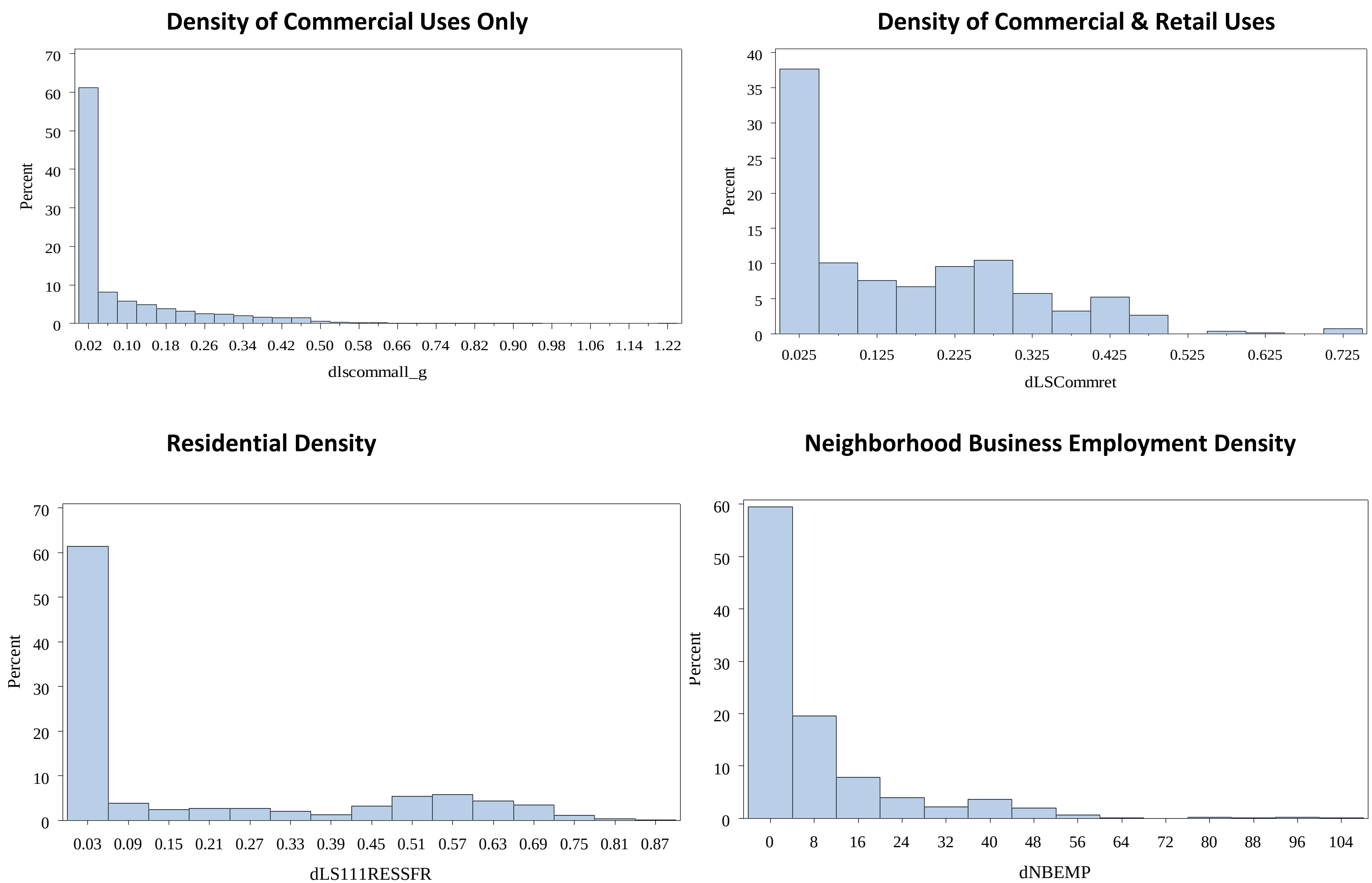


Figure 4-11: Density Distributions of Commercial Uses and Commercial with Retail Uses

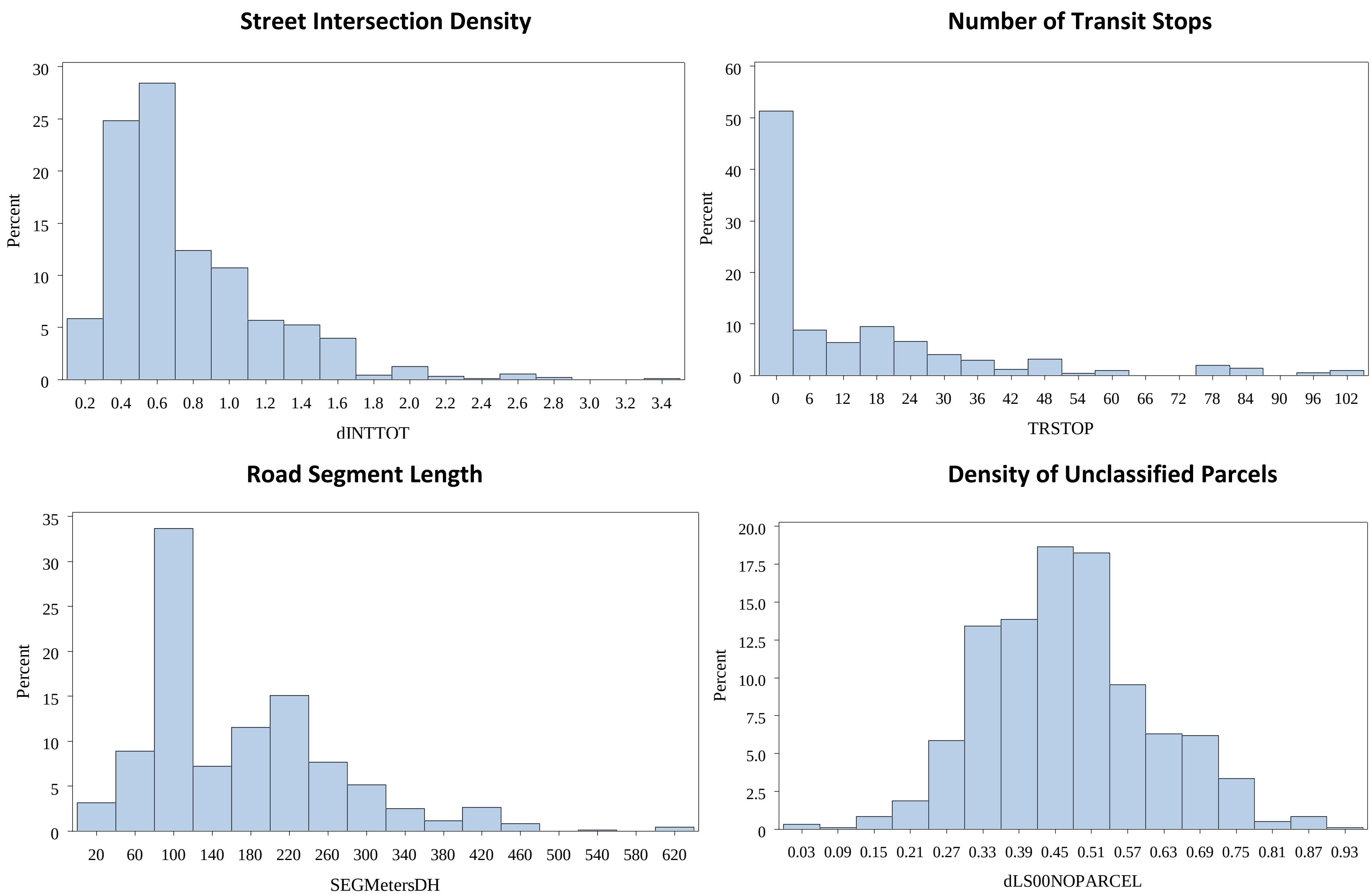


Figure 4-12: Road Characteristics

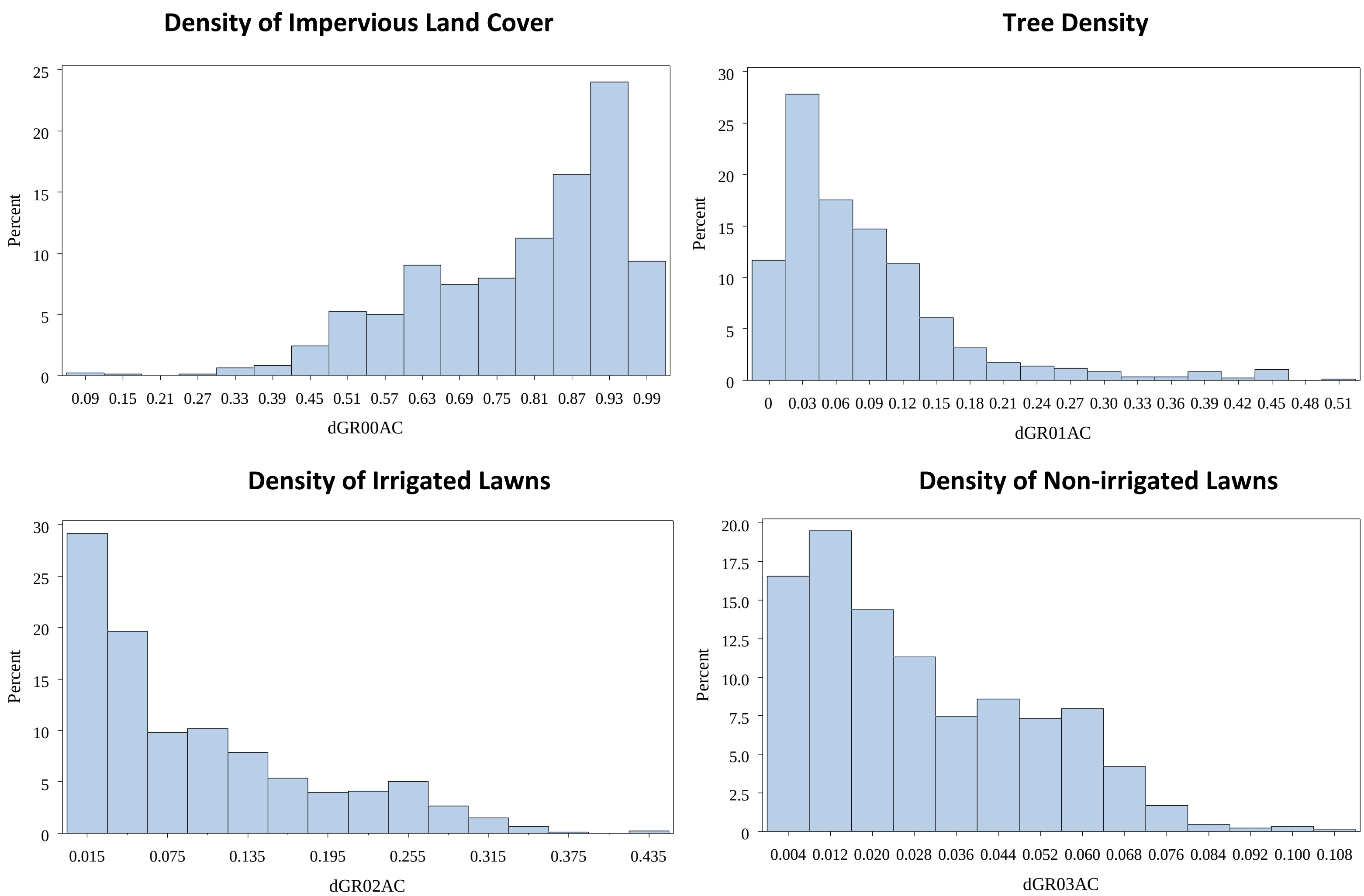


FIGURE 4-13: DIFFERENT LAND COVER DENSITY DISTRIBUTIONS

4.6.2 Land Use and MVPA Levels

The map in Figure 4-14 shows varying densities of the commercial and retail uses in the Expo study area. Darker colors reflect higher densities of this land use type and are visible along major roadways going North-South such as Western, Crenshaw, La Brea and La Cienega and East-West roadways like Slauson, Washington, Culver and even some pockets along Exposition Boulevard. Commercial and retail uses are hypothesized to be positively correlated with active travel.

Figure 4-15 is a diffusion interpolation map of the segment-level MVPA (min.) incidences. The underlying Expo roadway network was used as the basis where roadway segment length was used as a barrier to MVPA. Diffusion interpolation is a geostatistical method that uses distance as the 'cost' or barrier from the available data to calculate a raster figure of the cost of moving from one cell/grid to the adjacent one. Therefore, predictions in areas that have unknown/unidentified data points are interpolated and predicted from neighboring cells. I used the additive barrier default formula to calculate this distance. The result is a heat map reflecting areas that act as barriers to increased MVPA minutes (cooler colors) and those that facilitate MVPA (warmer colors).

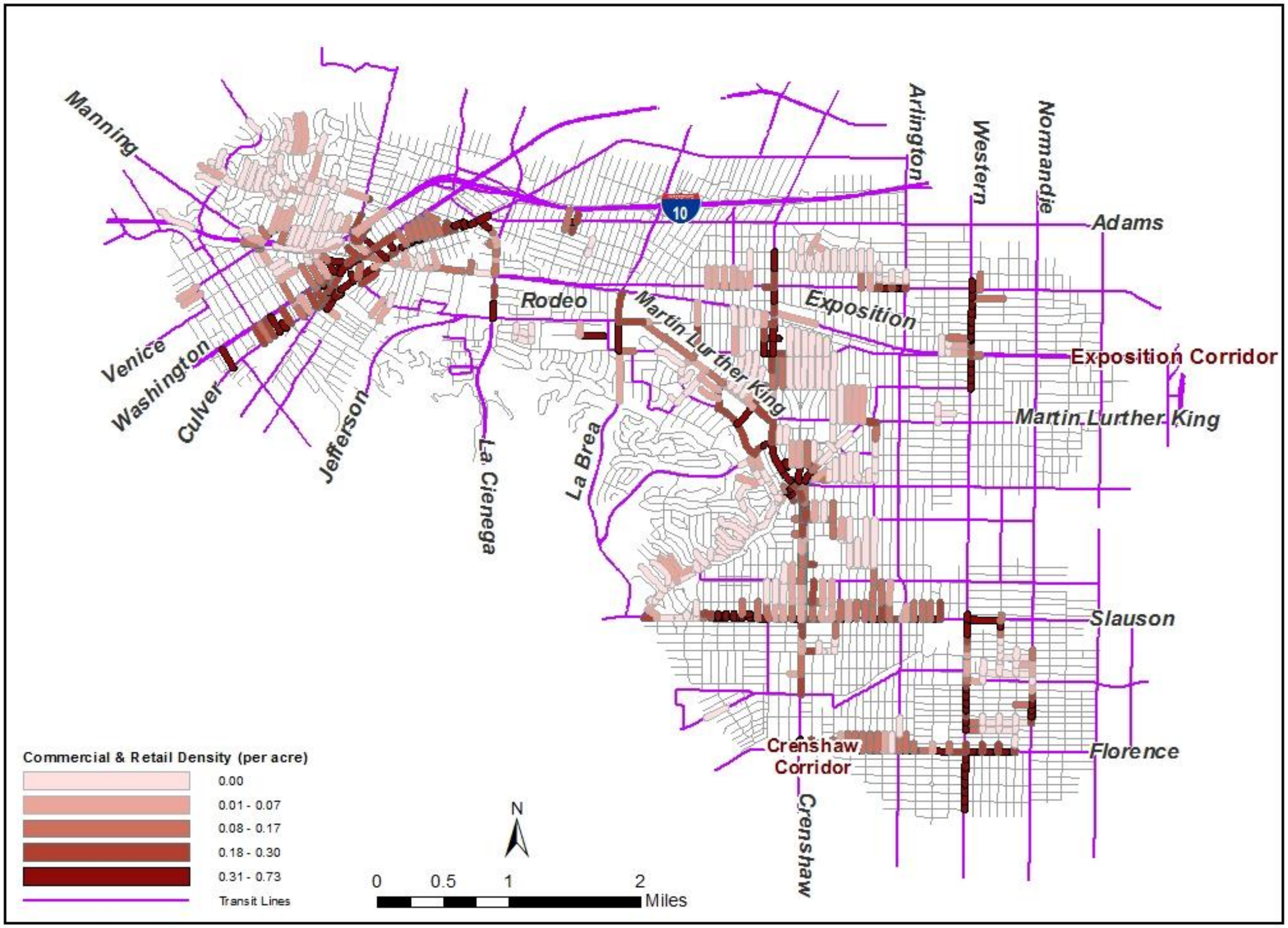
Based on this map, we can see that areas beyond our study area (half-mile road segments) which we did not have MVPA data for have interpolated results calculated from the existing data set. For example, the two areas south-west of Slauson and Florence streets have green and blue raster colors respectively indicating that they would be expected to be barriers to increased MVPA periods.

Further, the map shows the locations of the road segments that reveal increased periods of MVPA during walking. These are shown as pockets of oranges and reds (5-270 minutes of MVPA) especially in the center of the map along Martin Luther King Boulevard and adjacent to La Brea and even close to Exposition near La Brea. In addition, there are smaller pockets of yellows and oranges (5-36 minutes of MVPA) between Washington and Culver Boulevard to the Northwest of the map.

Comparing the maps in Figures 4-14 and 4-15 we can see a positive correlation between higher MVPA levels and higher commercial/retail densities. This is especially true along Martin Luther King Boulevard and between Washington and Culver Boulevard.

It is important to point out that this diffusion interpolation map was based on Phase 1 data of the Expo study (prior to the Expo Light Rail opening). Therefore, many road segments especially along the Exposition corridor were undergoing construction which may have directly affected the decreased levels of physical activity periods along them. In addition, the small sample size contributed to smaller episodes of total segment-level physical activity which resulted in smaller observed densities of the MVPA spectrum.

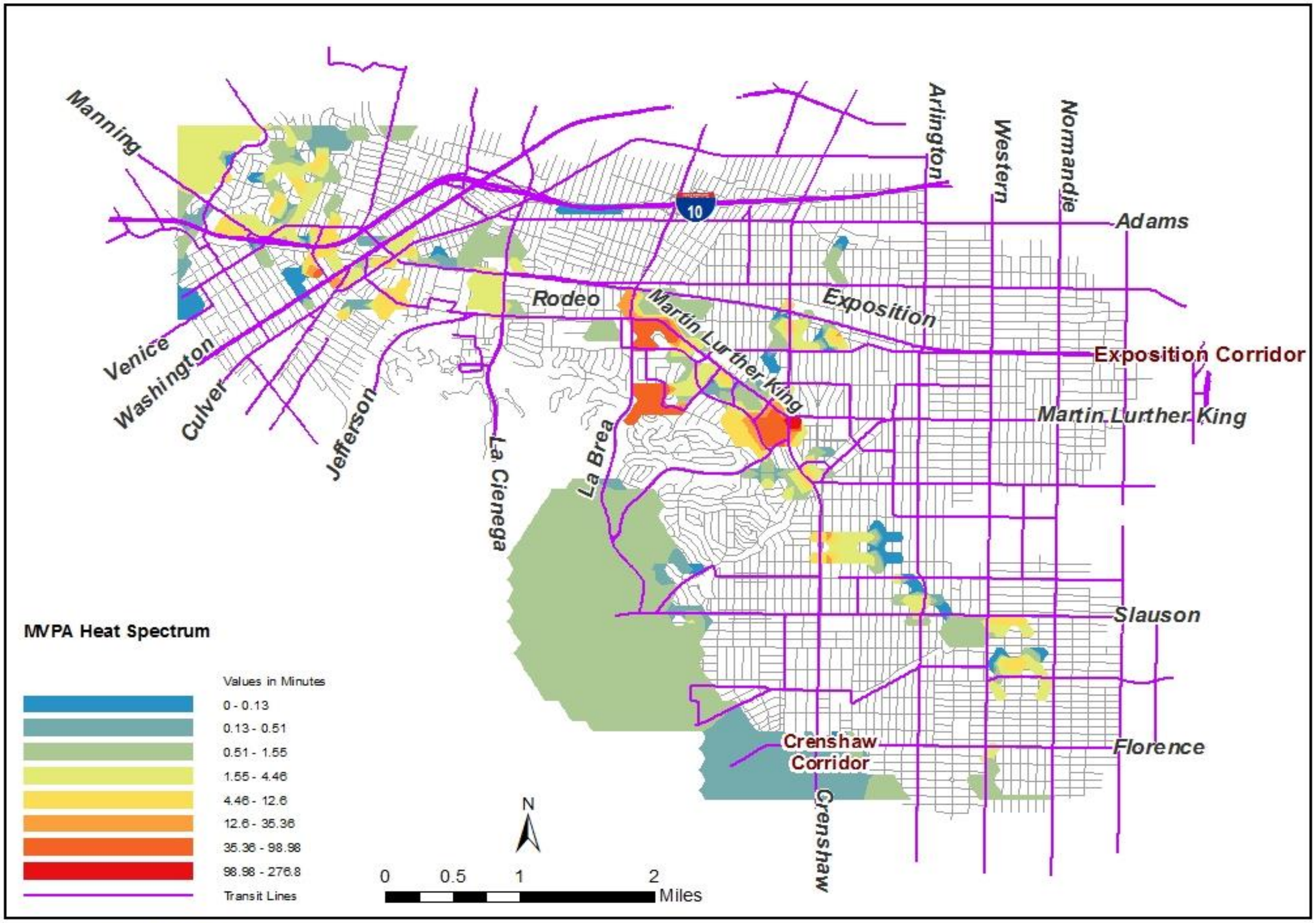
Commercial & Retail Density for 'Walked' Segments



Map Created by Gaby Abdel-Salam

Figure 4-14: Commercial & Retail Density

Segment-level MVPA Heat Map



Map Created by Gaby Abdel-Salam

Figure 4-15: Diffusion Interpolation of MVPA (Min.) & Road Segment Barrier
*Cooler-colored roads show less MVPA & warmer-colored roads show more MVPA levels.

4.6.3 Sample Point-Level GPS-accelerometer Patterns

The maps in this section are qualitative showing examples of the matched GPS-accelerometer data points along different road segments in the Expo study with a satellite image as the background. The data points of each map are also symbolized to reflect the different physical activity levels for two of the sample participants.

The map in Figure 4-16 shows a sample participants' spectrum for the MVPA levels as well as the location "where" MVPA occurs. Similarly, Figure 4-17 is a map for another sampled participant with his/her corresponding MVPA levels during a trip to access transit. The latter map shows that elevated MVPA levels (moderate) overlap with his/her route to transit. This observation has been confirmed by many recent studies which stress the role that public transit might play in encouraging active travel (Besser & Dannenberg, 2005; Morency, Trépanier, & Demers, 2011; Wasfi, Ross, & El-Geneidy, 2013).

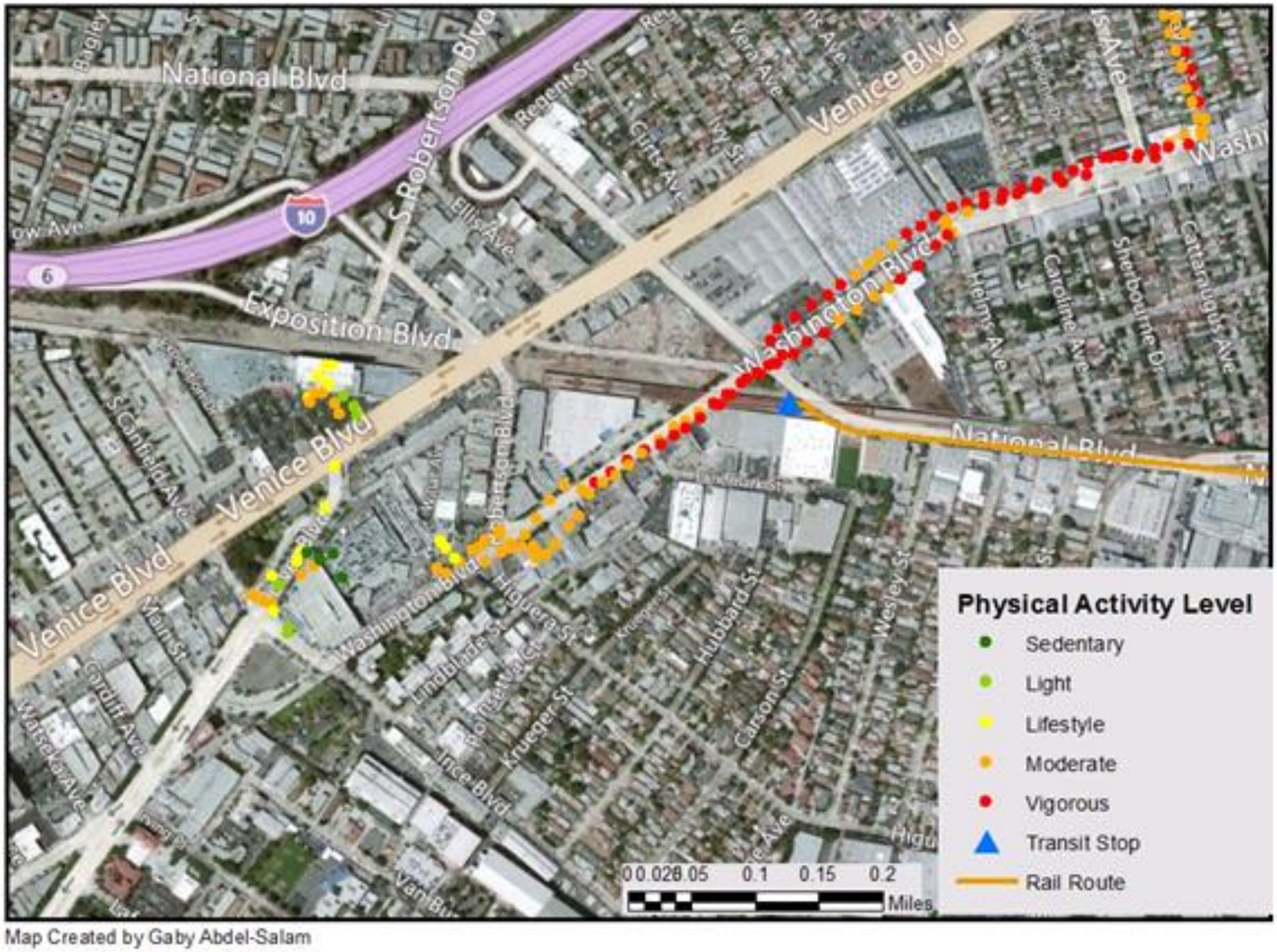


Figure 4-16: Locations Of Physical Activity (MVPA) Spectrum

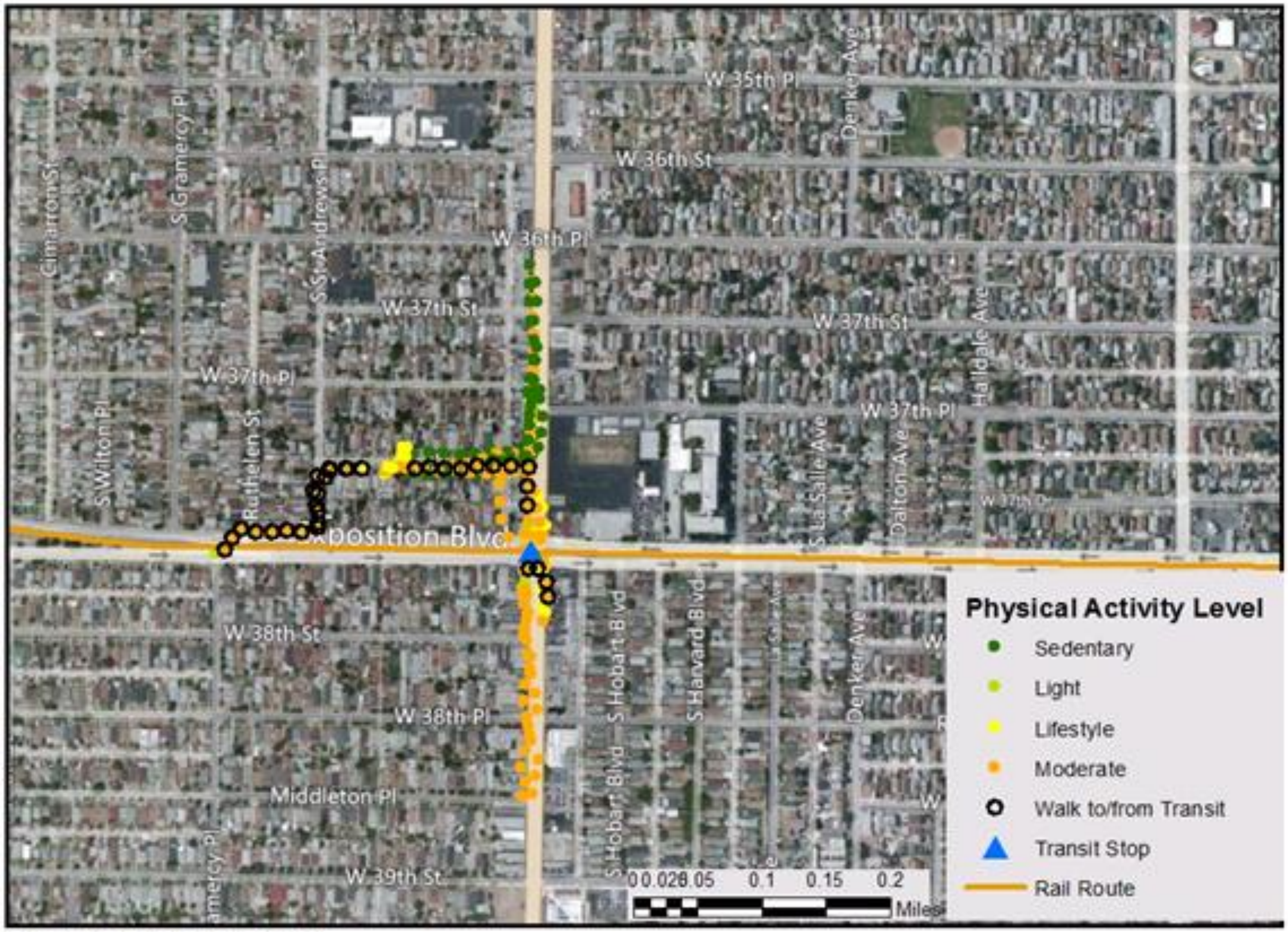


Figure 4-17: Physical Activity (MVPA) & Transit Access

4.6.4 Characteristics of Segments with Walking

Results in Table 4-2 below include the descriptive statistics and significance t-tests for the key variables in the data set. These variables were also used in analyzing the incidence of walking in the logit regressions of Tables 4-4(a) to 4-4(c). The table compares the means and standard deviations for 'all segments' (N = 5,649); these comprise all the extracted road segments in the 0.5-mile radius from EXPO participants' homes and 'segments with an incidence of walking' (N = 1,052).

A test for significant differences between the values of both columns was also performed and the results are reported on the right side of the table. The single sample t-test evaluates whether or not the population means of the variables of the 'segments with walking' are equal to the means of the variables of 'all segments' in the sample. Thus, it tests the null hypothesis (H_0) of equal means. For simplicity, only significant t-tests will be discussed below. Highlighted rows in Table 4-2 indicate higher means and significant t-tests for the 'Segments with Walking'.

Segments with walking tend to have more: commercial uses and commercial and employment densities. These segments also had higher densities of: unclassified parcels, impervious land cover, and higher percentile of impervious land cover. Further, segments with walking incidence had more: medium to high traffic, transit stops, transit stops standardized by road segment length and average total walking minutes relative to all segments. Residential density and uses and densities of: trees, irrigated and non-irrigated lawns on 'all segments' were greater than on 'segments with walking'. Finally, there were no significant differences between 'all segment' and 'segments with walking' among the street segment lengths, industrial uses, densities of: street intersections and 90th percentile green space.

TABLE 4-2: DESCRIPTIVE STATISTICS & T-TESTS FOR ROAD SEGMENTS AND WALKING

Variables		All Segments (N = 5,649)		Segments with Walking (N= 1,052)		T-Test
	Variable Code	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Sig.**
Industrial Uses	LSIndAll_G	0.13	0.60	0.12	0.38	
Residential Uses	LSResAll_G	1.73	1.50	1.33	1.37	***
Residential Density	dLS111RESSFR	0.26	0.27	0.20	0.25	***
Employment Density	dNBEMP	3.11	9.11	6.32	13.53	***
Commercial Uses	LSCommAll_G	0.30	0.49	0.58	0.69	***
Commercial Density	dLSCommAll_G	0.08	0.13	0.15	0.17	***
Density of Unclassified Parcels	dLS00Noparcel	0.44	0.16	0.46	0.14	***
Density of Non-Irrigated Lawn	dGR03AC	0.04	0.02	0.03	0.02	***
Density of Irrigated Lawn	dGR02AC	0.13	0.10	0.11	0.09	***
Density of Trees	dGR01AC	0.11	0.08	0.10	0.09	***
Density of Impervious Land Cover	dGR00AC	0.71	0.15	0.76	0.16	***
90th Percentile Impervious Flag	dGR00AC90flag	0.10	0.30	0.16	0.37	***
90th Percentile Green Flag	greenflag90	0.07	0.26	0.08	0.27	
Transit Stops Count	Trstop	3.35	8.55	7.72	13.71	***
Transit Stops (normalized by segment length)	Trstopseg	0.05	0.31	0.08	0.25	***
Density of Street Intersections	dINTTOT	0.85	0.75	0.86	0.58	
Segment Length	SEGMetersDH	141.10	103.10	140.70	89.15	
Medium-to-High Traffic Flag	CTMedHI	0.40	0.49	0.58	0.49	***
Minutes of Walking (incl. to transit)	MINWLKW2T	0.66	4.51	3.56	10.07	***

Significance: * p < .1, ** p < .05, *** p < .01

Similarly, the results in Table 4-3 display the descriptive statistics and significance t-tests for the key variables in the data set used in analyzing the minutes of moderate-to-vigorous physical activity (MVPA) in the negative binomial regressions of Table 4-5(a) and Table 4-5(b). The table compares the means and standard deviations for all segments that underwent any walking incidences (N = 933); these were extracted from the road segments in the 0.5-mile radius from Expo participants' homes and the segments with an incidence of walking where participants exerted high levels of MVPA on them (N = 557).

The results of the single sample t-tests between the means in each column are reported on the right-hand side of the table. This is a test of significant differences between all the segments in the sample with walking versus those that attracted higher total MVPA minutes that were exerted during walking episodes. For simplicity, only variables with significant t-tests will be discussed below. The highlighted rows in Table 4-3 indicate higher means and significant t-tests for the 'Segments with High MVPA during Walking'.

Segments with walking and had higher MVPA tend to have more: employment, commercial and retail densities. These segments also had a higher density of impervious land cover, higher levels of traffic, longer street segments and more transit stops. Further, on average, an additional 4.3 minutes of MVPA was observed on the road segments with higher MVPA during walking episodes over those with walking in general. Residential density and densities of: trees, irrigated and non-irrigated lawns were greater on 'all segments with walking'. Finally, there were no significant differences between 'all segments with walking' and those with higher MVPA in the overall highest green space levels and street intersection density variables.

TABLE 4-3: DESCRIPTIVE STATISTICS & T-TESTS FOR ALL SEGMENTS VS. HIGH MVPA

Variables		All Segments (N = 933)		Segments with High MVPA during Walking (N= 557)		T-Test
	Variable Code	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Sig.**
Residential Density	dLS111RESSFR	0.17	0.25	0.14	0.24	***
Employment Density	dNBEMP	7.84	13.59	9.96	15.09	***
Commercial & Retail Density	dLSCommret	0.15	0.15	0.17	0.15	***
Density of Unclassified Parcels	dLS00Noparcel	0.47	0.14	0.49	0.14	***
Density of Non-Irrigated Lawn	dGR03AC	0.03	0.02	0.03	0.02	***
Density of Irrigated Lawn	dGR02AC	0.10	0.09	0.08	0.08	***
Density of Trees	dGR01AC	0.09	0.08	0.08	0.08	***
Density of Impervious Land Cover	dGR00AC	0.79	0.16	0.81	0.16	***
90th Percentile Impervious Cover	dGR00AC90flag	0.22	0.41	0.28	0.45	***
90th Percentile Green space	greenflag90	0.06	0.24	0.05	0.23	
Transit Stops Count	Trstop	12.55	19.55	17.71	22.64	***
Density of Street Intersections	dINTTOT	0.77	0.43	0.75	0.44	
Segment Length	SEGMetersDH	165.75	95.97	183.40	98.25	***
Medium-to-High Traffic	CTMedHI	0.63	0.48	0.71	0.45	***
One or more Segments with Medium/High Traffic	CTFlag	0.93	0.26	0.95	0.21	***
MVPA (min.)	MVPA	7.13	12.02	11.43	14.01	***

Significance: * p < .1, ** p < .05, *** p < .01

4.6.5 Goodness of Fit Measures

Results in Tables 4-4(a) to 4-4(c) show binary logit regressions for N = 5,649 road segments for the dependent variable *WalkFlag* (all roadway segments within ½ mile of participant home locations). Binary logit models utilize a dependent variable that is dichotomous taking on 0 and 1 values. All models have consistent results and expected signs among the variables but Models 6 and 7 were deemed the most suitable since all the independent variables with the exception of segment length and residential density were significant at least at the 10% significance level ($p < 0.1$). These two independent variables were insignificant in the models. The *AIC* (Akaike Information Criterion) value was the lowest as well in Model 7 at 4904.94. We generally want to select the model with the smallest AIC value. This criterion is given by:

$$AIC = n \ln SSE - n \ln n + 2p$$

Where *n* is the sample size, *SSE* is the Error Sum of Square and *p* is the number of independent variables included. The first term: $n \ln SSE$ will always decrease as *p* increases and the second term: $n \ln n$ is fixed. Finally, the last term *2p* will increase with *p*. Under this criterion, models with small *SSE* values mean better explanation of the variance in the dependent variables and generally do well with this criterion unless the penalty (*2p*) is too large.

The *max-rescaled R-square* value is comparable to the adjusted R-square goodness of fit estimate in ordinary least squares (OLS) regression. For Model 6 and 7, the max-rescaled R-square equals 0.1455 and 0.1492 respectively; meaning that the variables in the model jointly explain 14.55% and 14.92% of the variations in *WalkFlag* respectively, a value considered very high in travel behavior models.

Table 4-5(a) and Table 4-5(b) display the results of the negative binomial regression of the built environment characteristics, adjusted for socio-demographic traits, on segment-level MVPA minutes. The *Dispersion Parameter* ϕ, listed in the bottom section of the tables, measures the ratio of the residual deviance to the degrees of freedom. The values for this parameter are as follows: ϕ = 1.32 in Model 1, ϕ = 1.14 in Model 2, ϕ = 1.13 in Model 3 and ϕ = 1.128 in Model 4. Even though this parameter decreases as more variables are added to the model, the values still exceed one. Since this parameter is greater than one, we can conclude that the data is overdispersed and that the conditional variance increases more rapidly than the mean of the outcome variable. Initially, a Poisson regression was attempted to fit the data, however, since the *Dispersion Parameter* was greater than one for all the models; a negative binomial regression was favored over the Poisson.

The results also indicate an improvement in the goodness of fit measures in Models 3 and 4 over the other two models. The *AIC* value is the lowest in Model 4 = 5,592.46 and in Model 3 the *AIC* value is 5,711.13.

The *Pearson χ2* statistic is another goodness of fit measure and it is defined as:

$$\chi^2 = \sum_i \frac{w_i (y_i - \mu_i)^2}{V(\mu_i)}$$

Where w_i is the weight of observation i and in cases where it was not specified (as it is in this regression) the default is $w_i = 1$, y_i is the outcome variable at observation i and the variance function is given by: $V(\mu) = \mu(1-\mu)$. Simply stated, it is the summed ratio of the squared difference between the observed and predicted values to the variance of the predicted values.

Finally, the last goodness of fit measure reported is the *Scaled Deviance Value/DF* (ratio of the Deviance to the degrees of freedom), generally, if the model is specified well; this ratio should be very close to one and larger ratios indicate an ill-fitting model or that the outcome variable is over-dispersed. The value for this statistic in Tables 4-5(a) and 4-5(b) is very close to unity for all four models ranging from 1.05 to 1.07 confirming the good fit of the independent variables in the model.

Based on the above model fit measures, the independent variables of Models 3 and 4 seem to have the best fit. The measures for the *AIC* and the *Scaled Deviance/DF* ratio in these two models reveal their superiority over the other two.

4.6.6 Factors Associated with the Probability Walking Occurring on Any Segment

The results of the binary logit models of Tables 4-4(a) to 4-4(c) reflect the associations of the various built environment, green space and traffic variables on the incidence of walking per road segment. The independent variables were added incrementally moving from Model 1 to Model 7 (Tables 4-4(a to c)). For simplicity, I will review the significant factors of Model 3 (Table 4-4(a)) since it has a high max-rescaled R-Square (0.1173) indicating a good model fit. The R-Square value means that approximately 12% of the variations in the dependent variable (WalkFlag) were explained by the independent variables in the regression model.

The model coefficients are in log form so to understand their true impact on the dependent variable; the estimates need to be exponentiated and the results are noted in the odds ratio

(O.R.) column. The results show that roadways with unclassified parcels[9] and higher levels of green spaces are twice as likely to attract walking and therefore are ideal pedestrians. Also, road segments with higher impervious land cover (and therefore very little green space) decreases the chance of walking per segment by 31%.

The highest effects on the incidence of walking is from the commercial density variable which is consistent with the literature that TPA chances increase for routes with higher concentrations of activity centers with commercial and retail uses. Also consistent with the literature, additional transit have a positive impact on the incidence of walking. However, increasing street intersection density seems to lower the probability of walking on a segment by 18% which might be because the participants prefer calmer streets with fewer intersections. Finally, street segments that have medium to high levels of traffic are associated with higher incidences of walking (43% higher) which might suggest that these road segments have attractive destinations that attract both motorists and pedestrians.

Similarly, Table 4-4(b) and Table 4-4(c) include Models 4 to 7 which are variations to Model 3 (Table 4-4(a)) discussed above. The coefficient estimates of the models are consistent in significance and signs. For simplicity, Model 6 and 7 will be discussed in depth since they have the highest max-rescaled R-Square 0.1455 and 0.1492 respectively; and the lowest AIC 4921.12 and 4904.94 respectively. This is an indication of the best overall model fit and they also have the most comprehensive list of the built environment, green space and traffic volume factors.

[9] This variable includes non-parcel space represented by roadways, medians and sidewalks.

The variables for commercial density (Model 6) and commercial and retail density (Model 7) confirm the importance of activity centers in attracting pedestrians; their impacts on the incidence of walking is the highest among the remaining built environment features.

Road segments with the highest percentile of green space and highest unclassified land use density are almost twice as likely to attract pedestrians. In contrast, those with the highest percentiles of impervious land cover reduce the probability of walking by 31-33%. Longer street segments and higher neighborhood employment densities attract more pedestrians. Increased levels of vehicular traffic also increase the odds of walking per segment. Whereas, the greater the density of street intersections; the lower the probability of walking on a segment (O.R. 0.82 in Model 6 & 0.85 in Model 7).

TABLE 4-4(a): BINARY LOGIT FOR INCIDENCE OF WALKING ON ANY SEGMENT

	Model 1				**Model 2**				**Model 3**			
Dependent Variable	**WalkFLAG**[10]				**WalkFLAG**				**WalkFLAG**			
Independent Variables[11]	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**
Intercept	-1.955	<.0001			-2.105	<.0001			-2.812	<.0001		
Commercial Only Density	4.253	<.0001	***	70.32	4.352	<.0001	***	77.59				
90th Green Space Percentile	0.459	0.001	***	1.58	0.448	0.001	***	1.57	0.733	<.0001	***	2.08
90th Impervious Percentile	-0.092	0.436		0.91	-0.073	0.535		0.93	-0.209	0.085	*	0.81
Transit Stops per Segment	0.223	0.014	**	1.25	0.25	0.007	***	1.28				
Transit Stops												
Segment Length					0.001	0.004	***	1				
Density of Unclassified Parcels[12]									1.243	<.0001	***	3.47
Density of Street Intersections									-0.292	<.0001	***	0.75
Medium/High Traffic Flag[13]												
Residential Density									-0.097	0.572		0.91
Employment Density												
Commercial & Retail Density									4.746	<.0001	***	115.1
At least one segment with Medium/High Traffic									0.7	<.0001	***	2.01
N	5649				5649				5649			
Max-rescaled R-Square	0.0918				0.094				0.1173			
AIC	5111.19				5105.28				5024.19			
Likelihood Ratio	329.834				337.74				424.83			
$\Pr > \chi^2$	<0.0001				<0.0001				<0.0001			

[10] Dichotomous variable (=1) for observed incidence of walking (including to transit) on any road segment by at least one participant during the period of analysis; (=0) otherwise.

[11] Built environment variables were calculated within 40-meter road segment buffers. Density variables were measured per acres of each buffer.

[12] Includes some roadway medians and sidewalks.

[13] Based on Caltrans AADT counts: Medium > 24,999 and High ≥ 50,000 vehicles.

TABLE 4-4(b): BINARY LOGIT FOR INCIDENCE OF WALKING ON ANY SEGMENT

	Model 4				Model 5			
Dependent Variable	**WalkFLAG**				**WalkFLAG**			
Independent Variables	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**
Intercept	-2.413	<.0001			-2.418	<.0001		
Commercial Only Density	3.797	<.0001	***	44.56	3.546	<.0001	***	34.66
90th Green Space Percentile	0.694	<.0001	***	2	0.693	<.0001	***	2
90th Impervious Percentile	-0.371	0.003	***	0.69	-0.371	0.003	***	0.69
Transit Stops per Segment								
Transit Stops	0.041	<.0001	***	1.04	0.04	<.0001	***	1.04
Segment Length	0	0.726		1				
Density of Unclassified Parcels	0.71	0.028	**	2.03	0.652	0.045	**	1.92
Density of Street Intersections	-0.194	0.008	***	0.82	-0.193	0.005	***	0.83
Medium/High Traffic Flag	0.357	<.0001	***	1.43	0.334	0	***	1.4
Residential Density	0.216	0.231		1.24	0.222	0.218		1.25
Employment Density					0.012	0	***	1.01
Commercial & Retail Density								
At least one segment with Medium/High Traffic								
N	5649				5649			
Max-rescaled R-Square	0.1423				0.1455			
AIC	4931.2				4919.19			
Likelihood Ratio	519.82				531.84			
$\Pr > \chi^2$	<0.0001				<0.0001			

Significance: * p < .1, ** p < .05, *** p < .01

TABLE 4-4(c): BINARY LOGIT FOR INCIDENCE OF WALKING ON ANY SEGMENT

	Model 6				Model 7			
Dependent Variable	**WalkFLAG**				**WalkFLAG**			
Independent Variables	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**	**Coef.**	$\Pr > \chi^2$	**Sig.**	**O.R.**
Intercept	-2.396	<.0001			-2.714	<.0001		
Commercial Only Density	3.535	<.0001	***	34.28				
90th Green Space Percentile	0.695	<.0001	***	2	0.739	<.0001	***	2.09
90th Impervious Percentile	-0.373	0.003	***	0.69	-0.401	0.001	***	0.67
Transit Stops per Segment								
Transit Stops	0.041	<.0001	***	1.04	0.044	<.0001	***	1.05
Segment Length	0	0.795		1	0	0.934		1
Density of Unclassified Parcels	0.649	0.047	**	1.91	0.516	0.112		1.68
Density of Street Intersections	-0.2	0.007	***	0.82	-0.161	0.028	**	0.85
Medium/High Traffic Flag	0.339	0	***	1.4				
Residential Density	0.223	0.217		1.25	0.084	0.631		1.09
Employment Density	0.012	0	***	1.01				
Commercial & Retail Density					4.438	<.0001	***	84.64
At least one segment with Medium/High Traffic					0.634	<.0001	***	1.88
N	5649				5649			
Max-rescaled R-Square	0.1455				0.1492			
AIC	4921.12				4904.94			
Likelihood Ratio	531.9				546.09			
$\Pr > \chi^2$	<0.0001				<0.0001			

Significance: * p < .1, ** p < .05, *** p < .01

4.6.7 Factors Associated with the MVPA Occurring on Any Segment

The results of the negative binomial regressions of Table 4-5(a) and Table 4-5(b) reflect the associations of the various built environment, green space and traffic variables on MVPA minutes per road segment. The independent variables were added incrementally moving from Model 1 to Model 4. All the regression estimates are adjusted for the participants' socio-demographic characteristics, income level and employment status. These adjustments were necessary since the public health and transportation planning literature established an imminent association between the individual's socio-demographic traits and the outcome variable MVPA (see Guo, 2009; Lin & Moudon, 2010; Saelens, Sallis, & Frank, 2003). Therefore, the adjustments were made to prevent confounding the regression results.

For simplicity, I will review only the significant factors of Models 3 and 4 (Table 4-5(b)) since they have the best model fit measures and the most comprehensive list of covariates. The intercept has an important meaning and it reflects the difference in MVPA minutes between males and females. Specifically, the negative binomial models were controlling for the female gender and therefore the intercept displays the results for the reference groups noted in the footnote on the bottom of Table (Table 4-5(b)). For example, in Model 3 the intercept coefficient shows that the odds for MVPA increases more than twofold (O.R. = 2.25) for white, employed males with a high school education or less and earning an annual income less than $15,000 compared to the other groups. The intercept in Model 4 however shows no difference in the odds of MVPA occurring for the same reference groups when the three land cover variables are added to the model.

Consistent with previous results, activity centers with more commercial and retail uses increase the odds of MVPA on road segments. The odds of observing more MVPA minutes increase almost twice in Model 4 (O.R. = 2.17) and more than three and a half times (O.R. = 3.53) and in Model 3 when road segments have greater commercial and retail densities.

Further, the odds for higher MVPA increases only marginally with more public transit stops, longer street segments and higher densities of neighborhood employment. The coefficients on these variables are still significant across both models but the magnitude of their effect on the odds of MVPA at the street segment level is very minimal.

The effect of increased segment-level vehicular traffic volume is only significant in Model 3 and not in Model 4. The odds of observing more MVPA in Model 3 increases by 24% on streets with medium to high traffic volumes but this effect is only significant at the 10% significance level. Moreover, there is an inverse relation between MVPA and street intersection density. Increasing the number of street intersections per acre of road segment reduces the odds of observing MVPA by about 40% in both models.

The highest effects on the odds of observing more MVPA minutes per road segment appear from the green space variables (tree and irrigated lawn density) in Model 4. These odds increase the chances of MVPA by more than 100 times indicating the importance of green spaces in attracting pedestrians and therefore higher physical activity levels. Similarly, the odds also increase more than 100 times when street segments have more impervious land cover which may include trails and sidewalks that cater to pedestrians.

TABLE 4-5(a): NEGATIVE BINOMIAL FOR MVPA (MIN.) ON SEGMENTS WITH WALKING

	Model 1				Model 2			
Dependent Variable	**MVPA (min.)**				**MVPA (min.)**			
Independent Variables[14]	Coef.	Wald χ^2	Sig.	O.R.	Coef.	Wald χ^2	Sig.	O.R.
Intercept	2.936	116.22	***	18.85	0.690	4.85	**	1.99
Commercial & Retail Density	1.194	14.16	***	3.30	1.720	29.60	***	5.58
Residential Density	-0.585	5.60	**	0.56	0.084	0.10		1.09
Total Transit Stops	0.031	172.27	***	1.03	0.021	85.93	***	1.02
Street Intersection Density	-0.885	70.01	***	0.41	-0.517	13.45	***	0.60
Segment Length					0.005	86.02	***	1.00
Unclassified Parcel					1.791	16.99	***	5.99
Medium-to-High Traffic					0.205	3.22	*	1.23
Employment Density								
90th Percentile Green Space					0.140	0.53		1.15
Impervious Land Cover Density								
Tree Density								
Irrigated Lawn Density								
N	954				954			
AIC	5,734.30				5,601.50			
BIC	5,831.51				5,718.16			
Pearson χ2	1,537.67				1,690.36			
Dispersion Parameter	1.320				1.140			
Scaled Deviance/DF	1.070				1.051			

Significance: * p < .1, ** p < .05, *** p < .01

[14] Regression estimates are adjusted for gender, age, race, employment status, income and education level.
Reference Groups: Male, White, Employed, Household Income < 15K, and Education Level: High School or Less.

TABLE 4-5(b): NEGATIVE BINOMIAL FOR MVPA (MIN.) ON SEGMENTS WITH WALKING

	Model 3				Model 4			
Dependent Variable	**MVPA (min.)**				**MVPA (min.)**			
Independent Variables	Coef.	Wald χ^2	Sig.	O.R.	Coef.	Wald χ^2	Sig.	O.R.
Intercept	0.810	7.25	***	2.25	-7.088	5.12	**	0
Commercial & Retail Density	1.262	15.05	***	3.53	0.774	3.45	*	2.17
Residential Density								
Total Transit Stops	0.021	86.94	***	1.02	0.021	80.01	***	1.02
Street Intersection Density	-0.530	14.14	***	0.59	-0.511	13.18	***	0.60
Segment Length	0.005	81.22	***	1.00	0.005	75.31	***	1.00
Unclassified Parcel	1.560	14.08	***	4.76	1.327	9.35	***	3.77
Medium-to-High Traffic	0.213	3.56	*	1.24	0.162	1.98		1.18
Employment Density	0.009	6.74	***	1.01	0.009	7.04	***	1.01
90th Percentile Green Space	0.147	0.60		1.16				
Impervious Land Cover Density					8.296	6.47	**	>100
Tree Density					7.845	5.61	**	>100
Irrigated Lawn Density					9.218	6.51	**	>100
N	954				954			
AIC	5,594.47				5,592.46			
BIC	5,711.13				5,718.84			
Pearson χ2	1,744.45				1,716.85			
Dispersion Parameter	1.130				1.128			
Scaled Deviance/DF	1.050				1.051			

Significance: * p < .1, ** p < .05, *** p < .01

4.7 Discussion & Policy Implications

The results of the single sample t-tests of Tables 4-2 and 4-3 indicate that roadway segments with walking incidences have higher densities of commercial and employment uses, more transit stops, higher traffic volumes and higher densities of impervious land cover but less areas with green spaces.

Similarly, results from the binary logit models of Tables 4-4(a to c) show that public transit stations and green spaces seem to contribute positively to chances of observing walking on roads which may be used as important policy tools to promote greater physical activity and non-motorized travel. The largest effects on observing greater incidences of walking in these models are due to commercial and retail activity centers. This is consistent with other studies that concluded the pertinent role that these activity centers play in attracting pedestrians and increasing the probability of walking per segment particularly along streets with more vehicular activity.

Likewise, the negative binomial regression models of Tables 4-5(a) and 4-5(b) produce similar results for the effects of the built environment on elevated physical activity levels (MVPA) due to active transport (TPA). The most significant effects of all land use variables on MVPA come from the three land covers: trees, irrigated lawns and impervious areas. The significance of these variables in the models sheds some light onto the importance of aesthetics, shade, green spaces, trails and sidewalks to individuals that exert higher levels of physical activity during their walks.

Combining the findings from the t-tests and all the regression models we can distill some important planning insights to advice policymakers. Both t-tests indicate that the routes taken by pedestrians in the sample have more commercial and retail uses, neighborhood employment locations, transit stops which have all been shown to be associated with greater chances of walking activity and higher MVPA levels in the regressions. The t-tests also illustrate that these routes are lacking in green spaces although they have been shown to have a positive association and the greatest effects on MVPA levels in the regressions. This noticeable absence in green spaces along the 'walked' routes point out to the need for more progressive policies to improve landscape designs. These policy reforms may be particularly pertinent in low-income communities whose residents may be more likely to use public transit and active transport than motorized vehicles. Increasing green spaces and improving the pedestrian pathways would likely encourage more individuals to walk to destinations and to transit which reduces the reliance on vehicular modes and overall contributes to the goals of SB 375.

4.8 Limitations

The analyses presented in this chapter have a few limitations. The data used in the analysis relied on a sub-sample of the participants of Phase 1 of the Expo study. Only participants that exhibited any form of TPA (walking to destinations or to transit) were extracted and examined further. This sample was only 54 participants out of the possible 117 of Phase 1 of the study. The remaining participants only used motorized modes for transport and were therefore excluded. Future research on Phases 2 and 3 may expand on the analyses highlighted in this chapter to examine the effects of introducing the Expo Light Rail on attracting increased levels of active transport and more pedestrians.

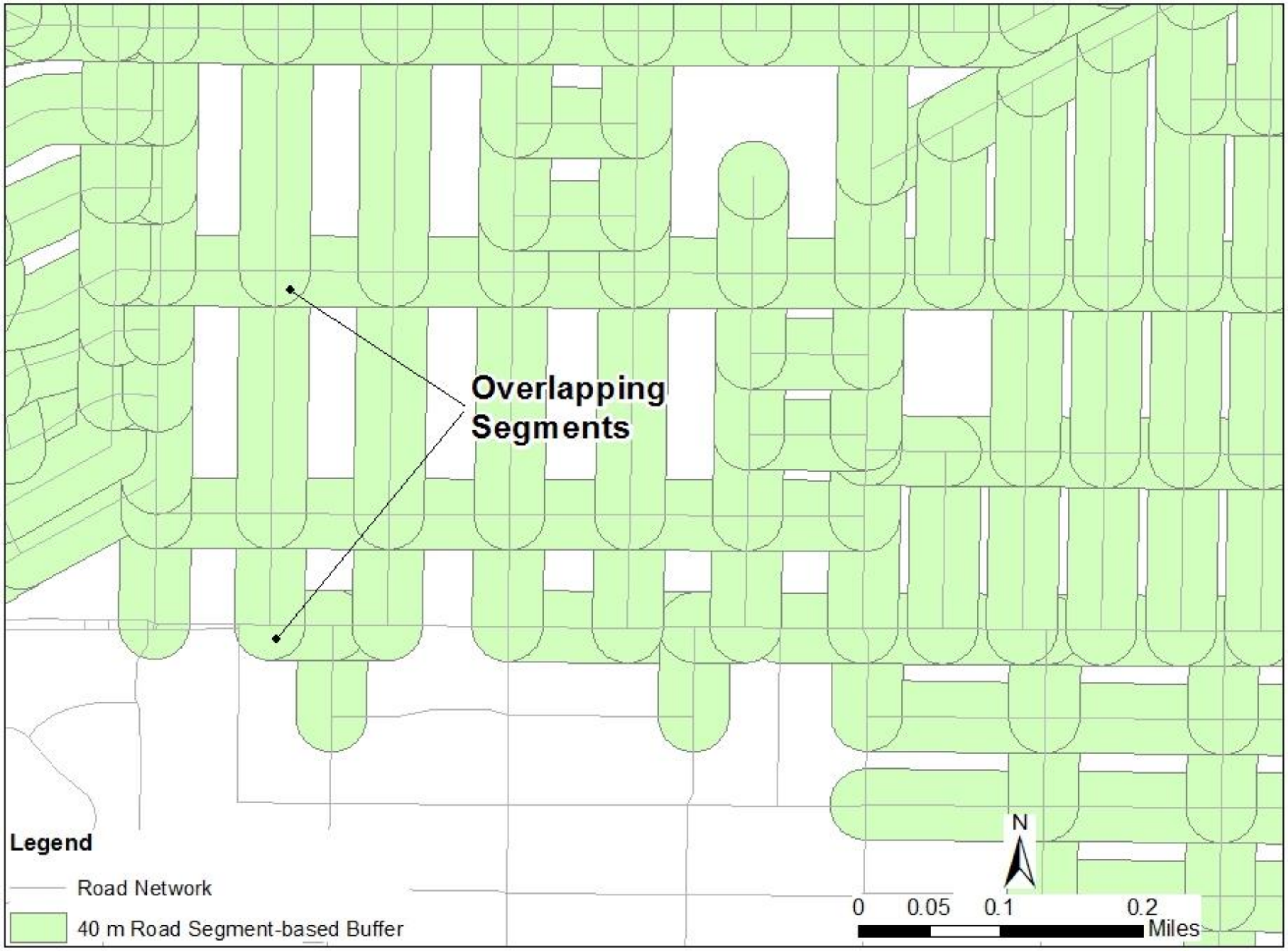
Another limitation of this study exists in the road segment quantification method. As explained before, the road segments have been loaded into ArcGIS from the 2010 TIGER street network file. Some road segments may have been used multiple times by the respondents in the sample while others may have not been traversed at all during the study time period. Considering the road segment sample selected within one-half mile from respondents' residences; a chance of spatial autocorrelation may occur between the incidences of observing TPA (dependent variables) per segment. Spatial autocorrelation occurs due to the close proximity of the road segments to one another and due to the fact that some may have been traversed numerous times. This causes the errors of the statistical model to be no longer independent which is a statistical assumption violation that introduces bias into the model (Lascala, Gerber, & Gruenewald, 2000). Future research may employ geostatistical modeling techniques such as spatial regression analysis that corrects for spatial autocorrelation to reduce bias.

Further, the road segments have two nodes or endpoints and therefore when the segment-based buffers were created; each buffer took on two endpoints as well. Therefore, the road segments may overlap at the endpoints thus increasing the chance for violating the *Independent Identical Distribution (i.i.d.)* assumption[15] (Guo, 2010) which indicates the possibility of heteroskedastic variances of the road segment attributes. This also creates a potential for 'double-counting' of the land uses at the overlapped endpoints. This may lead to an overestimation of the impacts of the land uses present. Figure 4-18 below depicts this situation within a sample area of the Expo study.

[15] This means that each observation is assumed to come from the same probability distribution as the remaining observations and that the probability of its occurrence is mutually independent of the occurrence of others.

The green space measures used in the models were calculated at a two feet resolution; this implies that some street amenities may not have been fully captured in the analysis. Future studies may benefit from performing detailed street audits and objective measurements using Google Earth that have both been proven to accurately assess and improve street-level and neighborhood conditions (Clarke, Ailshire, Melendez, Bader, & Morenoff, 2010; Hoehner, Brennan Ramirez, Elliott, Handy, & Brownson, 2005; Sugiyama, Neuhaus, & Owen, 2012).

The above issues are beyond the scope of this research since a more simplistic approach of calculating the road segments and the land use attributes was performed here. Future research may want to take these limitations into consideration and use more advanced GIS mapping software since the current ArcGIS tools cannot disentangle overlapping features of the street segments (Guo & Ferreira Jr, 2008).



Map Created by Gaby Abdel-Salam

Figure 4-18: Road Segments Overlapping at the Endpoints

4.9 Conclusion

The segment-level analysis performed in this chapter yields a finer spatial scope which helps in providing more precise effects of the built environment on active travel and MVPA. Findings from this research show that road segments used frequently for walking in low-income communities may attract more pedestrians if they have higher densities of green spaces and activity centers. Further, implementing smart growth tools along transit corridors may potentially lead to increases in non-motorized travel over private vehicle use which helps accomplish SB 375 goals.

Combining the findings from the t-tests and all the regression models we can distill some important planning insights to advice policymakers. The t-tests performed indicate that the routes taken by pedestrians in the sample generally have more commercial and retail uses, neighborhood employment locations, and public transit stops which have all been shown to be associated with greater chances of walking activity and higher MVPA levels in the regressions.

However, the same t-tests also illustrate the lack of green spaces along the same routes even though regression results show a positive association between green space and active transport especially during episodes of elevated physical activity levels. This noticeable absence in green spaces along the 'walked' routes point out to the need for more progressive smart growth policies that aim to improve landscape designs to promote alternative modes to vehicular travel. Moreover, these policy reforms may be particularly pertinent in low-income communities whose residents may be more likely to use public transit and active transport than motorized vehicles. Thus, by increasing green spaces and improving pedestrian pathways; we would likely encourage more individuals to walk to destinations and to transit which reduces their reliance on vehicular modes and contributes to the overall goals of SB 375.

CHAPTER FIVE

A Multilevel Model for the Effects of the Built Environment on Active Transport

Introduction

Active transportation or transport-related physical activity (TPA) has several environmental and health benefits. Switching to non-motorized modes of travel or TPA could mitigate pollution concerns, reduce Ghg emissions and decrease traffic congestions (de Nazelle et al., 2011; World Health Organization (WHO), 2011). In addition, physical activity in the form of TPA reduces health risks from chronic diseases, colon and breast cancer, diabetes, ischemic heart disease, obesity as well as mortality (L. D. Frank et al., 2008; Leitzmann et al., 2007; Oliver et al., 2010; Saarloos et al., 2009). The daily recommendations of moderate physical activity, of at least 30 minutes per day five days per week has been declared by the Office of the U.S. Surgeon General and the American College of Sports Medicine as the healthy standard that could be achieved by most adults in the U.S. (Centers for Disease Control and Prevention (CDC), 2014). In fact, achieving this minimum health goal of moderate physical activity relative to being sedentary is correlated with a 32% reduction in mortality risk (Leitzmann et al., 2007). Unfortunately, national statistics show that more than 50% of U.S. adults do not meet this health guideline (Leitzmann et al., 2007).

To mitigate this health risk, community advocates and health practitioners convened at a workshop in 2002. The workshop, hosted by the Centers for Disease Control and Prevention (CDC), was a collaboration of public health practitioners and community representatives. The main objective was to help scope a program to promote physical activity on the community level. Among the topics discussed was the link between public health and the built environment and ways to alter the latter to improve the former (Dannenberg et al., 2003). Smart growth tools utilized by transportation planners encompass these environmental, built environment and health concerns. The present trend has been to develop communities that encourage more TPA.

The segment-level analyses discussed in the previous chapter provided statistically significant results of the synergistic relationship between the built environment, transit use and physical activity. Alternatively, the models in this chapter expand upon this relationship by adding an extra element: the individual's characteristics through the use of multilevel or hierarchical modeling that includes information regarding a traveler and his social and physical environment.

This modeling technique is known as the ecological modeling approach. It utilizes multilevel analyses and incorporates concepts from many disciplines that influence policy, individuals, communities and the built environment (Giles-Corti, Timperio, et al., 2005; Sallis et al., 2006). These types of models are growing in popularity especially in the public health literature. As the name suggests, ecological models examine the relationship between individuals and their social, cultural and physical environments; and have been generally used as the basis of many studies to promote physical activity (Sallis et al., 2006).

Thus, evidence suggests that the built environment exerts a considerable effect on travel behavior. Many notable studies included the various facets of the built environment (or the 5 Ds) in their analyses of active travel usually concluding a positive relationship between the two

(LFrank et al., 2005; Handy et al., 2002; Rissel et al., 2012). The five Ds include increasing parcel **d**ensity, **d**iversity of mixed uses, reducing **d**istance to transit, improving connectivity and **d**esign and allowing for more **d**estination accessibility (Campoli, 2012). Many have also controlled for variations in individual traits by including the traveler's socio-demographic characteristics hoping to minimize bias (Guo, 2010; Prince et al., 2011; Saelens et al., 2003). However, very few accounted for the inherent hierarchy in their data structure which could potentially lead to increased risk of Type 1 error. The analyses I conduct in this chapter expand upon the current active travel-built environment literature. This is done through an ecological model framework of multilevel analyses that predicts TPA through impacts of land use at varying geographic extents while controlling for household- and individual-level covariates. A hierarchical model is introduced in this chapter that exploits the three data levels which define the respondents' characteristics and variables relating to his social and physical environments.

This chapter focuses on *walking* as the active mode of travel. Therefore, all analyses were performed at a reasonably fine 'walkable' extent. I used the traveler's home neighborhood environment (quarter- and half-miles from his/her residence and 40 meter segment-level buffers) as the geographic extents to test the effects of roadway segment built environment characteristics on active transport. I further utilize the matched accelerometer-GPS data in a momentary analysis of 15-second epoch location and physical activity tracking. Momentary activity tracking has been used in the public health literature due to the fine level of data it provides (Almanza et al., 2012; Chaix et al., 2013; Quigg et al., 2010).

Objectives

The objectives of this chapter are to (1) explore impacts of different built environment facets on active travel in the microenvironment when socio-demographic traits are controlled for in a multilevel modeling setting, (2) unravel the momentary-location variations in active travel behavior and the corresponding environment and (3) investigate variations in geographic extent in land use variable calculations and their effects on active travel behavior.

This study builds upon the active travel literature and expands upon it in several ways. The moment-by-moment unit of analysis measured by matched accelerometer-GPS data; provide more accuracy and richness to the data (Badland et al., 2010a). This data is then used in a three-level regression model (compared to two-levels performed in other studies (Goulias, 2002; Prince et al., 2011)) which yields more comprehensive results and more precise associations with the response variable at each hierarchical level.

Further, impacts from the built environment on active travel are modeled via three different geographic extents. The reasoning for this is twofold: (1) to compare various scopes of analysis within the microenvironment of an active traveler and (2) to advise policymakers of potential enhancements to key land use types at the respective geographic extent that may encourage more transport-related physical activity (TPA).

5.1 Background

The studies mentioned thus far in the travel behavior literature examined the impacts of changes in the built environment on one's active travel habits. Some authors focused on the geographic extent (Cerin et al., 2009; Handy et al., 2002) making the argument that a finer level of analysis at the block level and within a quarter-mile from homes yields more accurate results and that smaller, more diversified land uses encourage more active travel (Cervero, 1996; Lee & Moudon, 2006b). Others contrasted physical activity levels by stratifying neighborhoods by various walkability indices (Adams et al., 2011; Boarnet et al., 2011; Frank et al., 2005; Hankey, Marshall, & Brauer, 2012) and observing the respective positive impacts on physical activity from higher walkable environments.

Others utilized objective measurements by using GPS, accelerometers and Geographic Information System (GIS) mapping technology. Some studies used this technology to find differences in moderate-to-vigorous physical activity (MVPA) between transport modes (Adams et al., 2011; Badland et al., 2010a; Oliver et al., 2010). Others used objective measurements to contrast changes in physical activity among various urban settings such as green spaces versus other locations and geographic extents (Boarnet et al., 2011; Cooper et al., 2010; Houston, 2014; Rainham et al., 2012; Rodriguez et al., 2005). However, these studies utilized other modeling techniques that did not exploit the hierarchical nature of their data. Thus, disregarding any interactions occurring among the components of each level which may invalidate the statistical models (Goldstein, 1999).

On the other hand, a growing number of studies in the public health field examined associations of the built environment with physical activity in multilevel frameworks while using objective measuring devices. Many however, only used accelerometers to track MVPA levels without tracking the location where it occurred (Ding et al., 2012; Kneeshaw-price et al., 2013; Witten et al., 2012) or used a maximum of two class levels to model their data (Sundquist et al., 2011; Van Dyck et al., 2010). The two class levels identified in these studies pertained to the individual and his/her built environment without including the social context of the households' characteristics.

To my knowledge, only two papers utilized data from both accelerometers and GPS devices and used the multilevel modeling technique for the *PLACES* study in Chino, California. The first used a generalized linear mixed model to examine momentary exposure of children to green spaces and the probability of MVPA occurring at the epoch-level (Almanza et al., 2012). The second paper predicted land use types through four different joint parent-child MVPA levels (Dunton et al., 2013). These papers however, did not control for household level characteristics and only used two classes for their models (neighborhood- and individual-level).

The models that will be discussed in the current study aim to rectify the above mentioned gaps in the travel behavior literature. I propose modeling TPA occurrences in three different geographic settings and thus examining associations to the built environment in each context. I also utilize objectively measured: land use variables and momentary-location physical activity level-tracking data. Finally, I use an ecological model framework to predict MVPA occurrences in a three-level generalized linear mixed model controlling for neighborhood, household and individual characteristics.

5.2 Theoretical Motivation for Using Hierarchical Analysis

Thus far, the models mentioned earlier in chapter one have focused on the effects of variations in land use, street networks, and traffic variables on physical activity during episodes of active travel. Here, I introduce the potential impacts of the individuals' attributes on the relationship between the built environment and active travel.

Controlling for a "traveler's" socio-demographic characteristics avoids the risk of confounding the results. In fact, if the proxy used to describe travel behavior (dependent variable) demonstrates within group-level variations, which I expect to find in this case; the results would yield ill-fitting models and biased estimates. To reduce this risk, multilevel analysis will be utilized in this chapter which is a modeling approach that takes advantage of the “contextual behavior of individuals” (Clark & Linzer, 2013). As the name suggests, this modeling method considers and exploits the hierarchical nature which is intrinsic in the data.

This modeling technique has been interchangeably referred to as: random coefficient models, mixed models and hierarchical linear models (Goulias, 2002). There are slight differences among these methods but overall they encompass the following:

(1) The nested nature of the data structure

(2) The observed and unobserved heterogeneity that may be overlooked with other models

(3) The mean effects of the independent variables on the dependent variables

(4) The within group random deviations around the means of these independent variables

(5) Random errors of the regressions

As noted before, data that have an inherent nested or hierarchical nature are often better analyzed via multilevel modeling. In some instances, failure to account for this nature influences the estimated variances and can also increase the chance of Type 1 error (Bell, Ene, Smiley, & Schoeneberger, 2013).

Further, multilevel regression is useful in highlighting unapparent correlations among observations in various data levels. This technique is useful in exploiting detailed inferences among smaller class levels that have been previously aggregated to higher levels thus decreasing potential bias (Goulias, 2002). In this chapter, I employ the multilevel regression approach because the data utilized combines different facets and levels from the built environment or neighborhood attributes, household information and respondents' personal characteristics.

5.3 Expo Data Hierarchical Structure

The Expo data set discussed earlier in the previous chapter included seven-day trip logs and a baseline survey. Only information from the baseline survey was utilized which include household and personal socio-demographic characteristics.

The basic three-level structure of the data set is of the following type: respondents nested within households nested within neighborhoods. Figure 5-1 below shows a depiction of the variables used and the three-level nesting they belong to. The third and outermost level is reserved for the neighborhood attributes. Information for this level was obtained using buffers of three geographic extents: 0.25-mile and 0.5-mile from home and 40 meter buffered road segments as catchment areas around each of the participants' homes. This includes the percentages and types of each land use, traffic information and volumes, street characteristics and green space classifications. Next is the household level where general household characteristics are stored such as the number of children and vehicles in the household and annual income. Finally, the innermost and first level is the 15-second momentary activity level and holds the characteristics of the primary participant in the baseline survey who also wore the accelerometer and GPS devices. Note, the surveys also collected information regarding the other adults in the household and were each provided a person I.D., however, only one adult from each household was given both a GPS and accelerometer device for mobile tracking purposes. Therefore, the momentary-activity level referred to in this chapter pertains to the main respondent from the mobile tracking group.

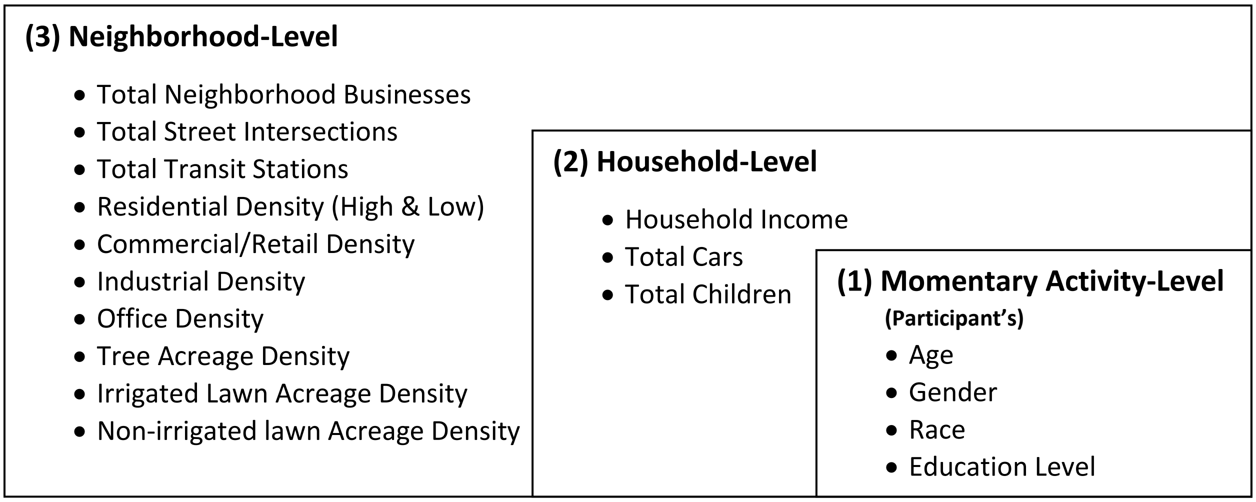


Figure 5-1: Hierarchy & Nesting of Data

5.4 Literature Review

Linkages have been made between active travel and the built environment in the transportation planning and public health fields. The emphasis in this chapter is on the utilization of the econometric technique: multilevel modeling, which exploits the nested data structure.

Goulias (2002) was the first study in the travel behavior literature to utilize multilevel modeling applied to activity-based concepts to examine travelers' time-allocations. Goulias used five waves from a longitudinal study *(Puget Sound Transportation Panel (PSTP))* and was the first to define four hierarchical classes for his data set: space, household, person and temporal levels (Goulias, 2002). The study however mainly relied on self-reported travel behavior over a two-day period and thus no objective measuring devices were utilized.

Many prominent studies have also used this modeling technique in the public health literature. In Ottawa, Canada one cross-sectional study collected self-reported surveys regarding travel habits, socio-demographic traits, height and weight among other things from respondents within 85 local neighborhoods. Built environment characteristics of the neighborhoods were also objectively measured and respondents were asked to wear accelerometers to track their physical activity bouts. The authors used a two-level binary logistic regression stratified by gender to measure the relative impacts of the built environment covariates on physical activity and the likelihood of being obese while controlling for the individual's characteristics (Prince et al., 2011). Results showed that men were twice as likely to be active (O.R. = 2.08), and that chances for women to be overweight increased almost twofold for each additional local specialty food store built (Prince et al., 2011). The association between green space and physical activity was unexpectedly negative (O.R. = 0.93) and was even found to be associated with a greater chance of obesity for men (O.R. = 1.10) but a much less likelihood for being overweight was found for women (O.R. = 0.66) (Prince et al., 2011). This study has two important strengths: it included a spectrum of green spaces (outdoors and indoors) and it controlled for seasonality since one's physical activity levels may change with the ambient temperature. However, one limitation was that the authors were unable to explain the difference in gender preferences to green space access due to study design limitations.

A parallel study that was completed in Sweden, *The Swedish Neighborhood and Physical Activity (SNAP) Study* also used multilevel regression to model active transportation and MVPA. Surveys from the *International Physical Activity Questionnaires (IPAQ)* were collected from 2,269 adults (Sundquist et al., 2011). Walking periods were validated by the seven-day accelerometer readings gathered. A number of land use variables were also objectively measured using GIS. These measures were used to calculate a walkability index specific for Stockholm City which was later subdivided into deciles: 1-4 was considered less walkable and 7-10 was classified as highly walkable. Additional hierarchical data included: neighborhood-level income and the respondent's physical characteristics.

The authors combined multilevel linear and logistic regressions to obtain mixed effects and mixed-distribution models. They found that the odds of walking for transport was 2.75 times greater than that for leisure activity for individuals residing in highly walkable neighborhoods (Sundquist et al., 2011). Alternatively, this translates into an additional 50 minutes of walking for transport per week or an extra 3.1 minutes of MVPA per day when individual-level and neighborhood-level variables are accounted for (Sundquist et al., 2011). The paper also highlights the potential bias in the parameter estimates when a regular logistic regression is employed instead of a multilevel analysis where the odds for active transportation were found to be inflated for highly walkable neighborhoods. The logistic model overestimated this association to O.R. = 1.92 versus a much lower but still significant O.R. = 1.77 for the odds of walking for transport using a multilevel modeling technique (Sundquist et al., 2011). Further, there were many limitations in this study. The instrumentation of the survey may have included recruitment bias due to the use of phones for survey collection. The study might also have response and self-selection bias: where only physically active individuals who choose to live in these neighborhoods completed the surveys and self-selected into neighborhoods because of a perceived walkability advantage.

Another European study, in the Netherlands, used multilevel regression to model the physical environment and links to both walking and bicycling. The response variable in this study was time spent walking or biking during commutes and leisure activities. Data was acquired through self-administered surveys from the National Institute for Public Health & the Environment 1987-92 and 1993-97 for a total of 11,541 adults (Wendel-Vos et al., 2004). The built environment variables were objectively measured in GIS including green space in 300-meter and 500-meter buffers from the respondent's postal code. The first level of the models estimated the associations of the socio-demographic characteristics with physical activity; the second level expanded the model by also including neighborhood-level characteristics. Parameter estimates were significant in the 300-meter buffer model for the effects of sport grounds on time-spent walking/biking both for leisure and commuting purposes. Further, the models also showed that the variance between individuals was much greater than that between zip codes (Wendel-Vos et al., 2004). However, the models did not show significant associations between the response variable and the green space variables. This was probably due to the larger unit of analysis that the green space variables were aggregated to (postal zip code level versus a smaller unit such as the block-group level).

Similarly, a more recent Canadian paper in Montreal used multilevel regressions but the outcome variable was total daily walking distance to access public transportation. This study utilized the 2003 Origin-Destination (OD) one-day travel survey for 6,913 adults commuting to work or school via different public transit modes (Wasfi et al., 2013). The respondents' routes were recorded and estimates of the total minutes of transport-related physical activity were obtained. Further, total distances were also calculated based on average walking speed assumptions. Model estimates showed that males walked more and that lower-income individuals walked on average 2.12 minutes less per day than those earning $80,000 or more annually (Wasfi et al., 2013). In addition, the contribution of commuter trains to total walking distances to access transit, was the highest among the different transit modes which translates to an additional 14.47 minutes daily relative to only 2.99 minutes of walking to access the city bus (Wasfi et al., 2013). Approximately 11% of the respondents in this study achieved the daily recommended 30 minutes of physical activity just by walking to and from transit to get to school or work (Wasfi et al., 2013). The multilevel modeling revealed that the Intra-class correlation coefficient (ICC) for census tracts = 6.67% which suggests significant variability in walking to transit among the different census tracts in Montreal and therefore validates the use of multilevel modeling.

There are a number of limitations that exist for this study. One being in its design, that it was based on only one-day of travel which excludes any daily variations. Further, the survey was collected in 2003 and the land use variables were aggregated from a 2006 database which may have differed over the three-year period. Also, episodes of walking during transit transfers were not accounted for and therefore total walking distance measurements may be significantly underestimated. While Wasfi et al. (2013) used GIS to display origin-destination pairs from survey results and to simulate shortest distance to transit stops, no other objectively measuring devices (e.g. Accelerometer & GPS) were utilized.

On the other hand, one study in Belgium (*Belgian Environmental Physical Activity Study – BEPAS)* used accelerometers and GIS to measure minutes of MVPA and to stratify neighborhoods by walkability ranking (Van Dyck et al., 2010). This study collected *IPAQ* surveys from 1,166 adults and had them wear accelerometers for a seven-day period. The authors then performed multilevel modeling (two-level: neighborhood & participant) to predict log-transformed MVPA, recreational walking behavior and active transport by mode. Their findings suggested that higher versus lower walkable areas contributed to more MVPA minutes (38.6 min./day vs. 31.8 min./day), encouraged more active transport and was related to overall reductions in motorized travel (Van Dyck et al., 2010). The authors also noticed that lower walkable environments were associated with higher levels of cycling as a mode of active transport and less dependence on vehicular modes (Van Dyck et al., 2010).

The above Belgian study represented an improvement over previous research that used multilevel analyses; however, it still has some limitations. The accurate measurement of MVPA by accelerometers is definitely an important contribution and avoids bias from self-reported physical activity occurrences. This study however, did not jointly track the locations where this activity occurred via GPS for example.

Similarly, a cross-sectional study in New Zealand (*Understanding the Relationship between Activity and Neighbourhoods - URBAN)* examined associations between the built environment and accelerometer-measured TPA in a multilevel setting. Built environment features describing access to destinations, residential density and street connectivity were objectively measured from 48 neighborhoods and 2,033 adults participated in the survey (Witten et al., 2012). The authors concluded that one standard deviation (S.D.) increase in the three built environment variables, was correlated with a 7% increase in accelerometer counts in the weekdays and 5-7% increase over the weekend (Witten et al., 2012). The large sample size and accelerometer use add to the strengths of this study however, again, physical activity location tracking was not carried out and therefore the richness of the data collected is lacking.

In this current study, I utilize measurements from both accelerometer and GPS devices to calculate and track momentary-activity levels. The benefits of such rich data are to: (1) reduce recall bias (since respondents are likely to complete survey questions regarding physical activity after some period has passed and not instantaneously) and (2) avoid self-reported bias (since respondents may want to impress interviewers by over-reporting their physical activity levels).

Only two recent papers used matched accelerometer-GPS data in their multilevel modeling of physical activity and built environment correlates. Both papers used data from the *PLACES* study in Chino, California. The objectives of the first paper was to examine momentary exposure of children to green spaces and to predict the probability of MVPA occurring at the epoch-level (Almanza et al., 2012). The authors classified Chino neighborhoods into percentiles of green space: high (90th percentile) versus low (10th percentile). Their findings suggest that the odds of MVPA occurring increases by 39% for residents living in neighborhoods with higher green spaces over lower ones (Almanza et al., 2012). The second paper predicted land use type using four different joint parent-child MVPA levels. The authors' findings suggest that the majority of sedentary instances occur in open spaces among parent-child pairs. These instances amounted to about eight minutes per day which if converted to MVPA could account for one-third of the recommended physical activity standards (Dunton et al., 2013).

The last two papers of momentary and location-based physical activity in the multilevel modeling literature have limitations in their data structure. Both used only two classes for their models (neighborhood- and individual-levels) instead of the three levels proposed in the current study. In addition, they have not controlled for different household-level covariates (only annual household income was accounted for) which might be a source of bias in their estimates. The expansion in the data stratification process used in my methodology provides a finer level of analysis and accounts for various level-specific variables. I use the matched accelerometer-GPS momentary-activity level data of the main survey respondent to more accurately predict the probability of MVPA. A more detailed explanation for the use of three levels in my models will be discussed later in section 5.8.

5.5 Contributions and Policy Implications

There is compelling evidence that suggests that a significant relationship exists between the built environment and active travel behavior or TPA. Many notable studies aimed to unravel such associations by including various facets of the built environment (or the 5 D's) in their analyses of active travel. Several controlled for the traveler's individual or socio-demographic traits that were considered to be a source of bias if omitted. However, most overlooked the 'natural' hierarchy in their data structure and used conventional modeling methods instead of techniques that exploit this hierarchy. This results in an increased chance of Type 1 error occurring in their estimates. One of the contributions of this current study is that I expand upon recent active travel-built environment literature through the use of multilevel regressions. I adopt the ecological modeling approach as the conceptual framework for the analyses which allows for better predictions of active travel through individual-, household- and neighborhood-level correlates.

In addition, the majority of the studies in the field were performed on one neighborhood geographic scale but the objectives of this study are to examine the effects of built environment correlates on TPA at varying geographic extents in the immediate microenvironment of respondents. The choice of the two of the three extents selected (quarter- and half-mile radii) are commonly used in the literature where smart growth tools have been found to be most effective (Cervero, 1996; Guerra, E. & Cervero, 2013). The selection of the third extent (segment-level) is to explore the effects on TPA at the street segment-level. Another reason for avoiding larger geographic scales is that many authors observed better estimates for TPA and the built environment at smaller geographic extents especially at the individual or household levels which they note to surpass those collected at the block-group level (Cerin et al., 2009; Handy et al., 2002).

Since the main focus in this study is on active travel, the geographic extents selected needed to be fine-grain and reasonable for walking. As mentioned earlier, I used three buffer sizes representing the traveler's microenvironment (quarter- and half-miles from his/her residence and segment-level buffers) to test the effects of the built environment on active transport. By measuring the built environment variables at these three levels of aggregation, I am able to compare the respective significance and magnitudes of the effect sizes on active travel behavior. Ultimately, the goal is to be able to advise policymakers on which key variables and at what study level walking behavior may be promoted and ultimately reduce the reliance on motorized vehicles.

Another contribution of this study is the use of matched accelerometer-GPS data in a momentary-activity tracking of MVPA in 15-second epochs. I utilize these in an ecological multilevel setting to explore potential key built environment correlates while simultaneously controlling for individual- and household-level characteristics. The detailed momentary location and activity tracking are expected to unravel variations in active travel behavior and the corresponding land use types.

Thus, the use of momentary location and activity tracking by both objective-measuring devices are a major contribution in the multilevel modeling literature of active travel. As noted earlier, the use of accelerometers only in this field have been prevalent, however, only two papers utilized data from both accelerometers and GPS devices in a multilevel setting. This current study will thus be contributing to the literature through the accuracy of the results due to more precise data collection techniques. Therefore, policymakers can be better and more accurately advised of true effect sizes.

5.6 Methodology

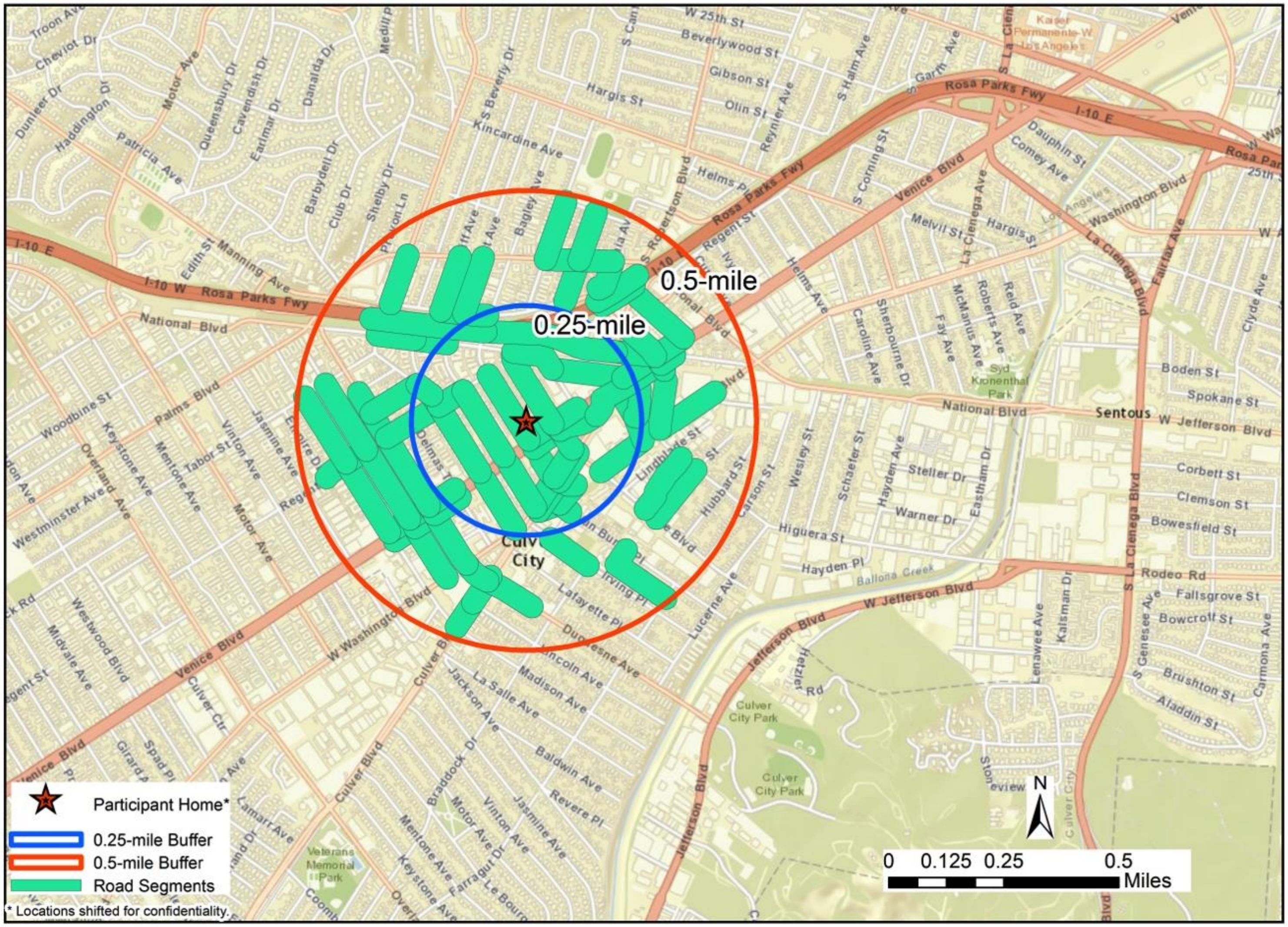
In this section, I discuss the types and sources for the data used, the creation of level identifiers, the creation of the land uses and the response variable, model specification, model building and reasons for using three-level versus a two-level analysis.

5.6.1 Data Sources

This data set includes 40 meter buffered street segments near Expo respondents' homes and land use catchment buffer areas of 0.25 mile and 0.5 mile radii to test the influence of the built environment in both settings on active travel. A sample map of the three geographic aggregation levels around a respondents' home is displayed in Figure 5-2.

In addition, built environment uses were classified and clustered as follows: high-density residential, low-density residential, commercial, office use and industrial. High-density residential included a number of different compact uses such as: duplexes, triplexes, two- or three-unit condominiums & townhouses, high-density single family residential; medium-rise apartments or condominiums, mixed & multi-family residential and mixed urban. Low-density residential included: low-density single-family residential and low-rise apartments, condominiums or townhouses. Commercial uses are comprised of: restaurants, commercial recreation, retail centers, modern or old strip developments and regional shopping centers. Office uses are made up of low- and medium-rise major offices as well as government offices. Finally, manufacturing, assembly, industrial services and mineral extraction make up the industrial uses. All the built environment data were obtained from the 2005 existing land use data from the Southern California Association of Governments (SCAG). The green space variables were measured at a two feet resolution and were obtained from 2002-2005 QuickBird remote sensing and aerial imagery (McPherson et al., 2011).

Further, this data set was augmented with matched accelerometer-GPS data points containing various household and individual traits. For example, household level information such as household size, number of children, annual income, etc. are included as well as person-level characteristics for the respondent that agreed to carry the accelerometer and GPS devices over a four to seven day period. The accelerometer readings recorded different physical activity levels in one-minute increments, while the GPS devices traced locations for the individuals in 15-second epochs. Momentary readings from both devices were matched into 15-second incremental observations (N = 14,265) uniquely identified by a variable that combines the household number and the date and time stamp.



Map Created by Gaby Abdel-Salam

Figure 5-2: Three Geographic Aggregation Levels

5.6.2 Key Level Identifiers

A main requirement for multilevel analysis is to create a unique identifier for each level that is modeled. Here, I propose using three levels as portrayed in Figure 5-1. Therefore, I needed three identifiers for: the neighborhood level, the household level and the participant's momentary activity level.

5.6.2.1 Neighborhood-level Identifier

To create the neighborhood-level identifier (I.D.), home locations were ranked according to the number of accessible amenities or local businesses in its relative microenvironment. To achieve this, home locations of Expo participants was geocoded in GIS to the street network file and a shapefile of the local neighborhood businesses was added to display locations of these amenities relative to the participants' homes. Data for the local businesses were obtained from InfoUSA via SCAG.

Two different buffer sizes were used to define the microenvironment of each home location. The two buffer extents were: 0.25 mile and 0.125 mile around each of the 68 home locations. Next, the local businesses within the buffers were selected in GIS and enumerated into a new field. This field is a count variable of the number of business within each buffer extent from the respective home locations. After preliminary analyses and review of current literature (Guerra, E. & Cervero, 2013) the 0.125 mile buffers were excluded in favor of the 0.25 mile buffers because the former distance was considered too small to observe any neighborhood variations and the latter was deemed appropriate as a 'walkable' distance to access local amenities and jobs. This reasoning for the choice of this geographic extent is confirmed in the smart growth literature where the design mainly caters to pedestrian-friendly environments and the accessibility of amenities within one-quarter to one-half miles from residence (Ewing, 1999; Guerra, E. & Cervero, 2013).

The newly created neighborhood businesses count variable based on data from InfoUSA is a measure for proximity to *Destinations* (one of the five Ds). This concept was adopted from a previous study in the South Bay area of Los Angeles. The authors examined the effects of concentrations of commercial and retail centers classified by InfoUSA, on increasing walking activity (Boarnet et al., 2010).

The *Destination* count variable was sorted in ascending order and another field was added as a ranking variable, *NBCount*. The residence with the highest number of local businesses in the 0.25 mile catchment area was given a ranking of one in this neighborhood I.D. field suggesting the highest level of accessibility to local amenities. Therefore, the residence with a neighborhood I.D. equal to 68 reflects the lowest accessibility to amenities and to businesses. This concept was adapted from WALKSCORE.com which ranks physical addresses with a score ranging from 0 to 100 according to pedestrian accessibility to local amenities. A zero ranking indicates "car dependence" or that the majority of errands would require car use and a ranking of 90 or above is labeled as a "walker's paradise" where one can achieve daily errands without the use of a vehicle (Carr, Dunsiger, & Marcus, 2011).

The map in Figure 5-3 below shows the approximate[16] locations of the Expo participant homes and the local businesses in the corresponding 0.25-mile radius catchment areas. This map was used to extract the neighborhood level identifier used in the ranking procedure of households by number of amenities or local businesses. Since the geographic extent of the buffers is quite

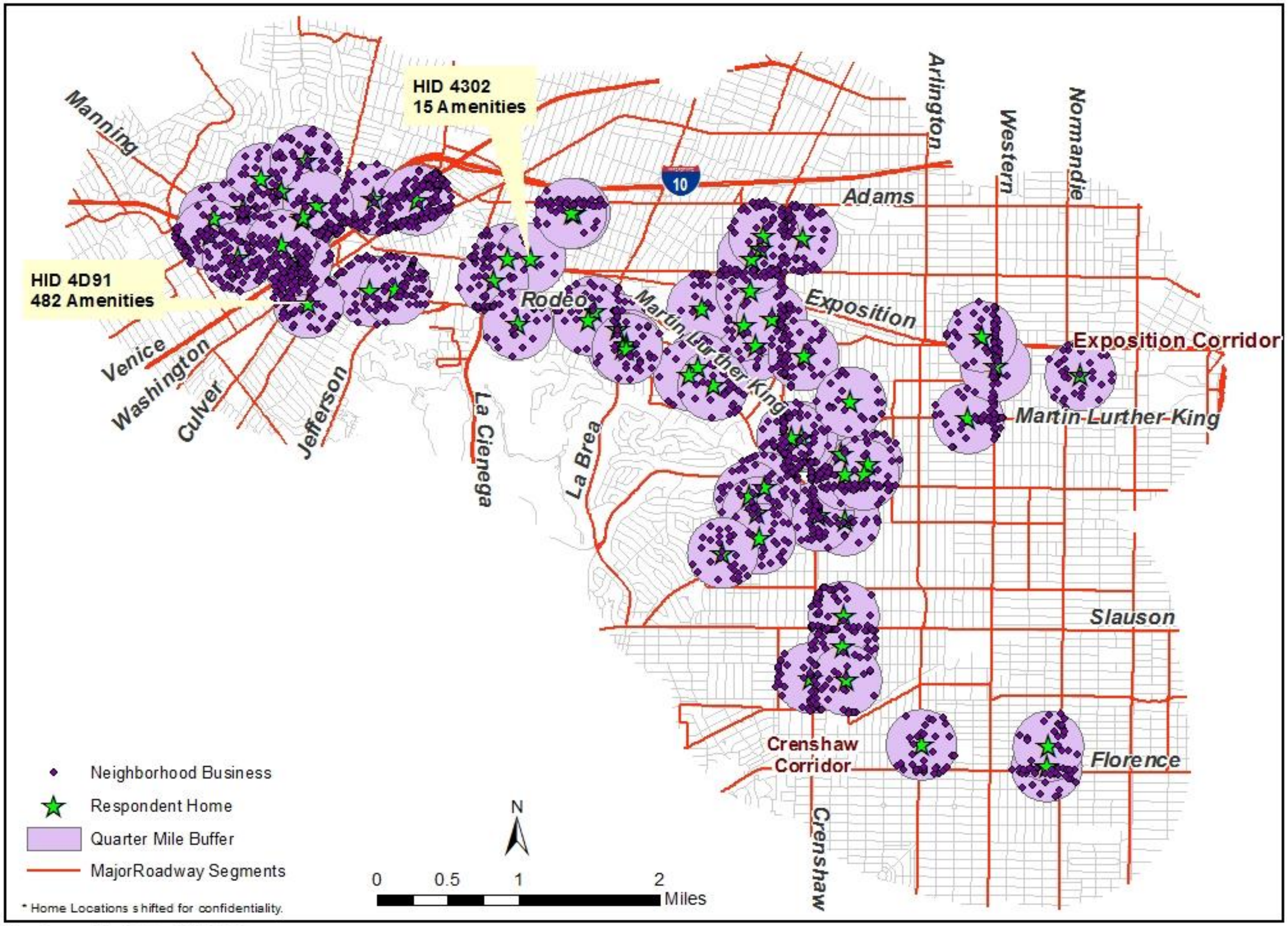
[16] Exact home locations are not displayed to protect the privacy of the survey participants.

small (0.25-mile) and many participants lived in close proximity to others in the survey, many of the buffers in the map overlapped. This however, did not affect the enumeration of the neighborhood businesses in the *NBCount* variable.

As a result, household number 4302 has the lowest accessibility to amenities; only 15 local businesses exist in its 0.25-mile catchment area. In contrast, household 4D91 has 482 local businesses in its 0.25-mile catchment area, the highest number of amenity accessibility in the sample. Figure 5-3 shows the approximate locations of these two households relative to the others in the sample.

One important condition for using a level-identifier is its unique quality. Unfortunately, a total of 20 households in pairs had the exact value for *NBCount*, indicating a duplicate for the number of local businesses. This caused a violation to the uniqueness of the neighborhood I.D. variable. To correct this, households with a duplicated *NBCount* value were selected, and for each identical pair their household I.D. (HID) was sorted alphabetically. The higher alphabetically ranking HID maintained the same value for *NBCount* and the lower ranking HID was selected and adjusted. The pre-sorting of similar HID pairs alphabetically was intended to permit the randomness of the ranking process to reduce potential bias.

In addition, another variable was created, *NBCountInt*, identical to *NBCount* except that it is an integer variable to allow for a 'mid-point' ranking with a 0.5 decimal classification. The HID with the lower alphabetical order had its *NBCount* value downgraded by 0.5 points and coded as such into the new variable *NBCountInt*. For example, households: A3F2 and ABE4 both had the value 27 for *NBCount,* ABE4 was selected to have its value changed to 26.5. This process ensured that each household has a unique ranking by the number of immediate accessible amenities. All neighborhood built environment variables were linked to this identifier.



Map Created by Gaby Abdel-Salam

Figure 5-3: Neighborhood Businesses within One-Quarter Mile from EXPO Households

5.6.2.2 Household-level Identifier

The household-level identifier (*HID*) was already present in the original data set. This variable was initially created during the design phase of the Expo study. The *HID* uniquely identifies each home and its respective coordinates or location. Variables at the household level such as number of children, total number of vehicles and annual household income are all linked to this identifier.

5.6.2.3 Momentary Activity-level Identifier

The momentary activity-level identifier was created using the matched accelerometer-GPS observations. The activity levels and locations of the primary respondent/participant in the sample were objectively measured via an Actigraph GT1M accelerometer (to track vertical and lateral movements in one-minute intervals) and the QT-1000x (QSTAR) GPS device (to track locations in 15-second epochs). Only Phase 1 data (before the Expo line introduction) is included here from both objective-measuring devices that were matched and checked for validity[17]. Data for 68 primary respondents (one from each household) were extracted from the 117 in the mobile-tracking group because they exhibited active transport through walking. The final sample used here is a moment-by-moment (15-second epochs) data set of 14,265 observations for all the 68 respondents uniquely identified by a variable that combines the household I.D. and time-date stamp. This unique identifier was sorted in chronological order and was used as the basis for creating the momentary activity-level identifier (a count I.D.) *AccGPS_ID*. This variable uniquely represents the time and date stamp of each of the primary

[17] Explained earlier in chapter three under the data quality control and validation section.

participant's momentary activity (15-second intervals) and the HID to which he/she resides (total *AccGPS_ID* = N = 14,265).

5.6.3 Creating the Land Use Variables

As noted earlier, one purpose of this study is to investigate the impacts of the built environment on active travel at varying extents. The various land use variables were selected according to the 5 D's of the built environment (Density, Destinations, Diversity, Distance & Design) that are commonly known in the planning literature (Boarnet et al., 2010; Condon, Cavens, & Miller, 2009; Durand, Andalib, Dunton, Wolch, & Pentz, 2011; Ewing, 1999). Land use was objectively measured using GIS in three different geographic extents: 40 meter buffered street segments, 0.25 and 0.5 mile radii from the participants' homes in the Expo study.

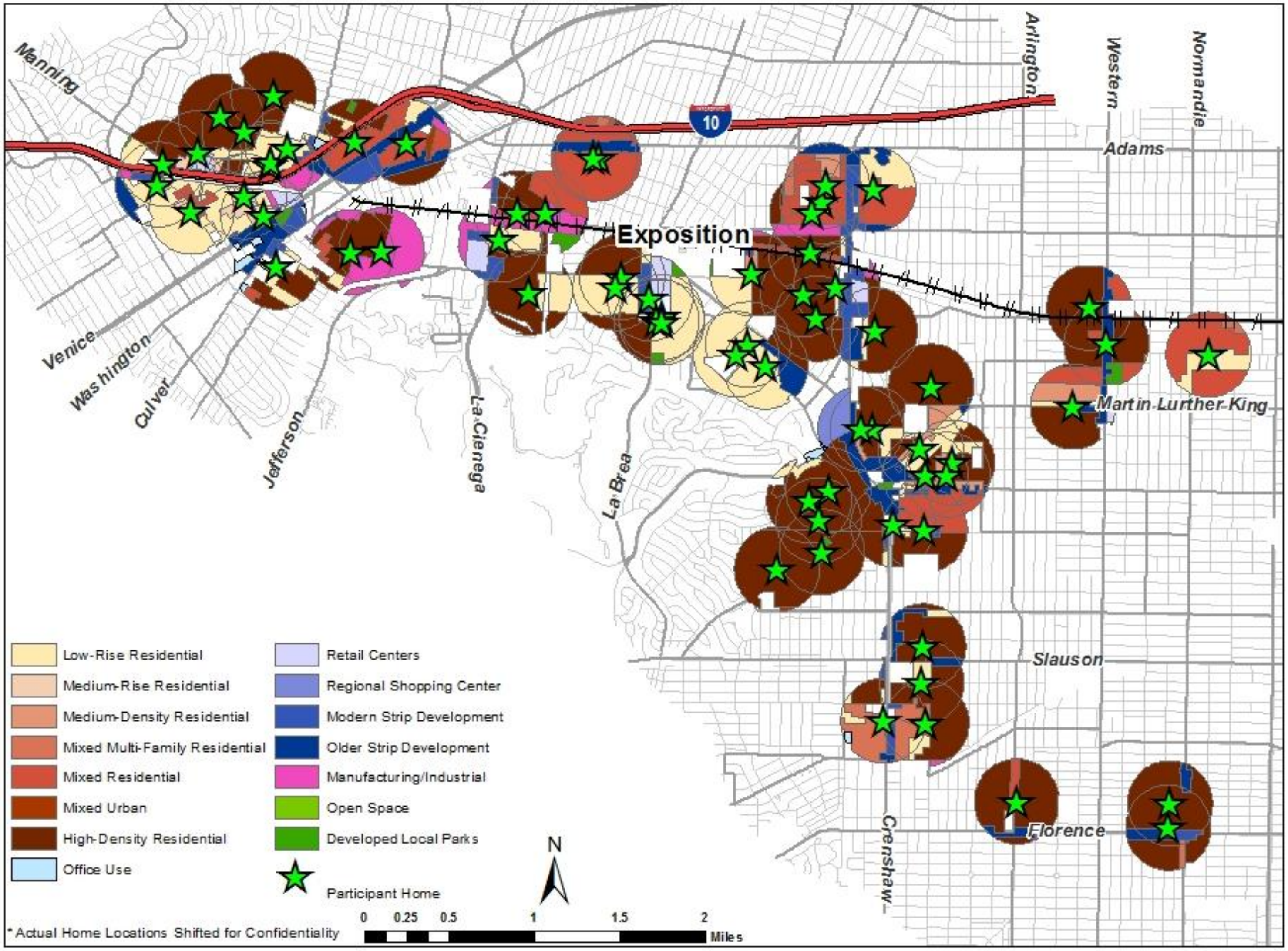
To achieve this goal, first, land use information (from SCAG shapefiles) were imported into GIS. This data comes subdivided into pre-defined polygons by land use type. Also included is information regarding land use code, types and year of collection. Next, the three different buffer sizes were created around the locations of each Expo household. The choice of the 0.25-mile and 0.5-mile buffer sizes was informed by the prevalent literature on a suitable walkable extent (Boarnet et al., 2010; Guerra, E. & Cervero, 2013; Lee & Moudon, 2006a). Generally, more statistical results were obtained at finer geographic extents (Almanza et al., 2012; Houston, 2014).

In contrast, the use of the 40 meter buffer size was inspired from the previous chapter's segment-level analysis that tests the effects of the built environment characteristics in the smallest possible setting; the immediate pedestrian environment.

Using the "geoprocessing" feature in GIS, I then combined each buffer with SCAG's land use database to create a new shapefile. This was done using the "intersect" tool. This resulted in the creation of new buffers which are now populated with the relevant buffer size information, coordinates (with Expo home as center) and the respective land use polygon information.

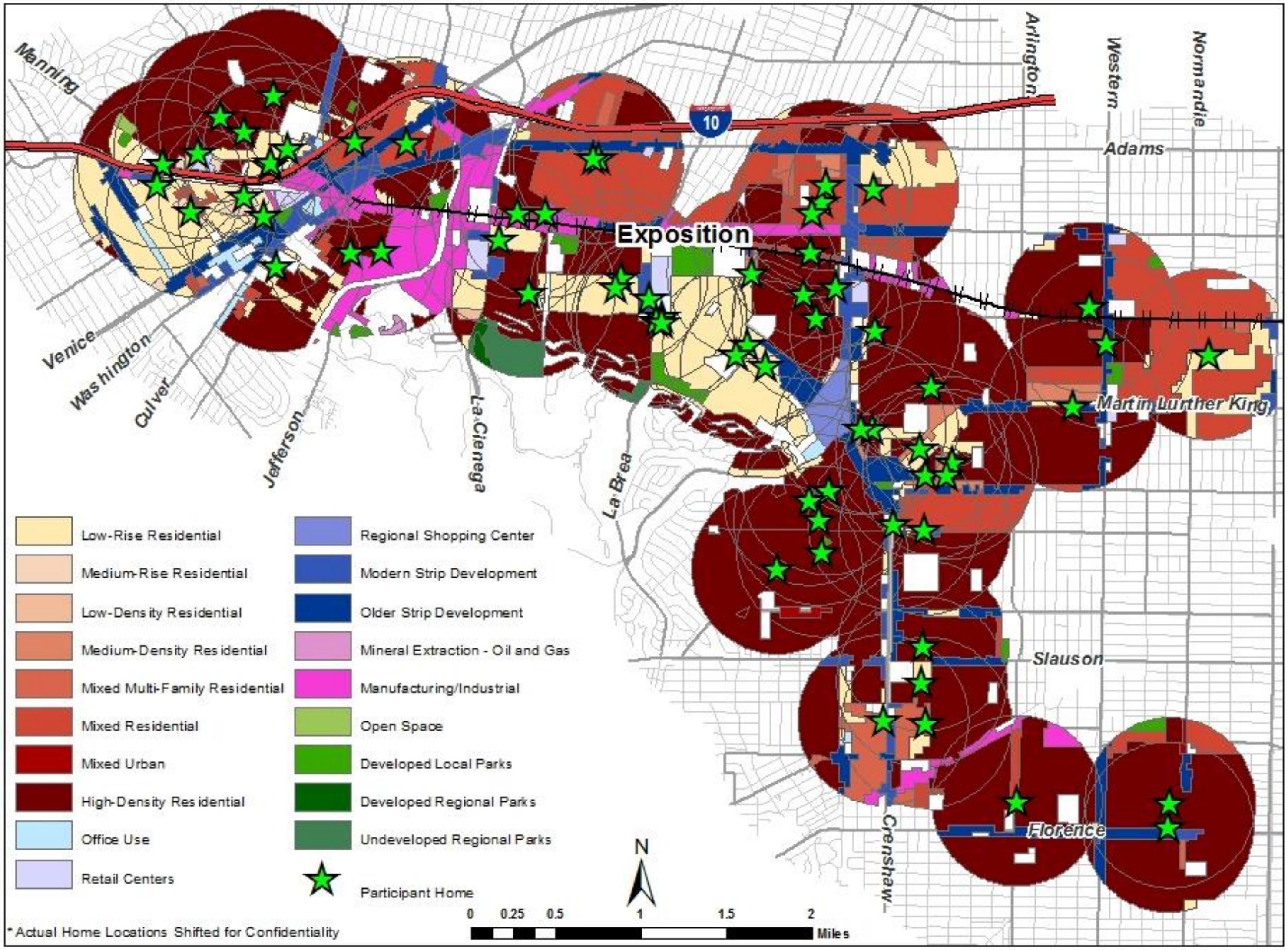
At this stage, all previously computed polygon area and perimeter measurements needed to be recalculated because of the intersection with the new buffer radii. The new calculations were done via the "calculate geometry" tool where "area" was selected in meters squared and square miles. In addition, another constant field representing total buffer area was added to the shapefiles: area (for 0.25 mile buffer) = 0.1963 sq. miles and the area (for 0.5 mile buffer) = 0.7853 sq. miles. Another field was created to represent the proportion of the area by land use type which was sorted in ascending order to ensure that none of the values exceed 1. The resulting shapefile is a household level file for all the 68 identified households. The database file (.dbf) for this new shapefile was then exported and merged by household I.D. with the original dataset. The result was a combined dataset (N = 14,265) where the unit of analysis is the 15-second epochs of matched accelerometer-GPS data points that included individual-, household- and neighborhood-level (built environment) variables.

The maps in Figures 5-4 and 5-5 display the buffered land use types with the EXPO home locations in the center for the quarter- and half-mile geographic extents respectively. The maps show that the majority of the industrial and manufacturing facilities exist along the Exposition corridor and that the majority of the commercial and retail uses exist near higher residential neighborhoods.



Map Created by Gaby Abdel-Salam

Figure 5-4: Types of Land Uses within One-Quarter Mile from EXPO Households



Map Created by Gaby Abdel-Salam

Figure 5-5: Types of Land Uses within One-Half Mile from EXPO Households

5.6.4 Grand-Mean Centering of Variables

With the hierarchical structure of multilevel analyses, a challenge arises in the interpretation of higher level variables and therefore requires some transformation. This is especially true for variables with a "raw metric" quality that lacks a "meaningful zero point" such as dichotomous variables (Enders & Tofighi, 2007). Variables that have a raw metric quality in this data set at the household (second) level are for example: number of vehicles or number of children living in the household. At the neighborhood (third) level, variables such as tree density or the density of commercial establishments also have a raw metric quality. The Binary variables, taking on values of 0 or 1, like household income that was coded into three dichotomous variables: low-income, middle-income and high-income; have inherently meaningful zero points and therefore no further actions were required to transform them.

The most common transformation techniques to create a meaningful zero interpretation for higher level variables are: group-mean and grand-mean centering (Bell et al., 2013). The former is especially appropriate when the variable has many clusters and the deviation of each data point from its within-cluster-mean is calculated for that variable. The latter is simpler and more suitable when the main focus of the regression analysis is on higher levels and it involves calculating the deviation of each observation from the overall mean of the relative variable. For a more detailed explanation of the differences between the two centering methods see Enders & Tofighi, 2007. I employ the grand-mean centering method throughout this chapter since the main interest is to analyze the associations between active transport and the built environment (third-level variables) given household and socio-demographic covariates.

5.6.5 Creating the Outcome Variable MVPAFlag

As mentioned earlier, the participants in this subsample were extracted from the larger mobile tracking group of phase one of the Expo study. These participants were selected because their activity patterns were consistent with walking behavior. As noted in chapter three, the matched accelerometer-GPS readings for each participant went through extensive automated then manual reviews to define the mode of travel using GIS. Since the main focus of this dissertation is active transport, only walking (including to transit) instances were used.

A categorical variable *ACTLevel* was previously created and was based on the activity level readings obtained from the participants' accelerometers. Consistent with the activity cut points used in the literature, participant activity levels were coded earlier by my colleague in the original data set into six categories as shown below in Table 5-1 which was based on Freedson, Melanson, & Sirard (1998).

TABLE 5-1: Activity Level and Respective Activity Counts

Activity Level	Counts per Minute (CPM)
Sedentary	0-99
Light	100-759
Lifestyle	760-1951
Moderate	1952-5724
Vigorous	5725-9498
Very Vigorous	> 9499

After selecting all valid instances of walking in Phase 1 and selecting the 14,265 momentary activity data points, I created the dichotomous variable *MVPAFlag* (= 1 if moderate-to-vigorous activity is observed; = 0 otherwise). This variable was based on the above categorical activity level variable and any activity level lower than "Moderate" meant that *MVPAFlag* = 0. Further, there were no instances in this sample with "Very Vigorous" activity levels. Therefore, *MVPAFlag* = 1 if "Moderate" or "Vigorous" activity levels were observed.

5.7 Model Specification

The dependent variable in the models *MVPAFlag*, is a dichotomous variable (= 1) reflecting instances of moderate-to-vigorous physical activity and (= 0) if less than moderate activity bouts are detected in any of the combined accelerometer-GPS readings. Considering a simpler model with one independent variable first X_{ij}, we can model our outcome variable *MVPAFlag* or Y_{ij}, an (n x 1) response vector representing the outcome variable, at two-levels in reduced form as:

$$Y_{ij} = \beta_0 + X_{ij}\beta_1 + u_j + e_{ij} \qquad \text{--------------------} \ (1)$$

Where [*i*] represents level one which is our respondent's level, [*j*] is level two or the household level, β_0 is an (n x 1) vector of the intercept values, X_{ij} is an (n x m) design matrix of the explanatory variables at level one, β_1 is an (m x 1) vector of the regression parameter estimates, u_j is an (n x 1) vector of random errors representing the random variation at level two and e_{ij} is an (n x 1) vector of random errors representing the random variation at level one.

The normal assumptions of equation (1) is that the expected values of the error terms is zero: $E[u_j] = E[e_{ij}] = 0$; and their variance and covariance is as follows: $\mathrm{var}(u_j) = \sigma_u^2$, $\mathrm{var}(e_{ij}) = \sigma_e^2$, $\mathrm{cov}(u_j, e_{ij}) = 0$ and $\mathrm{cov}(u_j, u_{j'}) = 0$ for $j \neq j'$. The last two terms mean an assumption of no covariance between the two-level error terms and no covariance for within same level errors respectively.

Expanding on the above equation, we can add a third term [*k*] representing the neighborhood level and adjusting equation (1) to allow for the inclusion of level-specific covariates. The equation becomes:

$$Y_{ijk} = \beta_0 + X_{ijk}\beta_1 + W_{jk}\beta_2 + Z_k\beta_3 + v_k + u_{jk} + e_{ijk} \quad \text{-------------------- (2)}$$

Equation (2) models the outcome variable Y_{ijk} at the three different levels, and can now include explanatory variables from all three levels such that X_{ijk}, W_{jk}, and Z_k are (n x m) matrices of covariates at levels one, two and three respectively and v_k, u_{jk} are level-three and level-two random intercepts and e_{ijk} is the random error of level-one and all have expected values equal to zero.

Since the outcome variable Y_{ijk} is dichotomous, we can therefore use a binary logistic regression to estimate the probability of *MVPAFlag* (Y_{ijk}) occurring for a single observation, thus, we model $p_{ijk} = \Pr(Y_{ijk} = 1)$ as in equation (3) below:

$$\log\left[\frac{p_{ijk}}{(1 - p_{ijk})}\right] = \beta_0 + \beta_1 x_{ijk} + \beta_2 w_{jk} + \beta_3 z_k + v_k + u_{jk} + e_{ijk} \quad \text{-------------------- (3)}$$

Equation (3) is estimable by full maximum likelihood (ML). All models were estimated by the PROC GLIMMIX command for non-linear response variables using SAS 9.2 software.

5.8 Model Building

For the analyses in this chapter, I follow the technique outlined in (Bell et al., 2013) with two adjustments: I only model the fixed effects not the random effects and use the command PROC GLIMMIX (for non-linear or binary response variables) instead of PROC MIXED (reserved for continuous variables). The basic model building process was as follows:

1. Model 1 is the unconditional model with no covariates only a random intercept is included. This model is estimated to obtain the intra-class correlation (ICC) or the variation in the response variable *MVPAFlag* between level two and level three observations.

2. Model 2 includes Model 1 covariates in addition to level one fixed effects. This model explains associations between the response variable *MVPAFlag* and level one (momentary activity-level) covariates.

3. Model 3 includes Model 2 covariates in addition to level two fixed effects. This model explains associations between the response variable *MVPAFlag* and level two (household-level) covariates.

4. Model 4 through 7 include Model 3 covariates in addition to level three fixed effects. These models explain the associations between the response variable *MVPAFlag* and level three (neighborhood-level) variables.

5.9 Using Three-Level vs. Two-Level Modeling

As noted earlier, many previous studies performed two-level analyses instead of the three-levels suggested currently in this study. The majority of them modeled the data into two strata: the neighborhood- and individual-levels (Ding et al., 2012; Prince et al., 2011; Van Dyck et al., 2010; Wasfi et al., 2013). I employed similar techniques to model the participants' microenvironment (neighborhood-level) but I further subdivided household from individual traits into the household-level and momentary-activity level of the individual which generates the additional third level. This was validated subjectively through the inherent hierarchy of the data and objectively by calculating the intra-class correlation coefficients (ICC). ICC values reported earlier indicated the suitability of modeling as three levels.

After estimating the unconditional model, the initial step in the previous section, we obtain the results from the 'covariance parameter estimates' table. This table provides three estimates: the neighborhood-level covariance σ^2_{NB}, the household-level covariance σ^2_{HH} and the model covariance residual estimate σ^2_{error}.

I used all three values to calculate the intra-class correlation coefficient which is also often referred to as the intra-cluster correlation coefficient. This is a measure of homogeneity within classes or clusters and ranges from 0 to 1. The ICC_{HH} (Household) coefficient reflects the correlation between two individuals within the same household and ICC_{NB} (Neighborhood) coefficient reflects the correlation between two households within the same neighborhood.

These coefficients are estimated from the unconditional Model One as follows:

Segment-level Model

$$ICC_{HH} = \frac{\sigma^2_{HH}}{\sigma^2_{NB} + \sigma^2_{HH} + \sigma^2_e} = \frac{0.03219}{[0.08078 + 0.03219 + 0.1327]} = \frac{0.03219}{0.24567} = 0.1310$$

$$ICC_{NB} = \frac{\sigma^2_{NB}}{\sigma^2_{NB} + \sigma^2_{HH} + \sigma^2_e} = \frac{0.08078}{[0.08078 + 0.03219 + 0.1327]} = \frac{0.08078}{0.24567} = 0.3288$$

Quarter- and Half-Mile Radius Models

$$ICC_{HH} = \frac{\sigma^2_{HH}}{\sigma^2_{NB} + \sigma^2_{HH} + \sigma^2_e} = \frac{0.03213}{[0.08267 + 0.03213 + 0.1327]} = \frac{0.03213}{0.2475} = 0.1298$$

$$ICC_{NB} = \frac{\sigma^2_{NB}}{\sigma^2_{NB} + \sigma^2_{HH} + \sigma^2_e} = \frac{0.08267}{[0.08267 + 0.03213 + 0.1327]} = \frac{0.08267}{0.2475} = 0.3340$$

The above calculations show that for the Segment-level Model, the ICC_{HH} = 13.10% and for the other two models, Quarter- and Half-Mile, the ICC_{HH} = 12.98%. These values reflect the similarity of participants within the same household or alternatively speaking, how much of the total variation in the *MVPAFlag* is explained by the households.

Similarly, for the Segment-level Model, the ICC_{NB}= 32.88% and for the other two models, Quarter- and Half-Mile, the ICC_{HH} = 33.40%. Again, these values show the similarity of households within the same neighborhood or alternatively speaking, how much of the total variation in the *MVPAFlag* is explained by the neighborhood.

Since all values for ICC > 0.01, clustering possibility is imminent and therefore multilevel regression is warranted at all geographic extents. Further, using both ICC_{HH} and ICC_{NB} values at the segment-level extent we can infer that 54.02% (100- (32.88 + 13.10) = 54.02%), the remaining variations in *MVPAFlag* exists at the respondents' or momentary-activity level.

Similarly, we can use the ICC_{HH} and ICC_{NB} values at the Quarter- and Half-Mile geographic extent to infer that 53.62% (100- (33.40 + 12.98) = 53.62%), the remaining variations in *MVPAFlag* exists at the respondents' or momentary-activity level.

Therefore, from the above results, we can conclude that a significant portion of the variance in MVPAFlag, the response variable, exists at the household and neighborhood levels. This indicates the validity of a three-level multivariate analysis over a two-level regression[18].

5.10 Model Results

In this section, the descriptive statistics of level-specific variables are presented as well as the results from the three class multilevel models. As explained earlier, the analyses were performed at three geographic extents: segment-level, quarter-mile and half-mile radii from the Expo study participants' residences. The reasoning for the varying geographic extents is to display detailed level results and show variations in the associations between key land use variables and the probability of observing momentary MVPA instances.

5.10.1 Descriptive Statistics

Tables 5-2, 5-3, 5-4(a) and 5-4(b) below show the descriptive statistics of the various variables used at each level including the grand-mean (GM) centered statistics for the mean (GM_Mean),

[18] Initially I performed a two-level (neighborhood and household) regression model yielding ICC_{NB} = 0.4587. This indicates that 45.87% of the variations in MVPA exist between neighborhood types suggesting that 54.13% of the variations in the response variable exist among households. However, by aggregating the smaller level (individual) to the higher household level we may inject bias into the models (Goulias, 2002) and prevent the parceling out of associations of level-specific variables with the response variable.

median (GM_Median) and standard deviation (GM_S.D.). Grand-mean centering of the variables is not necessary for binary (having values 1/0) or first-level variables. The tables include information regarding the number of observations (N) used, the mean, median and the standard deviation of each variable.

The histograms displayed in Figures 5-5 through 5-8 were created in GIS based on SCAG (2005) land use database. These graphs provide insight into the distributions of the three key land use variables at each respective household I.D. (HID) in the quarter-mile and the half-mile radii from the homes in the sample.

Figures 5-6 and 5-8 show the respective distributions of the key land use density variables. From the figures, we can see that the high and low residential density graphs are complementary to one another. Specifically, negative spaces that appear in the former graph are actually present as positive ones in the latter one. This is expected since both classification categories add up to 100% of all residential density. Also interesting to note, are the trends in the values of the commercial density; it corresponds with those of the high residential density graph. This suggests a synergy between denser neighborhoods and access to more commercial and retail amenities since denser neighborhoods are targeted for new retail centers because of potential higher demands.

The graphs in Figures 5-7 and 5-9 on the other hand, show the respective distributions of the same variables when they have been grand-mean centered. As described before, variables in the highest strata (beyond the first level) in multilevel modeling require the transformation of their values to a “meaningful zero” for ease of interpretation. The transformation was accomplished by grand-mean centering of the variables. Therefore, the horizontal zero line on each graph represents the mean density value of the land use type modeled. Positive values (above the zero line), represent higher than average densities for the respective land use type and negative values (below the zero line); correspond to lower than average densities of the relative land use type. For example, household number C21D has a lower than average "high" residential density (and therefore, higher than average "low" residential density). Similarly, household number D5F7 has much higher than average commercial density at both geographic extents.

Table 5-2: Descriptive Statistics First Level

Variable	Momentary-Activity Level (Participant Characteristics)								
	[Reference: Male, White, Employed, Education less than High School]								
	Age	Female (Binary)	Race Asian (Binary)	Race Hispanic (Binary)	Race Black (Binary)	Race Other (Binary)	Unemployed (Binary)	Education AA/Less (Binary)	Education BA/BS/Grad (Binary)
N	14,265	14,265	14,265	14,265	14,265	14,265	14,265	14,265	14,265
Mean	57.467	0.614	0.016	0.014	0.521	0.521	0.716	0.206	0.153
Median	56	1	0	0	1	1	1	0	0
S.D.	12.534	0.487	0.125	0.116	0.500	0.500	0.451	0.405	0.360
N									
GM_Mean									
GM_Median									
GM_S.D.									

Table 5-3: Descriptive Statistics Second Level

Variable	Household-Level			
	[Reference: Low Income < $35K Annually]			
	No. Children	No. Cars	Middle Income (Binary)	High Income (Binary)
N			14,265	14,265
Mean			0.100	0.074
Median			0	0
S.D.			0.300	0.263
N	14,265	12,396		
GM_Mean	-0.003	-0.031		
GM_Median	-0.300	-0.300		
GM_S.D.	0.543	1.057		

Table 5-4(a): Descriptive Statistics Third Level

Variable	Qtr-Mile Industrial Density	Qtr-Mile Office Density	Qtr-Mile Open Park Density	Qtr-Mile Commercial Density	Qtr-Mile High Residential Density	Qtr-Mile Low Residential Density	Tree Density	Irrigated Lawn Density	Non-Irrigated Lawn Density
	Neighborhood-Level								
N	14,265	14,265	14,265	14,265	14,265	14,265			
Mean	0.018	0.005	0.005	0.140	0.580	0.170			
Median	0	0	0	0.143	0.642	0.098			
S.D.	0.066	0.018	0.009	0.094	0.230	0.186			
N	14,265	14,265	14,265	14,265	14,265	14,265	14,265	14,265	14,265
GM_Mean	0	0	0	0	0	0	0.002	-0.002	0.004
GM_Median	-0.018	-0.005	-0.005	0.004	0.062	-0.072	-0.200	-0.300	0.000
GM_S.D.	0.066	0.018	0.009	0.094	0.230	0.186	0.580	0.739	0.148

Table 5-4(b): Descriptive Statistics Third Level

Variable	Medium/ Hi Traffic (Binary)	No. Street Intersections	No. Transit Stops	Hlf-Mile Industrial Density	Hlf-Mile Office Density	Hlf-Mile Open Park Density	Hlf-Mile Commercial Density	Hlf-Mile High Residential Density	Hlf-Mile Low Residential Density
	Neighborhood-Level								
N	14,265			14,265	14,265	14,265	14,265	14,265	14,265
Mean	0.601			0.035	0.008	0.019	0.116	0.587	0.145
Median	1			0.006	0	0.007	0.109	0.652	0.081
S.D.	0.490			0.069	0.014	0.032	0.056	0.199	0.147
N		14,265	14,265	14,265	14,265	14,265	14,265	14,265	14,265
GM_Mean		-0.021	-0.043	0	0	0	0	0	0
GM_Median		-1	-13	-0.029	-0.008	-0.012	-0.007	0.064	-0.064
GM_S.D.		1.727	23.065	0.069	0.014	0.032	0.056	0.199	0.147

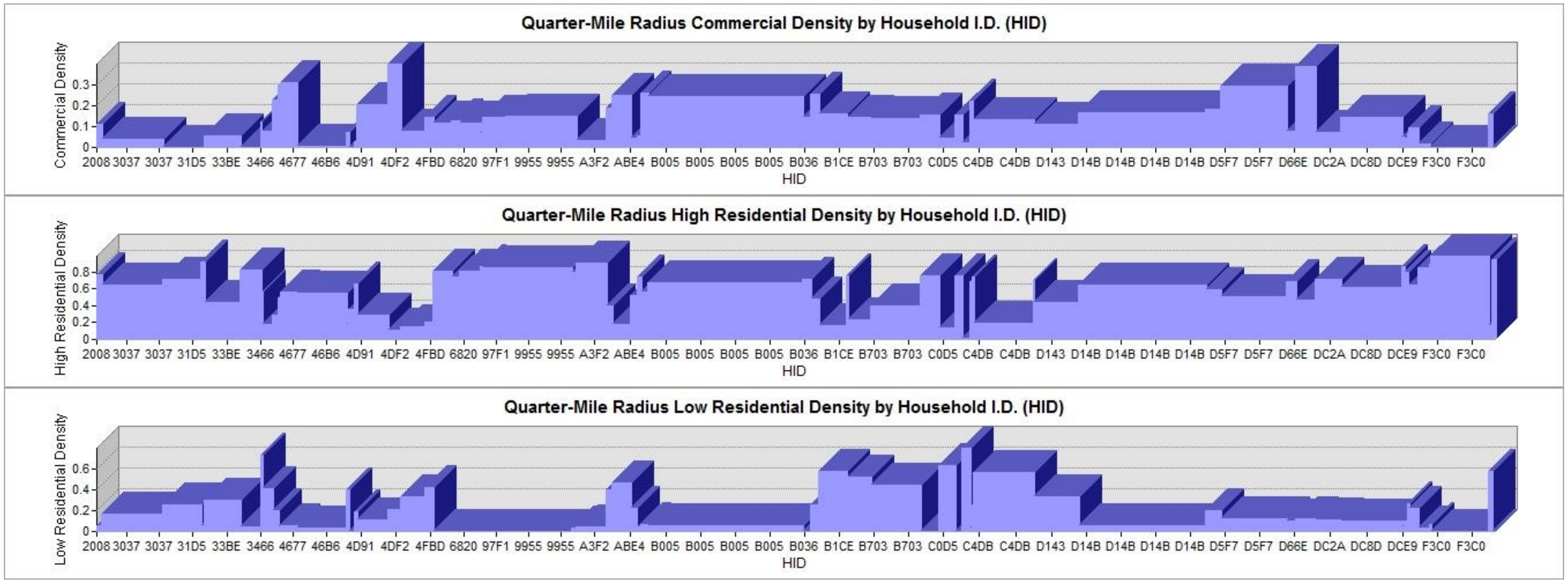


Figure 5-6: Land Use Densities One-Quarter Mile from Expo Households

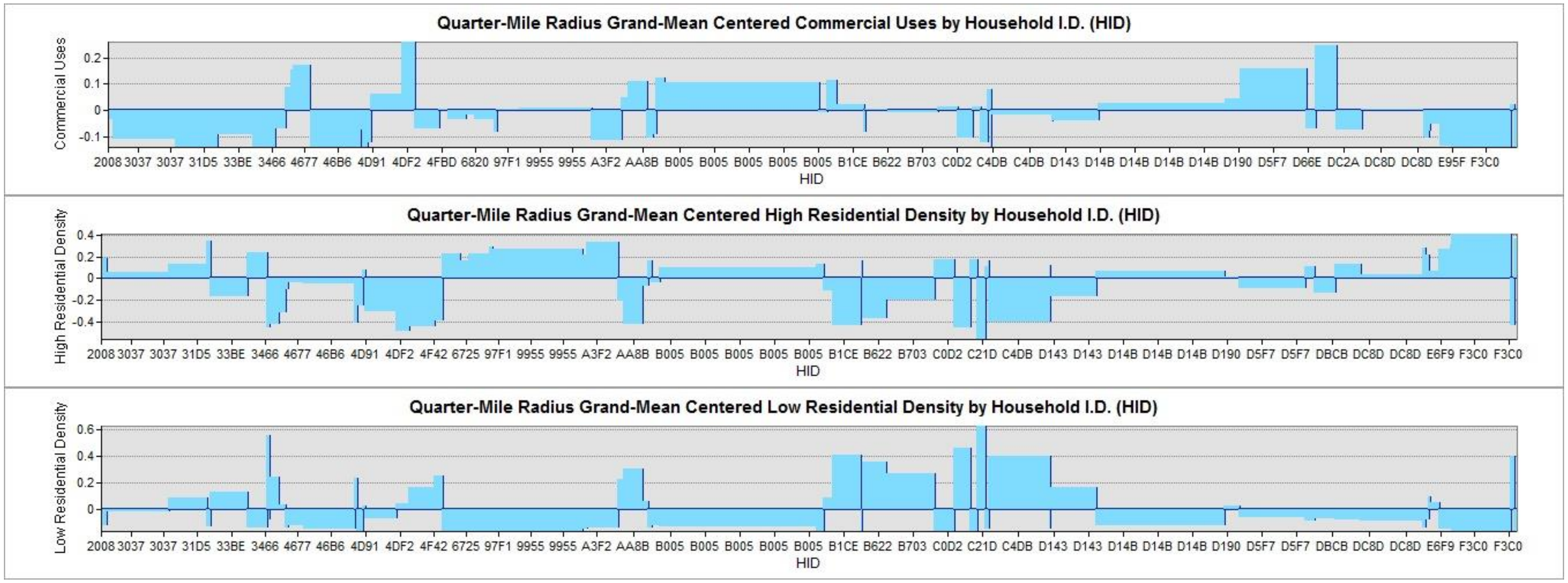


Figure 5-7: Grand-Mean Centered Land Uses One-Quarter Mile from Expo Households

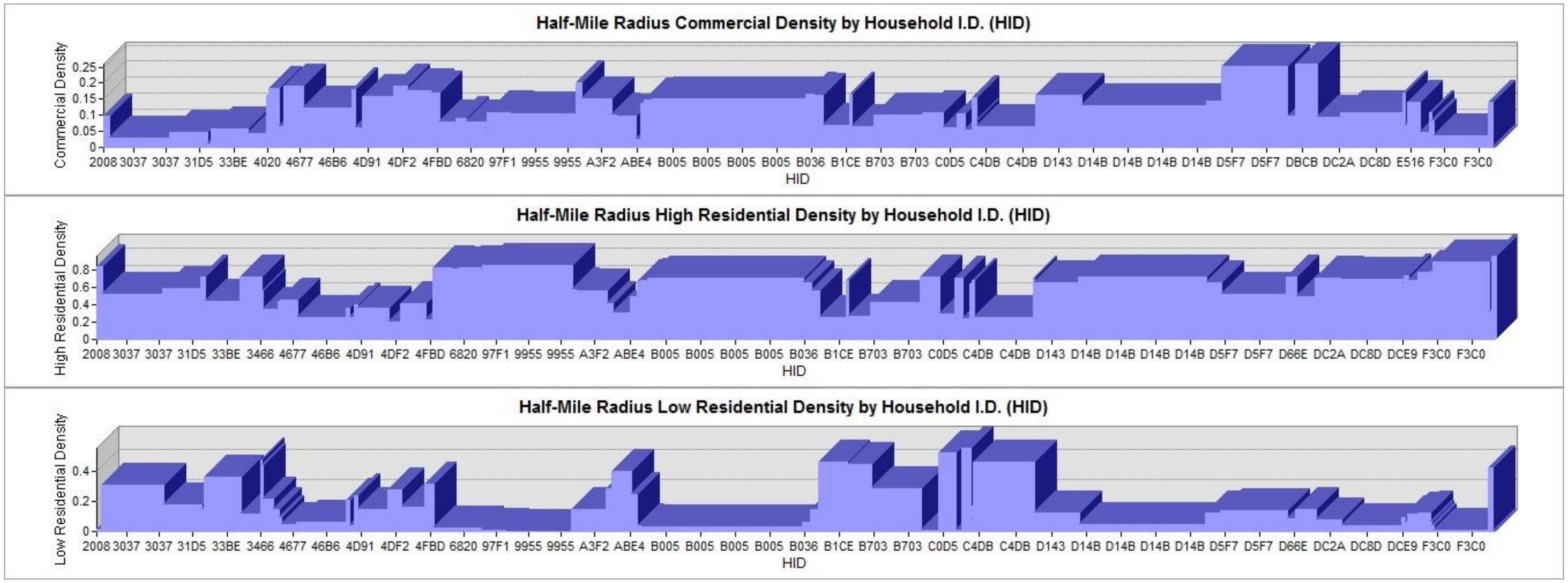


Figure 5-8: Land Use Densities One-Half Mile from Expo Households

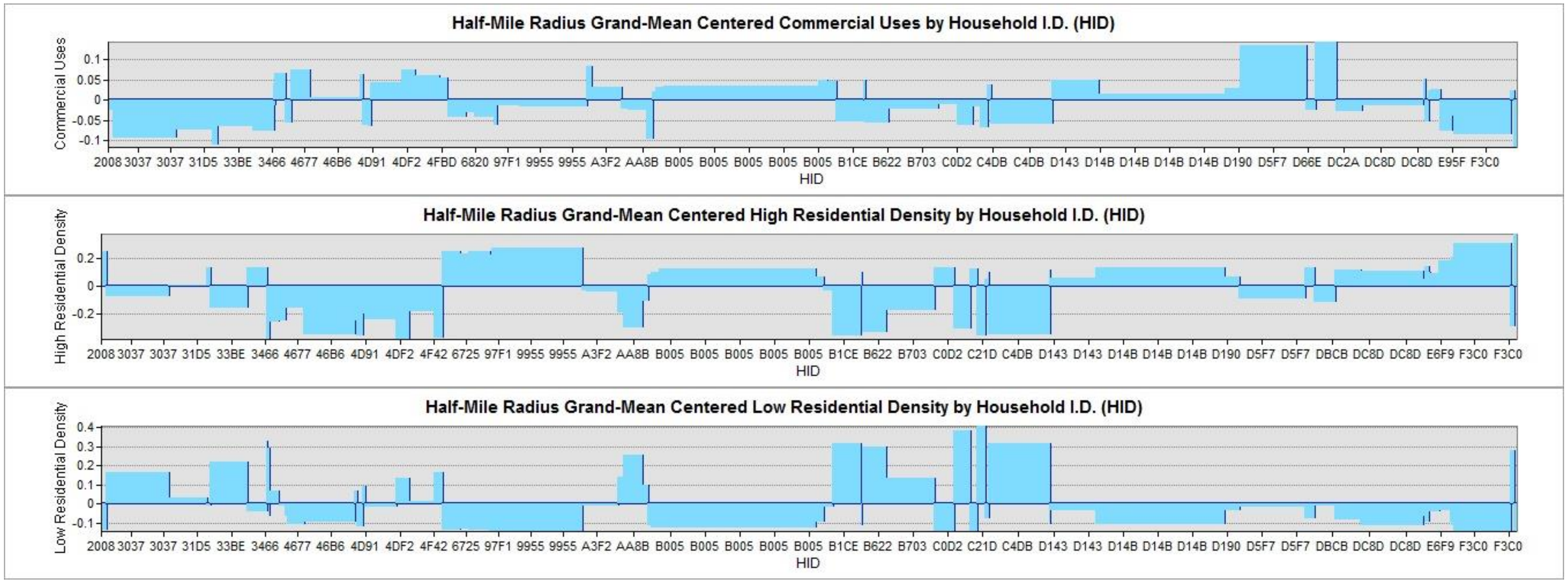


Figure 5-9: Grand-Mean Centered Land Uses One-Half Mile from Expo Households

5.10.2 Model Description

As noted previously, the outcome variable *MVPAFlag* is dichotomous and is therefore estimable by a multilevel logistic regression model using maximum likelihood. I used the generalized linear mixed model PROC GLIMMIX (with a binary distribution and a logit link function a.k.a multilevel logistic regression for binary outcomes) instead of the suggested PROC Mixed command in Bell et al. (2013). The latter is reserved for continuous outcome variables, whereas the former is utilized for non-linear or binary response variables. The PROC GLIMMIX command uses the Newton-Raphson optimization technique with ridging. The three-level generalized linear mixed model results are displayed below in Tables 5-5(a to b), 5-6(a to d) and 5-7(a to d). The results in all tables reflect predicted values for the probability of momentary MVPA occurring when covariates at the neighborhood-, household- and individual-level are accounted for. Tables 5-5(a to b) hold the multilevel regression results for built environment variables calculated within the 40 meter road buffers near the participants' homes. Tables 5-6(a to d) show results from the built environment variables calculated within a quarter-mile from residence. In addition to the quarter-mile land use features included in the tables, I also added segment-level built environment features in Models 6 and 7. Lastly, Tables 5-7(a to d) display results for the built environment characteristics calculated within a half-mile from Expo participants' residence, except for Models 6 and 7; I also added built environment features at both the half-mile and buffered segment-level extents. The analyses have been performed as such to compare variations in significance and effects of land use on active travel at different geographic extents.

In all the tables, the three-level multivariate analyses reported here only models fixed effects of the level-specific variables. No random effects for these three-level variables were estimated. Further, only the intercepts were included in the random effects portion of the models.

In all the tables, Model 1 is unconditional on any predictors and is estimated primarily to obtain the between class variances to validate the use of multilevel regression analysis. A large variance justifies the use of this modeling technique. In Model 2, the momentary-activity or respondent-level identifier was utilized that uniquely captures increments of the respondent's 15-second date and time stamp. Estimates from this model show the respective significance and associations between level-one respondent characteristics and the response variable, *MVPAFlag*. In Model 3, I incorporate the household identification code as a class identifier, added level-two household characteristics in addition to level-one variables defined in Model 2.

Model 4 in Table 5-5(b), integrates variables from all three levels. Two class identifiers (household I.D. and neighborhood I.D.) were used in this model which included level-three built environment characteristics in addition to the level-specific variables of Model 3.

Similarly, Models 4 to 7 in Table 5-6 (quarter-mile extent) and Table 5-7 (half-mile extent) integrate variables from all three levels. Two class identifiers (household I.D. and neighborhood I.D.) were used in these models which included level-three built environment characteristics in addition to the level-specific variables from Model Three.

A mix of land use measures were incorporated in Tables 5-6(a to d) and Tables 5-7(a to d). Model 4 includes density measures for the land use variables. Model 5 has area measures for the same land use variables. Models 6 and 7 have a mix of the density land use measures and five other segment-level density and indicator variables. The segment-level variables include: street connectivity, number of transit stops, traffic volume (medium/high) and three land cover variables (tree density, irrigated and non-irrigated lawn densities) representing green space.

Goodness-of-fit measures are reported at the bottom of each model in all three tables. These include: the pseudo-Akaike Information Criterion (AIC), the pseudo-Bayesian Information Criterion (BIC) and Likelihood Ratio. Only pseudo-criterion values were estimated since PROC GLIMMIX in SAS does not estimate the raw values for these measures.

In the segment-level extent models of Table 5-5(b), goodness-of-fit measures show tremendous improvements in their values for Model 4 over the previous three models. Generally, we want to choose the model with the lowest pseudo-AIC or pseudo-BIC value and a ten-unit or more reduction from one model to the next is considerable improvement. Here, we can compare the values from Model 3 and Model 4 which were 10,953.5 and 10,916.1 respectively. This validates that Model 4 is a better fit for the data suggesting the appropriateness of using a three-level over a two-level modeling technique for *MVPAFlag*.

The measures for the quarter-mile radius models in Tables 5-6(a to d) and half-mile radius models in Table 5-7(a to d) reflect promising results. We can see that both the pseudo-AIC and pseudo-BIC values drop significantly from the unconditional Model 1 as we add more variables from the three-levels. Comparing the density variable models, Model 4, in both geographic extents has the lowest pseudo-AIC and pseudo-BIC values. However, Model 6 has relatively low pseudo-AIC and pseudo-BIC measurements and reflects the largest number of significant variables from all three levels. Regression results are discussed in the next section.

Table 5-5(a): Segment-level Multilevel Analysis

	Model 1 (Unconditional)				Model 2 (Level 1)			
Dependent Variable	MVPAFlag				MVPAFlag			
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	0.501				0.594			
First Level: Individuals' Characteristics								
Female					-0.011	0.489		0.989
Age					-0.002	0.489		0.998
Asian					-0.391	0.111		0.676
Hispanic					-0.636	0.021	**	0.530
Black					-0.145	0.115		0.865
Other					0.078	0.424		1.081
Unemployed					0.229	0.010	***	1.257
Education AA or Less					0.153	0.202		1.166
Education Graduate Degree or less					0.031	0.749		1.031
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)								
High Income (75K or more/yr.)								
No. of Cars								
No. Children								
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections								
Total Transit Stops								
Total Commercial Uses								
Total Industrial Uses								
Medium-to-High Traffic								
Tree Density								
Irrigated Lawn Density								
Non-Irrigated Lawn Density								
ICC (Household)	0.032							
ICC (Neighborhood)	0.081							
Residual	0.133							
N	14,265				14,265			
AIC	11,975.2				11,972.2			
BIC	11,984.1				11,998.7			
Likelihood Ratio	11,967.2				11,948.2			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-5(b): Segment-level Multilevel Analysis

	Model 3 (Levels 1 & 2)				**Model 4 (All 3 Levels)**			
Dependent Variable	**MVPAFlag**				**MVPAFlag**			
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	0.783				0.764			
First Level: Individuals' Characteristics								
Female	-0.066	0.446		0.936	-0.061	0.488		0.941
Age	-0.005	0.099	*	0.995	-0.005	0.115		0.995
Asian	-0.375	0.114		0.688	-0.373	0.120		0.689
Hispanic	-0.644	0.017	**	0.525	-0.643	0.018	**	0.526
Black	-0.124	0.186		0.883	-0.110	0.245		0.895
Other	0.071	0.482		1.073	0.087	0.391		1.091
Unemployed	0.198	0.027	**	1.218	0.206	0.022	**	1.229
Education AA or Less	0.063	0.625		1.065	0.084	0.519		1.088
Education Graduate Degree or less	-0.097	0.325		0.907	-0.084	0.401		0.920
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)	0.322	0.024	**	1.380	0.313	0.029	**	1.368
High Income (75K or more/yr.)	0.175	0.195		1.191	0.158	0.244		1.172
No. of Cars	0.053	0.237		1.054	0.048	0.286		1.049
No. Children	-0.089	0.195		0.915	-0.088	0.207		0.916
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections					0.007	0.005	***	1.007
Total Transit Stops					-0.001	<.0001	***	0.999
Total Commercial Uses					0.019	<.0001	***	1.019
Total Industrial Uses					0.015	0.278		1.015
Medium-to-High Traffic					-0.020	0.069	*	0.981
Tree Density					0.017	0.103	*	1.017
Irrigated Lawn Density					-0.007	0.545		0.993
Non-Irrigated Lawn Density					-0.011	0.796		0.989
ICC (Household)								
ICC (Neighborhood)								
Residual								
N	14,265				14,265			
AIC	10,953.5				10,916.1			
BIC	10,988.0				10,967.9			
Likelihood Ratio	10,921.5				10,868.1			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-6(a): Quarter-mile Radius Multilevel Analysis

	Model 1 (Unconditional)				Model 2 (Level 1)			
Dependent Variable	**MVPAFlag**				**MVPAFlag**			
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	0.501				0.613			
First Level: Individuals' Characteristics								
Female					-0.318	0.643		0.728
Age					-0.011	0.633		0.989
Asian					-3.294	0.043		0.037
Hispanic					-5.171	0.043	**	0.006
Black					-1.058	0.151		0.347
Other					0.293	0.707		1.340
Unemployed					1.376	0.412	**	3.957
Education AA or Less					0.800	0.412		2.226
Education Graduate Degree or less					0.408	0.599		1.503
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)								
High Income (75K or more/yr.)								
No. of Cars								
No. Children								
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections								
Total Transit Stops								
Medium-to-High Traffic								
Tree Density								
Irrigated Lawn Density								
Non-Irrigated Lawn Density								
0.25-Mile Data								
Industrial Uses								
Office Uses								
Commercial/Retail Uses								
High-Density Residential Uses								
Low-Density Residential Uses								
Open Park Area								
ICC (Household)	0.130				0.000			
ICC (Neighborhood)	0.334				0.978			
Residual	0.133				0.133			
N	14,265				14,265			
AIC	11977.72				73523.45			
BIC	11984.34				73525.65			
Likelihood Ratio	11971.72				73521.45			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-6(b): Quarter-Mile Radius Multilevel Analysis

	Model 3 (Level 1 & 2)				Model 4 (All 3 Levels)			
Dependent Variable	**MVPAFlag**				**MVPAFlag**			
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	2.007				2.299			
First Level: Individuals' Characteristics								
Female	-0.699	0.347		0.497	-0.707	0.362		0.493
Age	-0.032	0.197		0.968	-0.035	0.204		0.966
Asian	-2.807	0.179		0.060	-3.471	0.112		0.031
Hispanic	-4.917	0.045	**	0.007	-5.665	0.026	**	0.003
Black	-0.929	0.245		0.395	-0.636	0.444		0.529
Other	0.194	0.819		1.214	0.428	0.629		1.534
Unemployed	1.164	0.119		3.201	0.883	0.274		2.418
Education AA or Less	0.072	0.949		1.074	-0.040	0.973		0.961
Education Graduate Degree or less	-0.508	0.549		0.602	-0.488	0.573		0.614
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)	2.417	0.043	**	11.212	2.615	0.035	**	13.67
High Income (75K or more/yr.)	0.969	0.393		2.635	0.431	0.736		1.54
No. of Cars	0.352	0.355		1.422	0.208	0.625		1.23
No. Children	-0.757	0.219		0.469	-0.760	0.222		0.47
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections								
Total Transit Stops								
Medium-to-High Traffic					-0.217	0.002	***	0.805
Tree Density								
Irrigated Lawn Density								
Non-Irrigated Lawn Density								
0.25-Mile Data								
Industrial Uses					-15.09	0.052	*	0
Office Uses					-0.144	0.996		0.866
Commercial/Retail Uses					-11.31	0.119		0
High-Density Residential Uses					-8.274	0.158		0
Low-Density Residential Uses					-11.34	0.097	*	0
Open Park Area					-32.00	0.362		0
ICC (Household)	0.000							
ICC (Neighborhood)	0.977							
Residual	0.133							
N	12,396				12,396			
AIC	63809.96				64000.36			
BIC	63812.11				64002.52			
Likelihood Ratio	63807.96				63998.36			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-6(c): Quarter-Mile Radius Multilevel Analysis

	Model 5 (All 3 Levels)				Model 6 (All 3 Levels)			
Dependent Variable	**MVPAFlag**				**MVPAFlag**			
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	2.243				1.810			
First Level: Individuals' Characteristics								
Female	-0.720	0.355		0.487	-0.546	0.462		0.579
Age	-0.034	0.214		0.967	-0.029	0.252		0.971
Asian	-3.430	0.116		0.032	-3.276	0.121		0.038
Hispanic	-5.619	0.028	**	0.004	-5.270	0.030	**	0.005
Black	-0.628	0.451		0.534	-0.543	0.504		0.581
Other	0.444	0.617		1.558	0.598	0.491		1.818
Unemployed	0.862	0.287		2.368	0.834	0.286		2.303
Education AA or Less	-0.032	0.979		0.969	0.133	0.905		1.142
Education Graduate Degree or less	-0.494	0.569		0.610	-0.564	0.496		0.569
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)	2.638	0.034	**	13.99	2.433	0.041	**	11.40
High Income (75K or more/yr.)	0.457	0.721		1.580	0.258	0.829		1.294
No. of Cars	0.212	0.619		1.236	0.305	0.443		1.356
No. Children	-0.750	0.230		0.472	-0.763	0.209		0.466
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections					0.033	0.055	*	1.034
Total Transit Stops					-0.005	0.001	***	0.995
Medium-to-High Traffic	-0.217	0.002	***	0.805	-0.132	0.110		0.877
Tree Density					0.220	0.015	**	1.246
Irrigated Lawn Density					-0.071	0.465		0.931
Non-Irrigated Lawn Density					-0.423	0.161		0.655
0.25-Mile Data								
Industrial Uses	-74.90	0.053	*	0	-15.69	0.019	**	0
Office Uses	-0.239	0.999		0.788				
Commercial/Retail Uses	-56.58	0.122		0	-12.17	0.077	*	0
High-Density Residential Uses	-41.20	0.160		0	-8.52	0.087	*	0
Low-Density Residential Uses	-56.60	0.099	*	0	-11.70	0.040	**	0
Open Park Area	-156.62	0.381		0				
ICC (Household)								
ICC (Neighborhood)								
Residual								
N	12,396				12,396			
AIC	63984.19				64038.60			
BIC	63986.35				64040.76			
Likelihood Ratio	63982.19				64036.60			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-6(d): Quarter-Mile Radius Multilevel Analysis

	Model 7 (All 3 Levels)			
Dependent Variable	**MVPAFlag**			
	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables				
Intercept	1.836			
First Level: Individuals' Characteristics				
Female	-0.551	0.465		0.576
Age	-0.030	0.256		0.970
Asian	-3.245	0.134		0.039
Hispanic	-5.224	0.036	**	0.005
Black	-0.543	0.511		0.581
Other	0.588	0.505		1.799
Unemployed	0.854	0.287		2.348
Education AA or Less	0.170	0.883		1.185
Education Graduate Degree or less	-0.545	0.524		0.580
Second Level: Households' Characteristics				
Middle Income (35K - 55K/yr.)	2.419	0.046	**	11.23
High Income (75K or more/yr.)	0.199	0.875		1.221
No. of Cars	0.310	0.444		1.364
No. Children	-0.775	0.210		0.461
Third Level: Neighborhoods' Characteristics				
Segment-level Data				
Total Street Intersections	0.033	0.055	*	1.033
Total Transit Stops	-0.005	0.001	***	0.995
Medium-to-High Traffic	-0.132	0.110		0.877
Tree Density	0.220	0.015	**	1.246
Irrigated Lawn Density	-0.071	0.465		0.931
Non-Irrigated Lawn Density	-0.423	0.161		0.655
0.25-Mile Data				
Industrial Uses	-15.194	0.049	**	0
Office Uses	4.020	0.882		55.71
Commercial/Retail Uses	-11.936	0.098	*	0
High-Density Residential Uses	-8.115	0.163		0
Low-Density Residential Uses	-11.203	0.099	*	0
Open Park Area				
ICC (Household)				
ICC (Neighborhood)				
Residual				
N	12,396			
AIC	64060.87			
BIC	64063.03			
Likelihood Ratio	64058.87			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-7(a): Half-Mile Radius Multilevel Analysis

	Model 1 (Unconditional)				**Model 2 (Level 1)**			
Dependent Variable		**MVPAFlag**				**MVPAFlag**		
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	0.501				0.613			
First Level: Individuals' Characteristics								
Female					-0.318	0.643		0.728
Age					-0.011	0.633		0.989
Asian					-3.294	0.043		0.037
Hispanic					-5.171	0.043	**	0.006
Black					-1.058	0.151		0.347
Other					0.293	0.707		1.340
Unemployed					1.376	0.412	**	3.957
Education AA or Less					0.800	0.412		2.226
Education Graduate Degree or less					0.408	0.599		1.503
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)								
High Income (75K or more/yr.)								
No. of Cars								
No. Children								
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections								
Total Transit Stops								
Medium-to-High Traffic								
Tree Density								
Irrigated Lawn Density								
Non-Irrigated Lawn Density								
0.5-Mile Data								
Industrial Uses								
Office Uses								
Commercial/Retail Uses								
High-Density Residential Uses								
Low-Density Residential Uses								
Open Park Area								
ICC (Household)	0.130							
ICC (Neighborhood)	0.334							
Residual	0.133							
N	14,265				14,265			
AIC	11977.72				73523.45			
BIC	11984.34				73525.65			
Likelihood Ratio	11971.72				73521.45			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-7(b): Half-Mile Radius Multilevel Analysis

	Model 3 (Level 1 & 2)				Model 4 (All 3 Levels)			
Dependent Variable	**MVPAFlag**				**MVPAFlag**			
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	2.007				1.719			
First Level: Individuals' Characteristics								
Female	-0.699	0.347		0.497	-0.787	0.348		0.455
Age	-0.032	0.197		0.968	-0.025	0.358		0.975
Asian	-2.807	0.179		0.060	-2.655	0.218		0.070
Hispanic	-4.917	0.045	**	0.007	-4.705	0.085	*	0.009
Black	-0.929	0.245		0.395	-0.166	0.879		0.847
Other	0.194	0.819		1.214	1.013	0.382		2.752
Unemployed	1.164	0.119		3.201	0.665	0.418		1.944
Education AA or Less	0.072	0.949		1.074	0.160	0.894		1.173
Education Graduate Degree or less	-0.508	0.549		0.602	-0.711	0.436		0.491
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)	2.417	0.043	**	11.21	2.230	0.077	*	9.296
High Income (75K or more/yr.)	0.969	0.393		2.635	0.632	0.632		1.882
No. of Cars	0.352	0.355		1.422	0.140	0.740		1.150
No. Children	-0.757	0.219		0.469	-0.709	0.255		0.492
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections								
Total Transit Stops								
Medium-to-High Traffic					-0.218	0.002	***	0.804
Tree Density								
Irrigated Lawn Density								
Non-Irrigated Lawn Density								
0.5-Mile Data								
Industrial Uses					-18.87	0.390		0
Office Uses					37.08	0.408		>100
Commercial/Retail Uses					-18.54	0.288		0
High-Density Residential Uses					-8.914	0.577		0
Low-Density Residential Uses					-13.66	0.453		0
Open Park Area					-13.38	0.493		0
ICC (Household)								
ICC (Neighborhood)								
Residual								
N	12,396				12,396			
AIC	63809.96				64000.36			
BIC	63812.11				64002.52			
Likelihood Ratio	63807.96				63998.36			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-7(c): Half-Mile Radius Multilevel Analysis

	Model 5 (All 3 Levels)				Model 6 (All 3 Levels)			
Dependent Variable		**MVPAFlag**				**MVPAFlag**		
	Coef.	Pr>|t|	Sig.	O.R.	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables								
Intercept	1.724				1.961			
First Level: Individuals' Characteristics								
Female	-0.776	0.352		0.460	-0.510	0.524		0.600
Age	-0.026	0.355		0.975	-0.026	0.347		0.975
Asian	-2.666	0.216		0.070	-3.066	0.145		0.047
Hispanic	-4.673	0.086	*	0.009	-4.949	0.067	*	0.007
Black	-0.139	0.899		0.870	-0.764	0.407		0.466
Other	1.037	0.375		2.821	0.459	0.651		1.583
Unemployed	0.665	0.418		1.944	0.473	0.554		1.605
Education AA or Less	0.153	0.898		1.166	0.146	0.901		1.158
Education Graduate Degree or less	-0.705	0.439		0.494	-0.278	0.747		0.758
Second Level: Households' Characteristics								
Middle Income (35K - 55K/yr.)	2.213	0.079	*	9.140	2.337	0.057	*	10.35
High Income (75K or more/yr.)	0.626	0.635		1.870	0.530	0.677		1.699
No. of Cars	0.137	0.745		1.146	0.145	0.726		1.157
No. Children	-0.708	0.256		0.493	-0.797	0.190		0.451
Third Level: Neighborhoods' Characteristics								
Segment-level Data								
Total Street Intersections					0.033	0.055	*	1.034
Total Transit Stops					-0.005	0.001	***	0.995
Medium-to-High Traffic	-0.218	0.002	***	0.804	-0.132	0.108		0.876
Tree Density					0.219	0.016	**	1.244
Irrigated Lawn Density					-0.072	0.458		0.930
Non-Irrigated Lawn Density					-0.417	0.167		0.659
0.5-Mile Data								
Industrial Uses	-24.396	0.380		0	-24.366	0.123		0
Office Uses	47.576	0.396		>100				
Commercial/Retail Uses	-24.157	0.276		0	-15.736	0.244		0
High-Density Residential Uses	-11.578	0.568		0	-11.208	0.294		0
Low-Density Residential Uses	-17.664	0.443		0	-17.823	0.187		0
Open Park Area	-17.224	0.494		0				
ICC (Household)								
ICC (Neighborhood)								
Residual								
N	12,396				12,396			
AIC	63984.19				64048.73			
BIC	63986.35				64057.05			
Likelihood Ratio	63982.19				64048.73			

Significance: * p < .1, ** p < .05, *** p < .01

Table 5-7(d): Half-Mile Radius Multilevel Analysis

	Model 7 (All 3 Levels)			
Dependent Variable		**MVPAFlag**		
	Coef.	Pr>|t|	Sig.	O.R.
Independent Variables				
Intercept	1.949			
First Level: Individuals' Characteristics				
Female	-0.814	0.326		0.443
Age	-0.027	0.330		0.974
Asian	-2.640	0.218		0.071
Hispanic	-4.851	0.073	*	0.008
Black	-0.410	0.667		0.664
Other	0.762	0.456		2.142
Unemployed	0.680	0.403		1.973
Education AA or Less	0.163	0.890		1.177
Education Graduate Degree or less	-0.644	0.477		0.525
Second Level: Households' Characteristics				
Middle Income (35K - 55K/yr.)	2.090	0.093	*	8.081
High Income (75K or more/yr.)	0.402	0.754		1.494
No. of Cars	0.144	0.729		1.155
No. Children	-0.717	0.246		0.488
Third Level: Neighborhoods' Characteristics				
Segment-level Data				
Total Street Intersections	0.033	0.055	*	1.033
Total Transit Stops	-0.005	0.001	***	0.995
Medium-to-High Traffic	-0.132	0.109		0.876
Tree Density	0.220	0.015	**	1.246
Irrigated Lawn Density	-0.073	0.455		0.930
Non-Irrigated Lawn Density	-0.417	0.166		0.659
0.5-Mile Data				
Industrial Uses	-10.907	0.555		0
Office Uses	54.121	0.156		>100
Commercial/Retail Uses	-11.667	0.399		0
High-Density Residential Uses	-2.098	0.867		0.123
Low-Density Residential Uses	-7.291	0.638		0.001
Open Park Area				
ICC (Household)				
ICC (Neighborhood)				
Residual				
N	12,396			
AIC	64064.67			
BIC	64066.83			
Likelihood Ratio	64062.67			

Significance: * p < .1, ** p < .05, *** p < .01

5.10.3 Multilevel Regression Results

In this section, the regression results displayed in Tables 5-5(a to b), Tables 5-6(a to d) and Tables 5-7(a to d) are discussed and the significant variables are outlined. For simplicity, only significant results in the models will be reviewed below.

Segment-level Models

The variables in Model 4 in Tables 5-5(a to b), combine all three levels of the data. In the first section, level one variables model the individual characteristics of gender, age, race, education and employment status. There were only two significant variables in level one: race and employment status. The parameter estimate on the race variable, Hispanic suggest that the odds of moderate-to-vigorous physical activity (MVPA) occurring decreases by 47.4% (O.R. = 0.526) if the individual is from the Hispanic race relative to the reference category White. Further, the estimate on unemployed suggests that the odds of MVPA increase by 22.9% (O.R. = 1.229) if the individual is unemployed.

The next section displays the second-level predictors of the household characteristics. These include the combined household income modeled as middle-income (for households earning $35,000 to $55,000 per year) and high-income (for households earning $75,000 or more annually) and the reference category is low-income (for households earning less than $35,000 per year). The household-level variables also include number of household cars and number of children at home (less than 17 years). These variables have been previously found in the literature to affect travel habits and therefore to also affect physical activity patterns and levels. The only significant variable at this level is for middle-income and its parameter estimates suggests that the odds of MVPA increases by almost 37% (O.R. = 1.368) if a household is in this category relative to a low-income household.

The next section reflects the third-level variables or the neighborhood built environment characteristics obtained using a street segment aggregation method. Variables in this section are key variables needed to find their associations with the odds of MVPA. These variables include: number of street intersections, number of transit stops, and number of commercial and industrial uses, a dichotomous variable if medium-to-high traffic is observed and three green space variables modeled within the Expo street buffers for: acreage of trees, irrigated and non-irrigated lawn all relative to acreage of impervious land. The results show that the odds for MVPA increases slightly (O.R. = 1.007) when an additional street intersection is added and decreases very marginally (O.R. = 0.999) with an additional transit stop. Similarly, an additional commercial use increases the odds of MVPA by almost 2% (O.R. = 1.019). Busier streets with medium-to-high traffic volumes appear to hinder MVPA by decreasing its odds of occurring by almost 2% (O.R. = 0.981). Finally, the green space variable acreage of trees is significant at the 10.3% significance level and its parameter suggests that an additional acre of trees nominally increases the odds of MVPA (O.R. = 1.017).

<u>Quarter-mile Radius Models</u>

The regression results for the geographic extent quarter-mile from home are displayed in Tables 5-6(a to d). Estimates from Model 2 reflect the probability of observing momentary MVPA and associations with individual-level traits. Results show that the odds of MVPA occurring is lower for Hispanic individuals (99.4% lower) than Whites (reference category) and that unemployed individuals are almost four times as likely to be physically active (O.R. = 3.957) than employed ones.

Similarly, in Model 3 variables from the individual and household levels are accounted for. However, only the race variable, *Hispanic* is significant again indicating that the odds of MVPA occurring is 99.3% lower for individuals of this race than Whites after household income, number of children and number of cars are accounted for.

In Models 4 and 5, density measures of land uses and quarter-mile radius land use areas (in square miles) are included respectively in addition to the individual- and household-level characteristics. The odds for MVPA occurring in Models 4 and 5 for Hispanics are (99.7%) and (99.6%) lower than Whites respectively. Middle income earners in both models are 13 times as likely to perform MVPA relative to individuals from the lower income group. The odds of observing any MVPA is almost zero in areas with increased industrial and sparser residential densities, indicating that these uses are deterrents to physical activity behavior. Further, the medium-to-high traffic variable, show that the odds of observing any MVPA is decreased by almost 20% (O.R. = 0.805) in busy corridors.

Models 6 and 7 have a mix of density measures of land uses calculated at the quarter-mile and segment-level. In addition, variables describing individual- and household-level characteristics are also included. The difference between the two models is that density of office use is excluded in Model 6. Goodness-of-fit measures for the two models indicate that Model 6 is smaller showing the superiority of this model over the other. In addition, this model shows the largest number of significant variables and therefore only estimates from it will be discussed next.

The odds of MVPA occurring is 99.5% lower for Hispanics than Whites and middle-income earners have 11 times the likelihood of being physically active relative to lower-income individuals. At the segment-level, better street connectivity and higher tree densities increase the odds of MVPA by 3.4% and 24.6% respectively. More transit stops lowers the odds of MVPA but the magnitude is marginal, less than 0.5%. At the quarter-mile extent, increasing the density of industrial, commercial/retail and both denser and sparser residential units contributes to almost no MVPA levels.

Half-mile Radius Models

The regression results for the geographic extent half-mile from home are displayed in Tables 5-7(a to d). Estimates from Model 2 and 3 are the same as discussed in the Quarter-Mile Radius Models section since the same models are estimated and differences only appear between the two extents in Models 4 to 7 at the neighborhood level.

In Model 4, only density measures for the half-mile land use variables are included along with the individual- and household-level characteristics. The odds for MVPA occurring is almost zero for Hispanics compared to Whites (O.R. = 0.009) and middle income earners are nine times as likely to be physically active (O.R. = 9.296). Further, the medium-to-high traffic variable, show that the odds of observing any MVPA is decreased by almost 20% (O.R. = 0.804) in busy corridors.

In Model 5, only half-mile radius land use variables measured in square mile are included along with the individual- and household-level characteristics. Again, the odds for MVPA occurring for Hispanics are even lower than for Whites (O.R. = 0.009). The estimate on middle income indicates that the odds for observing MVPA are nine times (O.R. = 9.140) as much for this group than individuals from the lower income group. Finally, from the third strata, the medium-to-high traffic variable, shows that the odds of observing any MVPA is decreased by almost 20% (O.R. = 0.804) in busy corridors. Estimates were very similar in Models 4 and 5 only the metric differed which apparently did not change the results.

Models 6 and 7 have a mix of density measures of land uses calculated at the half-mile and segment-level in addition to the individual- and household-level characteristics. The difference between the two models is that density of office use is excluded in Model 6. The goodness-of-fit measures of Model 6 are smaller showing the superiority of this model over the other and therefore only estimates from it will be discussed. The odds of MVPA occurring is 99.3% lower for Hispanics than Whites and middle-income earners have 10 times the likelihood of physical activity than lower-income individuals. At the segment-level, better street connectivity and higher tree densities increase the probability of observing MVPA by 3.4% and 24.4% respectively. Surprisingly, an additional transit stop lowers the odds of MVPA but the magnitude is less than 0.5%.

5.11 Discussion

Model estimates from the previous section show many pertinent results. The results are consistent across all geographic extents in the estimate magnitudes and expected signs for significant individual- and household-level characteristics. Focusing on built environment correlates, the estimates at the half-mile extent yield the least number of significant land use variables and generally have larger goodness-of-fit measures than models estimated at the other two extents. This might indicate that a half-mile radius from one’s home might be too large of an extent to observe occurrences of TPA during walking episodes while segment-level or even a quarter-mile extent might be more suitable as the walkable distances. These results are consistent with studies that found a positive correlation between TPA and green space variables when measured within 50-meter and 100-meter from home (Almanza et al., 2012; Rodriguez et al., 2014) and had a distant decay effect that resulted in reduced magnitudes of the effects on MVPA when distance from home was increased (Houston, 2014).

Using results from the mixed models of quarter-mile and the buffered segment-level land use variables, we can conclude that green spaces and well-connected street networks are most significant and positive factors in promoting more physical activity. In particular, tree density seems to be the largest correlate of increased odds of MVPA during walking instances. Thus, policymakers can target neighborhoods and street segments within them with smart growth tools that enhance pedestrian walkways and increase tree density along those paths for aesthetics and shade.

Interestingly, commercial density affects active travel differently at the three geographic extents when individual- and household-level traits are accounted for. For example, at the segment-level, the association is *positive* on the odds of MVPA occurring albeit it is a small magnitude (O.R. = 1.019). At the quarter-mile extent, increases in commercial density *reduces* the odds of observing any MVPA to almost zero (O.R. = 5E-06) and at the half-mile extent; commercial density is no longer significant in predicting any MVPA instances. This might be another indication of a distance decay effect of the impact of this use on MVPA. Additionally, individuals might be active around commercial uses but their level of activity might not exceed the lifestyle threshold (a level just below moderate activity) or a stroll pace. Further, research may be required to identify a more exact association between MVPA and commercial uses at varying geographic extents.

From a policy standpoint, interventions that target greener spaces and better pedestrian walkways appear to be most effective in attracting non-motorized travel across all geographic extents.

5.12 Limitations

Although the analyses in this study provide new contributions to the active travel-built environment literature, many limitations still exist. The data used here were from Phase 1 (before introducing the Exposition light rail) of the Expo Study. The response rate for the surveys was about 1% which is considered lower than average for travel surveys. However, unlike national travel surveys that are completed over one or two day periods; mobile tracking group participants from the Expo study were required to complete a seven-day trip log and carry the accelerometer and GPS devices the whole duration of their participation. This might have contributed to the low response rate. Of the 143 responses in the mobile tracking group, 117 households had complete and reliable entries; and only 68 of these households exhibited any form of active travel.

Further, the sample was not balanced in gender, race or income level. The majority of the sample was female (68%). Almost half of the households (45%) were low-income earners having an annual income of less than $35,000. African-Americans comprised 56%, Whites were 24%, Hispanics 8%, Asians 3% and other races were 6% of the sample. In comparison to Census data, African-Americans were over-represented and Hispanics were under-represented in the sample. This suggests that the study sample is not a random sample and does not fully represent the racial distribution of the actual population and this was indicated by the low response rate. Future research can employ other methods to incentivize more respondents into the study which may result in a more representative sample.

Effect sizes that were statistically significant may not yield real magnitude effects. Even though many neighborhood built environment variables were statistically significant at the segment-level, they did not produce practical significance. Most of the regression estimate values were very close to 1.0, which is a very small effect. The largest effect size however, that had both practical and statistical significance was observed for the tree density variable at the quarter- and half-mile extents where a 1% increase in tree density was associated with a 24% increase in the odds of MVPA.

The tree density variable captures both the number of trees and the size of the tree canopy. Older communities however, may have more mature trees with a larger tree canopy than areas adjacent to the newly constructed Expo Light Rail stations and redevelopment locations. Moreover, trip destinations may also affect this variable; a larger tree density may be observed near open spaces and parks whereas industrial areas may have less trees. As noted earlier, street-level audits may help address these concerns and more accurately measure green spaces.

The use of objective measurements via GPS sometimes resulted in some data loss. Initial acquisition time lapse from satellites may have influenced the time and date stamp. Also, data loss was particularly true for participants that used the underground rail which produced consecutive periods of missing GPS measures. In addition, areas with high structures (e.g. high-rise buildings) often caused interference and differential measurement error due to obstructions. Whenever needed, extensive manual reviews were performed to correct or eliminate (if required) such data points.

In addition, the possibility exists that participants may have intentionally altered their travel behavior since they were aware of being tracked. This might have injected bias into the data collected. Expo study administrators rectified this potential bias source by thoroughly explaining the instructions and objectives of the study. In addition, travel information was collected over a seven-day period which is a relatively long period for anyone to continue their altered travel habits.

The possibility of spatial autocorrelation in the road segment-level models may exist. As explained earlier, this is due to the fact that some road segments may have been used numerous times while others may not have been used at all. As a result the regression errors may be correlated and this violates the error independence assumption. Spatial regression models may mitigate this problem.

The green space measures used in the models were calculated at a two feet resolution; this implies that some street amenities may not have been fully captured in the analysis. Future studies may benefit from performing detailed street audits and objective measurements using Google Earth that have both been proven to accurately assess and improve street-level and neighborhood conditions (Clarke et al., 2010; Hoehner et al., 2005; Sugiyama, Neuhaus, & Owen, 2012).

Lastly, the multilevel analyses performed here only included fixed effects of the variables measured. The natural hierarchy of the data primarily influenced the choice of this analysis. In addition, the choice of modeling fixed effects only instead of adding both fixed and random effects was due to the relatively small study area of the Expo neighborhoods selected. Preliminary analyses showed that there were no apparent significant variations across the neighborhoods since the targeted areas were quite similar to one another. This leads to the assumption that effect sizes may be identical within the neighborhoods which validates the use of fixed effect modeling. Future research may explore the random effects when this longitudinal study is complete to explore differences among the various level-specific variables and temporal changes across all three phases.

5.13 Conclusion

Associations between the built environment and active travel behavior have been numerously examined in the planning and public health fields usually in the smart growth context. Using the ecological approach to modeling as the backdrop, many authors examined interactions of the ‘traveler’ to his/her physical, social and cultural environments (Giles-Corti, Timperio, et al., 2005; Sallis et al., 2006) and incorporated covariates at different levels into multilevel models (Ding et al., 2012; Van Dyck et al., 2010; Witten, Hiscock, Pearce, & Blakely, 2008).

This study examined the impacts of the built environment on active transport modeled as moderate-to-vigorous physical activity (MVPA) from walking behavior. I used an ecological model framework that exploits the natural "nesting" of the variables in the data. The hierarchies are defined as: individuals residing within households and households within neighborhoods.

I estimated multilevel binary logit regressions to model the probability of occurrence of momentary MVPA of walking (using paired accelerometer-GPS measurements) given the different level-specific variables. The objectives were to study the effects of the built environment correlates (at three geographic extents) on MVPA after individual- and household-level characteristics have been accounted for. Previous multilevel studies have either not used objectively measured active travel (Wasfi et al., 2013; Wendel-Vos et al., 2004); or only used accelerometers to measure physical activity episodes (Ding et al., 2012; Kneeshaw-price et al., 2013; Sundquist et al., 2011; Van Dyck et al., 2010; Witten et al., 2008). Papers that comprehensively measured physical activity levels and their respective locations (Almanza et al., 2012; Dunton et al., 2013) provided more accurate results since they used matched accelerometer-GPS epochs of MVPA. These papers however, did not control for household-level covariates (except for household income) through a three-level analysis as suggested in this current study. This creates a limitation to their analyses because of the potential of having biased estimates at the first level that were aggregated to higher levels.

The final models in my current study show promising results. In the smallest geographic extent (segment-level) from Expo residence, improvements in street connectivity, and increasing tree acreage positively increase the odds for walking MVPA. Increased traffic volumes and building more transit stops were found to be a deterrent to MVPA at this extent. Final models from the mixed quarter-mile and buffered segment-level proved to have more significant built environment correlates than models from the mixed half-mile models. This suggests that smaller geographic extents (less than half-mile radius) may be more suitable (Cervero, 1996; Guerra, E. & Cervero, 2013) to examine correlates of walking behavior as confirmed in the urban planning literature.

Specifically, the mixed extent models (using quarter-mile and segment-level data) show that the odds of MVPA occurring increases with more connected streets and planting more acreage of trees when individual-level and household-level characteristics are accounted for. However, increasing the number of transit stops, having more industrial, commercial and residential (high and low) densities contribute to lowered odds to zero occurrences of MVPA. These findings suggest that pedestrians that use walking as a form of transport or as an intermediary mode for transit access might benefit more from policies targeted at the neighborhood level. This also means that given the large effect size (O.R. = 1.246) of increasing green space (tree density) on *MVPAFlag*; policymakers may encourage the developments of local parks and pedestrian targeted environments in order to promote more active travel and simultaneously reduce motorized travel.

CHAPTER SIX

Expo Study Active Transport Routes: A Comparison of Observed GPS Routes and Shortest Distance Paths

Introduction

Current route choice modeling research has primarily focused on methods that entail a detailed account of all possible routes but insights of these studies cannot be directly applied to understand pedestrian path selection and the role of the environmental factors. Few research has been directed towards the understanding of pedestrian route choice and most agreed that shortest distance is the strongest correlate to the probability of selecting a path. Others who tracked their respondents' routes by GPS devices; concluded that the built environment is more significant than minimizing distance in pedestrian route choice. Destinations to commercial and retail centers were overwhelmingly found to be very important for utilitarian walks as well as connectivity and aesthetics along the routes.

The methods outlined in this chapter rely on revealed preference data from actual pedestrian paths measured by a GPS device in 15-second epochs. To test whether minimizing distance is a primary concern among pedestrians, I created GIS simulated shortest-distance paths and compared them to the respective observed routes on several benchmarks.

The respondents' heterogeneity in travel preferences was considered which adds insight into by-group variations in the participants' route selections. Further, pertinent information was also obtained regarding travel preference by time of day for the different participant groups.

In addition, comparisons and objective assessments of the built environment factors along each route type was completed. This analysis yield interesting policy related findings of the optimum mix of land uses that attracts more pedestrians.

The selection of travel routes is a dynamic process that occurs regularly. Individuals face a decision-making process daily when they select their paths of travel. This process is subject to the traveler's characteristics, preferences, destination, time constraints, familiarity with the transportation network, and the inherent mental map of his/her surroundings. Further, this process also creates differences in route selections among the travelers due to variations in the above constraints. Therefore, activity models that incorporate this traveler heterogeneity, more accurately simulate mobility and can be more easily translated into policy measures (Spissu, Meloni, & Sanjust, 2011).

The analyses in this chapter provide a practical method in which pedestrian observed GPS-tracked routes are examined and compared to analogous GIS simulated shortest paths. The methodology outlined here considers user heterogeneity and variations in their travel behavior, which has been for the most part, absent from many traditional route choice modeling techniques. The proposed methodology is also less cumbersome and involves less intensive computational steps than previous conventional route selection methods.

Traditionally, previous models adopted Wardrop's first principle that postulates the notion of user equilibrium. It states that equilibrium is achieved in traffic assignment when no traveler can decrease his travel time by unilaterally altering his route (Jan, Horowitz, & Peng, 2000; Zhu & Levinson, 2010). Therefore, this implies that travelers routinely use shortest paths (time saving) that are optimum for their respective origin-destination (O-D) combination. Shortest-path modeling however has been criticized for overlooking inherent user variations in tastes and travel preferences (Zhu & Levinson, 2010). One study concluded the stark difference between the shortest-paths and the observed GPS-recorded routes and declared that the location of where the O-D's occurred onto the road network was more pertinent in route selection (Jan et al., 2000). Another study also utilized GPS data to make a similar conclusion after tracking travelers for three weeks (Zhu & Levinson, 2010). Still, others concluded that drivers opted to minimize distance on GPS-recorded shorter trips but favored a travel time minimization alternative for longer trips (Spissu et al., 2011).

The precedence of these motorized route choice modeling studies is very useful but not all of their aspects can be extended to non-motorized route choice analysis. Road networks that are primarily designed for vehicular traffic can impact cyclists and pedestrians hindering their movement or in some cases even preventing it completely. Mixed messages to cyclists across different municipalities may provide inconsistencies in street evaluation projects. For example, in Phoenix, AZ the city allows cyclists to utilize sidewalks by placing signs to that effect whereas in Tempe, AZ this is prohibited (Howard & Burns, 2001). It is this discordance in where cyclists may travel that affects the cyclists' perceptions of safety, acceptance from other users of the road; and may even alter their routes completely.

Similarly, studies that aim to quantify the walkability of a place to promote more walking may not paint the whole picture. Many public health studies linked walking to averting heart disease, colon and breast cancer, type II diabetes (Hankey et al., 2012) and gestational diabetes prevention (Evenson & Wen, 2010). Several studies have realized the importance of pedestrian-oriented amenities in promoting walking (Ewing, 1999; Giles-Corti, Broomhall, et al., 2005; Kaczynski & Henderson, 2007; Sallis et al., 2006). However, others argue that a "paradigm shift" is in order in which the significance of mobility would be overshadowed by accessibility to goods, services and activity centers (Guhathakurta, Zhang, & Panguluru, 2013). Therefore, walkability scores that incorporate accessibility to amenities, transportation projects that provide comprehensive system-wide coverage enabling walking at origins, destinations and along the routes without interruptions in trails are optimum. Thus, route quality is of importance and is often evaluated through the presence of amenities or more broadly built environment features.

Pedestrians and drivers are both confronted with elements of the built environment, which therefore influences the decision for their path of travel. Relative to vehicular travel, walking provides the pedestrian with a more profound experience of the surrounding built environment (Foltête & Piombini, 2010). Many recent papers have focused on the effects of the neighborhood-level composition of land uses on walking (Durand et al., 2011; Kaczynski & Henderson, 2007; Van Dyck et al., 2010) however, pedestrian's experiences along routes might be different from these neighborhood-level correlates.

The street-level approach of pedestrian route choice analysis, provides more precise magnitudes of the effects of the built environment on this decision-making process. One reason for this is that the neighborhood-level effects are aggregated over a specific parcel area however the built environment effects for a pedestrian path involve linear measurements over the defined road segments traversed (Rodriguez et al., 2014). Inherently, neighborhood-level analyses are post hoc to a pedestrian's trip whereas pedestrian route choice modeling explores the underlying reasons for the selection of a particular route over another. This would include contrasting the built environment characteristics along each route and observing differences among them to clarify the reasons behind the route selection.

Objectives

The analyses in this chapter focus on pedestrian travel behavior interchangeably referred to as non-motorized or active travel. The methodology presented here provides another perspective to traditional pedestrian route choice modeling techniques in the literature by contrasting GPS-recorded walking routes to GIS simulated shortest roadway distance paths. Traveler heterogeneity is modeled via reported socio-demographic traits and the variations in their travel preferences are reflected through the observed routes taken. Further, comparisons between the GPS-tracked routes and the shortest paths were accomplished by: participant group types, deviations in various trip-level travel indices; time of day variations and differences among objectively measured built environment features along each set of routes. Ultimately, the goal is to report significant differences between the observed and shortest paths in the above analyses and to produce policy recommendations that promote active travel.

Research Questions

This study seeks to answer the following research questions:

1- How closely do shortest paths simulate GPS–recorded observed walking trips?

2- What significant socio-demographic differences exist among various participant groups in their walk route preferences?

3- Do the observed routes differ from shortest paths by time of day?

4- Are there variations in walking frequencies by time of day among certain participant groups?

5- Does the built environment affect the walking path choice of a pedestrian? If so, which built environment factors attract more pedestrians?

6- What are the policy implications of the proposed analyses?

In the remaining sections of this chapter, I present a review of the pertinent topics in the literature regarding route choice analysis and offer differences and limitations of previous studies. Next I depict the contributions that this chapter provides in expanding the current literature. The data utilized is briefly described but a detailed overview of the Expo data was presented in an earlier section of this dissertation.

The methodology for pedestrian route choice analysis is also explained. I describe in detail how the raw data was prepared and how the GPS points were transformed into the observed routes. I also present the process of creating the corresponding shortest paths using the Network Analyst tool in ArcGIS. Next, the different measures by which these two route types are to be compared are presented in the data analysis section. A detailed account of the results of these measures and travel indices are provided in the four sections that follow. The final sections present the policy implications of these analyses, limitations of the study, recommendations for future research and a conclusion for the research.

6.1 Literature Review

Route choice analysis has been limited to only motorized travel until recent years where travel models have expanded to also include pedestrian and cyclist routes as well. Pedestrian route choice modeling still remains an understudied topic but in recent years, a growing body of literature has examined this topic. The underlying theory is based on micro-economic principles in which the goal is to maximize the traveler's utility while minimizing costs in the form of time or distance traveled. The minimization of time is more pertinent to motorized travel studies because most transportation projects are geared towards providing a level of service which involves time savings (Guo & Loo, 2013).

In contrast, the minimization of distance is more relevant to non-motorized route analyses since it has been previously identified to be the main correlate with path selection. Shortest-distance routes that have the least mileage traversed and include biking facilities are routinely selected by cyclists (Howard & Burns, 2001). Likewise, pedestrians choose to minimize distance rather than to reduce travel time since the latter can only be attained by changing one's walking speed and this cannot be easily achieved (Guo & Loo, 2013). Typically, these studies have been concerned with comparisons of observed (actual) routes to shortest paths and examining any overlapping route features between them (Rodriguez et al., 2014; Stigell & Schantz, 2011).

Moreover, pedestrian path modeling has lately expanded to include the effects of the built environment on route selection (Guo & Loo, 2013; Guo, 2009; Sugiyama, Neuhaus, Cole, Giles-Corti, & Owen, 2012) focusing on connectivity, streetscapes, amenities and other route-specific attributes. This expansion was pertinent especially since previous pedestrian analyses that were centered on level of service transport projects had always considered such types of streetscape amenities to be hindrances to walking (Guo & Loo, 2013).

6.1.1 Route Choice Overview

Route choice analysis answers a basic question: what is the likely selected path for a certain mode given a set of origins and destinations and a defined road network? Therefore, route choice modeling defines a subset of paths connecting origins and destinations in a defined space obtained from a comprehensive universal choice set. In order to produce this desired path subset, researchers have either used deterministic or stochastic (probabilistic) processes.

The probability of selecting a specific route is calculated from the choice set. The formation of this choice set depends on the method utilized either deterministic or stochastic. The former always generates the same defined choice set for each origin-destination (O-D) pair upon which the model is created. This is an advantage of the deterministic method over the stochastic one which ensures that more concise results are obtained (Frejinger, Bierlaire, & Ben-Akiva, 2009).

Moreover, deterministic route choice modeling has been widely used and is well documented in the literature. It usually utilizes shortest path techniques to produce a route choice set. Shortest paths can either be based on minimizing distance or time but are seldom realistic as they rarely actual individual's travel habits. Thus, the drawbacks of this method are the strengths of the stochastic technique, which extends the model to include travel behavioral patterns and attitudes in addition to the traditional route choice modeling inputs.

The stochastic route choice method is therefore sensitive to individual travel choices. The stochastic technique builds upon the deterministic method to include an additional step. This step captures the probability of choosing the pre-defined path choice set from a master path subset. Therefore, the advantage of this probabilistic approach is that it facilitates the inclusion of the travel behavior aspect in the route choice model (Kaplan & Prato, 2012). On the other hand, the disadvantage lies in its computation complexity in which a finite universal route selection set needs to be pre-defined for each O-D pair.

In general, traditional route choice modeling techniques are not adequate for pedestrian analyses. These models usually produce routes that are not usually traversed by pedestrians (Foltête & Piombini, 2010). The deviation analysis proposed by Foltête & Piombini (2010) offer another alternative to the above methods that assumes that the shortest route resembles actual pedestrian trips since it maximizes the overall utility. Therefore, minimizing distance is the underlying metric used in such analysis. Deviations from the shortest path and examining route-specific attributes are also central in this method (Foltête & Piombini, 2010).

With the advent of technology applications, GPS-recorded routes became a valuable revealed preference method that combines aspects of both the deterministic and stochastic processes of route choice modeling. GPS datapoints are defined between O-D pairs and therefore the route choice is pre-determined. GPS data may also capture attitudes, perceptions, and the traveler's decision-making process through the revealed route he/she chooses. However, GPS-recorded routes still require extensive computer processing and data manipulation which entails map-matching to the underlying street network (Spissu et al., 2011).

The analysis included in this chapter is a combination of the last two methods. Actual pedestrian routes that were tracked by a GPS device were compared to GIS generated shortest paths. Then, a deviation analysis was performed which included comparing the two route types by different travel indices and the association of the results to socio-demographic characteristics.

6.1.2 Sources of Route Choice Data

An obvious component that affects the efficiency and accuracy of route choice modeling is the data used. Previous studies in this field utilized data from stated preference (sp) surveys such as Abdel-Aty, Kitamura, & Jovanis (1997) and Stinson & Bhat (2003). This type of data however, may not be optimum. The reason lies in the formation of such surveys where the researcher elicits the response of a user by questions about a hypothetical situation where travel intentions rather than actual travel experiences are recorded. Thus, a response bias is usually introduced especially when the respondent tries to answer the questions in anticipation of what the researcher would like to hear instead of his/her actual feelings towards the topic (Randall & Fernandes, 1991).

In contrast, travel data collected from revealed preference (rp) surveys depend on actual experiences and can be more reliable than sp data. An example of which are travel diaries such as the National Household Travel Survey (NHTS) where daily household trips are recorded by time of day, day of week, mode, purpose (work, school, recreation, etc.) and duration (Federal Highway Administration, n.d.). While this data is based on real travel, a major shortcoming due to recall loss occurs when respondents do not record their trips immediately and instead, base their travel diaries on memory alone.

Nevertheless, researchers continue to rely on rp data as it is more accurate than sp data. In some studies the researchers have even combined information from both sp and rp surveys such as Khattak, A. J., Poludoropoulou, A. & Ben-Akiva (1996) to determine travel patterns and to identify any correlations between the two types of data reporting. Similarly, Zeiler, Rudloff, & Bauer (2011) combined sp and rp data in a mixed logit model since standard logit models treat pedestrians as having the same preferences and travel choices and do not consider the heterogeneity among them. Though these studies provide some consistency over studies that used sp data they are less superior to methods that record actual travel such as through GPS-tracking.

The advent of technology has facilitated the use of yet another type of rp data based on global positioning system (GPS) tracking. This technology is used either as a stand-alone source of travel information or in conjunction with travel diaries which yields even richer spatio-temporal data. GPS data is considered an objective way of measuring travel as it minimizes some of the drawbacks mentioned above from other types of survey collection methods. Besides compiling information about the physical location of each sequential travel point, a time stamp is also attached to each GPS point from which we can estimate the relative speed and distance traveled.

The use of GPS data in the field of route choice modeling is relatively new and has mainly been applied to motorized travel such as in Papinski & Scott (2011). The authors used data from the Halifax STAR (Space-Time Activity Research) project in Nova Scotia, Canada for 237 home-to-work routes and compared these observed GPS routes to their shortest path counterparts. Contrary to their hypothesized claim that travelers select routes that minimize their general costs (travel time or distance), the authors found that the actual routes chosen were longer in distance and were probably selected on other attributes than those of the shortest paths (Papinski & Scott, 2011).

Likewise, another study for the Lexington, Kentucky area utilized GPS data to compare observed driver routes to their shortest-path counterparts. The researchers concluded that drivers routinely use the same route they are accustomed to which usually deviates more from the shortest paths as distance increases between origins and destinations (Jan et al., 2000). This study however, did not reveal the extent of which the two route types differed which may reveal important variations across travelers. Further, this study eliminated all short trips that were classified as being less than two miles for fear of being subject to bias. In contrast, I focus on short trips in this chapter since they are more convenient to pedestrians and reveal pertinent information.

A precursor to GPS-based route choice modeling is the proper matching of the data points to the relevant transportation network. This process allows both the visual representation and the spatial analysis of the routes taken. Chung and Shalaby (2005) used a Geographic Information System (GIS) map-matching algorithm in which they grouped the GPS points with the corresponding road segment and identified differences in the four travel modes (walking, cycling, bus and private vehicle) with a 79% accuracy on all segments. The authors note some previously documented limitations they faced that undermine the quality of the GPS data: (1) the warm start-cold start effect (where an initial stationary period is required to acquire a signal), (2) Signal interference from neighboring structures (this causes a temporary 'bouncing' of the points), (3) Interference from roofs of the vehicles (less common in private vehicles but was observed from bus travel especially if the respondent is located in the aisle and away from a window) (Chung & Shalaby, 2005). The general solution to most of these issues has been to eliminate the sequence of GPS points that might have been the cause of these errors after identifying them which may also affect the sample size.

6.1.3 GPS and Non-motorized Modes

Route choice modeling for non-motorized modes such as for cyclists and pedestrians is a fairly new field. Similar to vehicular routes, pedestrian paths can be studied from a behavioral perspective where the goal here is to maximize the individual's utility subject to limitations. However, these constraints could be quite different from those of a driver. For example, a driver might avoid a congested roadway in hopes of reaching his/her destination faster, this same factor however, may not be as important for a pedestrian. A pedestrian may choose a route on a busy street provided it has well-connected sidewalks and/or sufficient offset or buffer from oncoming traffic. A pedestrian may even prefer specific streets that attract other pedestrians to them (Guo & Loo, 2013).

Further, shortest route distance was found to be the strongest predictor of route choice. A recent study comparing pedestrian routes in Minneapolis and San Diego for adolescent girls used a logit model and concluded that the odds of selecting a route were greater when routes had the shortest distance (Rodriguez et al., 2014). The authors also found that greater percentages of green space, better connectivity and presence of destinations along the routes were also positively correlated with route selection (Rodriguez et al., 2014). Although the authors used an appropriate level of analysis to examine the built environment along routes, the study lacked any comparisons to other viable routes for the destinations outlined.

Another paper concluded the dominance of the built environment over distance in route choice analysis. The paper provided a review of current literature and concluded that 80% of recent studies argued that destinations, especially to commercial and retail uses along the routes were particularly important for utilitarian walking trips followed by connectivity (50% of the papers) and route aesthetics (35% of the papers) (Sugiyama, Neuhaus, Cole, et al., 2012).

Therefore, the built environment may be a strong determinant of pedestrian route choice analyses which implies the need for correct environmental assessments and their inclusion in pedestrian models. The ultimate objective of pedestrian route choice models and environmental measures, is to assist policymakers in their decisions regarding future transportation investments (Guo & Loo, 2013).

Environmental assessments can be made via a number of methods. The first methods can be tedious and labor intensive as they rely heavily on field audits and are usually performed by skilled personnel or by interviewing pedestrians on-site who are asked to complete stated or revealed preference surveys. The next method is contingent valuation, which is a preferred method when researchers need to assess intangible commodities such as road aesthetics with the intention of placing an economic value on them. Further, a more objectively measured method employs the use of GIS in assessing the various attributes of the built environment. Since this method relies on archival data from satellite images such as from Google Earth rather than on-site audits, it is an efficient, accurate and economical tool in characterizing micro-environmental attributes. For example, Clarke et al. (2010) compared the same neighborhood characteristics in a community in Chicago from two sources: field audits and objective measures using Google Street View. The authors used an inter-rater reliability test using the Kappa statistic and their findings show that their GIS methods yield very reliable results on most indicators. Those that were not measured precisely with Google Street View were either factors that changed over time (e.g. neighborhood and housing conditions) and indicators requiring a finer observation level such as by an auditor (e.g. presence of litter or broken glass) (Clarke et al., 2010). This statistic is used to measure the extent of conformity between two or more rating sources that may classify their responses in overlapping or non-overlapping groups (Gwet, 2002).

The influence of the built environment was considered in another study that used walk score to quantify its impacts on the optimum walking route. Their goal was to evaluate the different route attributes to achieve a minimum total cost per route. The authors developed an Analytical Hierarchy Process that ranks the route attributes relative to others and concluded that in case of relaxed time constraints, the most walkable routes may be quite different from the shortest paths (Guhathakurta et al., 2013). This study however only takes into account the objective measurements of walkability but subjective measures were missing in the data. Another limitation was in the weights the authors used for each attribute. They decided to associate these weights with the percentage recurrence of each variable in the literature, which may be inaccurate and no validity measures were established (Guhathakurta et al., 2013).

6.2 Contributions of the Chapter

The methodologies presented in this research provide a simplified approach to analyze pedestrian routes. The proposed method is a revealed preference procedure since it involves the objective measurement of actual pedestrian paths via GPS 15-second readings. This method still requires some extensive computational processes but is deemed less cumbersome and more intuitive than other conventional route modeling approaches.

Heterogeneity among travelers that stems from normative and socio-economic factors was also accounted for in the analyses presented in this chapter. The respondent sample were divided into participant groups and their travel route choices, associations with route-specific attributes and travel time of day preferences were examined. Results yield important by-group variations in the participants' route selections.

A detailed time of day analysis was also completed and produced important inferences regarding pedestrian preferences of the observed routes relative to the shortest paths in most time of day periods. This was supplemented by a comparison of the travel benchmarks in AM Peak Time versus the remaining time periods.

Finally, segment-level exposures of the built environment features along the routes were analyzed and contrasts between the observed and shortest routes were made. Policy makers may utilize the results of this analysis to implement different streetscape features and redevelop land parcels with land uses and green space improvements that ultimately attract more pedestrians.

6.3 Methodology

There were generally five initial steps needed to compare the observed GPS routes to the shortest distance paths. The first step involves the preparation of the spatial data to be analyzed. This includes the automated and manual coding of the data points into stationary (location) or mobile (travel mode specific) categories. If the data point was mobile it was further classified as travel by motorized or a non-motorized mode as discussed earlier. Since the primary focus here is on pedestrian travel; only trips using the walking mode (non-motorized) were analyzed. The second step was to identify the starting and ending points of the walk trips and the data points in between were also selected as the trip duration for each survey participant. The third step involves the spatial joining of the GPS data points to the underlying road network to enable the creation and *geoprocessing* of the observed GPS routes and the analyses by network analyst. Fourth, shortest distance routes were then created by the network analyst ArcGIS tool. Last, the comparison of the observed GPS and shortest paths were completed via several route-specific descriptive statistics and of built environment characteristics along the routes, various deviance indices, sign tests and paired sample t-tests and participant-specific socio-demographic traits that are assumed to have an influence on route choice.

6.3.1 Data Preparation

The details of the data sources for this research can be found in chapter three of this dissertation. Similar to the other chapters in this thesis, the data used here was based on Phase 1 data of the Expo study which is comprised of matched GPS-accelerometer data, trip diaries, transportation network information and built environment data.

As previously mentioned, participants in the mobile tracking group were asked to fill out a daily trip log and were each provided GPS (QSTAR model QT-1000x) and accelerometer devices. Information obtained from these measures is intended to minimize recall and self-reported survey bias errors. The initial step required was to define the daily trip rates, their start/end points and respective times as well as the durations. Participants were instructed on how to classify a valid trip, in particular, trips that can be defined as round trips from and to the same origins; were to be further categorized into two or more trips depending on the complexity of the tours or trip-chaining they entail. Trip logs were also matched and compared to the data from the GPS devices.

Further, the objectively measured GPS trips required additional processing to define valid trips. A review of the GPS data validation and automated/manual coding of participant trips was discussed in a previous section. Following the validation and coding steps, the resulting data set consisted of sequential data points with a unique household I.D., a time-date stamp, latitude/longitude readings, location or travel mode specification, distance traversed and speed in 15-second epochs.

This data set was saved as a .dbf file and exported to ArcMap 10.0. In ArcMap, I used the *Display XY Data* feature to link the coordinate readings from the database table to the X and Y coordinate fields in GIS. The result provided 15-second point-by-point participant GPS routes for the duration of his/her valid GPS readings that typically spanned three to four days of travel.

Active travel periods were extracted from the automated and manual coding of the GPS data classification steps. Only the non-motorized trips from Phase I of the Expo data were analyzed. A total of 68 participants were identified to have used *walking* as a mode of travel. This however does not preclude the participant from accessing multi-modes but since the main concern involves active travel, the classification was expanded to also include ingress/egress trips to public transit stations. All motorized trips were therefore excluded from the analysis. Out of the 68 identified participants, four households had insufficient data to qualify for an active trip. This included: GPS data points having incorrect coordinate readings (non-sequential or bouncing), inadequate time duration of a trip (beneath trip duration threshold) and/or the stacking of the GPS data points into a very fine location less than a pre-designated threshold

0.005 miles (about 9 meters). This resulted in the exclusion of these four participants' data points from further analyses. This consequently means that the final number of participants is 62. Collectively, the 62 participants completed 388 walking trips over the survey time period. The number of active trips ranged from a minimum of one trip to a maximum of 32 trips over the four to seven day period per participant.

6.3.2 Defining Trip Start and End Points

Up to this point, the GPS data points were saved chronologically by the time-date stamp but were not classified at the trip level. In order to aggregate the data points to the trip level, I needed to first define the start and end points of the trip and their durations per participant. This was also completed in GIS. Since the GPS data was in chronological order, the first observation was always designated as the start of a new trip. I followed similar techniques in the literature on motorized trip classification to define the end of the walk trip. The minimum threshold of two-minutes for a stop was maintained throughout the trip classification process which is usually characterized by the multiple stacking of GPS points into a very fine space or no movement for at least two minutes as described in Beckx, Panis, Janssens, & Wets, 2010; Schuessler & Axhausen (2009).

6.3.3 Observed GPS Routes

To create the observed GPS routes, all the GPS data points in the previous step were selected for each trip and then spatially joined to the underlying road network. This process was performed in the GIS user interface ArcMap 10. I selected all the observations between and including the trip start and end points and created individual shapefiles to represent each trip for the participants. To reduce potential data processing time and to maintain organized GIS files, I created a different GIS map for each survey respondent with all his/her walking trips over the travel survey period. For example, if respondent “A” participated in the survey for four days; his/her GIS map held the shapefile with his/her socio-demographic traits and respective detailed GPS datapoints over those four days. This post-processing organization proved to be important as it also decreased the processing time needed to create the shortest routes when network analyst was later employed in the GIS interface.

As a result, a total of 62 maps, one for each participant, were created jointly holding the sum of 388 walking trips. The GPS data at this point includes consecutive 15-second readings collected via the QSTAR GPS device which includes the: date and time stamp, latitude and longitude readings and the activity data obtained after matching it with the accelerometer data (speed, lateral and vertical movement translated to stationary location/travel mode). In addition, other participant-specific socio-demographic characteristics and unique household I.D. (attached to the primary participant in the mobile tracking group) were also linked to the GPS and accelerometer data.

6.3.3.1 Map-Matching of the GPS Data

For each observed trip, the GPS data points were then linked to the underlying road network through a geometric map-matching procedure. This involved using the *intersect* tool under geoprocessing to associate the raw GPS points to the road network shapefile. This also allowed each sequential data point to be linked to a specific road segment which meant that these data points now share the road segment information including: road segment unique I.D., street name, coordinates, and the route type (based on MAF/TIGER Feature Classification Codes (MTFCC)).

6.3.3.2 Trip Identifier

Each set of observed GPS points now represent a new route which is also considered one trip. An additional field was added to the participant GPS attribute table which uniquely identifies each trip. A coding convention was established for the trip identification process, namely, I used an “I” or “E” as the preceding letters in the Trip I.D. to designate it as an “Ingress” or “Egress” respectively, followed by the participant’s household I.D. Note, that only one household member was allowed to participate in the mobile tracking group which essentially means that each household I.D. also uniquely identifies the survey respondents’ information. The Trip I.D. also included a trip sequential numeral at the end. For example, Trip I.D. “E20082” meant that the trip was an egress for household I.D. number 2008 and that it was his/her second trip in the sequence. The decision on classifying the trip as an ingress or egress was made serially whereby the first trip was always denoted as ingress and the following one was egress and they alternated, so odd numbered trips were always ingresses and even numbered ones were always egresses. This convention was adopted irrespective of the actual trip description, that is, regardless of the fact that the participant was going or returning from a certain destination this labeling technique was adopted. Note however, if a participant used the same route on a different day, it was still given a different Trip I.D.

6.3.3.3 Distance between GPS Points

The other field that was added to the GPS attribute table is "Seglength" which represents the distance between each two consecutive GPS points. This is also equal to the distance traversed on the road segment every 15 seconds. The road segment attribute table already included a variable on distance. However, upon inspecting this field further, I found that it represented the total length of the whole segment which was not sufficient for this analysis; since the traveler may have crossed only a portion of the road segment and not the entire length. Alternatively, I needed a variable that captures the short distance between successive data points. In most cases, a number of GPS points were assigned to the same road segment due to their extreme proximity to one another. This was true especially during walking periods.

To calculate the distance variable between the GPS points, I first combined all the observed GPS trips into one file to allow for easier data manipulation. In SAS 9.2, I imported all the observed GPS routes from ArcMap as database files and then appended them sequentially into one SAS dataset. After adding all 62 participant trips, this dataset was N = 13,510 observations consisting of 395 unique total trips/routes. Upon inspection of these trips further, seven were excluded (reasons will be discussed later) and the final number of trips analyzed were 388. I then exported this dataset to Excel to calculate the distance variable.

I used the *Pythagorean Theorem* as a method to calculate the distance between the GPS coordinates as explained in (Groundspeak, 2014). A prior condition was that the projection of the points needed to be in *UTM*. This was indeed the case since the default projection for all the maps was selected as *NAD 1983 UTM*. Next, following the procedure outlined in the above source, the initial GPS point with X and Y coordinate values were designated as the *Northing* and *Easting* values respectively such that the first coordinates were X_1 and Y_1 respectively. The immediately successive coordinates for the next GPS point was therefore considered as X_2 and Y_2 respectively and this same point then becomes the initial point for the one succeeding it and so on.

Using this convention, I then used the following formula to calculate the distance between each set of successive coordinates:

$$\text{Distance} = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$$

Next, this measurement was then converted into miles since the default calculations in the GIS maps were in meters. Note, the *Great Circle Calculation* method for distance was not used, which is considered a more accurate method of calculation. The reasons for this are: the simplistic nature of the *Pythagorean* method over the Great Circle Calculation method and because of the close proximity of the GPS coordinates in our sample. As mentioned earlier, the readings were obtained in 15-second intervals, which is an indication that the successive distance is relatively short. Typically, an acceptable interval distance to be calculated by the *Pythagorean* method must be less than 0.10 miles; otherwise, distances exceeding this threshold would need to factor in the curvature of the Earth (Groundspeak, 2014).

6.3.4 Simulated Shortest Paths

A major obstacle researchers and practitioners face in route choice simulation, is defining a complete finite set of routes from a given origin and destination. However, this is very difficult since it involves a detailed enumeration of all the possible routes between these two locations.

This ultimately depends on the traveler's decision, travel habits, his inherent mental map of his surroundings, and time constraints. With the advent of GPS and GIS technology, the route choice set has become more defined and limited to the observed routes that are actually selected by the traveler. By defining the observed routes, further analyses can be performed by researchers who aim to understand travel behavior more by examining how these routes closely resemble the simulated routes based on shortest distance or time.

Using the observed set of GPS routes from the Expo travel survey, I examined the travel behavior of pedestrians through a comparison with GIS simulated shortest paths. I used the ArcMap tool Network Analyst to create the simulated shortest routes then generated several measures and indices to compare the two types of routes using SAS 9.2.

6.3.4.1 Using Network Analyst

In order to create the simulated shortest paths, I first added the extension Network Analyst from the tool bar in ArcMap. In ArcCatalog, I created a new network dataset based on the TIGER street network feature class. This process designates the TIGER street file as the target road network to be used. I did not specify any elevation fields and accepted the defaults, however, these values could be altered if required. In the evaluator window, I selected *Meters* as the preferred attribute measurement and accepted the remaining default fields. I did not specify network directions since my analysis pertains to pedestrian travel, which permits more flexibility in the direction of travel. After clicking finish in the New Network Dataset wizard and pressing okay to build the network, ArcCatalog notifies you that the new network was created. In ArcMap, the new network dataset is now ready to be added and updated to generate other layers under it. For detailed information on how to use Network Analyst, please refer to (ESRI, 2010).

6.3.4.2 Creating the Shortest Paths

I used the start and end points specified earlier in section 6.2 as the respective origin and destination pairs per shortest path trip. I used *distance* as the impedance factor for the shortest route creation which is used more in the literature as opposed to shortest travel time that is mainly reserved for motorized travel (Buliung, Larsen, Faulkner, & Stone, 2013). The shortest distance routes were then created based on the origin and destination pairs of the observed GPS routes. This was also critical to ensure that these were the same starting and ending points that describe each trip to allow easy comparisons between the two route types.

The result produced an individual route shapefile with five components: stops, point barriers, routes, line barriers and polygon barriers. For simplicity, I did not add any more impedance since pedestrians may not face the same restrictions and have more flexibility than car users (Buliung et al., 2013). Thus, the only components that were critical were the stops (origin, destination and sometimes short tours less than the two-minute threshold), and routes. Both held data about the cumulative trip length or distance.

Next, I used the *intersect* feature under *Geoprocessing* to populate the shortest route with the road network information. Again, this allowed the routes to include road segment specific data that included the street name, segment I.D., route type, etc. I then added the two additional fields: “Seglength” and Trip I.D. for segment-level distance and unique trip identification respectively. The variable “Seglength” was specified as double with precision 15 to allow for decimal point calculation and the default was the metric system, which was later converted into miles. I used the calculate geometry tool to obtain the length per segment for each route. To clarify, this is the distance traversed and may or may not be the total length of the respective road segment. I also used the same convention for Trip I.D. which ensured that each shortest route was linked to the observed GPS route via the same Trip I.D.

6.3.4.3 Shortest Paths Dataset

The last step performed after creating all the shortest routes was to aggregate all the created paths with their relevant information in one dataset. I imported all the simulated shortest distance routes as a database file and then appended them sequentially into one SAS dataset. Similar to the observed GPS routes, there was a total of 62 participants, with 388 unique trips/routes and the resulting number of observations was N = 1,632. To reiterate, the number of observations represents the relevant road segment(s) that the route has traversed, therefore, this dataset and the observed GPS route dataset are both aggregated to the road segment level. The number of observations here is noticeably smaller than the observed GPS route dataset, which potentially confirms the hypothesis that the observed GPS routes were much longer than the shortest distance routes.

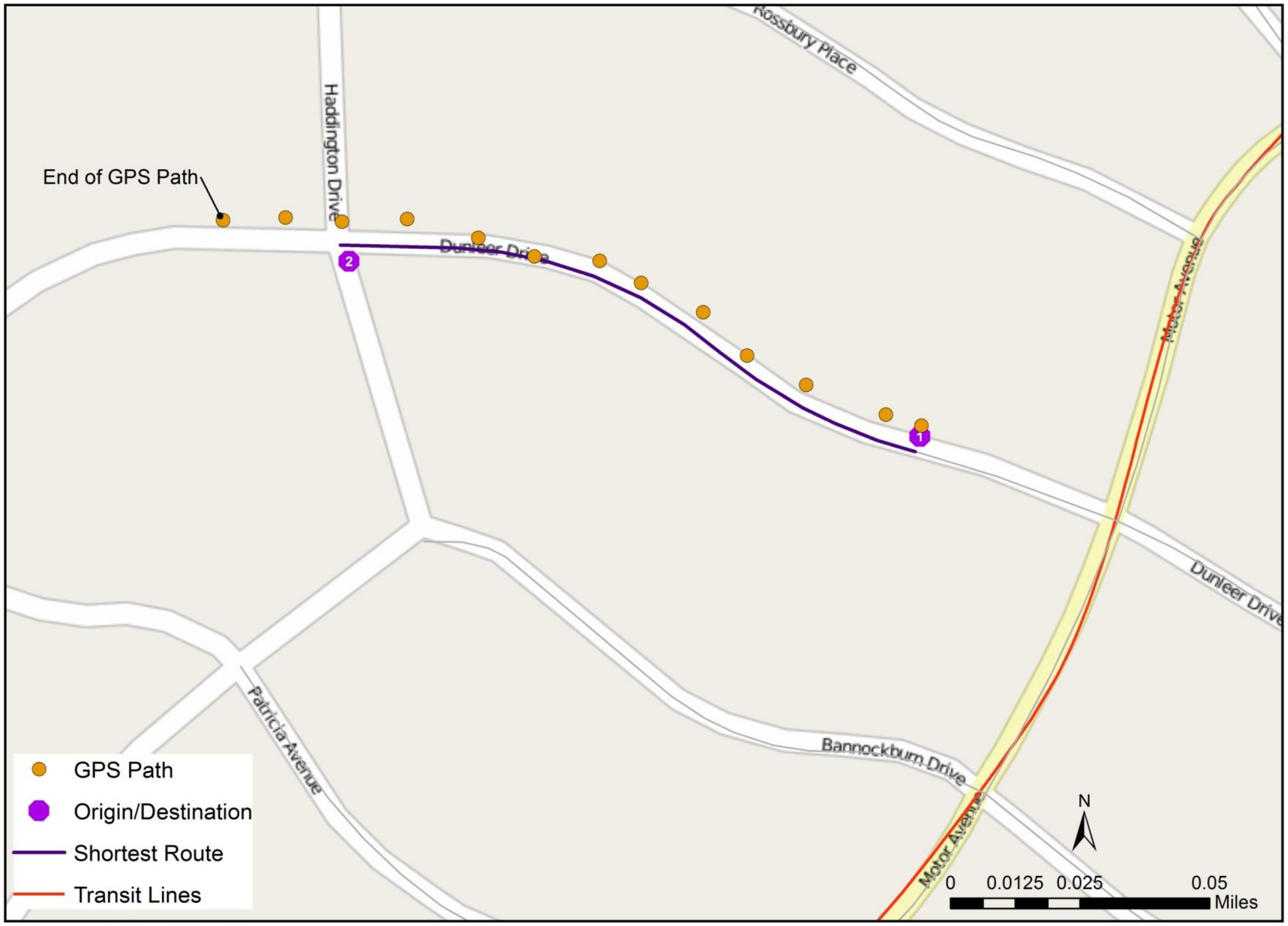
6.3.4.4 Assumptions and Observations

While using Network Analyst to create new routes, there were some issues observed and therefore I made the following adjustments and assumptions to overcome any problems:

- When creating new routes, the “create network location” tool would not “snap” to the correct observed location, instead, the shortest path stops (origin/destination) snapped automatically to the network nodes. I was able to manually shift these stops as close as possible to the observed point to approximate its location on the underlying network. Figure 6-1 below shows this phenomenon. The figure shows that the shortest path stop number 2 ended at the intersection of the road network at Haddington and Dunleer drives even though the observed GPS route ends further to the west. This approximation method was applied to all the routes to maintain a level of consistency.

- This was also true for the observed GPS points that did not necessarily snap to the network since some pedestrian paths or trails might not be represented or updated in the underlying road network file. The map in Figure 6-2 shows how the participant selected a trail that was not apparent on the street network and that was incidentally shorter than the shortest path. However, since the same road network dataset was used to create all the observed paths, a level of consistency was achieved to minimize potential errors from inaccurate GPS positioning.
- Some trips included stops along the way. This was apparent from the raw GPS points, which may have been characterized with a change in direction or a short stay where the participant usually remained stationary for just under two minutes. This did not qualify the points to be reclassified as a new trip and therefore were assumed to be a tour within the respective trip. Figure 6-3 gives an example of a participant making a walking trip with four tours. An adjustment was made in Network Analyst, which identified the tour as a stop and thereby changing the shortest path accordingly. Overall, there was a minimum of one tour/stop observed and a maximum of five tours for any one pedestrian trip in this dataset.

- Other raw GPS points involved a round-trip that appeared to be a leisurely walk around a neighborhood park or around the block. This was denoted in Network Analyst as a very short distance between the origin and destination. As a result, a manual adjustment was made where additional interim stops were added in periods exceeding the 15-second sequence (but less than the two-minute threshold) or where the observed GPS points seemed to have changed direction. This correction to the shortest route allowed for more accurate representation of the actual route taken by the participant and therefore, fewer deviations were noticed in the travel measures.
- Finally, one participant selected two different routes for his ingress and egress between the same origin and destination. The path created by Network Analyst however was equivalent for both observed trips. This shows a consistency for the algorithm in Network Analyst for the shortest path as long as the origin and destination is the same, it does not provide a set of routes only the shortest path of all as in Figure 6-4.



Map Created by Gaby Abdel-Salam

Figure 6-1: Shortest Path Stops Snap to Network Nodes

Figure 6-2: Pedestrian Trail Not Represented On The Road Network

Map Created by Gaby Abdel-Salam

Figure 6-3: One Trip with Four Tours

Map Created by Gaby Abdel-Salam

Figure 6-4: Two Observed Routes for the Same Shortest Path

6.4 Data Analysis

After creating the observed and simulated routes, I needed to generate a set of indices on which the comparison between the two types of routes can be made. These comparisons were completed in the form of descriptive statistics, Wilcoxon and paired sample t-test and GIS maps.

As I mentioned earlier, both the observed routes and the shortest path datasets were generated at the road segment level. To generate the comparative measures, I needed to aggregate both datasets to the trip level using the trip identifier. This step was important so that a one-to-one match is achieved between each trip from the observed GPS routes and the shortest paths. The result produced N = 388 observations or trips.

6.4.1 Distance Deviation Index

The distance deviation index was based on the “Seglength” variable mentioned earlier for both the observed and shortest paths. This variable was based on the coordinate *Pythagorean* distance calculation method for observed routes and for the shortest paths, it was calculated using the calculate geometry feature in ArcMap on the road segment-level. Both, route types were aggregated to the trip-level using the PROC SQL module in SAS using the unique Trip I.D. which produced trip-level summaries of all consecutive road segment lengths covered in each route type. The trip distance summaries were then converted into miles.

Now each Trip I.D. has two associated overall trip distance measures: a trip distance for the observed GPS route and another for the shortest path. One way to compare the two measures is to find the deviance between them per trip, which is the distance deviation index (DistDI). This was calculated as follows:

$$DistDI = \frac{d_{GPS} - d_{SP}}{d_{SP}}$$

Where, "d_{GPS}" and "d_{SP}" denote the trip-level distance for the observed and the shortest routes respectively. A limitation to this index is that in some cases, the trip shortest paths were actually longer than the observed or actual routes taken. This was represented as a negative number for the "DistDI" measure. Some instances where this occurred are presented later in the Limitations section.

6.4.2 Road Segments Deviation Index

Similarly, the segment deviation index (SegDI) was created to represent the discordance between the two route types in the number of road segments traversed. The goal of creating this index was to uncover any underlying associations with the built environment or route-specific characteristics. This measure was calculated as follows:

$$SegDI = \frac{TripSegN_{GPS} - TripSegN_{SP}}{TripSegnN_{SP}}$$

Where, "$TripSegN_{GPS}$" and "$TripSegN_{SP}$" denote the trip-level number of road segments for the observed and the shortest routes respectively.

6.4.3 Route Directness Measure

The route directness measure is the reciprocal of a circuitry factor and it defines the distance that a pedestrian travels relative to the shortest distance between any two nodes. I applied the definition for this measure from (Dill, 2004). The ratio is as follows:

$$RD = \frac{observed\_dist}{network\_dist}$$

Where the "observed_dist" was approximated by the actual distance between the GPS coordinates travelled per road segment and the "network_dist" was approximated by the shortest distance on the same road segment. Therefore, it is the ratio of the distances on the observed to the shortest paths aggregated to the trip-level. This ratio may be compared to the Pedestrian Route Directness (PRD) factor calculated by Dill (2004). A unity value for this ratio indicates a direct route. The suggested values for a pedestrian friendly neighborhood lies between 1.2 to 1.5 (Dill, 2004). Values exceeding 1.8 are generally described as circuitous.

6.4.4 Adjusted Route Directness Measure

The adjusted route directness measure is similar to the RD measure above with one difference, it was normalized with the number of road segments per trip. This adjustment was made to account for the increased number of road segments traversed during an observed relative to a shortest path trip. Therefore, the new measure was calculated as follows:

$$newRD = \frac{(observed\_dist / TripSegN_{GPS})}{(network\_dist / TripSegN_{SP})}$$

6.4.5 Overlap Index

The overlap index (OI) was modeled after the measure explained in Buliung et al. (2013). This measure displays the degree of overlap or matching of the number of road segments covered per trip in the observed and shortest routes.

To obtain an accurate measure, I needed first to remove the duplicated road segments from each dataset type (observed and shortest path datasets). I used the road segment identifier (unique for each road segment) and the *Nodupkey* function in SAS to extract a unique set of road segments per trip for each route type. As a result, the number of observations for the observed routes dataset was reduced from N = 13,510 to N = 2,384 for all observed trips, and the number of unique road segments was now 884. Similarly, the number of observations for the shortest path datasets decreased from N = 1,632 to N = 1,617 for all shortest path trips and the number of unique road segments was now 602.

Next, I combined the two dataset types into one with the non-duplicated road segment I.D's per trip and added two dichotomous variables to differentiate between the route types: "GPSFlag" and "SPFlag". The values of the flags equal to one if the route type was an observed (GPS) route or shortest path (SP) respectively and equal to zero otherwise. This dataset was now at the road segment-level uniquely identified by the segment I.D. Note, that some observations had missing data depending on the route type since there was a prevalent discordance between the two route types in the number of road segments traversed in each trip.

I created another variable, "SegOverlap" which was also dichotomous and equal to one if the two flags created GPSFlag and SPFlag were both equal to one and equal to zero otherwise.

That is, “SegOverlap” now signifies the extent of overlap per road segment for both route types. Then I aggregated this variable to the trip-level creating the “SegOverlapbyTrip” variable. Next, I created a count variable “OBSbyTrip” which captures the number of observations by trip I.D. in this combined dataset. This was used to calculate the percent of overlapping road segments per trip “PerSegOverlapbyTrip” as follows:

$$OI = \frac{SegOverlapbyTrip}{ObsbyTrip}$$

6.4.6 Travel Time Deviation Index

The basis for the travel time deviation index was the GPS data from the observed routes at the road segment-level. There were five general steps followed to calculate this index:

1. Calculate the overall trip duration in hours for the actual observed trips.
2. Use the trip duration and the previously calculated trip distances to calculate the observed average trip speed of the pedestrian.
3. Use the observed average trip speed as the assumed average speed per shortest path trip.
4. Calculate the overall trip duration in hours for the shortest path trips based on the previously calculated shortest path distance and speeds per trip.
5. Calculate the deviation index between the trip durations for each route type that corresponds to the trip-level travel time deviation index.

Initially, I needed to aggregate the GPS data from the 15-second interval readings to the hour-level and then summing these values by trip I.D. to obtain the observed trip duration in hours. At this point, the number of observations was collapsed from N = 13,510 to N = 388, the total number of trips. Next, I used these observed trip duration values to calculate the trip-level average walking speeds in miles per hour for each participant for each of his/her trip. Note, this means that there was also a variation among the average walking speeds per trip for the same participant.

The observed average walk speeds were assumed to be the same for the shortest path trips and were therefore used to calculate the trip duration of a shortest path trip however; the trip distance used was obtained from the shortest path dataset. The resulting values provided the trip durations or the trip-level travel times of the shortest paths.

Finally, I computed the travel time deviation index (TTDI) as follows:

$$TTDI = \frac{Tripduration\_h_{GPS} - Tripduration\_h_{SP}}{Tripduration\_h_{SP}}$$

Where the variables "Tripduration_h_{GPS}" and "Tripduration_h_{SP}" correspond to the trip-level mean values for the trip duration in hours for the observed and the shortest path trips respectively.

6.4.7 Built Environment Exposure Measure

The intention of creating this measure was to capture the impact of the surrounding built environment on a pedestrian's trip and therefore to shed some light on the choice of the path of travel selected. However, by merely including the values of the percentages or densities of the various land uses along a particular road segment is not sufficient since a pedestrian might not traverse the whole length of the segment during his/her trip. Therefore, it was important to create an exposure variable that approximates the participant's partial segment exposure to the land uses along the route. This was calculated as follows for the observed and shortest paths respectively:

$$SegExpGPS = \frac{SeglengthGPS}{SegTotlengthGPS}$$

And

$$SegExpSP = \frac{SeglengthSP}{SegTotlengthSP}$$

The variables: "SeglengthGPS" and "SeglengthSP" are the per road segment distance in miles traversed in a trip for the observed (GPS) and the shortest paths (SP) respectively. The variable "SegTotlength" variable is the total length of the network road segment in miles for each respective route type. A value equal to one for this ratio indicates that the traveler traversed the whole road segment during a trip and values closer to zero show only partial traveling for the same road segment.

The next step was to use these exposure ratios to calculate the relative exposure percentage for every land use type. This was accomplished by multiplying the exposure ratios by the land use densities and percentages present along the road segment. The results obtained yielded a set of percentage values for the various land use types for each of the observed and shortest paths. For example, the variable capturing the percentage of commercial and retail uses along a road segment was multiplied by each segment exposure ratio as follows:

$$LSCommretGPS = LSCommret * SegExpGPS$$

And

$$LSCommretSP = LSCommret * SegExpSP$$

Where the "LSCommretGPS" and "LSCommretSP" are the built environment exposure percentages for the commercial and retail uses for the observed and shortest routes respectively which were then aggregated to the trip-level. This process was repeated for the various built environment uses at the road segment-level and a list of the resulting percentages after factoring the exposure effects are displayed in the results section.

6.5 Results by Socio-Demographic Traits

6.5.1 Sample Characteristics

The participant sample used in this chapter was a subset of the participants in the Expo study mobile tracking group. Out of the 143 participants from Phase 1 of the Expo study that agreed to be in the mobile tracking group and carry portable GPS and accelerometer devices when they traveled; a total of 62 were considered to use walking as their main travel mode. These individuals relied on walking for transport for the majority of their trips but may still have utilized public transit and/or cars. These participants were considered the pedestrian sample.

All participants provided information regarding their socio-demographic characteristics. Basic socio-demographic traits are displayed in Table 6-1 below which was analyzed at the participant level. The majority of the sample was female (68%). The average age of the participants was 53.4 years. The youngest participant was 21 the oldest 79 years but almost 37% of the sample was over 60 years old. About 48% of the participants making the walking trips were unemployed. Not all households owned bikes in the sample but on average, those who did had at least one bike (Mean No. of Bikes = 1.12), and the maximum number of bikes owned was four. The average number of cars per household was higher (1.25) and some households had up to three cars where others did not have any. Further, the mean duration of the walks was 23.02 minutes and on average, over 27 minutes per day were expended during moderate-to-vigorous-physical-activity (MVPA) for the sample. This physical activity duration is very good especially because it is very close to the daily recommendations for an average adult. The daily recommendations of physical activity requires at least 30 minutes of moderate activity for five days or more per week or 20 minutes of vigorous activity for at least three days per week (Rodriguez, Khattak, & Evenson, 2007).

Table 6-1: Participant Characteristics

Socio-Demographic Traits	**N**	**Mean**	**S.D.**	**Max.**	**Min.**
Female	62	0.68	0.47	1	0
Age	62	53.40	14.52	79	21
Young (≤ 60 yr.)	62	0.63	0.49	1	0
Unemployed	62	0.48	0.50	1	0
No. of Bikes	60	1.12	0.961	4	0
No. of Cars	60	1.25	1.00	3	0
MVPA (min.)	62	27.38	19.78	96	0.5

Figure 6-5 shows more socio-demographic traits for the pedestrian sample. Over half the sample declined to answer the question regarding their education status. A total of 10% of the sample earned a high school degree or less and 13% obtained up to an Associate's degree. In addition, 22% held at least a Bachelor's degree.

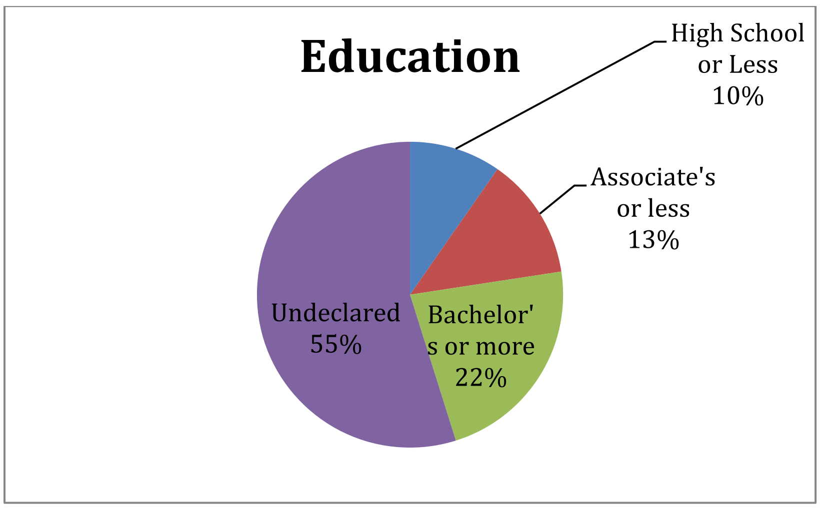


Figure 6-5: Participants' Education Distribution (N = 62)

The pie chart of Figure 6-6 displays the Annual Income percentages of the Expo sample. Over half of the sample (56%) refused to declare their income. The lowest income group earning $35,000 or less per year comprised 26% of the sample followed by 8% in the middle income group ($35K to $75K) and 10% in the highest income category earning $75,000 or more per year.

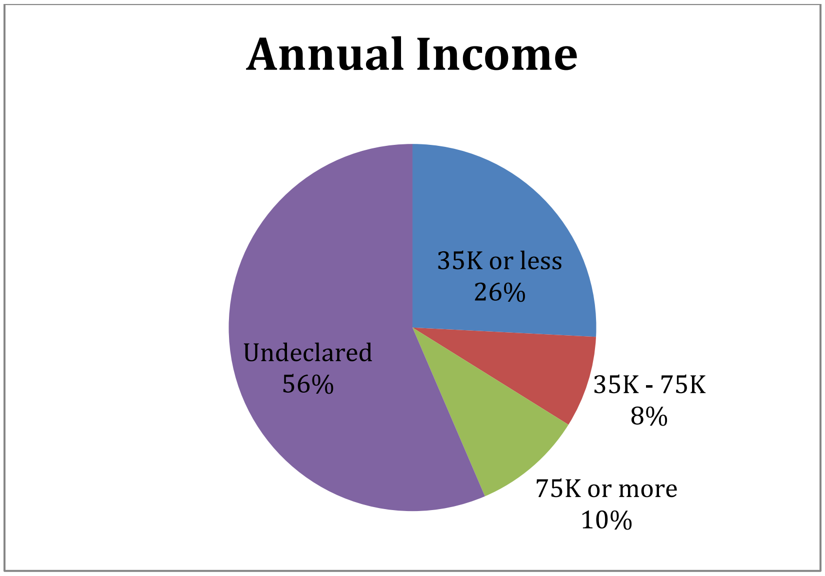


Figure 6-6: Participants' Annual Income Distribution (N = 62)

The racial composition of the sample is displayed In Figure 6-7. The majority of the participants in the sample were African American (58%). Whites made up 26% of the sample. Hispanics and Asians comprised 2% and 3% of the sample respectively; 6% claimed 'Other' for their race and another 5% declined to state their race.

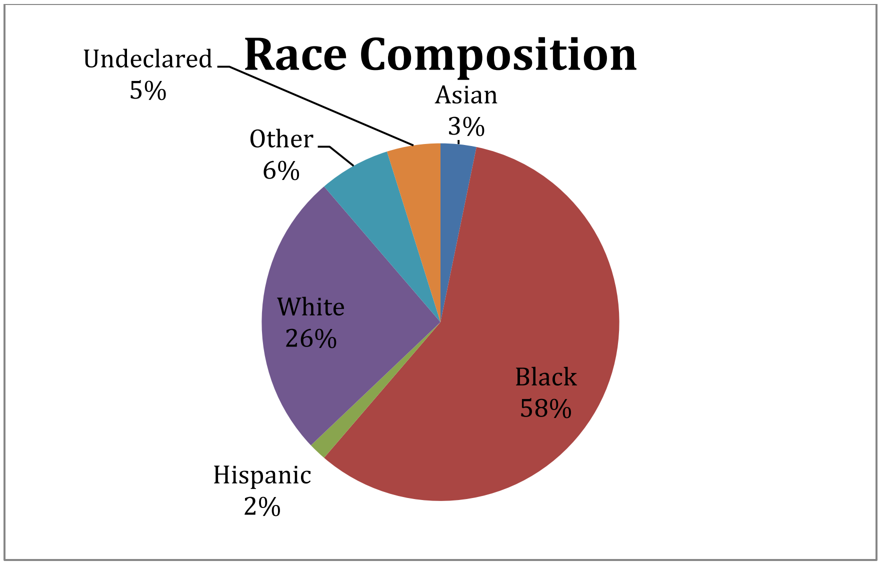


Figure 6-7: Participants' Race Distribution (N = 62)

6.5.2 Trip Characteristics by Participant Socio-Demographic Characteristics

There have been several studies that suggest that the location of pedestrian walking behavior varies by trip purpose. One study found that most leisure walks (82.9%) occur within participant's neighborhood which included areas within 1.24 miles (straight-line distance) from the respondent's home (Suminski, Wasserman, Mayfield, Kheyfets, & Norman, 2014). Others suggest that most walking occurs within 1 km (0.62 miles) from home regardless of purpose (L. Frank, Bradley, Kavage, Chapman, & Lawton, 2008; C. Lee, 2004). Still others believe that a half-mile radius is the optimal catchment area size for utilitarian walks (Howard, E. J., Kang, B., Hurvitz, Moudon, & Saelens, 2014) or is optimum when studying the impacts of transit use on walking trips (Brown & Werner, 2007; Schlossberg & Brown, 2004) or examining the potential extent of daily walking activities (Badland et al., 2013; Canepa, 2007; Guerra, E. & Cervero, 2013).

As mentioned earlier in the methodology, the road network used in this research was based on the TIGER street shapefile for the Expo study obtained from the U.S. Census. Half-mile buffers were created centered on the respondents' home locations and the streets within these buffers were extracted and added into a subset road network shapefile.

Realizing that some participants may walk more than a half-mile distance, I wanted to expand the walking distance threshold to one-mile, the distance an average person would normally walk in about twenty minutes. Table 6-2 shows the breakdown of the respondent groups who walked less than or equal to a mile in any one day and the trip duration that lasted for twenty minutes or less. The mean trip distance was 0.89 miles. The results from the table show that 89% of the female respondents walked for twenty minutes or less and that 93% of this group walked a mile or less per day. About 95% of African Americans in the sample walked for the same duration and distance. For the Whites in the sample, 87% and 91% walked for up to twenty minutes per day and up to one mile respectively. Almost all the remaining races walked for the same duration and distance per day. Similarly, 98% of the respondents in the low-income category, earning up to $35,000 annually also walked less than one mile and 99% walked twenty minutes or less per day. Further, the majority of the unemployed (about 90%) and those 60 years or under and walked the same duration (93%) and distance (96%) in any one day. Overall, the majority of the participant groups completed shorter trips in distance and duration.

Table 6-2: Socio-demographics and Short Trip Characteristics (N = 62)

Participant Group Type	**Trip Characteristics**	
	Trip Duration (< 20 min.)	Trip Distance (< 1 mile)
Female	89%	93%
African American	95%	95%
White	87%	91%
Asian	100%	100%
Hispanic	100%	100%
Other	95%	100%
Low Income (≤ 35K/yr.)	98%	99%
Young (< 60 yr.)	93%	96%
Unemployed	90%	92%

6.5.3 Trip Characteristics of Observed and Shortest Routes

Table 6-3 compares the trip distance, the number of road segments and the trip duration for each route type. The table shows that on average, the GPS observed routes were longer (0.34 miles) than the shortest paths (0.23 miles). This confirms the hypothesis that pedestrians do not necessarily select the shortest routes for their destination. Further, the maximum distance traveled by the participants by foot for any given trip was 4.35 miles compared to only 2.91 miles for the shortest paths for the same trip.

In addition, the average number of road segments was 32 segments along the observed routes compared to only 4.12 segments along the shortest paths. This further adds to the premise that the GPS routes are much longer than the shortest paths. The minimum and maximum observed were 4 and 300 road segments respectively compared to only 1 and 30 segments respectively along the shortest routes.

Further, the observed route trips were generally longer than those simulated with the shortest paths. The average trip duration observed was 8.4 minutes (0.14 hrs.) long compared to an average of 6.6 minutes (0.11 hrs.) along the shortest paths. The longest trip duration along the observed routes was 2.19 hours and the shortest trip was a little over a minute long. In contrast, the longest trip duration along the shortest paths was 2.30 hours and the shortest trip was negligible (0.0001 hours).

Table 6-3: Observed vs. Shortest Path Descriptive Statistics

Variable	GPS/Observed Route					SP/Simulated Route				
	Mean	S.D.	Max.	Min.	N	Mean	S.D.	Max.	Min.	N
Trip Distance (miles)	0.34	0.54	4.35	0.01	388	0.23	0.29	2.91	0	388
# Road Segments	32.33	41.77	300	4	388	4.12	4.07	30	1	388
Trip Duration (hrs.)	0.14	0.21	2.19	0.02	388	0.11	0.16	2.30	0	388

The trip speed was assumed to be the same for both route types. The mean observed walking speed in the sample was 2.29 mph. This is lower than the pedestrian speed range noted in one study that was estimated to be in the range of 3 - 3.6 mph for adults residing in North America and Australia (Wasfi et al., 2013). Other studies in the literature however, concluded that an average pedestrian speed is about 2.27 mph when we factor the landscape obstructions such as hills and arterial roads that would naturally impede a pedestrian's walk (Canepa, 2007). Further, the observed walks included brief stops that may have contributed to the slower average speed while other studies may not necessarily have included time for stops. Therefore the pedestrian speed observed in this chapter is considered reasonable especially seeing that the average age in the sample was about 55 years, slower walking speeds would be expected anyway.

6.6 Road Type Analysis

The classification of the road network in this dataset has been obtained from the TIGER street network shapefile administered by the U.S. Census Bureau. Each road is classified according to the MTFCC code (MAF/TIGER Feature Class Codes) that is linked to road-specific characteristics such as speed, capacity, number of street segments, etc. This code also designates the road type represented by its specified characteristics.

The comparative statistics of Table 6-4 provide the contrasting trip-level road type percentages for the observed and shortest paths per trip. The majority of roads used as a pedestrian path by participants were local roads. There were more of these road types in the observed GPS routes (97.47%) than the shortest paths (95.43%), and the overall local roads in the study area comprised about 98% of all the roadways. There were fewer parking lots along the observed routes (1.36%) than the shortest paths (1.63%) and no ramps detected along the shortest paths. The percentage of secondary and primary roads along the shortest paths (2.63% and 0.31% respectively) was much greater than along the observed routes.

The graph of Figure 6-8 provides an overview of the different road types and the comparative percentages for each route. These results indicate that pedestrians use local roads more than the remaining road types that may be busier and have higher traffic volumes. This may be contrary for drivers that may be inclined to take more primary and secondary roads.

Figure 6-9 shows a map for the Expo road network. The observed GPS routes were shown to be mostly along local or neighborhood roads. Further, the majority of the road network of the study area was comprised of local roads. This could be the main reason that the participants in this study used more local roads than all other road types.

Further, this may also mean that the settings of Network Analyst accommodate drivers more than it does non-motorized travelers. The algorithm of the impedances for the shortest paths may need to be adjusted to cater to other modes of travel than vehicles.

Table 6-4: Comparison of Observed & Shortest Paths by Road Types

Road Type Description	MTFCC Code*	GPS/Observed Route			SP/Simulated Route		
		Percent	Cumulative Percent	N	Percent	Cumulative Percent	N
Primary	S1100	0.24	0.24	388	0.31	0.31	388
Secondary	S1200	0.73	0.97	388	2.63	2.94	388
Local	S1400	97.47	98.45	388	95.43	98.37	388
Ramp	S1630	0.20	98.64	388	—	—	388
Parking Lot	S1780	1.36	100	388	1.63	100	388

*MAF/TIGER Feature Class Codes
Source: (Census, 2014)

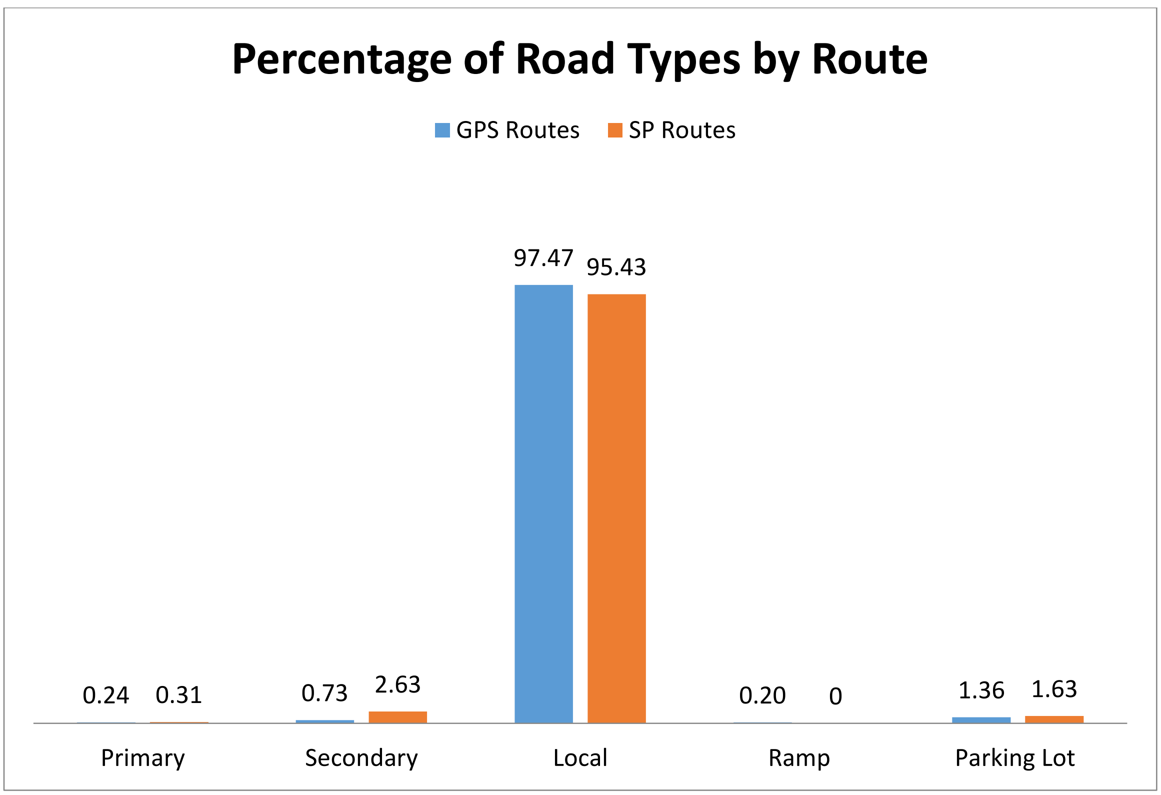
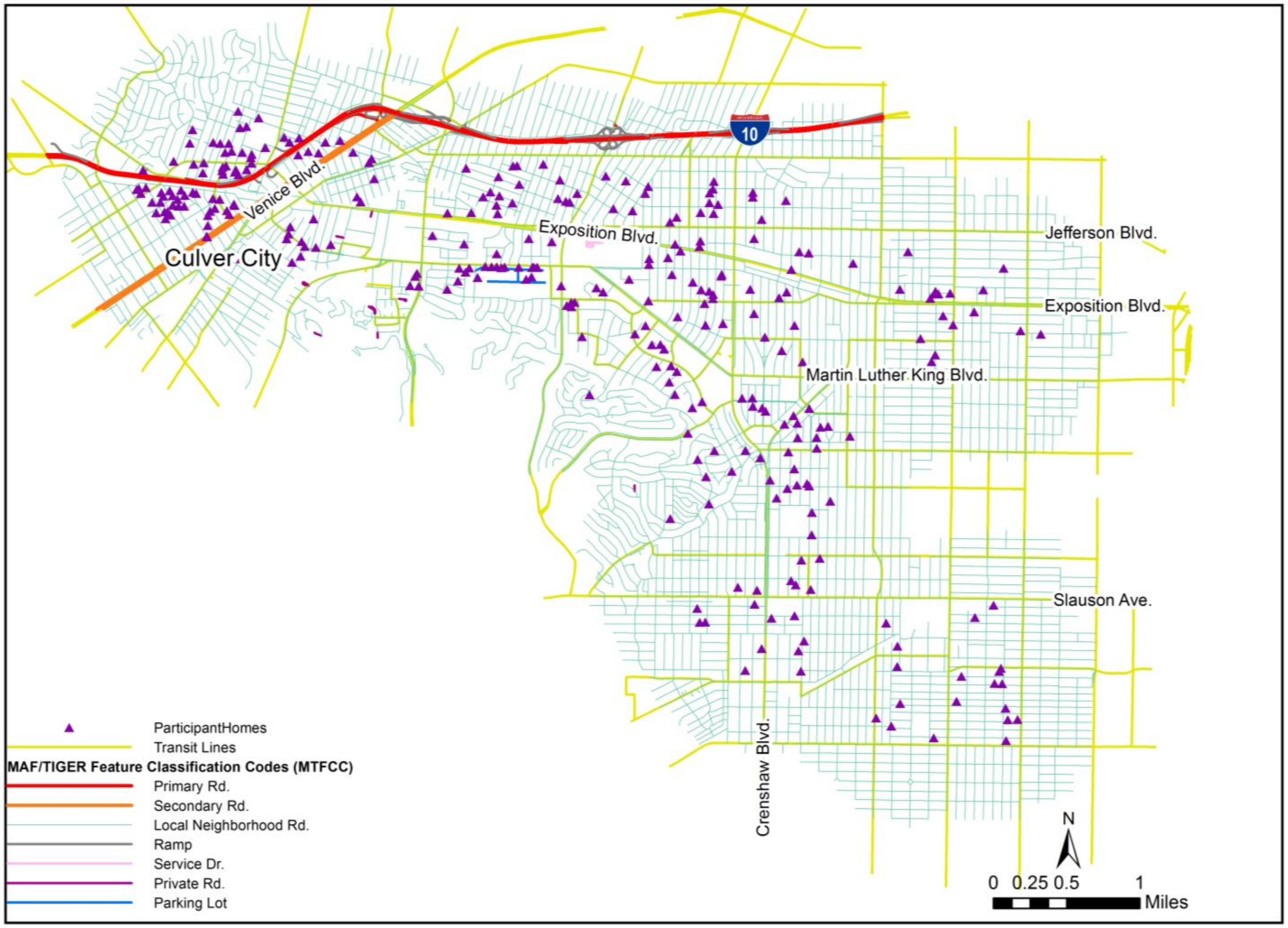


Figure 6-8: Road Types by Route



Map Created by Gaby Abdel-Salam

Figure 6-9: Road Type Classification of Network

6.7 Travel Indices Results

The results of the indices described in the previous section are displayed and explained below. The route deviation indices results for distance and road segments are described first in addition to a comparison of trip-level duration time of the trips along the observed and shortest paths. Next, the segment-level and trip-level findings of the exposure to various built environment factors are discussed and a sign test of significance was also performed at the trip-level. The following section provides a trip-level comparison of the different road types highlighting the percentages of each category along the observed and shortest paths. The travel time deviation index (TTDI) is discussed next followed by a time of day analysis and comparison of the different indices for the observed and shortest paths.

6.7.1 Distance and Segment Deviation Indices Results

The discordance among the observed GPS and shortest paths in distance and number of road segments per trip is shown via the distance deviation (DistDI) and road segments deviation (SegDI) indices. The mean values for DistDI and SegDI are 3.98 and 9.73 respectively, indicating a large discordance in distance and number of street segments between the observed and simulated shortest paths.

Figure 6-10 shows a trip-level representation of the variations between the two route types. Generally, the two indices follow the same patterns since route distance is directly correlated to the number of road segments.

Further, the DistDI Index for the most part follows the zero line indicating an overlap between the observed routes and the shortest paths. The discordance however is apparent the longer the trip distance is, typically right after the 2.5-mile mark. The pattern becomes irregular but generally the deviation increases greatly for longer trips. Values for DistDI that exceed zero indicate that the observed routes were longer than the shortest paths per trip. The larger the positive values, the greater the discordance between the observed and shortest routes.

A similar pattern can be observed for the road segments deviation Index. The index hovers just above the zero line of overlap between the observed GPS routes and the shortest paths but increases rapidly beyond the same distance threshold of 2.5 mile. Again, positive values indicate that the observed routes included a larger number of road segments than those of the shortest paths.

The observed routes depicted in Figure 6-11 probably resemble leisure walks. The respondent took a stroll around the neighborhood returning close to the point of origin. The simulated shortest paths however; indicate a much shorter distance traveled. As a result, the observed routes are much longer than the shortest paths.

In contrast, incidences where the observed trips were shorter than the simulated shortest paths are provided below in the GIS maps of Figures 6-12(a & b). This phenomenon could be observed for trips that the pedestrians used trails and pathways for to cut through lots which were not represented on the road network used to calculate the shortest paths.

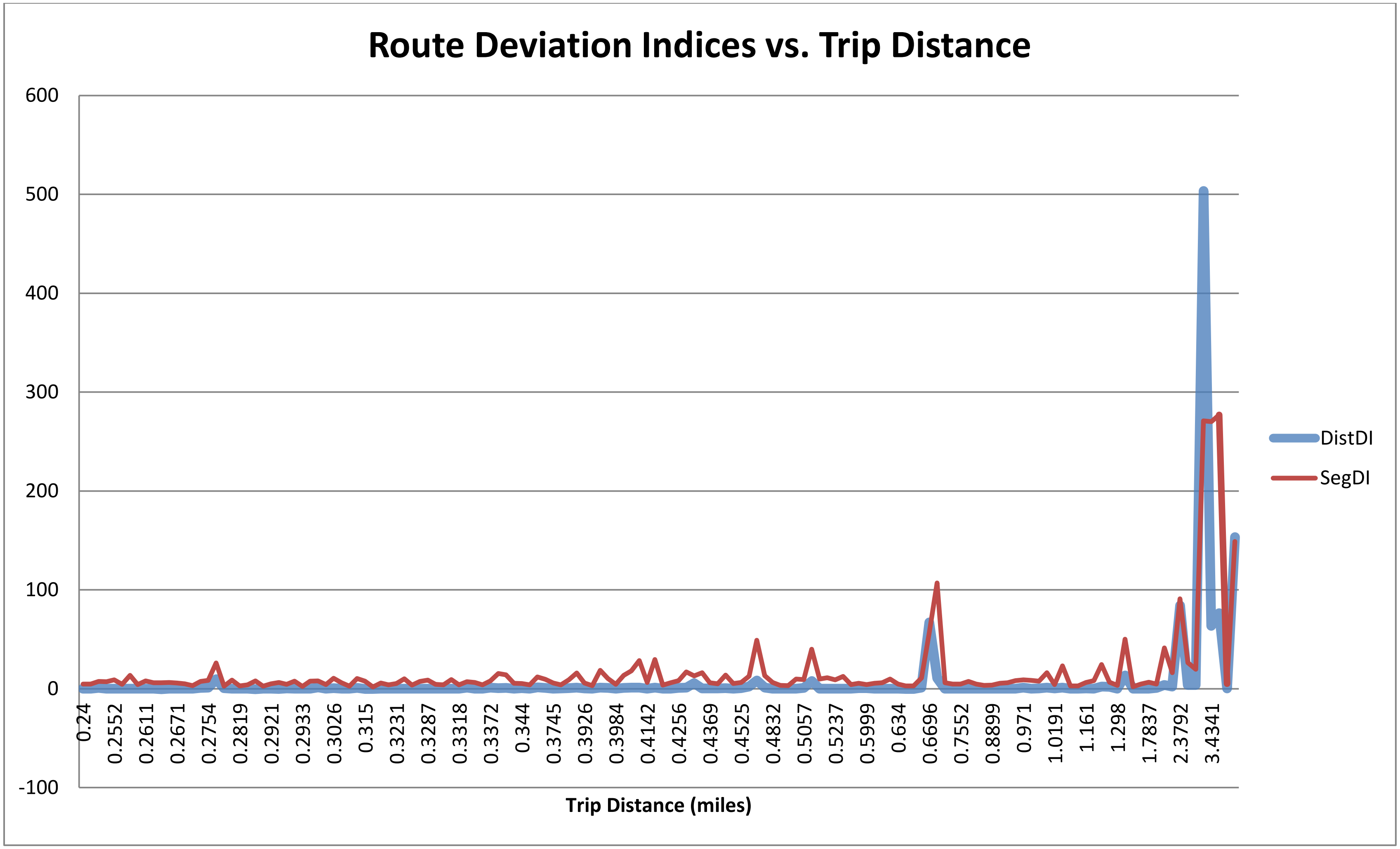


Figure 6-10: Route Deviations between Observed & Shortest Paths

Figure 6-11: Leisure Walks with a Positive TTDI Value

Figure 6-12(a): Examples of Shortest Paths Longer than the Observed Routes

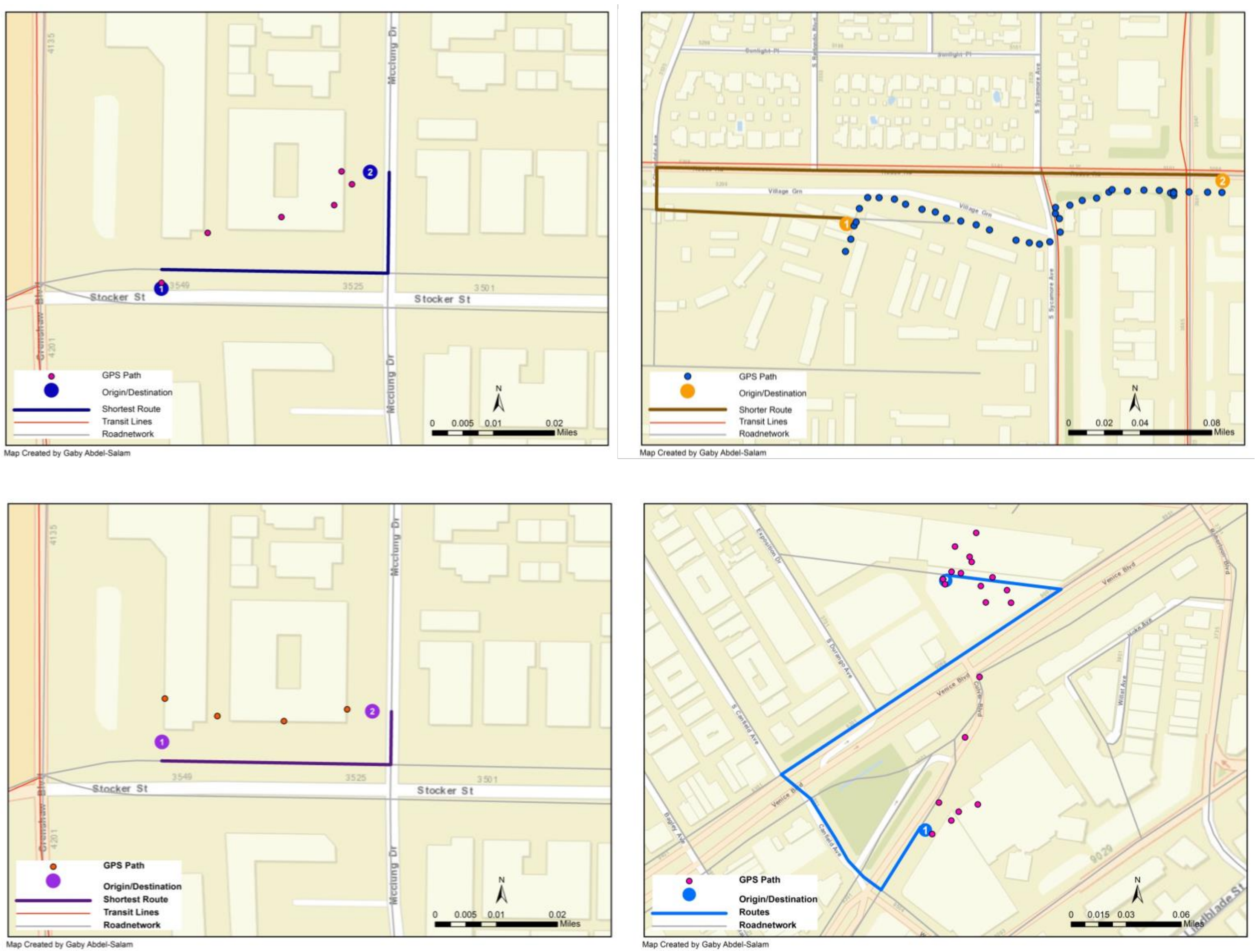


Figure 6-12(b): Examples of Shortest Paths Longer than the Observed Routes

6.7.2 Route Directness Indices Results

There were two route directness indices created, the first (RD) is a trip-level ratio of the observed (GPS) distance to the shortest path distance calculated by Network Analyst; the second index (newRD) was calculated similar to the first however, the distance variable is normalized by the number of road segments traversed per trip. This correction lessens the outlier values that may appear if the pedestrian walked for long distances (e.g. leisure) but the corresponding route calculated by Network Analyst for the shortest path was very short.

The median values for the RD and newRD indices were 1.055 and 0.199 respectively. I used the median value instead of the mean since the median is not sensitive to outliers. Further, 75% of the trips had an RD value less than 1.5. This indicates that the majority of trips were fairly direct and about 25% were more circuitous (RD > 1.5). Walking trips with values exceeding 1.63 have been previously observed to occur along neighborhoods with cul-de-sacs and curvilinear streets as opposed to grid patterns (Dill, 2004). The Expo study area however, generally has less of these more circuitous street designs.

Figure 6-13 displays the trip-level values for both indices against distance travelled in miles. Most trips appear to be direct until about 2.40-mile mark (red line). The index seems to increase beyond this threshold with increasing trip distance. These results echo those obtained from the previous section with the distance and segment deviation indices. All indices seem to be within expected range for shorter trips but their deviation increases the longer the trip is.

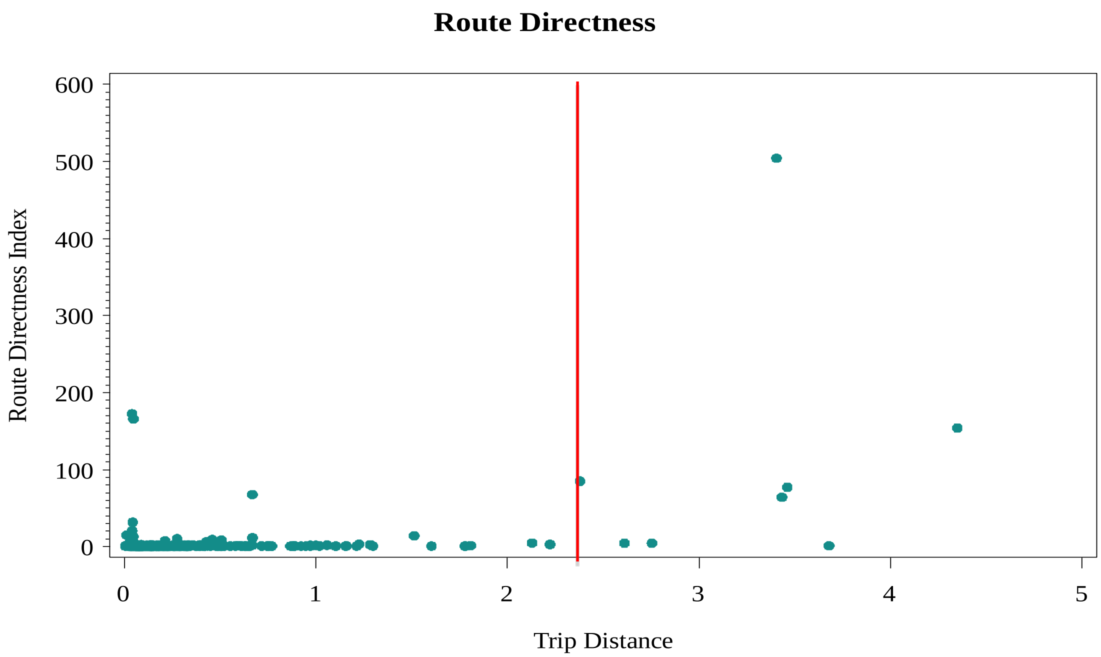


Figure 6-13: Route Directness Index (RD) vs. Trip Distance

6.7.3 Overlap Index Results

The overlap index (OI) discussed in the literature is a means to identify how well the selected (observed) route follows the simulated shortest path. It is also a ratio of the shared road segments from the observed and shortest paths to the segments of the observed GPS paths. This index ranges from zero to one where a unity indicates perfect overlap and values closer to zero mean that the observed route was different from the shortest path. The result of the univariate analysis for this index shows that the median value was 0.5 and the mean was 0.55. Generally, the overlap index followed a normal distribution. Almost 25% of the trips had an OI value equal to one (perfectly overlapping with the shortest path). The distribution plot of Figure 6-14 below shows a normal distribution of the index. The plot shows that about 44% of the trips had at least a value of 0.6 for the OI index suggesting that 44% of the observed trips mostly overlapped with the shortest paths.

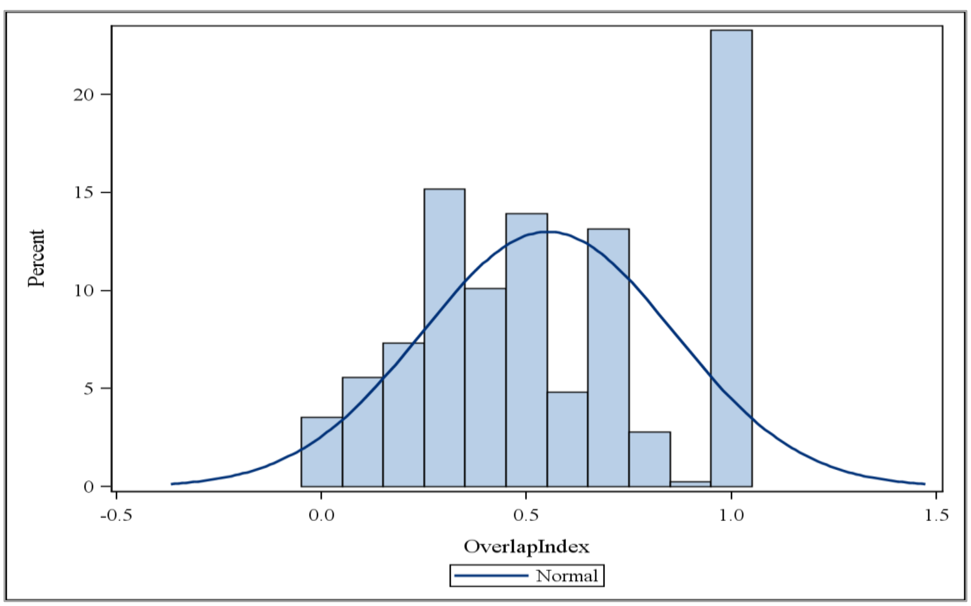


Figure 6-14: Distribution of the Overlap Index

6.7.4 Travel Time Analysis

The analyses discussed in this section pertain to travel time. First, the travel time deviation index (TTDI) is discussed then I discuss a time period analysis comparing the mean values of the travel indices by time of day, and finally I examine the time period breakdown by participant socio-demographic group type and compare differences among them.

6.7.4.1 Travel Time Deviation Index Results

The analysis of this section was performed at the trip-level. Values on this index closer to zero indicate a complete match between the trip durations of both route types. Negative values suggest that the SP trips were longer than the observed GPS-recorded routes and positive values indicate the opposite, that trips along the observed routes were longer than the shortest paths.

The distribution plot of TTDI in Figure 6-15 shows that this index is skewed to the right and that the positive values are outliers that would affect the mean TTDI value. Therefore, only the median TTDI will be discussed next because it is not sensitive to such outlying values.

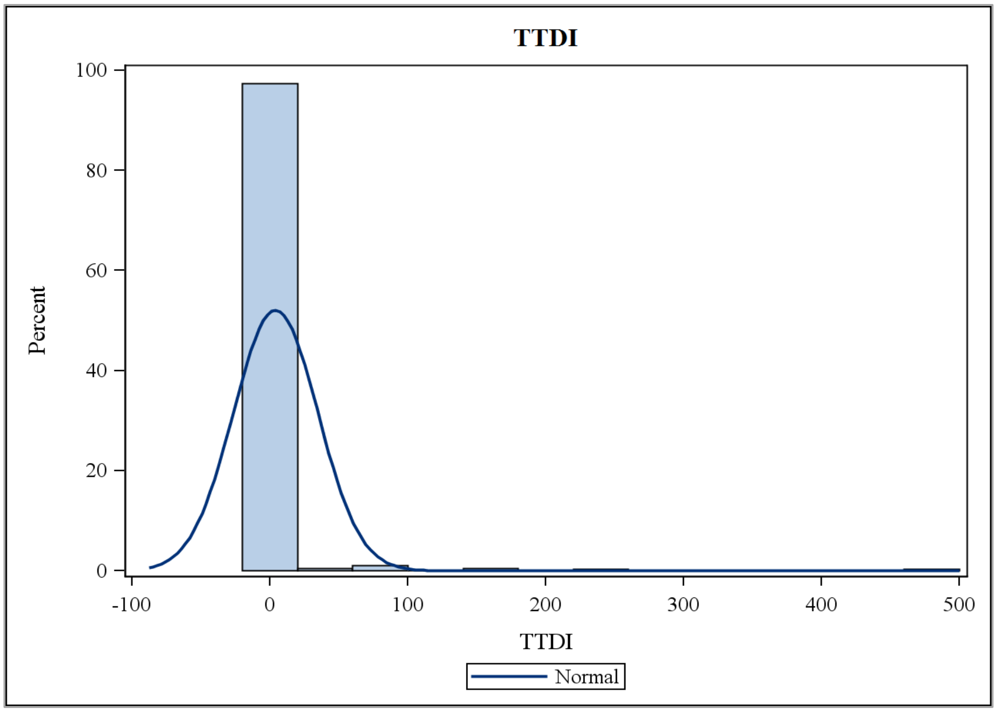


Figure 6-15: Distribution of the Travel Time Deviation Index

The median TTDI value was 0.05 suggesting more overlap between the observed and shortest paths in trip duration. Only 10% of the trips had values greater than or equal to 1.25 on this index, signifying that the trip durations of the observed routes were longer than that of the shortest path trips.

Further, 25% of the trips had a TTDI value less than or equal to -0.04 suggesting that the shortest path trips were longer than the actual walking trips. Overall however, the majority of the trips seem to have values very close to the zero threshold of overlapping durations and that only some deviated from this threshold.

6.7.4.2 Time of Day Analysis

The analysis in this section examines the travel indices by time of day. The objective is to check for any differences in the indices across the different time periods. As mentioned earlier, the GPS data contained a date and time stamp that was used here to classify the various time periods of the day. The time periods were classified as follows:

1. Early Morning (12:01 AM – 6 AM)
2. AM Peak (6:01 AM – 9 AM)
3. AM Off Peak (9:01 AM – 12 PM)
4. PM Off Peak (12:01 PM – 4 PM)
5. PM Peak (4:01 PM – 7 PM)
6. Evening (7:01 PM – 12 AM)

The results in Table 6-5 show a comparison of the different mean travel indices, observed and shortest path distances for the time of day periods. Figures 6-16(a) and 6-16(b) show this time of day breakdown and the corresponding values for the travel indices. Overall, 30% of the walking trips have been completed in the PM Off Peak (N = 114) this could be an indication that these trips were leisure walks occurring after working hours. Similarly, 29% of the walking trips occurred in the PM Peak time (N = 112) suggesting a more utilitarian trip function. The remaining trip time frequencies were as follows 1% for Early Morning (N = 4), 12% for AM Peak (N =47), 22% for AM Off Peak and 7% for Evening walking trips.

The greatest value for DistDI was observed for the AM Peak time (DistDI = 19.33). As a reminder, this value indicates that the observed trip distance exceeds that of the shortest path. This finding is also confirmed by the mean observed trip distance of 0.34 miles compared to only 0.23 miles for the same trip's shortest path distance. Since this time period starts at 6:01 AM, the large gap between the observed and shortest paths' distances may indicate that some of these trips may have been an early stroll prior to leaving for work. Strolls may not necessarily follow the shortest path since they would be likely classified as leisure walks. The smallest DistDI values occurred at the Early Morning (DistDI = 0.75) followed by the Evening time period (DistDI = 0.81) indicating more correspondence with the shortest routes.

Similarly, the patterns observed for TTDI mirror those of DistDI since trip duration is a function of distance traveled and therefore they are highly correlated. The least trip duration deviance occurred in the Early Morning time (TTDI = 0.75) followed by the Evening period (TTDI = 0.84). The largest gap between the observed and shortest paths trip durations occurred in the AM Peak time (TTDI = 19.10) as seen in the DistDI index.

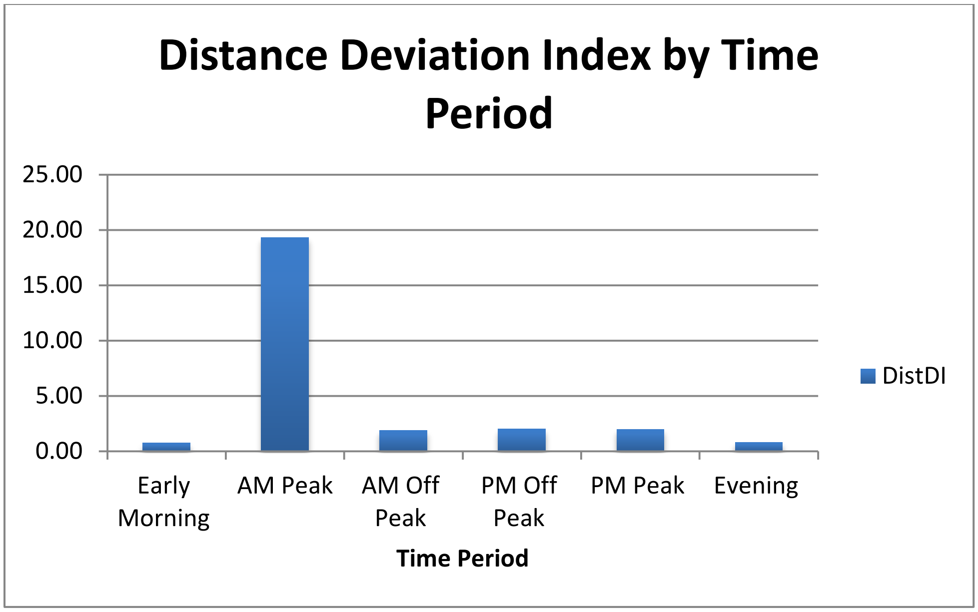
The RD index is less than a unity value (most direct route) for all time periods signaling that the observed GPS paths for all time periods combined were on average more direct than SP paths. The results from Table 6-3 indicate that on average, the observed routes were longer than SP paths. Therefore, we can conclude that shorter paths are not always the most direct. However, the PM Peak RD value was closer to unity (RD = 0.49) than the other periods indicating that the actual GPS paths in the PM Peak time were less direct than other time periods. One possible reason for this may be trip-chaining after work such as to convenience stores or for picking up children from daycare/school prior to returning home. Results from a previous study comparing socio-demographic differences in vehicular travel shows that trip-chaining behavior is highest among single adults with younger dependents before and after work and occur mostly in peak time periods (Li, Guensler, & Ogle, 2005). However, this may not apply to pedestrian travel and more data on trip purpose would be required to confirm these assumptions. The smallest RD value was observed for the Early Morning period (RD = 0.16) but only 1% of all the trips were completed in this time period which precludes further analysis. However, if a higher percentage of trips were made in the Early Morning time that have a small RD value; this may suggest safety concerns and that the respondents opted for more direct routes. Further crime analysis however would be necessary to confirm this inference if more trips were observed in this time period. This is also mirrored by the low DistDI value of 0.75 for the same time period suggesting shortest paths are preferred in this time of day. The value for RD in the AM Peak time is also very small 0.37 indicating that the respondents prefer the most direct routes possibly to reach their work destinations.

Overall, the mean OI values were close to the median (0.5) for AM Off Peak (OI = 0.53) time. The smallest OI value was observed for AM Peak (OI = 0.40) suggesting the least overlapping road segments with those of the shortest paths for this time period. In contrast, the largest OI values were detected for the Early Morning (OI = 0.64) followed by the PM Peak (OI = 0.6) and Evening (OI = 0.59) time periods indicating that respondents may be more willing to take shorter routes for this time period possibly for safety concerns (Early Morning and Evening times) or in the case of PM Peak time to return home as quickly as possible from work.

In almost all the time periods, the respondents' mean observed distances were longer than the distance of the shortest paths. The exception was for the Evening time period where the two distances were exactly equal (0.19 miles) reaffirming the preference for the shortest paths in this time period.

Table 6-5: Time of Day Comparison of the Mean Values of the Travel Indices

Time of Day	N	DistDI	TTDI	RD	OI	Trip Distance (mi) (Observed)	Trip Distance (mi) (Shortest Path)
Early Morning	4	0.75	0.75	0.16	0.64	0.38	0.22
AM Peak	47	19.33	19.10	0.37	0.40	0.73	0.28
AM Off Peak	84	1.88	1.86	0.34	0.53	0.37	0.29
PM Off Peak	114	2.01	1.84	0.43	0.57	0.25	0.19
PM Peak	112	1.98	2.60	0.49	0.60	0.28	0.23
Evening	27	0.81	0.84	0.33	0.59	0.19	0.19



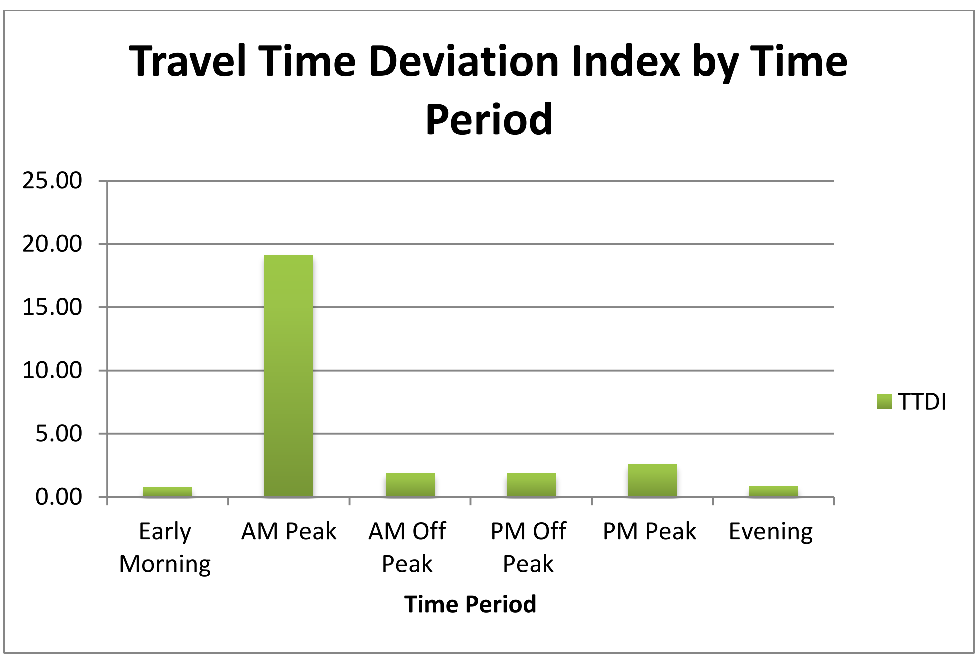
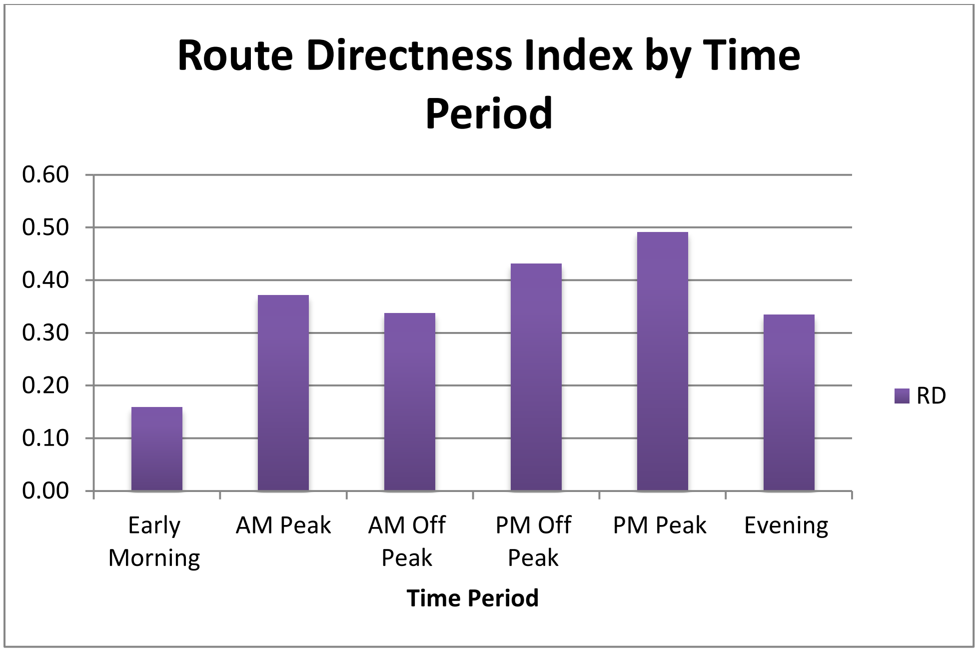


Figure 6-16(a): Travel Indices by Time Period



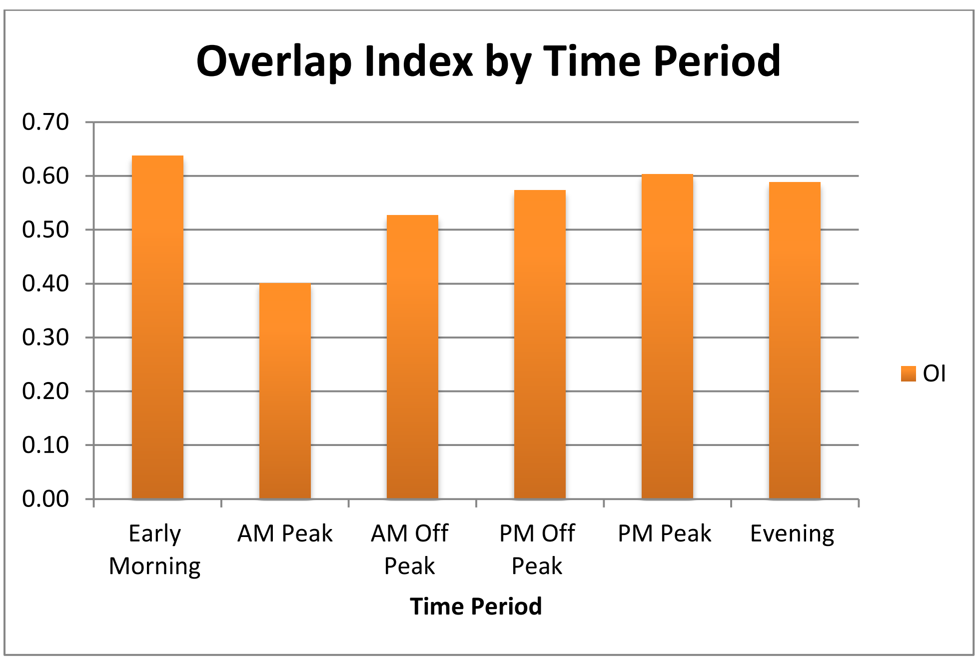


Figure 6-16(b): Travel Indices by Time Period

6.7.4.3 Peak Time Analysis

In this section, I contrasted the mean travel indices for the AM Peak time against those of the remaining time periods. The mean differences were computed and a t-test for significance was estimated. For simplicity, only significant differences will be discussed. The results are shown in Table 6-6. There were a total of 47 trips in the AM Peak Time compared to the majority of 341 occurring in the remaining time periods.

The average OI value for the AM Peak time was smaller and statistically significant than that for the remaining time periods (p-value = 0.0003) suggesting less overlap of the observed routes and the shortest paths. Similarly, the analysis in the previous section shows that the least road segment overlap occurred in the AM Peak time where the participants did not use the shortest routes.

The observed mean distance for the AM Peak time was also longer than all other times and was statistically significant at the 1% significance level (p-value = 0.006). However, it was also shown earlier that the average observed distance for the GPS routes were longer than that of the shortest paths except for the Evening time period where the two distances were equal. The Evening time period falls under the “Remaining Time Periods” category.

The remaining differences for the travel measures between the Peak and all other time periods were not statistically significant.

Table 6-6: Comparison of Travel Indices for AM Peak Time vs. All Remaining Time Periods

Measure	AM Peak Time			Remaining Time Periods						
	Mean	S.D.	N	Mean	S.D.	N	Diff[†]	t-value	p-Value	Sig.
DistDI	19.33	77.36	47	1.86	13.71	341	-17.47	-1.54	0.1292	
TTDI	19.10	75.96	47	2.00	15.91	341	-17.10	-1.54	0.1307	
RD	0.37	0.52	47	0.42	2.29	341	0.05	0.31	0.7538	
OI	0.40	0.29	47	0.57	0.30	341	0.17	3.83	0.0003	***
Observed Distance	0.73	1.04	47	0.29	0.40	341	-0.44	-2.87	0.0060	***
Shortest Path Distance	0.28	0.35	47	0.23	0.28	341	-0.06	-1.09	0.2796	

[†]Mean difference between the AM Peak Time and all other periods for GPS and SP indices
Significance: * p < .1, ** p < .05, *** p < .01

6.7.5 Travel Measures by Participant Group

The analysis in this section provides a comparison of the different travel time periods by the respondents' socio-demographic group in addition to comparing the travel measures for the same participant groups.

6.7.5.1 Travel Time Period by Socio-demographic Group

Various socio-demographic traits were collected for the respondents in the sample and a by-group analysis was completed to examine any variations in frequencies among them for the different time periods discussed earlier. The results displayed in Table 6-7 show the cross tabulation of the participant groups and the trip frequency distribution over the six time periods. The analysis was performed at the trip-level.

The last column provides the total percentage of the participant group type in our sample. As mentioned before, females comprised 56% of the respondent sample, the majority were of African-American descent (65.47%), those in the lowest income category made up 43.82% of the sample, 70.88% were at most 60 years of age and 68.30% unemployed.

Some interesting insights can be obtained from the participant group frequencies. Very few walks occurred in the Early Morning and AM Peak periods and most were concentrated beyond the AM Off Peak threshold. More than half of the females in the sample preferred afternoon walking trips ((14.95 +15.72)/55.93) in the PM Peak and PM Off Peak times. Similarly, 63% ((21.91 + 19.59)/65.47) of the African-Americans in the sample also preferred the same two time periods for traveling. Those in the sample who were 60 or under also preferred the late morning (12.41%) and afternoon time periods (PM Off Peak = 21.13 & PM Peak = 22.68%). The unemployed group preferred PM Peak (17.27%) and off peak times (AM Off Peak = 14.95%) and (PM Off Peak = 23.97%) suggesting more flexible schedules. Finally, the low-income earners in the sample mostly travelled in the afternoon PM Off Peak (16.75%) and PM Peak (14.69%) times.

Table 6-7: Distribution of Trip Frequencies by Participant Group & Time of Day (N = 388)

Participant Group Type	**Time Period**						**Total**
	Early Morning	**AM Peak**	AM Off Peak	PM Off Peak	**PM Peak**	Evening	
Female	0.77	9.02	11.86	14.95	15.72	3.61	55.93
African American	0.26	5.93	14.43	21.91	19.59	3.35	65.47
White	0.77	4.38	5.67	3.09	6.44	2.84	23.19
Asian	0	0	0	0.26	0	0.26	0.52
Hispanic	0	0	0.77	0.52	0	0	1.29
Other	0	0.52	0	2.06	1.8	0.52	4.90
Low Income	0.26	2.84	6.7	16.75	14.69	2.58	43.82
Young (≤ 60 yrs.)	0.26	7.99	12.41	21.13	22.68	6.19	70.88
Unemployed	0	8.76	14.95	23.97	17.27	3.35	68.30

6.7.5.2 Comparison of Travel Indices by Socio-demographic Group

In this section, comparisons of the descriptive statistics are reported for the different travel measures by each participant group. The results are shown in Table 6-8. Each participant group is compared to the remaining respondents in the sample and the mean differences between them are displayed. A Satterthwaite test was performed and the corresponding t-value, p-value and significance of the mean differences were also reported. The Satterthwaite test is similar to the sample t-test but differs in the approximation method of calculating the degrees of freedom and assumes unequal variances for both populations compared. Only significant measures will be discussed.

The overlapping index (OI) was the only travel index that was both significant and had a large magnitude among the female respondents. The mean OI among females was 0.484, a value close to the median (0.5) for this index and is greater than the OI value of the male respondents by 0.156. This indicates a fairly large overlap of the road segments in the observed and shortest paths. Therefore, females seem to use shorter paths than males.

The values for the DistDI, TTDI and OI for the Unemployed group were statistically significant and less than the remaining respondents' measures. The mean DistDI value for the Unemployed participants is 5.338, a value less by 4.3 units compared to the other participants indicating that this group follows the shortest paths more than the others in the sample. Similarly, the mean TTDI value is 5.482, less than the other respondents by 4.443 units suggesting that their trip durations match up more with that of the shortest paths than the employed respondents. The mean OI magnitude was smaller for the Unemployed group showing less overlap with the road segments of the shortest paths. Generally, this participant group used shorter paths than their employed counterparts.

Those classified as young were at most 60 years old and also had three statistically significant measures: RD, observed and shortest path distances. This group had a greater mean RD (0.232) value than older adults above 60 suggesting a preference for more direct routes. The observed and shortest path distances for the walking trips were also longer for this age group by 0.283 and 0.204 miles respectively, however, the DistDI was non-significant implying that the deviance between the two route types was negligible.

Interesting results were obtained for the low-income earners. To recap, this group comprised 44% of the sample and all the travel measures were statistically different than those of the remaining respondents. All the mean values of the travel measures were larger for this group with the exception of the OI measure. This group made longer observed trips that deviated from the shortest paths, which translated into longer trip durations that were also different from that of the shortest path. Further, the routes chosen by this group were more direct (possibly due to the respondent cutting through lots) but overlapped less with the shortest paths.

Respondents with an African-American background took routes that were significantly different from those of the remaining races. Further, their routes also overlapped less with the shortest paths (Diff = -0.165). The average observed and shortest path distances for the trips were 0.283 and 0.211 miles respectively for this group however, the DistDI was non-significant implying that the distance deviance between the two route types was negligible among the different race groups.

Table 6-8: Travel Indices by Participant Group

Measure	Female vs. Male						
				Satterthwaite Test			
	N	Mean	S.D.	Diff†	t-Value	p-Value	Significance
DistDI	217	3.810	35.060	0.373	0.13	0.899	
TTDI	217	3.763	34.379	0.705	0.23	0.816	
RD	217	0.245	0.308	0.378	1.53	0.127	
OI	217	0.484	0.278	0.156	5.08	<.0001	***
Observed Trip Distance	217	0.362	0.543	-0.052	-0.95	0.344	
SP Trip Distance	217	0.241	0.271	-0.018	-0.60	0.549	
	Unemployed vs. Employed						
DistDI	265	5.338	36.046	-4.300	-1.85	0.065	*
TTDI	265	5.482	36.654	-4.443	-1.89	0.060	*
RD	265	0.492	2.587	-0.254	-1.57	0.117	
OI	265	0.578	0.307	-0.080	-2.45	0.015	**
Observed Trip Distance	265	0.362	0.599	-0.074	-1.48	0.139	
SP Trip Distance	265	0.238	0.320	-0.014	-0.52	0.606	
	Young vs. Old (> 60 yrs.)						
DistDI	275	3.241	31.536	2.519	0.8	0.4228	
TTDI	275	3.204	30.938	2.985	0.88	0.378	
RD	275	0.232	0.263	0.618	1.67	0.099	*
OI	275	0.566	0.303	-0.044	-1.28	0.202	
Observed Trip Distance	275	0.283	0.441	0.191	2.66	0.009	***
SP Trip Distance	275	0.204	0.226	0.101	2.57	0.011	**
	Low Income vs. All Other Income Groups						
DistDI	170	0.314	1.167	6.514	2.4	0.0171	**
TTDI	170	0.313	1.151	6.693	2.43	0.016	**
RD	170	0.216	0.155	0.348	1.80	0.074	*
OI	170	0.629	0.277	-0.136	-4.53	<.0001	***
Observed Trip Distance	170	0.195	0.171	0.256	5.33	<.0001	***
SP Trip Distance	170	0.173	0.155	0.107	4.01	<.0001	***
	African American vs. All Remaining Races						
DistDI	254	2.692	17.875	3.713	0.92	0.3607	
TTDI	254	2.883	20.171	3.447	0.86	0.393	
RD	254	0.503	2.649	-0.266	-1.59	0.112	
OI	254	0.610	0.309	-0.165	-5.48	<.0001	***
Observed Trip Distance	254	0.283	0.490	0.161	2.64	0.009	***
SP Trip Distance	254	0.211	0.293	0.065	2.15	0.033	**

†Mean difference between participant group type and all others

Significance: * p < .1, ** p < .05, *** p < .01

6.8 Results of Built Environment Effects

Exposure to the built environment (BE) measures was estimated over three steps: computing all the road segment-level land use percentages, calculating the proportions of which the pedestrians experienced along their routes and then finally aggregating these proportions over each trip to yield a trip-level exposure estimate. The fine-grain nature of the initial road-segment analysis was selected for two reasons: to provide more accurate exposure percentages and to ultimately present a comparison of the BE exposure variables between the observed and shortest paths. Further, there were two levels of analysis for the BE exposure effects completed. Mean percentages of the BE factors at the road segment-level were obtained then aggregated over the length of each walking trip to yield trip-level mean values for these land use effects.

As explained earlier, there were a total of 388 shortest path trips corresponding to 388 observed GPS trips. Both sets of trips originate and end at the same locations. Aggregating to the trip-level was essential to compare the built environment exposure effects experienced over the whole trip, obtain a difference in means and perform a sign test of significance. The results for the trip-level built environment exposure effects are displayed in Table 6-9.

Table 6-9 shows a list of the different land uses, their mean trip-level percentage values per route type (Mean), the standard deviation (S.D.), the number of observations (N), the difference in means between the two route types (Diff), and the results from the sign test of significance (statistic, p-value and significance).

Overall, the results indicate the difference in means of the built environment exposure variables were statistically significant as shown by the p-values obtained from the sign test. As explained earlier, the “Diff” values are the difference in means for the two route types (GPS and SP) corresponding to the relevant built environment type. This was calculated as a straight subtraction between the two mean values. Next, I used the sign test to compare this difference in means to relay whether or not the mean difference was significant between the two route types. The sign test is a non-parametric test and it is similar to the paired samples t-test but it is used when the intention is to relax the following two assumptions: the ordinal nature and the normal distribution for the values of the difference in means. This test shows the magnitude and the sign of the difference as well as the significance of the results.

The mean values of Table 6-9 were generally greater and statistically significant for the observed trips relative to the shortest path trips with the exception of the exposure to residential uses and irrigated lawns, which was found to be greater along the shortest paths. The difference in the trip-level means for the irrigated lawns and residential uses was statistically significant and smaller for the observed routes probably due to the pedestrians’ preference to walk on trails and avoid walking on personal properties of others.

These results also indicate that pedestrians after completing their walks were exposed to more: commercial uses, retail, industrial and green spaces (non-irrigated lawns and trees) along their routes in comparison to the same uses along the shortest paths for the same trips. In addition, the pedestrians also traverse more unclassified parcels and impervious land per walking trip that may include sidewalks and street medians and experience higher street connectivity along their chosen observed routes. Perhaps the greatest values for the difference in means were detected for the exposure to neighborhood businesses followed by transit stops along the observed routes compared to the shortest path trips. These two land use types may even contribute the most to the choice of the route selected if the underlying purpose of the trips was for local shopping or to connect to public transit. Further analyses however, would be necessary to confirm this hypothesis that cannot be concluded here because trip purpose was unknown and data on it would need to be collected.

Table 6-9: Trip-level Built Environment Characteristics

Variable	GPS/Observed Route			SP/Simulated Route			Sign Test			
	Mean	S.D.	N	Mean	S.D.	N	Diff[†]	Statistic (M)	p-Value	Significance
Neighborhood Business	123.8771	208.1287	388	23.8187	85.0903	388	100.05845	106.5	<.0001	***
Impervious Land Cover	11.4768	17.5978	388	6.9083	10.1376	371	4.5684	96.5	<.0001	***
Trees	1.3020	2.3690	388	0.8977	1.2412	371	0.4044	27.5	0.005	***
Irrigated Lawn	1.5030	3.7489	388	2.0639	2.7398	371	-0.5608	-76.5	<.0001	***
Non-Irrigated Lawn	0.4459	0.9382	388	0.3989	0.5693	371	0.0470	-19.5	0.0484	**
Unclassified Parcels	6.7608	10.7250	388	4.3190	6.0659	388	2.4417	67	<.0001	***
Street Intersections	9.5577	16.6572	388	7.6581	12.0501	388	1.8997	37	0.0002	***
Transit Stops Count	41.7558	79.9355	388	8.8635	25.1465	388	32.8924	133	<.0001	***
Commercial Uses	2.5629	3.7958	388	0.5926	1.5463	388	1.9703	140.5	<.0001	***
Commercial & Retail Combined	2.1211	3.3039	388	0.5325	1.4613	388	1.5886	132	<.0001	***
Residential Uses	4.4337	10.1797	388	4.9184	6.7985	388	-0.4847	-58	<.0001	***
Industrial Uses	0.2990	1.4186	388	0.1726	1.0543	388	0.1264	24	0.0034	***

[†]Mean difference between GPS and SP

Significance: * p < .1, ** p < .05, *** p < .01

6.9 Policy Implications

Results from the AM Peak time indicate that the pedestrians opted for more circuitous routes that deviate more from the shortest paths. This might have been because the respondents were having morning strolls or taking longer routes to avoid highly impacted streets with motorized traffic. Policy measures that include traffic calming instruments, pedestrian buffers or medians may be beneficial to attract more pedestrians that otherwise would be intimidated by greater vehicular speeds especially in the morning peak time.

The time of day analyses for the travel indices provided some insights to pedestrian's route choice. The values for the travel indices for the Early Morning period were the lowest for DistDI, TTDI, RD and the highest for OI suggesting that the pedestrians may prefer using shortest routes for this time period compared to other times. Although the Early Morning trips were only 1% of all the observed walking trips, the time period they fall under render it a level of sensitivity. The tendency to use more of the shortest paths in this time period seem to suggest the participant's preference for functionality over leisure walks or may even hint at other underlying concerns for safety. Smart growth measures may target this travel time by ensuring that frequently traveled routes are: well-lit, have accessible emergency phone booths, absent of any obstructions, cleaner and less cluttered so that these trails may instill a sense of security among the pedestrian travelers especially for such early trips. However, because of the small number of trips in this time period, further investigation would be required to fully understand pedestrian travel behavior in this time period.

The analyses by built environment characteristics also yielded some interesting insights of the impacts of the various land uses on pedestrian route choice. The observed routes provided greater exposures to commercial centers, local business establishments and tree density, implying the general preference of the pedestrians to walk along these destinations. Therefore, further developments of these environmental correlates would be highly recommended especially along more connected streets with more transit stops which have also been proven to attract more pedestrians (Brown & Werner, 2008; Werner et al., 2010).

6.10 Limitations and Future Research

The analyses in this chapter offered an objective methodology of comparing GPS-tracked routes to GIS simulated shortest paths. Although the data collection method outlined in this chapter more accurately represent the actual walking routes completed, there were still some limitations that existed.

In some instances of the network analyst shortest path creation, the O-D stop locations did not correctly snap to the exact location of the actual GPS point. A correction was performed in which the stops were moved manually to the nearest node on the road network. Although this injects some error in the calculation of the optimum route, the error was reduced by the manual correction. Future research may utilize a more comprehensive road network that may lessen this offset error.

In other instances, the pedestrians used walkways or trails that were not represented on the road network and therefore affected the algorithm for the shortest path. These rendered the shortest path to be actually longer than the observed routes. The road network GIS shapefile was not populated with these trails and mainly caters to vehicular travel. Therefore, it is recommended that the street network files be expanded to include pedestrian trails, pathways and underground passages that were not originally defined. This may include the use of aerial imagery such as Google Street View or Pictometry or in some cases it may even require extensive field visits and on-site surveys. Doing so provides a better representation of the pedestrian paths, which can be simulated more accurately by network analyst.

Results indicate that a very small proportion (0.24%) of the roadways traversed by the respondents were primary roads. The only primary road in the Expo sample is the 10-freeway where pedestrians are prohibited from using. Further investigation should be performed to identify whether the GPS points were erroneously associated to this freeway or if indeed this was a legitimate walk trip along the freeway. More likely, the respondent may have been walking in the vicinity of the freeway, such as under the freeway overpass bridge and the GPS points were instead matched to the freeway. In the case of a legitimate walk along the freeway, a possible scenario where this situation may arise is if the previous set of GPS points indicated a vehicular mode where the respondent was driving and the car breaks down. In this situation, the driver may become a pedestrian possibly pacing or pushing his inoperative vehicle to the freeway shoulder to await roadside assistance. However, since only the walking trips were extracted, further research is required to identify the before and after GPS patterns to understand this situation and/or to correct it.

Another limitation was due to trips occurring at the edges of the road network. This occurred when clipping the road network file to include the roads that lie within the half-mile radius from the respondents' home locations, the shortest paths for the walking trips at the edges were affected and altered to reflect only the roads with complete information. Originally, there were a total of 395 trips, seven (1.8%) of which were at the edges. A decision to delete these trips was made to reduce the bias induced from the incomplete road network. Thus, the resulting number of walking trips decreased to 388 for the 62 respondents.

Distance was the only parameter selected as impedance for the shortest paths. Future pedestrian analyses may include streetscape obstructions and social barriers in the calculation of the shortest paths. The streetscape barriers include disconnected walkways, unmaintained or damaged sidewalks, great street inclines, and physical disorder. Social barriers may include indifference towards TPA; gang enforced boundaries, and high crime rates. Attitudes and perceptions of the respondents were collected for the Expo study but were beyond the scope of this dissertation. Previous research that utilized the Expo data applied a Perception-Intention-Adaptation framework to isolate the most significant attitudinal factors that affect public transit use. After controlling for built environment and transit access variables; the authors found that the most influential factors were: attitudes toward transit and safety apprehensions (Spears, Houston, & Boarnet, 2013). Future research may benefit from including safety concerns and crime perceptions.

Lastly, due to the low response rate (1%) for the first phase of the Expo study, the pedestrians in the sample were only 62 participants. This limits the generalizability of results to other areas. However, results provide some insights regarding the travel habits of specific sub-groups such as the low-income households as they were well represented in the sample and their proportions were equivalent to the Census population percentages. The methods in this chapter were intended as a simplified version of comparing actual routes to simulated shortest paths and offer a variety of different benchmarks on which this comparison was performed. Further research may benefit from these analyses and may apply them in other geographic areas for larger sample sizes.

6.11 Conclusion

The methodologies outlined in this research provide a practical approach of representing actual pedestrian routes. This approach utilizes less computation-intensive methods than traditional pedestrian route choice modeling techniques. GPS travel data for a sample of sixty-two respondents residing in Los Angeles, CA was assembled to form observed or actual routes that were consequently compared to GIS simulated shortest paths. A distinction should be made between subjective shortest paths; that depend on the individual's perceptions and mental maps and the shortest roadway paths created by GIS. All the simulated routes in this research were objectively calculated shortest roadway paths.

Comparisons between each set of the observed and shortest paths were performed through the contrasting of different travel indices, socio-demographic group types, time of day analyses and exposure to built environment factors.

The majority of the pedestrian sample selected walking trips that were shorter in distance and in duration but were still observed to be longer than the respective shortest path. These typically were less than a mile long and lasted twenty minutes or less. In fact, the average observed walking trip was 0.34 miles long and the average travel time was 8.7 minutes long. These were much longer than the corresponding mean shortest path distance (0.23 miles) and travel times (6.34 minutes) thus confirming the hypothesis that pedestrians deviate from the shortest path. In addition, results from the distance deviation and segment deviation indices showed that route deviations from the shortest path increased with greater distances especially beyond 2.5-mile threshold. Coincidentally, this threshold was very close to the 2.4-mile mark where trips appeared to be more circuitous and these represented 25% of the overall trips.

The analyses of the travel indices by participant socio-demographic characteristics provided some important insights. Routes taken by females appeared to overlap more with shortest paths than males. Unemployed respondents generally opted for shorter paths than those who were employed.

Travelers who were 60 years or younger in the sample preferred the most direct routes but did not differ from older adults in their deviations from the shortest path. This directness may be because this age group encompasses working adults (under 65 years) and therefore they may have more time constraints in terms of work and other related obligations.

Further, the African-Americans who comprised the majority (65%) of the sample, took routes that overlapped the least with the shortest paths than the remaining races. However, there were no significant differences in the deviations by distance and travel time.

Interestingly, the low-income earners who made up 44% of the sample were the only income group that was significantly different in their travel preferences than the remaining groups. This group made longer distance trips that lasted longer and overlapped the least with the shortest paths. Overall, their routes were quite different and deviated from the corresponding shortest paths. Nevertheless, they opted to take more direct routes than other income groups. This point raises an interesting inference that the shortest paths may not necessarily be the most direct ones. This can be confirmed by the many instances where the observed routes were found to be more direct and in fact shorter in distance than the shortest paths such as in pedestrian walkways or trails.

The time of day analyses confirmed the deviation of the observed routes from the shortest paths. With the exception of the Evening time, all the remaining time periods depicted average distances along the GPS-tracked routes that were much longer than that of the shortest path. The largest deviance in distance occurred in the AM Peak time, which may have been due to morning strolls prior to leaving to work and which do not necessarily follow the shortest path. Moreover, the Early Morning time period was the most consistent across all travel indices. This period witnessed the most direct routes, had the least deviances in distance and travel time, and produced the most overlapping segments with the shortest paths. However, only a small percentage of trips occurred at this time period which prevents any tangible conclusions.

The participant travel frequencies by time of day revealed more aspects of travel time period preferences. Generally, most of the walks took place beyond the AM Off Peak threshold. Adults under 60 years preferred the late morning and afternoon time periods until the PM peak time. Similarly, unemployed participants preferred PM Peak and both off peak times suggesting more flexible mornings. Finally, low-income respondents preferred walks in the afternoons.

Further, significant differences from the shortest paths were also detected after examining the built environment features along the observed routes. The actual routes offered the pedestrians exposure to a multitude of different land uses. These routes had larger percentages of commercial centers, local businesses and green spaces. They also had greater connectivity levels and were associated with more transit stops.

CHAPTER SEVEN. REFERENCES

Abdel-Aty, M. a., Kitamura, R., & Jovanis, P. P. (1997). Using stated preference data for studying the effect of advanced traffic information on drivers' route choice. *Transportation Research Part C: Emerging Technologies*, *5*(1), 39–50. doi:10.1016/S0968-090X(96)00023-X

Adams, M. a, Sallis, J. F., Kerr, J., Conway, T. L., Saelens, B. E., Frank, L. D., … Cain, K. L. (2011). Neighborhood environment profiles related to physical activity and weight status: a latent profile analysis. *Preventive Medicine*, *52*(5), 326–31. doi:10.1016/j.ypmed.2011.02.020

Almanza, E., Jerrett, M., Dunton, G., Seto, E., & Pentz, M. A. (2012). A study of community design, greenness, and physical activity in children using satellite, GPS and accelerometer data. *Health & Place*, *18*(1), 46–54. doi:10.1016/j.healthplace.2011.09.003

Badland, H. M., Duncan, M. J., Oliver, M., Duncan, J. S., & Mavoa, S. (2010a). Examining commute routes: applications of GIS and GPS technology. *Environmental Health and Preventive Medicine*, *15*(5), 327–30. doi:10.1007/s12199-010-0138-1

Badland, H. M., Duncan, M. J., Oliver, M., Duncan, J. S., & Mavoa, S. (2010b). Examining commute routes: applications of GIS and GPS technology. *Environmental Health and Preventive Medicine*, *15*(5), 327–30. doi:10.1007/s12199-010-0138-1

Badland, H., White, M., Macaulay, G., Eagleson, S., Mavoa, S., Pettit, C., & Giles-Corti, B. (2013). Using simple agent-based modeling to inform and enhance neighborhood walkability. *International Journal of Health Geographics*, *12*, 58. doi:10.1186/1476-072X-12-58

Beckx, C., Panis, L. I., Janssens, D., & Wets, G. (2010). Applying activity-travel data for the assessment of vehicle exhaust emissions: Application of a GPS-enhanced data collection tool. *Transportation Research Part D: Transport and Environment*, *15*(2), 117–122. doi:10.1016/j.trd.2009.10.004

Bell, B. A., Ene, M., Smiley, W., & Schoeneberger, J. A. (2013). A Multilevel Model Primer Using SAS® PROC MIXED. In *SAS Global Forum 2013* (pp. 1–19).

Besser, L. M., & Dannenberg, A. L. (2005). Walking to Public Transit Steps to Help Meet Physical Activity Recommendations. *American Journal of Preventive Medicine*, *29*(4), 273–280. doi:10.1016/j.ampre.2005.06.010

Boarnet, M. G., Forsyth, a., Day, K., & Oakes, J. M. (2011). The Street Level Built Environment and Physical Activity and Walking: Results of a Predictive Validity Study for the Irvine Minnesota Inventory. *Environment and Behavior*, *43*(6), 735–775. doi:10.1177/0013916510379760

Boarnet, M. G., Joh, K., Siembab, W., & Fulton, W. (2010). Retrofitting the Suburbs to Increase Walking: Evidence from a Land-use-Travel Study. *Urban Studies*, *48*(1), 129–159. doi:10.1177/0042098010364859

Boarnet, M., Houston, D., & Spears, S. (2013). *The Exposition Light Rail Line Study: "Before-After" Opening Travel Impacts and New Resident Sample Preliminary Analysis. Working Paper.* Retrieved from http://community-wealth.org/sites/clone.community-wealth.org/files/downloads/paper-boarnet.pdf

Boone-Heinonen, J., Gordon-Larsen, P., Guilkey, D. K., Jacobs, D. R., & Popkin, B. M. (2011). Environment and Physical Activity Dynamics: The Role of Residential Self-selection. *Psychology of Sport and Exercise*, *12*(1), 54–60. doi:10.1016/j.psychsport.2009.09.003

Brown, B. B., & Werner, C. M. (2007). A new rail stop: tracking moderate physical activity bouts and ridership. *American Journal of Preventive Medicine*, *33*(4), 306–309. doi:10.1016/j.amepre.2007.06.002

Brown, B. B., & Werner, C. M. (2008). Before and After a New Light Rail Stop: Resident Attitudes, Travel Behavior, and Obesity. *Journal of the American Planning Association*, *75*(1), 5–12. doi:10.1080/01944360802458013

Buliung, R. N., Larsen, K., Faulkner, G. E. J., & Stone, M. R. (2013). The "path" not taken: exploring structural differences in mapped- versus shortest-network-path school travel routes. *American Journal of Public Health*, *103*(9), 1589–96. doi:10.2105/AJPH.2012.301172

Campoli, J. (2012). *Made for Walking*. Cambridge, Massachusetts. Retrieved from https://www.lincolninst.edu/pubs/dl/2150_1509_MFW_Web_Chapter.pdf

Canepa, B. (2007). Bursting the Bubble: Determining the Transit-Oriented Development's Walkable Limits. *Transportation Research Record*, *1992*(1), 28–34. doi:10.3141/1992-04

Carr, L. J. ., Dunsiger, S. I., & Marcus, B. H. (2011). Validation of Walk Score for estimating access to walkable amenities. *British Journal of Sports Medicine*, *45*(14), 1144–8. doi:10.1136/bjsm.2009.069609

Centers for Disease Control and Prevention (CDC). (n.d.). *Physical Activity and Health*. Retrieved from www.cdc.gov/nccdphp/sgr/pdf/execsumm.pdf?

Centers for Disease Control and Prevention (CDC). (2014). How much physical activity do adults need? Retrieved from http://www.cdc.gov/physicalactivity/everyone/guidelines/adults.html

Cerin, E., Conway, T. L., Saelens, B. E., Frank, L. D., & Sallis, J. F. (2009). Cross-validation of the factorial structure of the Neighborhood Environment Walkability Scale (NEWS) and its abbreviated form (NEWS-A). *The International Journal of Behavioral Nutrition and Physical Activity*, *6*, 32. doi:10.1186/1479-5868-6-32

Cerin, E., Leslie, E., du Toit, L., Owen, N., & Frank, L. D. (2007). Destinations that matter: associations with walking for transport. *Health & Place*, *13*(3), 713–24. doi:10.1016/j.healthplace.2006.11.002

Cervero, R. B. (1996). Urban Design Issues Related to Transporation Modes, Designs and Services for Neo-traditional Developments. In *Urban Design, Telecommuting and Travel Forecasting Conference, Williamsburg, VA.*

Chaix, B., Méline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C., … Kestens, Y. (2013). GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference? *Health & Place*, *21*, 46–51. doi:10.1016/j.healthplace.2013.01.003

Choi, S., Wang, P., Delgado, E., & Ryu, S. H. (2007). Monitoring Small Area Growth with GIS : An Application to the City of Los Angeles. In *ESRI International User Conference.*

Chung, E.-H., & Shalaby, A. (2005). A Trip Reconstruction Tool for GPS-based Personal Travel Surveys. *Transportation Planning and Technology*, *28*(5), 381–401. doi:10.1080/03081060500322599

Clark, T. S., & Linzer, D. A. (2013). *Should I Use Fixed or Random Effects ?* (pp. 1–39). Retrieved from http://userwww.service.emory.edu/~tclark7/randomeffects.pdf

Clarke, P., Ailshire, J., Melendez, R., Bader, M., & Morenoff, J. (2010). Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. *Health & Place*, *16*(6), 1224–9. doi:10.1016/j.healthplace.2010.08.007

Condon, P. M., Cavens, D., & Miller, N. (2009). *Urban Planning Tools for Climate Change Mitigation*. Cambridge, MA: Lincoln Institute of Land Policy.

Coombes, E., Jones, A. P., & Hillsdon, M. (2010). The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social Science & Medicine (1982)*, *70*(6), 816–22. doi:10.1016/j.socscimed.2009.11.020

Cooper, A. R., Page, A. S., Wheeler, B. W., Griew, P., Davis, L., Hillsdon, M., & Jago, R. (2010). Mapping the walk to school using accelerometry combined with a global positioning system. *American Journal of Preventive Medicine*, *38*(2), 178–83. doi:10.1016/j.amepre.2009.10.036

Dannenberg, A. L., Jackson, R. J., Frumkin, H., Schieber, R. a, Pratt, M., Kochtitzky, C., & Tilson, H. H. (2003). The impact of community design and land-use choices on public health: a scientific research agenda. *American Journal of Public Health*, *93*(9), 1500–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1448000&tool=pmcentrez&rendertype=abstract

De Nazelle, A., Nieuwenhuijsen, M. J., Antó, J. M., Brauer, M., Briggs, D., Braun-Fahrlander, C., … Lebret, E. (2011). Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environment International*, *37*(4), 766–77. doi:10.1016/j.envint.2011.02.003

De Nazelle, A., Rodríguez, D. a, & Crawford-Brown, D. (2009). The built environment and health: impacts of pedestrian-friendly designs on air pollution exposure. *The Science of the Total Environment*, *407*(8), 2525–35. doi:10.1016/j.scitotenv.2009.01.006

Dill, J. (2004). Measuring Network Connectivity for Bicycling and Walking. In *Joint Congress of ACSP-AESOP*.

Ding, D., Sallis, J. F., Conway, T. L., Saelens, B. E., Frank, L. D., Cain, K. L., & Slymen, D. J. (2012). Interactive effects of built environment and psychosocial attributes on physical activity: a test of ecological models. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*, *44*(3), 365–74. doi:10.1007/s12160-012-9394-1

Duncan, D. T., Aldstadt, J., Whalen, J., & Melly, S. J. (2012). Validation of Walk Scores and Transit Scores for estimating neighborhood walkability and transit availability: a small-area analysis. *GeoJournal*. doi:10.1007/s10708-011-9444-4

Duncan, M. J., Badland, H. M., & Mummery, W. K. (2009). Applying GPS to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport / Sports Medicine Australia*, *12*(5), 549–56. doi:10.1016/j.jsams.2008.10.010

Dunton, G. F., Liao, Y., Almanza, E., Jerrett, M., Spruijt-Metz, D., & Pentz, M. A. (2013). Locations of joint physical activity in parent-child pairs based on accelerometer and GPS monitoring. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*, *45*(Suppl 1), S162–S172. doi:10.1007/s12160-012-9417-y

Durand, C. P., Andalib, M., Dunton, G. F., Wolch, J., & Pentz, M. a. (2011). A systematic review of built environment factors related to physical activity and obesity risk: implications for smart growth urban planning. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *12*(5), e173–82. doi:10.1111/j.1467-789X.2010.00826.x

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods*, *12*(2), 121–38. doi:10.1037/1082-989X.12.2.121

ESRI. (2010). Network Analyst Tutorial. Retrieved from http://help.arcgis.com/en/arcgisdesktop/10.0/pdf/network-analyst-tutorial.pdf

Evenson, K. R., & Wen, F. (2010). National trends in self-reported physical activity and sedentary behaviors among pregnant women: NHANES 1999-2006. *Preventive Medicine*, *50*(3), 123–8. doi:10.1016/j.ypmed.2009.12.015

Ewing, R. (1999). *PEDESTRIAN- AND TRANSIT-FRIENDLY DESIGN : A Primer for Smart Growth* (Smart Grow.). American Planning Association. Retrieved from http://www.epa.gov/smartgrowth/pdf/ptfd_primer.pdf

Foltête, J.-C., & Piombini, A. (2010). Deviations in pedestrian itineraries in urban areas: a method to assess the role of environmental factors. *Environment and Planning B: Planning and Design*, *37*(4), 723–739. doi:10.1068/b35015

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation*, *35*(1), 37–54. doi:10.1007/s11116-007-9136-6

Frank, L. D., Kerr, J., Sallis, J. F., Miles, R., & Chapman, J. (2008). A hierarchy of sociodemographic and environmental correlates of walking and obesity. *Preventive Medicine*, *47*(2), 172–8. doi:10.1016/j.ypmed.2008.04.004

Frank, L. D., Schmid, T. L., Sallis, J. F., Chapman, J., & Saelens, B. E. (2005). Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. *American Journal of Preventive Medicine*, *28*(2 Suppl 2), 117–25. doi:10.1016/j.amepre.2004.11.001

Freedson, P. S., Melanson, E., & Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine and Science in Sports and Exercise*, *30*(5), 777–781. Retrieved from http://cat.inist.fr/?aModele=afficheN&cpsidt=2252600

Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, *43*(10), 984–994. doi:10.1016/j.trb.2009.03.001

Frumkin, H. (2002). Urban sprawl and public health. *Public Health Reports (Washington, D.C. : 1974)*, *117*(3), 201–17. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1497432&tool=pmcentrez&rendertype=abstract

Giles-Corti, B., Broomhall, M. H., Knuiman, M., Collins, C., Douglas, K., Ng, K., … Donovan, R. J. (2005). Increasing walking: how important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine*, *28*(2 Suppl 2), 169–76. doi:10.1016/j.amepre.2004.10.018

Giles-Corti, B., Timperio, A., Bull, F., & Pikora, T. (2005). Understanding Physical Activity Environmental Correlates : Increased Specificity for Ecological. *Exercise and Sport Sciences Reviews*, *33*(4), 175–181.

Goldstein, H. (1999). *Harvey Goldstein: Multilevel Statistical Models* (Education,., pp. 1–163). London. Retrieved from http://www.ats.ucla.edu/stat/examples/msm_goldstein/goldstein.pdf

Goulias, K. G. (2002). Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures. *Transportation*, *29*, 31–48.

Greenwald, M. J., & Boarnet, M. G. (2001). Built Environment as Determinant of Walking Behavior Analyzing Nonwork Pedestrian Travel in Portland , Oregon. *Transportation Research Record*, *1780*(01), 33–42.

Groundspeak. (2014). Calculating Distance Between Two Sets of Coordinates. Retrieved from http://support.groundspeak.com/index.php?pg=kb.page&id=211

Guerra, E. & Cervero, R. (2013). Is a half-mile circle the right standard for TODs? *Access*, 17–22.

Guhathakurta, S., Zhang, G., & Panguluru, M. K. (2013). Walk Route: A New Methodology to Find the Optimal Walking Route in the City of Atlanta. In S. Geertman, F. Toppen, & J. Stillwell (Eds.), *Planning Support Systems for Sustainable Urban Development* (Vol. 195, pp. 309–325). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-37533-0

Guo, Z. (2009). Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston. *Transportation Research Part D: Transport and Environment*, *14*(5), 343–352. doi:10.1016/j.trd.2009.03.007

Guo, Z. (2010). *Causality vs . Correlation : Rethinking Research Design in the Case of Pedestrian Environments and Walking*. *NYU Wagner Research Paper, (2010-11)*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1573754

Guo, Z., & Ferreira Jr, J. (2008). Pedestrian environments, transit path choice, and transfer penalties: understanding land-use impacts on transit travel. *Environment and Planning B: Planning and Design*, *35*(3), 461–479. doi:10.1068/b33074

Guo, Z., & Loo, B. P. Y. (2013). Pedestrian environment and route choice: evidence from New York City and Hong Kong. *Journal of Transport Geography*, *28*, 124–136. doi:10.1016/j.jtrangeo.2012.11.013

Gwet, K. (2002). Kappa Statistic is not Satisfactory for Assessing the Extent of Agreement Between Raters. In *Statistical Methods for Inter-Rater Reliability Assessment* (pp. 1–5).

Handy, S. (1993). Regional Versus Local Accessibility : Implications for Nonwork Travel, (234), 58–66.

Handy, S., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: views from urban planning. *American Journal of Preventive Medicine*, *23*(2 Suppl), 64–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12133739

Hankey, S., Marshall, J. D., & Brauer, M. (2012). Inactivity , Air Pollution , and Ischemic Heart Disease Mortality. *Environmental Health Perspectives*, *120*(2), 247–253.

Hoehner, C. M., Brennan Ramirez, L. K., Elliott, M. B., Handy, S. L., & Brownson, R. C. (2005). Perceived and objective environmental measures and physical activity among urban adults. *American Journal of Preventive Medicine*, *28*(2 Suppl 2), 105–16. doi:10.1016/j.amepre.2004.10.023

Horowitz, A. J. (1982). A Comparison of Socioeconomic and Structural Determinants Trip Tour Length. *Papers in Regional Science*, *50*(1), 185–195.

Houston, D. (2014). Implications of the modifiable areal unit problem for assessing built environment correlates of moderate and vigorous physical activity. *Applied Geography*, *50*, 40–47. doi:10.1016/j.apgeog.2014.02.008

Houston, D., Wu, J., Yang, D., & Jaimes, G. (2013). Particle-bound polycyclic aromatic hydrocarbon concentrations in transportation microenvironments. *Atmospheric Environment*, *71*, 148–157. doi:10.1016/j.atmosenv.2013.02.002

Howard, C., & Burns, E. K. (2001). Cycling to Work in Phoenix Route Choice , Travel Behavior , and Commuter Characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, *1773*(01), 39–46.

Howard, E. J., Kang, B., Hurvitz, P., Moudon, A. V., & Saelens, B. E. (2014). Modeling GPS-Based Walking Activity and Its Association with Objectively Measured Built Environment. In *Transportation Research Board 93rd Annual Meeting (No. 14-5591)*.

Institute for Local Government. (2014). The Basics of SB 375. Retrieved from http://www.ca-ilg.org/post/basics-sb-375

Jan, O., Horowitz, A. J., & Peng, Z. (2000). Using Global Positioning System Data to Understand Variations in Path Choice. *Transportation Research Record: Journal of the Transportation Research Board*, *1725*(1), 37–44.

Jones, A. P., Coombes, E. G., Griffin, S. J., & van Sluijs, E. M. (2009). Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems. *The International Journal of Behavioral Nutrition and Physical Activity*, *6*, 42. doi:10.1186/1479-5868-6-42

Kaczynski, A. T., & Henderson, K. a. (2007). Environmental Correlates of Physical Activity: A Review of Evidence about Parks and Recreation. *Leisure Sciences*, *29*(4), 315–354. doi:10.1080/01490400701394865

Kaplan, S., & Prato, C. G. (2012). Closing the gap between behavior and models in route choice: The role of spatiotemporal constraints and latent traits in choice set formation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(1), 9–24. doi:10.1016/j.trf.2011.11.001

Khattak, A. J., Poludoropoulou, A. & Ben-Akiva, M. (1996). Modeling Revealed and Stated Pretrip Travel Response to Advanced Traveler. *Transportation Research Record: Journal of the Transportation Research Board*, *1537*, 46–54.

Kneeshaw-price, S., Saelens, B. E., Sallis, J. F., Glanz, K., Frank, L. D., Kerr, J., … Cain, K. L. (2013). Children ' s Objective Physical Activity by Location : Why the Neighborhood Matters. *Pediatric Exercise Science*, *25*, 468–486.

Krenn, P. J., Titze, S., Oja, P., Jones, A., & Ogilvie, D. (2011). Use of global positioning systems to study physical activity and the environment: a systematic review. *American Journal of Preventive Medicine*, *41*(5), 508–15. doi:10.1016/j.amepre.2011.06.046

Lachapelle, U., Frank, L., Saelens, B. E., Sallis, J. F., & Conway, T. L. (2011). Commuting by public transit and physical activity: where you live, where you work, and how you get there. *Journal of Physical Activity & Health*, *8 Suppl 1*(Suppl 1), S72–82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21350265

Lachowycz, K., & Jones, a P. (2011). Greenspace and obesity: a systematic review of the evidence. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *12*(5), e183–9. doi:10.1111/j.1467-789X.2010.00827.x

Lascala, E. a, Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident; Analysis and Prevention*, *32*(5), 651–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10908137

Leal, C., & Chaix, B. (2011). The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *12*(3), 217–30. doi:10.1111/j.1467-789X.2010.00726.x

Lee, C. (2004). Physical Activity and Environment Research in the Health Field: Implications for Urban and Transportation Planning Practice and Research. *Journal of Planning Literature*, *19*(2), 147–181. doi:10.1177/0885412204267680

Lee, C., & Moudon, A. V. (2006a). Correlates of Walking for Transportation or Recreation Purposes, *Journal of Physical Activity & Health*, *3, S*77–S98.

Lee, C., & Moudon, A. V. (2006b). The 3Ds+R: Quantifying land use and urban form correlates of walking. *Transportation Research Part D: Transport and Environment*, *11*(3), 204–215. doi:10.1016/j.trd.2006.02.003

Leitzmann, M. F., Park, Y., Blair, A., Ballard-Barbash, R., Mouw, T., Hollenbeck, A. R., & Schatzkin, A. (2007). Physical activity recommendations and decreased risk of mortality. *Archives of Internal Medicine*, *167*(22), 2453–60. doi:10.1001/archinte.167.22.2453

Li, H., Guensler, R., & Ogle, J. (2005). Comparing Women's and Men's Morning Commute Trip Chaining in Atlanta, Georgia, by Using Instrumented Vehicle Activity Data. In *Research on Women ' s Issues in Transportation* (pp. 14–20).

Lin, L., & Moudon, A. V. (2010). Objective versus subjective measures of the built environment, which are most effective in capturing associations with walking? *Health & Place*, *16*(2), 339–48. doi:10.1016/j.healthplace.2009.11.002

Litman, T. (2012). *Short and Sweet: Analysis of shorter trips using National Personal Travel Survey Data*. Retrieved from http://www.vtpi.org/short_sweet.pdf

McPherson, E. G., Simpson, J. R., Xiao, Q., & Wu, C. (2011). Million trees Los Angeles canopy cover and benefit assessment. *Landscape and Urban Planning*, *99*(1), 40–50. doi:10.1016/j.landurbplan.2010.08.011

Metro. (2014). Metro Expo Line. Retrieved from http://www.metro.net/projects/exposition/

Mitra, R., & Buliung, R. N. (2012). Built environment correlates of active school transportation: neighborhood and the modifiable areal unit problem. *Journal of Transport Geography*, *20*(1), 51–61. doi:10.1016/j.jtrangeo.2011.07.009

Morency, C., Trépanier, M., & Demers, M. (2011). Walking to transit: An unexpected source of physical activity. *Transport Policy*, *18*(6), 800–806. doi:10.1016/j.tranpol.2011.03.010

Moudon, A. V., & Lee, C. (2003). Walking and bicycling: an evaluation of environmental audit instruments. *American Journal of Health Promotion : AJHP*, *18*(1), 21–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13677960

Oliver, L., Schuurman, N., & Hall, A. W. (2007). Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *International Journal of Health Geographics*, *6*, 41. doi:10.1186/1476-072X-6-41

Oliver, M., Badland, H., Mavoa, S., Duncan, M. J., & Duncan, S. (2010). Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. *Journal of Physical Activity & Health*, *7*(1), 102–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20231761

Owen, N., Cerin, E., Leslie, E., duToit, L., Coffee, N., Frank, L. D., … Sallis, J. F. (2007). Neighborhood walkability and the walking behavior of Australian adults. *American Journal of Preventive Medicine*, *33*(5), 387–95. doi:10.1016/j.amepre.2007.07.025

Papinski, D., & Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. *Journal of Transport Geography*, *19*(3), 434–442. doi:10.1016/j.jtrangeo.2010.09.009

Prince, S. a, Kristjansson, E. a, Russell, K., Billette, J.-M., Sawada, M., Ali, A., … Prud'homme, D. (2011). A multilevel analysis of neighbourhood built and social environments and adult self-reported physical activity and body mass index in Ottawa, Canada. *International Journal of Environmental Research and Public Health*, *8*(10), 3953–78. doi:10.3390/ijerph8103953

Quigg, R., Gray, A., Reeder, A. I., Holt, A., & Waters, D. L. (2010). Using accelerometers and GPS units to identify the proportion of daily physical activity located in parks with playgrounds in New Zealand children. *Preventive Medicine*, *50*(5-6), 235–40. doi:10.1016/j.ypmed.2010.02.002

Rainham, D. G., Bates, C. J., Blanchard, C. M., Dummer, T. J., Kirk, S. F., & Shearer, C. L. (2012). Spatial classification of youth physical activity patterns. *American Journal of Preventive Medicine*, *42*(5), e87–96. doi:10.1016/j.amepre.2012.02.011

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics*, *10*(11), 805–817. doi:10.1007/BF00383696

Rissel, C., Curac, N., Greenaway, M., & Bauman, A. (2012). Physical activity associated with public transport use--a review and modelling of potential benefits. *International Journal of Environmental Research and Public Health*, *9*(7), 2454–78. doi:10.3390/ijerph9072454

Rodríguez, D. A, Cho, G.-H., Evenson, K. R., Conway, T. L., Cohen, D., Ghosh-Dastidar, B., … Lytle, L. a. (2012). Out and about: association of the built environment with physical activity behaviors of adolescent females. *Health & Place*, *18*(1), 55–62. doi:10.1016/j.healthplace.2011.08.020

Rodriguez, D. A., Brown, A. L., & Troped, P. J. (2005). Portable Global Positioning Units to Complement Accelerometry-Based Physical Activity Monitors. *Medicine & Science in Sports & Exercise*, *37*(Supplement), S572–S581. doi:10.1249/01.mss.0000185297.72328.ce

Rodriguez, D. A., Khattak, A. J., & Evenson, K. R. (2007). Can new urbanism encourage physical activity: Comparing a new urbanist neighborhood with conventional suburbs, (April 2012), 37–41.

Rodriguez, D. A., Merlin, L., Prato, C. G., Conway, T. L., Cohen, D., Elder, J. P., … Veblen-Mortenson, S. (2014). Influence of the Built Environment on Pedestrian Route Choices of Adolescent Girls. *Environment and Behavior*, 1–36. doi:10.1177/0013916513520004

Saarloos, D., Kim, J.-E., & Timmermans, H. (2009). The built environment and health: introducing individual space-time behavior. *International Journal of Environmental Research and Public Health*, *6*(6), 1724–43. doi:10.3390/ijerph6061724

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*, *25*(2), 80–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12704009

Sallis, J. F., Cervero, R. B., Ascher, W., Henderson, K. a, Kraft, M. K., & Kerr, J. (2006). An ecological approach to creating active living communities. *Annual Review of Public Health*, *27*, 297–322. doi:10.1146/annurev.publhealth.27.021405.102100

SCAG. (2014). Regional Tranportation Plan 2012-2035 Sustainable Communities Strategy Towards a Sustainable Future. Retrieved from http://rtpscs.scag.ca.gov/Pages/Draft-2012-2035-RTP-SCS.aspx

Schlossberg, M., & Brown, N. (2004). Comparing Transit-Oriented Development Sites by Walkability Indicators. *Transportation Research Record*, *1887*(1), 34–42. doi:10.3141/1887-05

Schuessler, N., & Axhausen, K. W. (2009). Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, *2105*(-1), 28–36. doi:10.3141/2105-04

Spears, S., Houston, D., & Boarnet, M. G. (2013). Illuminating the unseen in transit use: A framework for examining the effect of attitudes and perceptions on travel behavior. *Transportation Research Part A: Policy and Practice*, *58*, 40–53. doi:10.1016/j.tra.2013.10.011

Spissu, E., Meloni, I., & Sanjust, B. (2011). Behavioral Analysis of Choice of Daily Route with Data from Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board*, *2230*(-1), 96–103. doi:10.3141/2230-11

Stigell, E., & Schantz, P. (2011). Methods for determining route distances in active commuting – Their validity and reproducibility. *Journal of Transport Geography*, *19*(4), 563–574. doi:10.1016/j.jtrangeo.2010.06.006

Stinson, M. A., & Bhat, C. R. (2003). Analysis Using a Stated Preference Survey. *Transportation Research Record: Journal of the Transportation Research Board*, 1828(1), 107-115.

Sugiyama, T., Neuhaus, M., Cole, R., Giles-Corti, B., & Owen, N. (2012). Destination and route attributes associated with adults' walking: a review. *Medicine and Science in Sports and Exercise*, *44*(7), 1275–86. doi:10.1249/MSS.0b013e318247d286

Sugiyama, T., Neuhaus, M., & Owen, N. (2012). Sustainable Environmental Design in Architecture, *56*. doi:10.1007/978-1-4419-0745-5

Suminski, R. R., Wasserman, J. a., Mayfield, C. a., Kheyfets, a., & Norman, J. (2014). Walking During Leisure-Time in Relation to Perceived Neighborhoods. *Environment and Behavior*. doi:10.1177/0013916513520605

Sundquist, K., Eriksson, U., Kawakami, N., Skog, L., Ohlsson, H., & Arvidsson, D. (2011). Neighborhood walkability, physical activity, and walking behavior: the Swedish Neighborhood and Physical Activity (SNAP) study. *Social Science & Medicine (1982)*, *72*(8), 1266–73. doi:10.1016/j.socscimed.2011.03.004

Troped, P. J., Oliveira, M. S., Matthews, C. E., Cromley, E. K., Melly, S. J., & Craig, B. a. (2008). Prediction of activity mode with global positioning system and accelerometer data. *Medicine and Science in Sports and Exercise*, *40*(5), 972–8. doi:10.1249/MSS.0b013e318164c407

Troped, P. J., Wilson, J. S., Matthews, C. E., Cromley, E. K., & Melly, S. J. (2010). The built environment and location-based physical activity. *American Journal of Preventive Medicine*, *38*(4), 429–38. doi:10.1016/j.amepre.2009.12.032

U.S. Department of Health & Human Services. (2014). Physical Activity Guidelines. Retrieved from http://www.health.gov/paguidelines/guidelines/

U.S. Environmental Protection Agency. (2013). Smart Growth. Retrieved from http://www.epa.gov/smartgrowth/about_sg.htm

Van Dyck, D., Cardon, G., Deforche, B., Sallis, J. F., Owen, N., & De Bourdeaudhuij, I. (2010). Neighborhood SES and walkability are related to physical activity behavior in Belgian adults. *Preventive Medicine*, *50 Suppl 1*, S74–9. doi:10.1016/j.ypmed.2009.07.027

Wasfi, R. a, Ross, N. a, & El-Geneidy, A. M. (2013). Achieving recommended daily physical activity levels through commuting by public transportation: Unpacking individual and contextual influences. *Health & Place*, *23C*, 18–25. doi:10.1016/j.healthplace.2013.04.006

Wendel-Vos, G. C. W., Schuit, a. J., De Niet, R., Boshuizen, H. C., Saris, W. H. M., & Kromhout, D. (2004). Factors of the Physical Environment Associated with Walking and Bicycling. *Medicine & Science in Sports & Exercise*, *36*(4), 725–730. doi:10.1249/01.MSS.0000121955.03461.0A

Wener, R. E., & Evans, G. W. (2007). A Morning Stroll: Levels of Physical Activity in Car and Mass Transit Commuting. *Environment and Behavior*, *39*(1), 62–74. doi:10.1177/0013916506295571

Werner, C. M., Brown, B. B., & Gallimore, J. (2010). Light rail use is more likely on "walkable" blocks: Further support for using micro-level environmental audit measures. *Journal of Environmental Psychology*, *30*(2), 206–214. doi:10.1016/j.jenvp.2009.11.003

Wheeler, B. W., Cooper, A. R., Page, A. S., & Jago, R. (2010). Greenspace and children's physical activity: a GPS/GIS analysis of the PEACH project. *Preventive Medicine*, *51*(2), 148–52. doi:10.1016/j.ypmed.2010.06.001

Witten, K., Blakely, T., Bagheri, N., Badland, H., Ivory, V., Pearce, J., … Schofield, G. (2012). Neighborhood built environment and transport and leisure physical activity: findings using objective exposure and outcome measures in New Zealand. *Environmental Health Perspectives*, *120*(7), 971–7. doi:10.1289/ehp.1104584

Witten, K., Hiscock, R., Pearce, J., & Blakely, T. (2008). Neighbourhood access to open spaces and the physical activity of residents: a national study. *Preventive Medicine*, *47*(3), 299–303. doi:10.1016/j.ypmed.2008.04.010

World Health Organization (WHO). (2011). *Health co-benefits of climate change mitigation-Transport sector: Health in the green economy* (pp. 1–156). Geneva, Switzerland. Retrieved from http://www.who.int/hia/examples/trspt_comms/hge_transport_lowresdurban_30_11_2011.pdf

Zeiler, I., Rudloff, C., & Bauer, D. (2011). Modelling random taste variations on level changes in passenger route choice in a public transport station. In R. D. Peacock, E. D. Kuligowski, & J. D. Averill (Eds.), *Pedestrian and Evacuation Dynamics* (pp. 185–195). Boston, MA: Springer US. doi:10.1007/978-1-4419-9725-8

Zhu, S., & Levinson, D. (2010). Do people use the shortest path ? An empirical test of Wardrop ' s first principle . In *91st annual meeting of the Transportation Research Board, Washington. Vol. 8* (pp. 1–22).