ABSTRACT

Title: TRIP GENERATION MODEL FOR PEDESTRIANS BASED ON NHTS 2001

Nam Seok Kim, Master of Science, 2005

Directed By: Dr. Kelly J. Clifton, Civil and Environmental Engineering

Since non-motorized transportation has been excluded from the main stream of the general transportation planning, there is no standard technique for estimating non-motorized transportation demand. This thesis aims to develop a trip generation model for pedestrians. This research uses Poisson regression to model as walk trips can be regarded as event counts. The National Household Travel Survey (NHTS) 2001 is the primary source of data used, supplemented by MD Property View data. At the individual level, regression models for general walk trips and walk trips to work are estimated. Then, the walk trips to work are validated by Census data at the tract level. Finally, using that walk trips to work have similar tendency to total walk trips, this thesis estimates the number of total pedestrians a day for 506 tracts in Baltimore metropolitan region.

TRIP GENERATION MODEL FOR PEDESTRIANS BASED ON NHTS 2001

By

Nam Seok Kim

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:

Assistant Professor	Kelly J. Clifton, Chair
Professor	Paul M. Schonfeld
Associate Professor	Qing Shen

Dedication

I dedicate this work to my mother who passed away 2 years ago.

No words can describe how much I love her.

Acknowledgements

To my family, I would like to express my deepest thanks to my father and the rest of my family. Without their support, this thesis would never have been possible.

To Dr. Kelly J. Clifton, I would like to especially thank my advisor.

To all of experts and friends who have helped me throughout this research in numerous ways, especially, Dr. Paul M. Shonfeld, Dr. Qing Shen, Kyongpyo Kang, Dr. Jungyul Sohn, Hyunsoo Joo, Arnab Chakraborty, Ravi Amin, Seokwoo Jin, Woon Kim, and Sunho Kwak.

Table of Contents

List of Tables

List of Figures

Chapter 1: INTRODUCTION

1.1 Introduction

Since transport modeling has been developed in the 1950s, techniques have been focused on motorized transportation modes such as private car and public transportation (Bates, 2000). Past studies have brought effective traffic management, and rapid progress of the demand forecasting methods. However, the present transportation system simultaneously caused side effects such as problem with air quality, lack of transportation mode choices, and traffic accidents among vulnerable road users.

To overcome these harmful effects, public policies have been implemented to reduce the demand of motorized transportation. Ewing (1997) diagnosed problems and suggested potential solutions. He emphasizes five strategies: land planning, travel demand management (TDM), transportation system management (TSM), enhanced transit services, and pedestrian and bicycle-friendly design. Theses are specific ways to mitigate the demand for motorized travels. Government also needs a new paradigm of transportation planning to minimize problems caused from the car-oriented transportation system. The Federal government has established two pieces of legislation: The Intermodal Surface Transportation Efficiency Act (ISTEA)[1] and Clean Air Act (CAA). These two acts were a result of the new paradigm of transportation planning and were in clear recognition that the car oriented transportation system could not promote sustainable development. In the national

[1] Congestion Management/Air Quality(CMAQ) category

bicycling and walking study (1994), FHWA sets a goal to increase the current percentage of non-motorized trips from 7.9% to 15.8%. Thus, it was justifiable that non-motorized transportation such as biking and walking should be promoted as an alternative.

Often ignored in traditional transportation demand models, a pedestrian oriented demand modelng could be useful in understanding the issues related to walking and other non-motorized modes thereby contributing to research and development of alternative transportation modes. An appropriate pedestrian demand model is an essential tool for pedestrian planning. According to Ewing (1997), pedestrian and bicycle-friendly design can be constructed only if demands are estimated. According to Raford (2004), the prediction of pedestrian demand makes calculation of exposure[2] of pedestrian risk possible. In addition, Ewing and Cervero (2001) clarified relationships among built environment, health and walking as fundamental physical activity. According to them, the built environment can be a barrier to physical activity, including walking. Through estimating walking demand, the places that impede walking can be examined and appropriately redeveloped.

To sum up, demand models have been focused on the automobile. The techniques used in planning for pedestrians are underdeveloped. Since public policies attempt to increase walking for mitigating congestion and promoting health, efforts to develop and improve pedestrian demand models should be undertaken.

---

[2] The term "exposure" results from the field of epidemiology and is defined as "the rate of contact with a potentially harmful agent or event" (Raford, N.,2004)

1.2 Objective of This Study

The substance of this study is an attempt to fill methodological gaps through the development of a pedestrian trip generation model. The first purpose of this study is to develop a trip generation model for pedestrians at the individual level by using NHTS 2001 data and supplementary land use data. This trip generation model is applied on an area-wide basis, which is sensitive to land use variables. More specifically, the individual level model estimated using NHTS data on Baltimore metropolitan region is aggregated to tract level and validated using Census "walk to work" data.

The second purpose of this study is to apply the model, which only deals with work trips, to total home based walk trips including non-work trips. Thus, two types of home based walk trips are considered: commuting trips and total home based walk trips. Once the commuting trips and the total trips by walking are modeled at the individual level, both models are expanded at the census-tract level. Substituting independent variables at the tract level, commuting trips by walking per-tract are predicted. Then, the predicted trips per tract are compared with the observed Census commute data. Finally, using the error between the predicted and observed value, the general walking trip demand is estimated at the tract level.

Chapter 2 renews approaches to walking demand modeling and meaningful relationships between several factors such as socio-economic variables and built environment and walking. Based upon this review, a conceptual model is developed to shows the relationship between walking frequency and socio-economic variables and factors of the built environment. Chapter 3 presents the data used in the model estimation including NHTS (National Household Travel Survey), Census, and land use data, and specifies methodology used in this study. In the Chapter 4, the models for walk trips are estimated and validated through Census data. Then, the total walk trips are estimated at the tract level. Some findings and future studies are discussed in the Chapter 5.

Chapter 2: LITERATURE REVIEW

Chapter 2 reviews a traditional vehicle trip generation model and several pedestrian models. In Section 2.1, the traditional trip generation model is discussed because pedestrian modeling is referred to as the vehicle trip generation. In Section 2.2 and Section 2.3, an overview of pedestrian researches is presented. Pedestrian researches are divided into two categories. Some are focused on estimating the pedestrian demand (Pushkarev and Zupan, 1971; Behnam and Patel, 1977; Davis, King and Robertson, 1991, Matlick, 1996; Ercolano, Olson, Spring, 1997; Targa and Clifton, 2005). Others are efforts to estimate LOS(Level of Service) or walkability indexes rather than explicit demand models (Turner Fairbank Highway Research Center, 1998; Dixon, 1996; Landis, et al., 2001; Bradshaw, 1993), or specify the relationship between walking and the built environment (Targa and Clifton, 2005; Ewing and Cervero, 2001; Kitamura, Laidet, and Mokhtarian., 1997; Rutherford, McCormack, and Wilkinson. 1996; Cervero, 1989).

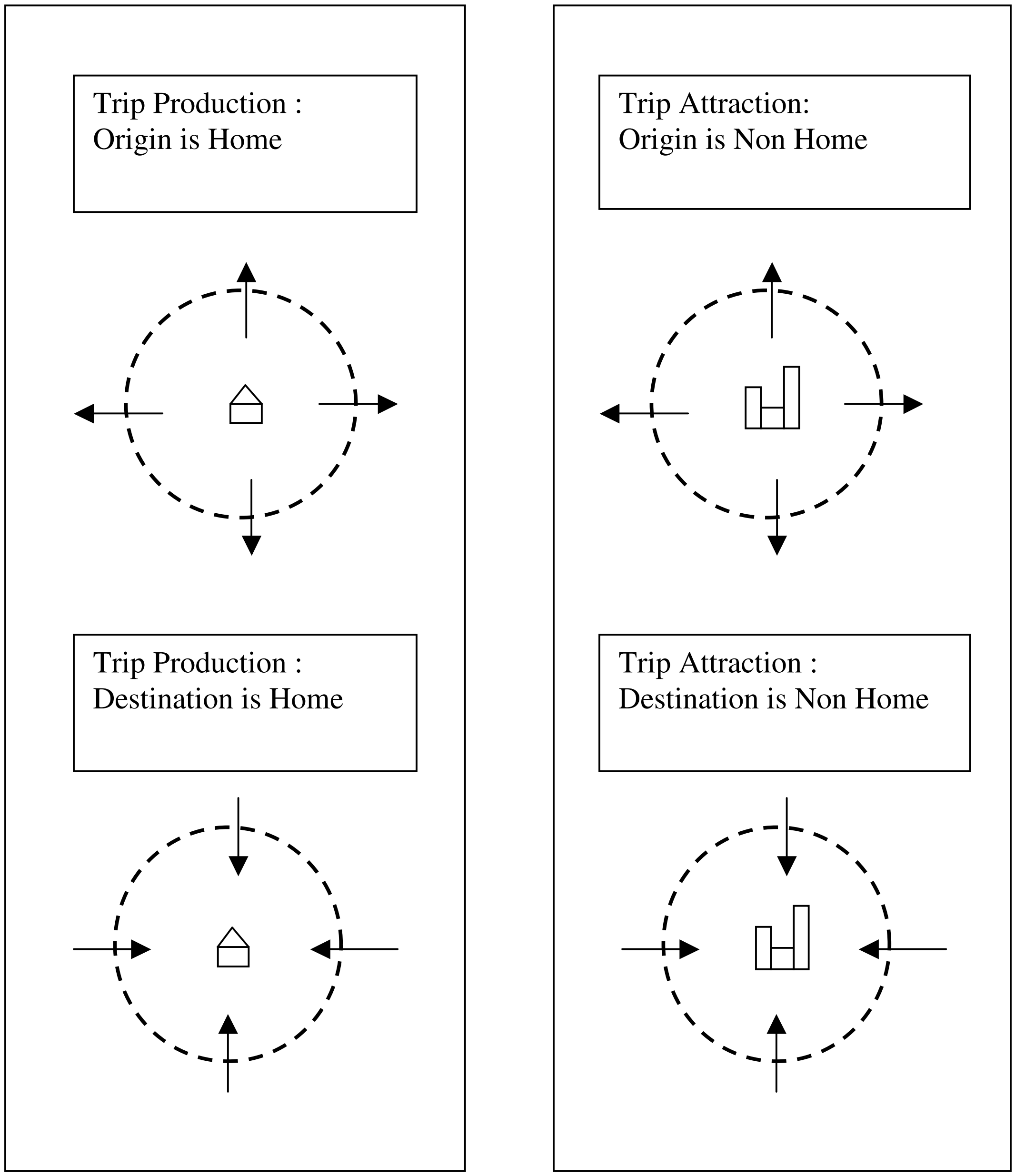
2.1 Traditional Modeling Technique for Pedestrians

Several transportation planning textbooks discuss two common approaches for trip generation: Regression model and categorical analysis (Meyer and Miller, 2001, Hutchinson, 1974, and Oi and Whuldiner, 1962). However, since pedestrian trips are not considered, both methods are focused on vehicle trips. Historically, both techniques relate trip making to the number of vehicles in a household. Oi and Shuldiner clearly summarized the other factors that influence vehicle trip making.

They enumerate the six important factors: household size, distance from CBD, residential density, income, occupation of the household head, and social area indexes. These factors have been commonly used as the independent variables to explain vehicle trip making. However, it is not proven that the same factors are associated with pedestrian trip making.

In the categorical analysis, trip observations are aggregated in terms of the common socioeconomic characteristics rather than spatially grouping household or person (i.e. by TAZ or by Census tract). As the regression modeling, the vehicle ownership is the critical variable. The other factors, which are used in the regression model, can be also applied to the categorical analysis.

The traditional concept of the origin and destination is, in general, borrowed for the pedestrian demand models. Typically, trip generation, the first stage in a four steps travel demand model, is to evaluate the number of trip ends that occurred in each sub-area such as Traffic Analysis Zone (TAZ). According to 'Principles of Urban Transport Systems Planning' (Hutchinson, 1974), "Two types of trip-generation analysis are carried out, referred to the terms, trip production and trip attraction. For the trip production, trip ends that are based at a residence are called home-based trips. For the trip attraction, trip ends are based on non-home end such as employment, retail services, recreation places, and so on." Figure 2.1 shows the relationship between origin and destination (O/D) and production and attraction model. Since trip production is associated with the home, if the home is either origin or destination, the trips should be considered in trip production model. From this separate process, the traditional trip-generation phase is assumed that trip production equivalently occurs with trip attraction.



(a) Trip Production (b) Trip Attraction

Figure 2.1 The Relationship between O/D and Production and Attraction

On the other hand, although it is possible that two phases of traditional four step demand estimation method, trip generation and mode choice, are separately analyzed, this thesis focuses on trip generation of pedestrian trips, which implicitly includes a mode choice. Since the purpose of the mode choice is to predict the percentage of the transportation mode's selection, the final output is the number of people with respect to each mode. However, the mode choice of pedestrians almost depends on the relationship among other modes. In addition, the in-vehicle time and out-of-vehicle time is only adaptable for vehicle trips. Thus, the most important factors of the utility function of logit model, which is the common approach in mode choice phases, cannot be used.

2.2 Pedestrian Modeling and Scope of Study

Several studies have tried to explain the nature of walking. Recently, Teknomo (2002) reviewed the origin of pedestrian modeling and categorized into two distinct branches. The first considers walking in confined circumstances such as building entrances, airports, and train stations. The second considers behavior in larger areas such as neighborhoods and shopping districts. However, as the simulation and computation technology have progressed, more advanced models have been developed. Michael (2001) specifically categorized pedestrian modeling as following five big approaches: (1) simple statistical regression, (2) spatial interaction theory, (3) accessibility approach, (4) fluid-flow analysis, and (5) micro simulation model. Since the objective of this thesis is to develop a pedestrian demand modelat the tract level, most of these approaches were not applicable to this thesis. Among the five approaches, simple statistical regression, which is called "Pedestrian Sketch Plan Methods" by FHWA (1999), most addresses this study purpose.

2.3 Pedestrian Sketch Plan Method

'Bicycle and pedestrian trip generation workshop' (FHWA, 1996) and 'Guidebook on Method to Estimate Non-Motorized Travel' (FHWA, 1998) describe pedestrian sketch plan methods. Pedestrian Sketch Plan Method is developed to estimate pedestrian volumes under existing and future condition. As a starting point of this method, Pushkarev and Zupan (1971) and Behnam and Patel (1977) estimated pedestrian demand in high density areas by using existing land use data and pedestrian counts. They counted the number of pedestrians and surveyed the characteristics of their trips, including trip times and distances. Pushkarev and Zupan used regression analysis to predict total pedestrian volumes per block. Explanatory variables included commercial land uses, distance to transit stops, and sidewalk. Behnam and Patel showed similar regression models to the research conduced by Pushkarev and Zupan. Pedestrian volume per hour per block is used as the dependent variable. The independent variables included commercial space, office space, cultural and entertainment space, manufacturing space, residential space, parking space, vacant space, and storage and maintenance space. Based on future land use variables, Behnam and Patel predicted future pedestrian volumes in Milwaukee CBD.

Ercolano, Olson, and Spring (1997) used peak vehicles per hour, transit vehicle/ridership, and non-motrozied model share to estimate the pedestrian travel demand at the peak hour in suburban areas. Ercolano, Olson, and Spring applied the pedestrian travel demand at the peak hour to determine the location of pedestrian crossings, sidewalks, and signal re-timings.

Matlick (1996) also modeled the pedestrian demand by using household population, transportation mode share, and activity center data. Matlick's model is used to determine the priority areas or corridors for improvement of pedestrian facilities.

Targa and Clifton (2005) showed a walk trip generation model by Poisson models. It is notable that Targa and Clifton (2005) showed the significant results for the same topic by using Poisson regression. Regarding walk trips as event counts, they mainly used 2001 NHTS (National Household Travel Survey) and categorized the independent variables into four characteristics: individual, attitudes/ perceptions, urban form/land use attributes, and neighborhood socio-demographics.

From the literature described above, Table 2.1 presents the synthesis of this method. The pedestrian sketch plan method is developed according to different time periods (hourly, daily), spatial levels (facility level, local level, and regional level), and estimation methods (linear regression, Poisson regression, and simple computation). The trip generation phase in this study estimates the walk trip frequency from NHTS. For the spatial level, the Baltimore metropolitan region is considered as a study area for its proximity, availability of data and unique ability to reflect metropolitan characteristics including a city, urban, suburban, and rural area. The mixed urban form addresses the purpose of this study which is to model the prototype of pedestrian generation. For the temporal difference, this study uses one-day time period as NHTS travel day data are based on one-day trips. For the estimation method, since a few studies discussed this issue, it is discussed in Section 3.5 using the NHTS data.

Table 2.1 Synthesis of the Sketch Plan Model

Researchers	Level of Study Area	Time period	Data needs		Technique
			Pedestrian volume	Land use and socio economic data	
Pushkarev and Zupan, 1971	Block (Midtown Manhattan)	Hourly	Pedestrian counts (aerial photography)	Square mile of office, retail, and restaurant space	Linear regression
Behnam and Patel, 1977	Block (CBD of Milwaukee)	Hourly	Pedestrian counts (real counts)	Commercial space, Office space Cultural and entertainment space, Manufacturing space, Residential space, Parking space Vacant space, Storage and maintenance space	Linear regression
Davis, King and Robertson, 1991	Crosswalk level(Washing ton D.C)	5, to 10 minute time segments during Peak hours	Pedestrian counts (real counts)	Vehicle traffic counts	Simple Equation
Matlick, 1996	Corridor-level (Seattle, Washington)	Daily	Transportation mode share information (Census) / NPTS	Housing types, density, persons per household unit, and hotels Retail, recreation, social facilities, schools, employment, and churches.	Linear regression

Table 2.1 Synthesis of the Sketch Plan Model (Continued)

Ercolano, Olson, Spring, 1997	City level (Plattsburgh, New York)	Hourly (peak hour)	Vehicles per hour from traffic counts and mode share from Census	Vehicle traffic counts	Computa-tion using spreadsheets
Targa and Clifton, 2005	City level (Baltimore City)	One day	The number of walk trips from NHTS 2001	Car ownership in household, type of housing unit, household income, age, sex, driver status, education status, attitudes/ perceptions of pedestrians, household density, street connectivity, land use diversity, proportion of commercial units	Poisson regression

2.4 Empirical Evidence

There have been several pedestrian studies including "Sketch Plan Method" discussed in Section 2.3 based on empirically examining the effect of land use characteristics and socio economic characteristicson walking trip frequency. This section summarizes empirical findings of the pedestrian studies.

According to Cervero and Radisch (1995), the effect of neighborhoods on travel demand was almost firstly researched by Levinson and Wynn (1963). They found that neighborhood density is closely associated with decreasing vehicle trips. In the high density city, decreasing vehicle trip frequency means increasing transit trips and non-motorized trips.

Ewing and Cervero (2001) summarized empirical findings and provided synthesis of the relationship between travel and built environment. Their synthesis focused on the effect of walking trips on four kinds of category: prototypical neighborhoods, activity center, land use variables, and transportation network variables. According to them, walking trips are associated with transit-oriented neighborhood, the distance between commercial districts and residential areas, higher density areas, land use mixing areas, and multi-story buildings. Even though several empirical studies do not use a trip generation method but mode choice technique, several findings supports that pedestrian demands are associated with land use characteristics.

Moudon, Hess, Snyder, and Stanilov (1997) showed effects of site design on pedestrian travel in mixed-use, medium-density environments. They selected 12 neighborhood centers or sites in the Puget Sound area in Washington by some criteria: residential density, income, automobile ownership, and intensity and type of commercial development. 6 urban areas out of 12 neighborhood sites show 37.7 pedestrians per hour per 1,000 residents on average, while other 6 suburban areas show 12.5 pedestrians per hour per 1,000 residents. They found a “clear break” of pedestrian volumes per hour per 1,000 residents as 16 to 22 pedestrians.

Targa and Clifton (2005) found that lower vehicle ownership, college dorm home type, and lower household income are associated with higher walking frequency. In addition, denser urban area, higher street connectivity, and more mixed land use generate more walk trips.

2.5 Summary

Most literature comcludes that walk trips are closely related to socio-economic data and land use variables. However, they differ in the level of study area and time period. Nevertheless, some variables such as density, mixed land use, and car-ownership are considered steadily. However, there are limitations to collect those data. Pushkarev and Zupan used aerial photography data collection techniques. It is difficult to apply to city or regional level analyses. In addition, since they focused on the high-density CBD site, it is limited to apply to other areas. Similarly, the model of Behnam and Patel is also limited in low density areas.

Since Ercolano, Olson, and Spring do not use a regression technique, it is impossible to predict the pedestrian change with respect to other factors (land use and socio-economic data). In other words, the model only depends on other mode share. Since utility function including the travel time and the travel distance should be estimated to calculate mode share percentage, this technique is also limited to estimate the pedestrian demand.

The limitation of these two previous studies was the data collection. Since the real count data usually reflect both general and unique characteristics of the area where the data are collected, it is hard to apply the model estimated in one place to the other places. The unique characteristics are usually unknown. This study tries to overcome this disadvantage focused on general measures such as socioeconomic factors and land use factors from various urban forms. In addition, the study area of previous studies is blocks, corridors, and a city. This study tries to model the same topic for metropolitan level (Baltimore city and 5 neighboring counties). The considerable quantity of data from NHTS Baltimore add-on reflects the general characteristics of pedestrians.

In general, the conceptual model, which is based on the empirical studies, is that walking frequency is the function of socio-economic data and land use variables. Socio-economic data consists of age, income, race, education, and car ownership. Land use variables include population density, household density, non-residential unit density, and mixed land use. On the other hand, it was found by Targa and Clifton (2005) that the Poisson regression model can be an appropriate model for walk trips. Therefore, it is assumed that walking trip frequency is followed by Poisson distribution.

Chapter 3 METHODOLOGY

Chapter 3 presents the methodology developed to study the factors influencing walk trips leading to the generation of an appropriate model. This chapter is organized into three main sections. The first section covers overall research design. The second elaborates on sources of data and characteristics of the data. The third discusses the statistical method.

3.1 Research Design

The major objective of this thesis is to develop commuting walk trip generation models and apply it on area-wide basis, which is sensitive to land use variables. To do this, a model was estimated using NHTS data for Baltimore region at the individual level. This model was then applied to tract level and validated using Census "Work to walk" data.

Before a model is developed at the individual level, exploratory analysis are tested to examine the relationship among variables and both types of walking frequencies (i.e. the number of home based walk trips and the number of home based commuting walk trips). Then, some variables are chosen for modeling because this thesis aims to develop the simple model that can be available to apply at area-wide level. Figure 3.1 shows the flow chart for this analysis.

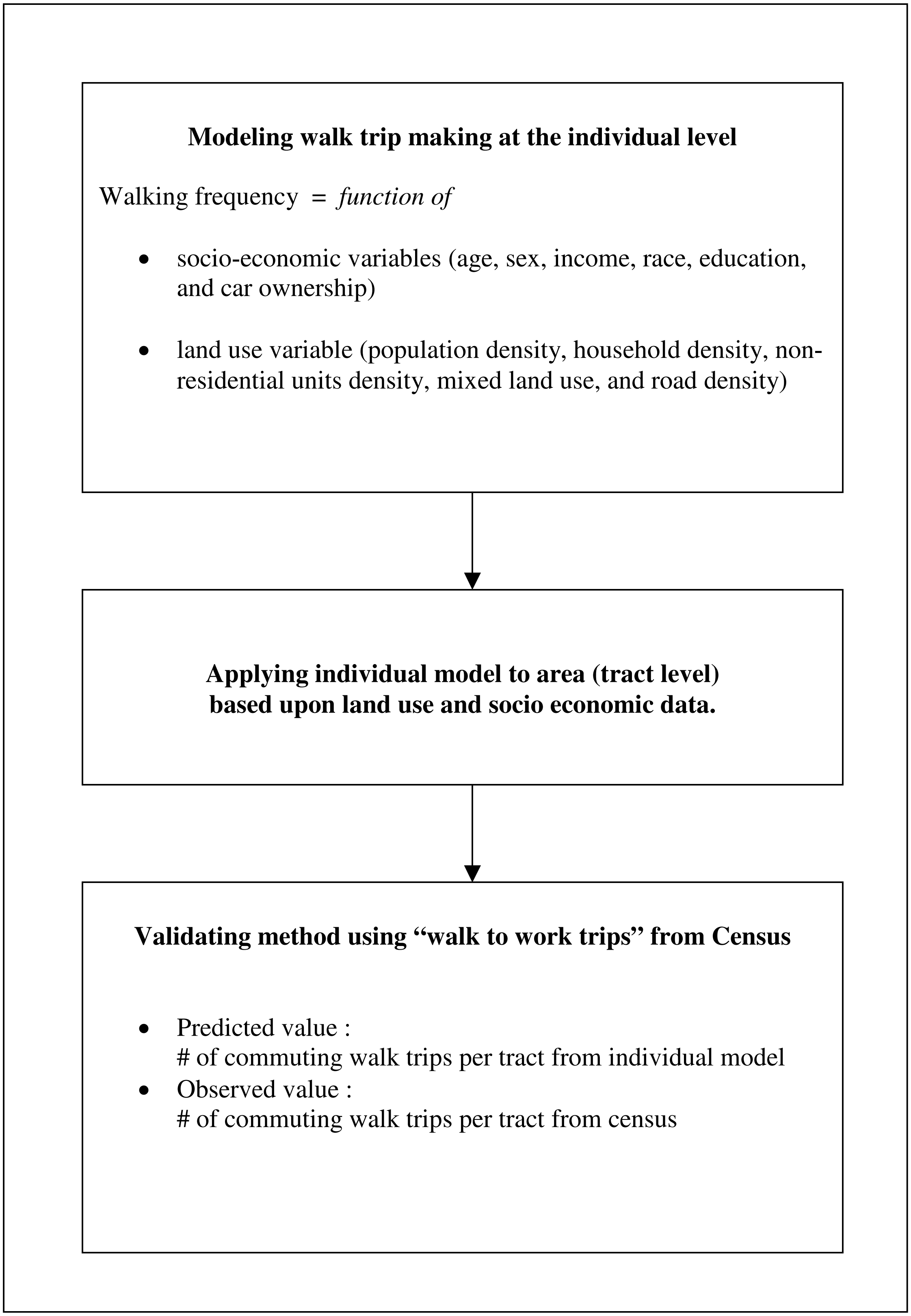


Figure 3.1 Analysis Flow

3.2 Dependent Variables from NHTS 2001

3.2.1 Description of NHTS 2001

In terms of the conceptual model in Figure 3.1, three kinds of data are needed: walking trip frequency, socio economic data, and land use data. National Household Travel Survey 2001 (NHTS 2001) satisfies the data needs of this research since NHTS 2001 provides not only trip frequency with transportation modes, but also socio-economic data and land use data at Census tract level. NHTS collected data by using Computer-Assisted Telephone Interviewing (CATI) technology. Each household member in the sample household records all travel for a "Travel Day" which is defined specifically 24-hours. NHTS 2001 was collected by interviews from April 2001 through May 2002. NHTS 2001 contains 4 kinds of data sets: household characteristics, person characteristics, vehicle characteristics and travel information[3]. NHTS 2001 has 66,000 household samples including 40,000 households of nine add-on areas. Among these add-on areas, the Baltimore metropolitan region is chosen for the study area because the sample size of the region is enough to be modeled at the individual levels. Specifically, the Baltimore metropolitan area consists of Baltimore City and 5 neighboring counties (i.e. Anne Arundel County, Baltimore County, Carroll County, Harford County, and Howard County.). Figure 3.2 shows the study area. 3,519 households are included in the Baltimore add-on. 3,519 households generated 27,366 trips in a travel day. The 27,366 trips were combined with 7,825 household members (i.e. person level).

[3] NHTS Data description is provided on the its website: http://nhts.ornl.gov/2001/index.shtml

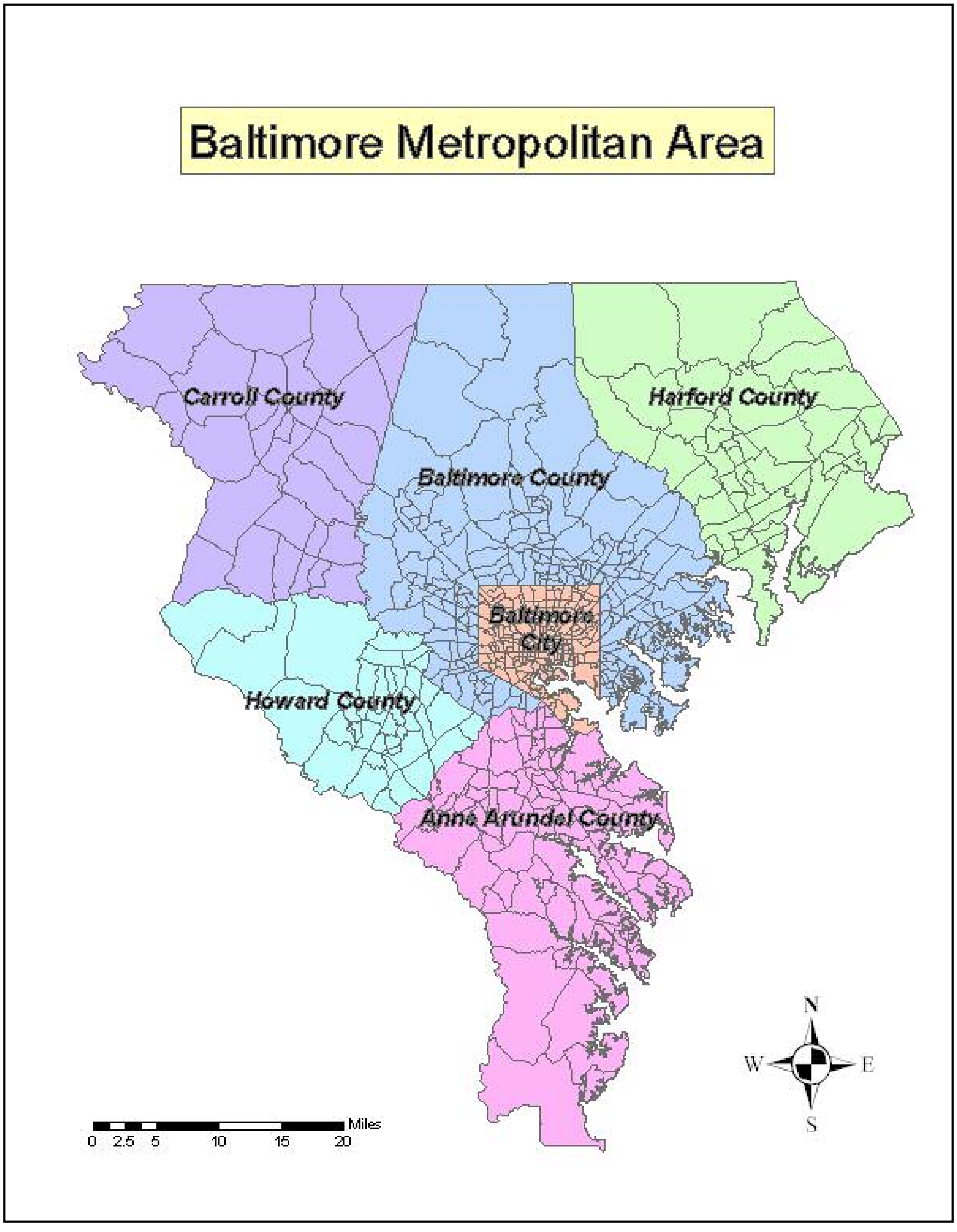


Figure 3.2 Study Area (Baltimore Metropolitan Region)

3.2.2 Definition of Walk Trips

According to NHTS instruction (appendix N), "a trip is whenever you travel from one address to another." The trip includes "walks, jogs, bike rides, and short drives." However, "the data do not include stops just to change the type of transportation."(i.e. the walk trips from parking lots to work places or the walk trips from home to transit stops[4]) Figure 3.3 presents an example of trips on a travel day. In the example above, the mode of "trip 1" is coded as "subway" although a walk trip is included in "trip 1".

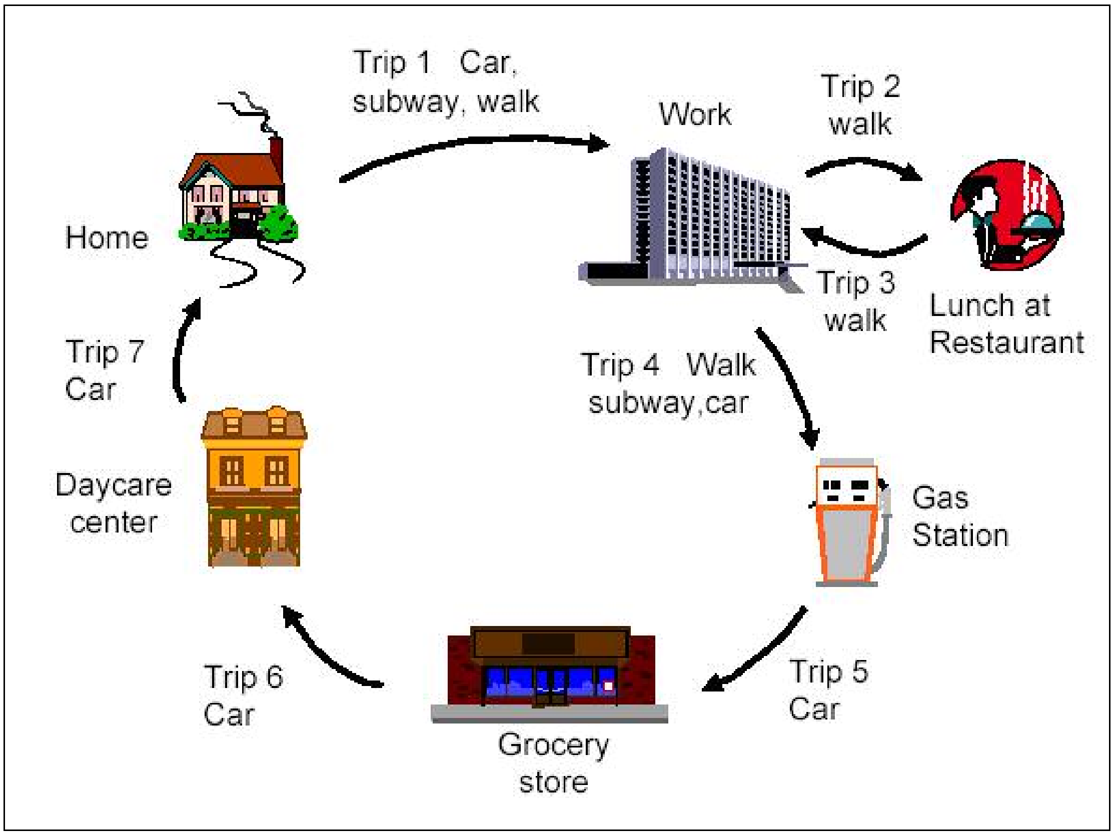


Figure 3.3 Example of Trips on a Travel Day

Source: NHTS instructions (Appendix N, page N-12)

---

[4] The modes to access/egress from transits are indicated in dataset. However, those are not regarded as a trip but as just the part of a trip.

All trips made for a specific reason, such as to go to work or school all kinds of trips are sorted out in terms of two criteria: home based trips and purpose of trips. Figure 3.3 shows two dependent variablesdefined as the total walk

3.2.3 Dependent Variables

All kinds of trips are sorted out in terms of two criteria: home based trips and purpose of trips. Figure 3.4 shows two dependent variables defined as the total walk trips (TW) home based commuting walk trips (HBCW). The total walk trips indicate the home based walk trips regardless the origin of the trip and the trip purpose. The home based commuting walk trips are the walk trips to work from or to home. These two dependent variables are used to develop walking frequency model at the individual level.

In the travel dataset of NHTS 2001, the home based trips are classified in terms of the origin and destination of each trip. Then, walk trips are extracted using travel mode. Finally, walk trip frequency for specific purposes (i.e. Work trips) is summarized in terms of the person’s identification number and home based trips.

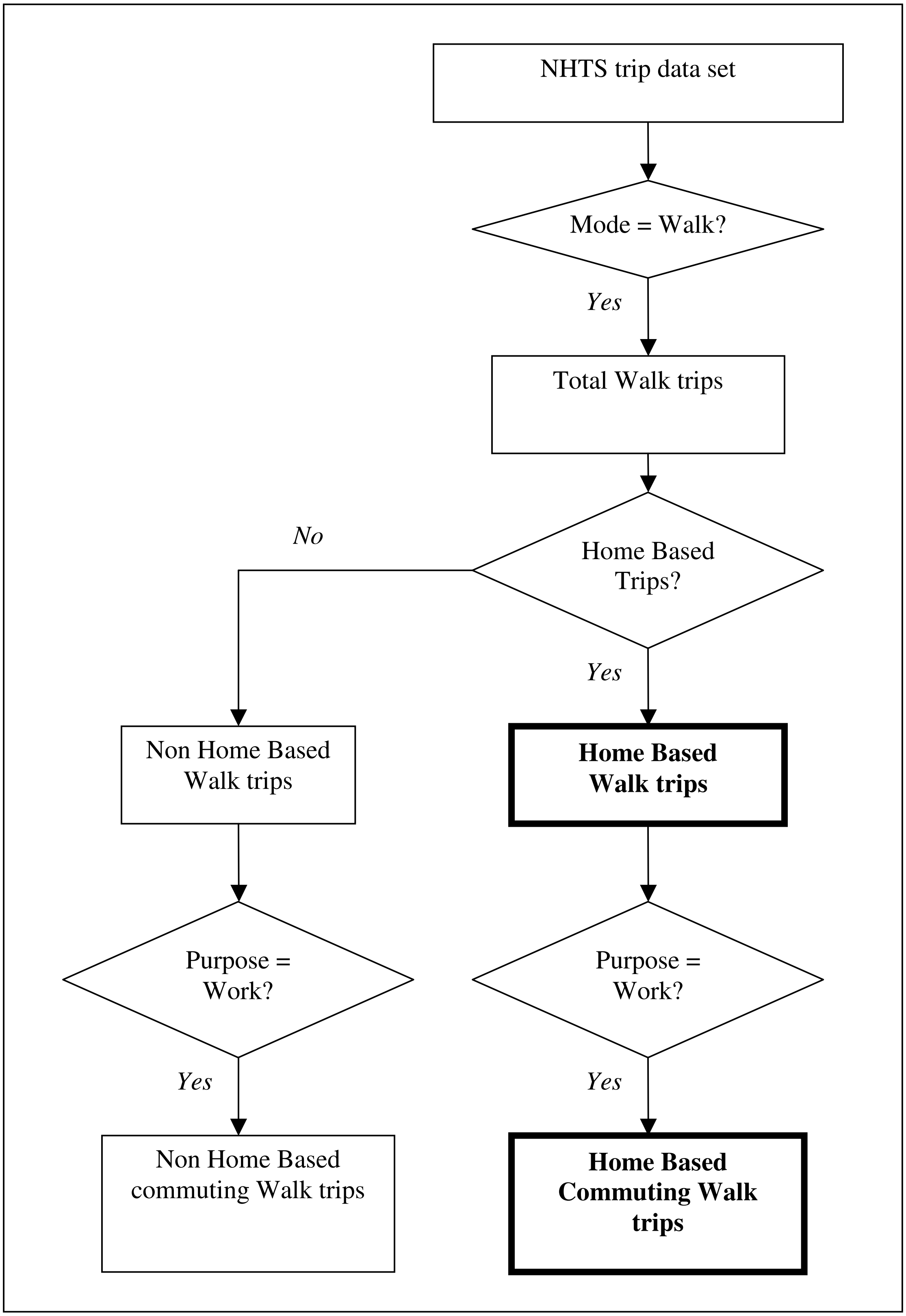


Figure 3.4 Classification of the Possible Dependent Variables

3.3 Socio-Economic Variables

The two dependent variables defined in the last section are assigned to socio-economic variables. Socio-economic data are collected from two sources: NHTS 2001 and Census data. NHTS data are used for developing a model at the individual level. Using personal identification numbers, the trip frequency of each person is assigned to personal and household characteristics of NHTS 2001. On the other hand, Census "Walk to work" data are used for the validation of the model at the tract level. Thus, the tract number is used to assign the trip frequency to the census data. Socio-economic variables consist of population, age, income, race, education, car-ownership, and driver status as shown Table 3.1.

The education and income are dummy variables. The break value (education is college graduation and income is 40,000) are determined by the simple correlation with the dependent variable as changing the values.

Table 3.1 Socio-economic Data from NHTS

Variables	Description
Personal Characteristics	
AGE	Age of respondents
SEX	Sex of respondents; 1 if male, 0 otherwise
RACE	Race of respondents; 1 if white, 0 otherwise
DRIVER	Driver status; 1 if driver, 0 otherwise
WORKER	Worker status; 1 if worker, 0 otherwise
EDUCATION	Education status; 1 if less than college graduation, 0 otherwise
Household Characteristics	
WORKER_HH	Number of workers in household
VEHICLE_HH	Number of vehicles in household
DRIVER_HH	Number of Drivers in household
ADULT_HH	Number of Adults in household
P_WORKER_HH	Percentage of household worker (the number of workers /household size)
P_VEHICLE_HH	Percentage of household vehicle (the number of vehicles/household size)
P_DRIVERS_HH	Percentage of household drivers (the number of drivers/household size)
P_ADULTS_HH	Percentage of household adults (the number of adults/household size)
P_VHE_ADU	Percentage of household adults-drivers (the number of drivers/household adults)
INCOME	Household income; 1 if less than 40,000, 0 otherwise

3.4 Land Use Measurements

As indicated previously, land use patterns might affect walking trip. NHTS 2001 provides basic land use variables such as population density and household density. However, to examine the influence of the detailed land use variables on walk trips, MDproperty View[5] is used as a source for additional land use variable (i.e. floor space of single family dwelling units, floor space of multiple family dwelling units, and mixed land use)

3.4.1. Land Use Variables from NHTS 2001

Since the household dataset of NHTS 2001, which is not NHTS Baltimore add-on but the national level of NHTS 2001, provides land use variables neighboring the household: housing units density (housing units per square mile), percent renter-occupied housing, population density (persons per square mile), and employment density (jobs per square mile). These land use variables are incorporated in NHTS Baltimore add-on data set. Then, all variables are estimated at the Census tract level.

3.4.2. Land Use Variables from Other Sources

Since walk trips might be affected by not only density related variable from NHTS 2001 but also the other variables such as degree of mixed use, floor space, amount of commercial and road network density, Maryland property view 2001(MDproperty View) and Census TIGER/Line data are used to create additional land use variables.

[5] http://www.mdp.state.md.us/

MDproperty View, a GIS (Geographic Information System) based parcel data issued per county, are specified as the followings:

- Complete properties in Baltimore metropolitan region;
- Location of most of buildings presented by centroid of parcel;
- Land use categorization; and
- Characteristics of property, including floor space.

MDproperty View consists of two data sets: Assessments and Taxation data (A&T data) and Computer Assisted Mass Appraisal (CAMA). The former is holistic taxation data set that includes all kinds of land use while the latter are focused on residential units.

Ten variables are generated by MDproperty View : density of residential units, density of single family dwelling units, density of multiple family dwelling units, floor space of residential units, floor space of single family dwellingunits , floor space of multiple family dwelling units, number of non-residential units, density of non-residential units, floor space of non-residential units, mix land use. These variables are manipulated by GIS and estimated at the Census tract level. More specifically, residential and non-residential units are laid on the Census tract polygon data. Then, according to specific criteria such as single family residential units, multiple family residential units, and commercial units, parcels are counted for "units" or summarized for "floor space" variables. "Density" variables are calculated by dividing number of units by tract area. However, it is not possible to collect transportation related variables such as transit stops and sidewalks through MDproperty view.

For compensating transportation variables, TIGER/Line data and MD transit view data are used. Firstly, TIGER/Line data (i.e. road line data) is used for estimation of total road length in each tract since the sidewalk layer was available. It is assumed that roads which are not included non-accessing road such as interstate highway almost mean sidewalks. More specifically, Census Feature Class Codes (CFCC)[6] is used for removing limited access for pedestrians (A10 to A18 in the code table). If CFCC is A10 to A18 in attribute table of Tiger Line data, they are removed. Then, all roads existed are summarized at the tract level. Secondly, MD transit view is used to calculate the number of transit stops at the tract level. Since there are several tracts where rail stations are not located, the rail station is assigned to dummy variables at the tract level. Finally, Table 3.2 summarized that land use data including additional road network density and transit stops.

[6] http://www.topodepot.com/Docs/Doc_Tiger.htm

Table 3.2 Land Use Data

Category	Variable name	Variable type
Land use factors	Population Density(POP_DEN)	Continuous; Population per tract / area
	Residential Density(RES_DEN)	Continuous; Residential units per tract / area
	Employment Density(EMP_DEN)	Continuous; Employment per tract / area
	Mixed land use(MIX)	Continuous; Non residential units/Residential units
	Floor Space of Single Family Dwelling Unit(FS_SFDU)	Continuous; Summation of floor space per tract
	Floor Space of Multiple Family Dwelling Unit(FS_MFDU)	Continuous; Summation of floor space per tract
	Floor Space of Non residential Unit(FS_NRDU)	Continuous; Summation of floor space per tract
	Density of Single Family Dwelling unit(SFDU_DEN)	Continuous; Single Family Dwelling Units per tract area
	Density of Multiple Family Dwelling Units(MFDU_DEN)	Continuous; Multiple Family Dwelling Units per tract area
	Density of Non residential Units(NRDU_DEN)	Continuous; Non residential Units per tract area
	Density of Road(ROAD DENSITY)	Continuous; total length of road per tract area
	The number of bus stop	Continuous; the number of bus stop per tract
	Rail station	Dummy; existence of rail station

3.5 Method of Model Estimation

The previous sections dealt with the types of data and their manipulation of both the dependent variable and independent variables. This section specifies the model estimation methodology including statistical issue such as Poisson regression and the problem of linear regression.

3.5.1 Analysis

Based on Figure 3.1, this subsection specifies the analysis including validation. First of all, twokinds of dependent varia bles in Figure 3.4 are brought: the home based walk trips and the home based commuting walk trips. The first stage for analysis is to develop the model for total walk trips and home based commuting walk trips at the individual level. Then, since the Census data provides the number of people who choose to walk to work at the Census tract level, it is validated with the home based commuting walk trips. Thus, home based commuting walk trips are aggregated at Census tract level and compared with CTPP data. Then, two adjusting factors are suggested: returning trips factor (1.86) and Census weighted factor (1.20). The returning trip factor is calculated for adjusting the discrepancy between the number of trips and the number of people in NHTS. Conceptually, people who commute by walking might return by walking. Thus, the number of home based commuting walk trips should be two times number of commuters by walking. However, based up on the NHTS data, the returning ratio by commuters that walking is not 2 but 1.86 on average (For the detail, see **4.3.1**).

After the returning factor is multiplied to the Census data, the predicted trips are estimated by substituting independent variables. Finally, MAFE (Mean Absolute Forecast Error) is calculated to show the gap between the predicted value and the observed value from Census

On the other hand, the Census weight factor is calculated for weighting NHTS data. After the number of commuters is adjusted to the number of trips by the returning factor, NHTS data and Census data are compared. NHTS walking trip data underestimates the number of commuting walking trips when compared to Census data by as much as 1.20 (For the detail, see **4.3.2**).

It is notable that signs of variables in both models (commuting trips and general trips) have same tendencies. It is the rationale of this research to apply the work trip model to the total trip model. It is possible that the model for the total walk is changed by other factors. However, it is tested that the models used by other factors are not better in terms of goodness of fit. Therefore, there are two assumptions. The first assumption is that factors estimated by the commuting walk trips is equivalent to those by the total walk trips. The second assumption is that models for both total walk trips and home based commuting walk trips consist of same independent variables.

3.5.2 Analysis for Dependent Variables

Since the distribution of walk trip frequency is different from vehicle trip_frequency, the methods of analysis account for this difference. Figure 3.5 (a) and (b) show trip frequency of vehicle trips and pedestrian trips respectively. Traditionally, it is assumed that vehicle trip generation is estimated by linear regression with independent variables such as car ownership, household size, and land use factors. Although the shape of Figure 3.5(a) is not a normal distribution, when the dependent variable is manipulated (i.e. rearranged in terms of home based trip), the shape is adjusted to be close to a normal distribution.

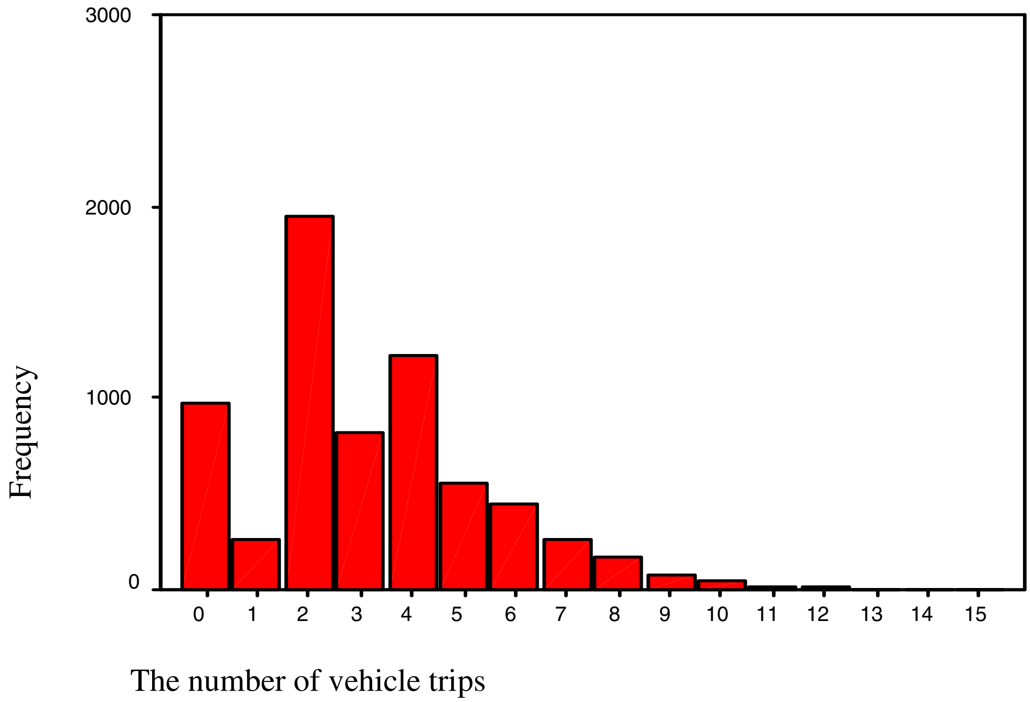


Figure 3.5(a) Vehicle Trip Frequency per a Day

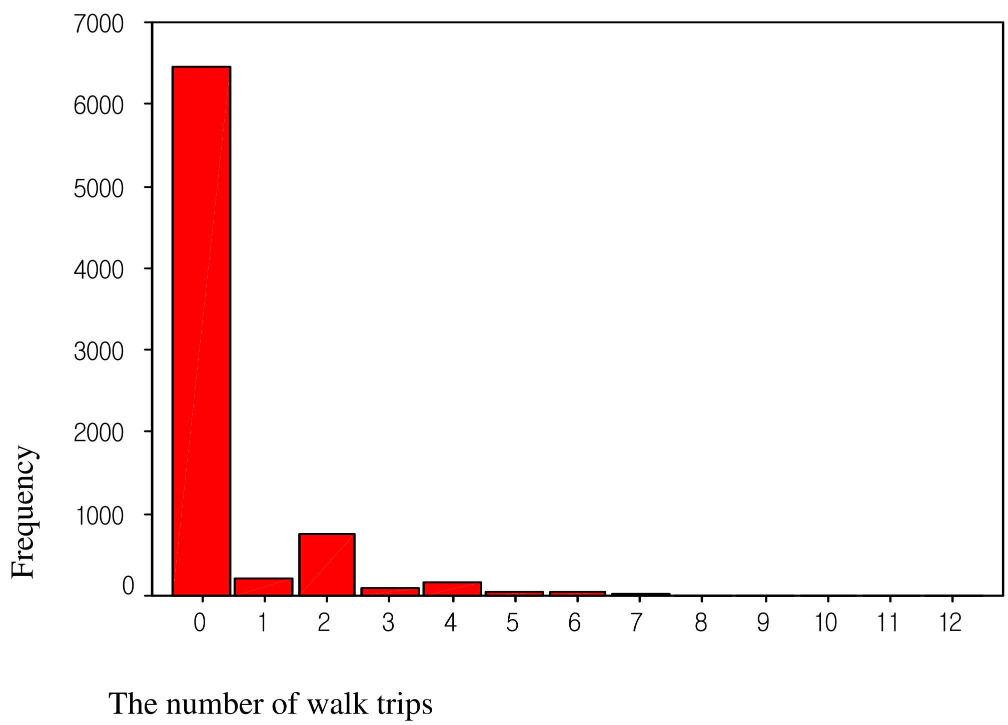


Figure 3.5(b) Walk Trip Frequency per a Day

(Source: National Household Travel Survey 2001)

On the other hand, although independent variables are similarly used in walk trip generation, the shape of the distribution of the dependent variable might not be normally distributed as shown Figure 3.5(b), even though the dependent variable is adjusted. In particular, the number of zero walk trips is over dispersed. Thus, walk trip frequency should be regarded as event count and Poisson regression can be applied. Poisson regression is discussed in next subsection.

3.5.3 Statistical Methodology

In this subsection, a statistical model is discussed by using the variables in this chapter. First of all, the limitations of OLS (Ordinary Least Square) estimation are briefly discussed. The characteristic of statistical models is to use sample data to generate a mathematical relationship (Taylor and Young, 1988). Traditionally, the walk trip generation equations were developed through the OLS method as shown in Equation 3.1.

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + \ldots..+b_kX_{kj} + U_i \qquad \textbf{(Equation 3.1)}$$

The column vector Y_i represents the trip rate as the dependent variable of the ith observation and matrix X_{ij} represent the independent variables such as age, income, and the number of vehicles in households. The column vector bj represents the parameters. The term U_i is a random error term. It is definitely possible to use linear regression if U_i followsa normal distribution.

However, the dependent variables, the walk trip frequency, can be regarded as discrete response variables that represent the number of occurrences of some event within a given domain. Thus, Poisson regression can also be used without loss of generality because it is assumed that the number of events that occur to each case in a given observation to be governed by a rate of event occurrence. As shown in Figure 3.5 (b), the number of walk trips obviously does not follow a normal distribution.

The Poisson regression model (PRM) assumes the dependent variable, which follows a Poisson distribution with parameter μ_i, is controlled by independent variables. Especially, Equation 3.2 shows the density for the dependent variable.

$$f(y_i \mid x^i, \beta) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$ **(Equation 3.2)**

where, $\mu_i = e^{C + \sum \beta_k X_{ik}}$, $y_i = 0, 1, 2\ldots.$

μ_i is an exponential function of the covariates that is conditional on the covariates for each case. μ_i can be expressed into a logarithm as Equation 3.3.

$$\log(\mu_i) = C + \beta_1 \times X_1 + \beta_2 \times X_2 + \cdots + \beta_k \times X_k$$

(Equation 3.3)

where, $y_i = 0, 1, 2\ldots.$

Thus, the log of μ_i is assumed to be a linear function of the independent variables. μ_i itself means both conditional mean and variance (DeMaris, 2004). Since Equation 3.2 is exponential, μ_i is always positive. The procedure for formulating the log-likelihood function for a Poisson regression estimator is summarized in Appendix A.

3.6 Limitations and Summary

Cervero and Radisch (1995) claim that "the absence of rich land-use and urban design data" is a primary limitation for the measuring the effects of built environment on trip making. Even though, at present, this lack of land use data is compensated by the development of GIS and data manipulation skills, the development of detailed land use data is still a time-consuming process. For example, to collect sidewalk data, aerial photographs were needed for 6 counties. Secondly, travel diary data for some census tracts are very limited. In particular, since the walk mode share is very limited in the NHTS at the tract level. It is a critical problem at the validation step because Census "Walk to work" data has the actual number of pedestrians to work, but more than 100 tracts, which were aggregated from NHTS at the tract level, has no commuting walk trips. Although the estimated average trip frequency is reliable, it is expected that trip frequency of each tract has higher error terms. Thirdly, lack of climate data should be inserted since the effect of weather on non-motorized transportation usage might be very critical (In particular, bike trips are more affected on the weather).

This chapter described the data and methodology utilized in carrying out this research. Data are collected from NHTS 2001, MD property view, and Census. For the statistical methodology, the distribution of the dependent variable is reviewed. From the interpretation of the distribution, Poisson regression is used to develop the trip generation model.

Chapter 4: ANALYSIS

Chapter 4 presents the model estimation for the pedestrian trip generation. As discussed in previous chapters, Poisson regression is employed to capture the trip behavior of pedestrians.

Section 4.1 shows exploratory analysis for two dependent variables: total walk trips and home base commuting walk trips. Exploratory analysis includes summarization of variables and Pearson correlation matrix. Section 4.2 suggests two regression models for the home based commuting walk trip and the total walk trip. Two models are estimated by Poisson regression. Section 4.3 compares the numbers of predicted trips and observed trips, with mean absolute forecast error (MAFE). Section 4.4 presents the summary and limitations.

4.1 Interrelations between Walk Trip Frequency and Independent Variables

This section explores the interrelations between the number of walk trips and independent variables. Table 4.1 summarizes simple statistics of these variables. DRIVER, WORKER, WHITE, EDUCATION, INCOME are dummy variables. The percentage of household worker (P_WORKER_HH), the percentage of household vehicle (P_VEHICLE_HH), and the percentage of vehicles per adults in household (P_VEH_ADU) basically have the range from 0 to 1. However, it is possible that P_VEHICLE_HH and P_VEH_ADU are greater than 1 (e.g., the number of vehicle in household is greater than household size or the number of adults).

Table 4.1 Summary of Variables

Variable	Obs	Mean	Std. Dev.	Min	Max
Home based commuting walk trip (HBCW)	4042	0.032	0.261	0	6
Home based walk trips (TW)	7825	0.434	1.117	0	12
AGE	7825	39.787	23.024	0	96
DRIVER	7825	0.681	0.466	0	1
WHITE	7825	0.753	0.431	0	1
EDUCATION	7825	0.473	0.499	0	1
INCOME	7825	0.486	0.500	0	1
P_WORKER_HH	7825	0.519	0.357	0	1
P_VEHICLE_HH	7825	0.721	0.478	0	9
P_VEH_ADU	7624	0.932	0.527	0	9
SUBURB	7825	0.284	0.451	0	1
URBAN	7825	0.077	0.266	0	1
CITY	7825	0.308	0.462	0	1
POP_DEN (Population Density)	7825	7731	8270	50	30000
RES_DEN (Residential Density)	7825	2646	2291	25	6000
EMP_DEN (Employment Density)	7725	0.492	3.504	0.001	85.500
MIX(Mixed Use)	7825	2203020	1726317	0.000	9801411
FS_SFDU(Floor Space of Single Dwelling Unit)	7825	24839.800	45998.360	0.000	424940.00
FS_MFDU(Floor Space of Multiple Dwelling Unit)	7816	2167.636	6356.531	0.000	49749.300
FS_NRDU(Floor Space of Non residential Unit)	7825	2293	7352	0.000	111372

Table 4.1 Summary of variables (Continued)

Variable	Obs	Mean	Std. Dev.	Min	Max
SFDU_DEN(Density of Single Dwelling unit)	7825	2070.214	3246.078	0.000	14568.270
MFDU_DEN(Density of Multiple Dwelling Units)	7825	79.328	382.193	0.000	6165.650
Density of Non residential Units(NON_RES_DEN)	7825	209.798	464.159	0.000	3880.464
Road Density (ROAD DENSITY)	7825	24.006	19.614	2.517	78.037
Transit stops	7825	8.14	11.38	0.00	91.000

To further analyze interrelations among dependent variables (the home based commuting walk trip and the total walk trip) and explanatory variables, correlation among variables is tested. The Pearson correlation matrix is shown in Table 4.2 and Table 4.3 after some variables are considered as the followings:

- The non residential density and road density are the most associated with the home based commuting trip frequency (0.20 and 0.18 respectively). These two variables are slightly correlated with each other (0.57).
- The road density (ROAD DENSITY) is highly correlated with other density related variables: population density (0.91), residential density (0.91), employment density (0.88), density of single family residential dwelling (0.74), and single family residential floor space (0.71), while the non residential density is slightly less correlated with population density (0.46), residential density (0.54), and employment density (0.57).
- The density of multiple residential dwelling units shows weak relationship with the dependent variable, although the sign is positive as expected.

- The density variables for single family units, multiple family units and non residential units are better than floor space variables in terms of the relationship with home based commuting trips.
- Three floor space variables (Floor Space of Single Family Unit, Floor Space of Multiple Family Unit, and Floor Space of Non residential Unit) are dropped because corresponding density variables are correlated with them and more associated with the dependent variables.
- The number of multiple family dwelling units presents the odd sign (Negative). Although multiple dwelling units are one of the major generators of pedestrians, it is dropped because of the sign of the correlation test. Although it is not included in the model, it is acceptable because it is only for the case of the commuting walk trips.
- Among socioeconomic variables, the driver status (dummy variable) and the percentage of drivers in household (only adults: over 16years old) show negative association with the home based commuting trips (– 0.13 and – 0.15 respectively). The correlation between these two variables is acceptable (0.40).

Table 4.2 Pearson Correlation Coefficient Matrix for Home Based Commuting Walk Trips (Obs. 3,916)

	HBCW	AGE	DRIVER	EDUCA-TION	INCOME	WHITE	P_VHE_ADU	NON-RES DENSITY	ROAD DENSITY	MIX
HBCW	1.00									
AGE	-0.05	1.00								
DRIVER	-0.13	0.10	1.00							
EDUCATION	-0.01	0.00	-0.19	1.00						
INCOME	0.08	-0.03	-0.20	0.30	1.00					
WHITE	-0.06	0.02	0.27	-0.14	-0.22	1.00				
P_VHE_ADU	-0.15	0.07	0.40	-0.12	-0.27	0.31	1.00			
NON-RES DENSITY	0.20	-0.12	-0.09	-0.11	0.12	-0.05	-0.21	1.00		
ROAD DENSITY	0.18	-0.10	-0.25	0.00	0.23	-0.32	-0.38	0.57	1.00	
MIX	0.14	-0.02	-0.05	-0.03	0.08	-0.10	-0.11	0.38	0.18	1.00

Table 4.3 Pearson Correlation Coefficient Matrix for Home Based Total Walk Trips (Obs. 7,524)

	TW	AGE	DRIVER	EDUCA -TION	INCOME	WHITE	P_VHE _ADU	NON-RES DENSITY	ROAD DENSITY	MIX
TW	1.00									
AGE	-0.06	1.00								
DRIVER	-0.09	0.46	1.00							
EDUCATIO N	-0.04	0.40	0.16	1.00						
INCOME	0.07	0.09	-0.12	0.25	1.00					
WHITE	-0.05	0.04	0.23	-0.09	-0.24	1.00				
P_VHE_AD U	-0.18	-0.09	0.31	-0.17	-0.31	0.38	1.00			
NON RES DENSITY	0.20	0.00	-0.03	-0.04	0.11	-0.09	-0.23	1.00		
ROAD DENSITY	0.25	0.00	-0.17	0.05	0.25	-0.41	-0.43	0.54	1.00	
MIX	0.08	-0.02	-0.05	0.00	0.07	-0.13	-0.11	0.24	0.13	1.00

Finally, though the correlation tests for all variables (land use variables, socio-economic variables, and transportation facilities), only nine variablesout of twenty three are chosen.

In both the cases (for the commuting walk trip and the total walk trip), severe correlation is not observed. However, the road density is associated with 'white' people, vehicle ownership per adult household members, and non residential units. The road density is representative of the density variables such as population density, and residential density because showsthe most associat ion amongst variables.

4.2 Model Estimation

This section estimates tworegression models using selected variables. Two models are estimated by Poisson regression at the individual level for the home based commuting trips and for the total walk trip. As indicated previously, Targa and Clifton (2005) showed a pedestrian trip generation model based on Poisson model. However, since their purpose is mainly to show the effect of the neighboring land use characteristics on the walk trips, there were statistically some insignificant variables even at 10% level of significance. As this study aims to develop the comprehensive walk trip generation model, some variables are dropped from the models in the last section.

Table 4.4 shows the parameter estimation of the home-based commuting walk trips at the individual level and summarize corresponding coefficients, z-statistics, and p-values. Likelihood ratio, Prob> chi2, and log likelihood in the table 4.4 indicate $-2[L(c) - L(\beta)]$, p-value of the overall model, and $L(\beta)$ respectively.

Likelihood ratio (-2[L(c) – L(β)] = 373.03) shows the null hypothesis that all the parameters other than the alternative-specific constant are zero is rejected at the 0.01 level of significance. Pseudo R^2 is one of the common goodness-of-fit index in maximum likelihood estimation. It is defined as 1 – (L(β) / L(0)). Basically, "pseudo R^2 is analogous to R^2 used in regression." (Ben-Akiva and Lerman, 1985). Thus, multiple linear regressions can be compared with Poisson regression in terms of R^2 and Pseudo R^2. The R^2 by linear regression is 0.066 as shown in Appendix B. Comparing to the result of linear regression, Poisson regression is likely to be more adaptable to the distribution of the number of commuting walk trips.

Table 4.4 Model Estimation [7] for the Home Based Commuting Walk Trip at the Individual Level.

Variables	Coefficient	z - value	P>z
AGE	-0.0131	-2.49	0.013
DRIVER	-0.4248	-2.23	0.026
EDUCATION	-0.4049	-2.65	0.008
INCOME	0.2980	1.95	0.052
WHITE	0.2361	1.49	0.136
P_VHE_ADU	-1.1646	-5.92	0.000
NON-RES DENSITY	0.0004	4.34	0.000
ROAD DENSITY	0.0256	7.02	0.000
MIX	0.0413	4.72	0.000
Constant	-2.2324	-6.30	0.000
Number of observations = 3,916 Likelihood Ratio = 373.03 Prob > chi2 = 0 Log likelihood = -788.74831 Pseudo R^2 = 0.1912			

[7] The method of estimation is maximum likelihood technique. As a statistical analysis tool, STATA is mainly used and TSP is used for verification.

The signs of all variables are reasonable (expected?). Since the model estimation of home based commuting walk trip is based on "only worker" in the sample, lower age among workers is associated with walking frequency to work (z-value = -2.49) Drivers, someone whose education level is less than graduation from college, and family members that own the higher percentage of vehicle per adults (P_VHE_ADU) are not likely to walk to walk (z-value = -2.23, -2.65, and -5.92, respectively). It is notable that if higher family income shows positive relationship with commuting trips by walk, although p-value of the income variable statistically does not indicates the significance at the 95% confidence level (z-value = 1.95, p-value = 0.052). The road density used instead of residential density related variables is the most significant variable. However, since non-residential dwelling units and mixed land use are associated with the frequency of walking trip to work without severe correlation with the road density, they are included in the model.

Table 4.5 presents the coefficient estimation for the total walk trips. Likelihood ratio ($-2[L(c) - L(\beta)] = 1507.78$) shows the null hypothesis that all the parameters other than the alternative-specific constant are zero is rejected at the 0.01 level of significance. It is possible to compare with the result of linear regression in Appendix B. Although four variables out of 9 variables in linear regression are not statistically significant at the 95% confidence level, the signs of coefficients are same as expected. However, overall goodness-of-fit of the Poisson model is much better than one of the linear regression in terms of R^2 and Pseudo R^2. (R^2 of linear regression is 0.084 and Pseudo R^2 of Poisson regression is 0.113.)

Table 4.5 Model Estimation for the Home Based Total Walk trip at the Individual Level.

	Coefficient	z-value	P>z
AGE	-0.0053	-4.88	0.000
DRIVER	-0.0323	-0.64	0.524
EDUCATION	-0.3082	-6.39	0.000
INCOME	0.0267	0.59	0.553
WHITE	0.4472	8.83	0.000
P_VHE_ADU	-0.7691	-13.65	0.000
NON-RES DENSITY	0.0001	4.08	0.000
ROAD DENSITY	0.0228	19.93	0.000
MIX	0.011266	3.78	0.000
Constant	-1.20822	-14.2	0.000
Number of observations = 7,524			
Likelihood Ratio = 1513.37			
Prob > chi2 = 0.00			
Log likelihood = -5872.16			
Pseudo R^2 = 0.1141			

Although the signs of coefficients in Table 4.5 are same as that in Table 4.4, two variables (DRIVER, INCOME) are not statistically significant at the 95% confidence level. Since non-workers are included in the model for total walk trips, sample size is increased by 7,524. It might cause correlation among the independent variables. However, the model validation in the next section is based on the commuting walk trips. Thus, since the model for the home based commuting walk trips should be more accurate than one for the total walk trips, two insignificant variables still remain.

Three land use variables associated with higher frequency of walk trip include non-residential density, road density, mixed land use (z-values are 4.08, 19.93, and 3.78, respectively). The road density is most significant variable in terms of z-value

(19.93) as the model for commuting walk trips. The coefficient of percentage of car ownership of household is also understandable (z-value = – 6.37).

Two models estimated in this section shows reasonable association of walk trips with socio-economic variables and land use variables. In the next section, the model for commuting walk trips is validated with the number of commuters from Census.

4.3 Model Validation

Using the estimated models and Census data, the predicted trip frequency and observed trip frequency are compared. However, there is a critical difference between the predicted values based on Model estimated and the counted data from Census. That is, the dependent variable (the home based commuting walk trip) in the estimated model is trip rate (the number of trips / person) and Census data is based on the number of people per tract (the number of people / tract). Thus, the number of walking commuters are converted to the number of trips per tract, and the dependent variable (the home based commuting walk trip) predicted with Census based independent variables is multiplied by population at the tract level. Finally, MAFE (Mean Absolute Forecast Error) is used to compare them.

Likewise, the general walk trip (the total walk trip) frequency is forecasted and validated. The trip frequency is aggregated at the tract level and multiplied by tract population. Then, similarly, the predicted trips are estimated and compared with the number of trips. Finally, MAFE are calculated.

4.3.1 Validation for the Home Based Commuting Walk Trips (HBCW)

Census provides the number of "people" who choose walking as a mode to commute rather than number of "trips" in estimated models. To convert walking "commuters" to the "trips", "returning trip factor" (RF) is suggested in this study. Table 4.10 shows the number of trips and the number of people. It is assumed that the even number of trips such as 2 trips, 4 trips and 6 trips in a day are combinations of the first home based trip and its returning trip. Based on Table 4.10, Equation 4.1 shows calculation for the returning trip factor.

Table 4.6 Trip Frequency with the Number of People for HBCW

Number of trips	Number of people	Percent
0	7690	98.3
1	35	.4
2	93	1.2
3	1	.0
4	4	.1
5	1	.0
6	1	.0
	7825	100.0

$$RF = \frac{\sum Num\_people \times Num\_trips}{\sum Num\_people} = \frac{35 + 93 \times 2 + 3 + 4 \times 4 + 5 + 6}{35 + 93 + 1 + 4 + 1 + 1} = 1.8592$$

(Equation 4.1)

As shown in Equation 4.1, the returning factor (RF) is 1.86. Once it is calculated, it is multiplied by the number of commuters from Census at tract level. Then, the number of commuters by walking is converted to the number of commuting walk trips.

On the other hand, the values of the independent variables at the tract level are input in Model 5 as shown in Equation 4.2

$$
\begin{aligned}
E(Y_i) &= The\_predicted\_commuting\_walk\_trip\_rate \\
&= Exp(-2.2324 - 0.013 \times AGE_i - 0.425 \times DRIVER_i - 0.405 \times EDUCATION_i \\
&+ 0.298 \times INCOME_i + 0.236 \times WHITE_i - 1.165 \times P\_VEH\_ADU_i \\
&+ 0.0004 \times NON\_RES\_DEN_i + 0.0256 \times ROAD\_DEN_i + 0.0413 \times MIX_i
\end{aligned}
$$

(Equation 4.2)

where, E (Y_i) is a row vector of the predicted number of home based commuting walk trips.

After the predicted commuting walk trip rate per tract is estimated, the population of the tract is multiplied. Then, the number of predicted commuting walk trips is estimated and compared with Census data. The mean absolute forecast error (MAFE) between predicted commuting walk trips based on NHTS and commuting walk trips based on Census is presented in Equation 4.3.

$$
MAFE = \frac{1}{n} \sum \left| \frac{the\_predicted\_trips - the\_observed\_trips}{the\_observed\_trips} \right|
$$

(Equation 4.3)

where, *n* is the number of tracts (if the observed trips = 0 in the tract, they are droped), *the predicted trips* is the number of trips per tract, and *the observed trips* is the number of trips per tract from Census data.

MAFE for the number of commuting walk trips are 77.5%. The result of MAFE indicates that there are slightly high errors between the estimated trip frequency from NHTS and Census data. This unexpected great discrepancy may be caused by insufficient sample size and geographically biased sample distribution for the number of commuting walk trips. Although the sample is sufficient at the individual level, the samples do not reflect the average trip behavior at each tract level when the model based on individual level is applied to the tract level (e.g., sample size is 251 trips for home based commuting walk trips and the number of tracts is 560).

On the other hand, as another way to show the error between the observed and the predicted values, a plot would be useful. Figure 4.1 shows the plotting observed versus estimated values. Points neighboring the 45 degree line indicate the lower MAFE.

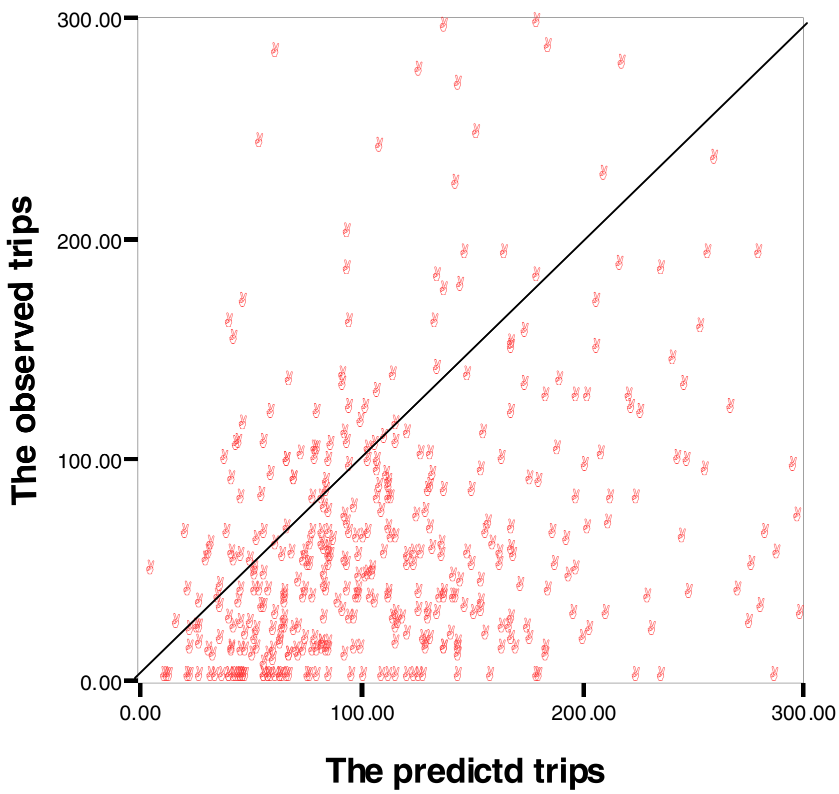


Figure 4.1 Plot of the Observed vs. the Estimated Trips of HBCW

4.3.2 Estimation for the Total number of Walking Trip per Tract

In the previous subsection, the expected value for home based commuting walk trips were compared with Census data. As indicated previously, Census data provides only the work-related travel variables. Thus, in this subsection, using the gap between the mean of observed value from Census and the mean of predicted value from NHTS, the total number of walk trips from NHTS is compared with the predicted walk trips. Even though the total number of walk trips based on NHTS is not the actual counts like Census data, it may be more reliable than the case of the number of commuting walk trips because the sample size for general trips is greater than for the commuting trips. Census weighted factor is used to estimate the total number of walk trips per tract as shown Equation 4.4.

$$\frac{\sum HBCW\_Trip\_Rate \times tract\_population}{Number\_of\_Tract} = CF \times \frac{\sum HBCW\_people \times RF}{Number\_of\_Tract}$$

(Equation 4.4)

where,

RF is the returning trip factor (1.86);

CF is the Census weight factor (1.22);

$\frac{\sum HBCW\_Trip\_Rate \times tract\_population}{Number\_of\_Tract}$ is the average number of HBCW trips from **NHTS**; and

$\frac{\sum HBCW\_people \times RF}{Number\_of\_Tract}$ is the adjusted average HBCW trips from **Census**.

The model 6 for the total walk trip is recalled for inputting the values of independent variables at tract level. Equation 4.5 shows the general walking trip formulation for estimating the expected value of total number of walking trip per tracts. the mean absolute forecast error (MAFE) is 97.2% and Figure 4.2 show the plot for the total walk trips.

$$E(Y_i) = The\_adjusted\_total\_walk\_trip\_rate$$
$$= Exp(-1.2082 - 0.005 \times AGE_i - 0.032 \times DRIVER_i - 0.308 \times EDUCATION_i$$
$$+ 0.0267 \times INCOME_i + 0.447 \times WHITE_i - 0.769 \times P\_VEH\_ADU_i$$
$$+ 0.0001 \times NON\_RES\_DEN_i + 0.023 \times ROAD\_DEN_i + 0.011 \times MIX_i$$

(Equation 4.5)

where, E (Z_i) is a row vector of the predicted number of walk trips.

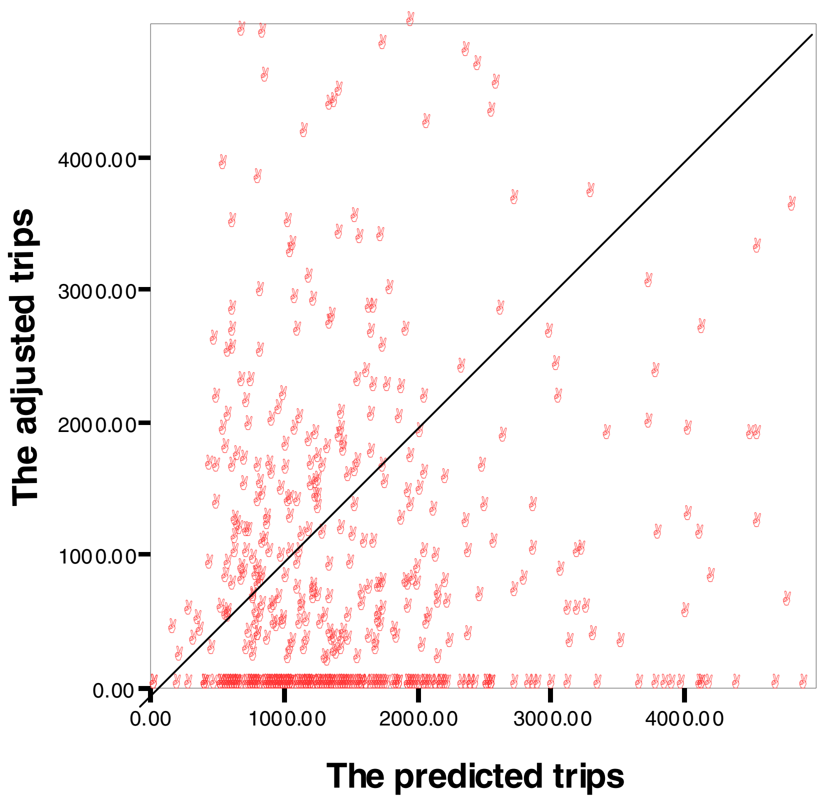


Figure 4.2 Plot of the Observed vs. the Estimated Trips of TW

Although the mean of estimated value from NHTS is not the actual count, the MAFE for the total walk trip is as reliable as the case of the home based commuting walk trip. Unlike the observed values (Census data), the adjusted trips have several zero values. As expected, MAFE for total walk trip is higher than the case of the home based commuting walk because the adjusted trips are estimated from the number of home based commuting walk trips.

4.4 Summary and Extension

Chapter 4 presents the estimation for the pedestrian trip generation with Poisson regression. Six models are estimated with different independent variables. That is, Model 1 and Model 2 are regression models that consider all variables for the home base commuting walk trips and the general walk trips, respectively. Model 3 and Model 4 are estimated to explain walking characteristics. Model 5 and Model 6 are estimated to validate the model and predict the number of trips.

Since these trip generation models are developed at the individual level, the number of predicted trips at the Census tract show remarkable error (MAFE for the home based commuting walk trip: 77.50%, MAFE for the total walk trip: 97.2%). The error appears due to the number of samples for the home based commuting walk trips. Since 251 trips for the home based commuting walk trip are distributed to 560 tracts, several tracts do not have home based commuting walk trips.

Chapter 5: CONCLUSIONS AND FUTURE STUDY

Since the non-motorized transportation system has been promoted and concerned by the government, it is important for transportation engineers, urban planners and policy analysts to investigate critical factors which influence on the individual's walking choice. Although several studies showed reasonable relationships between non-motorized transportation and land use factors or socio-economic factors, most of them have not developed appropriate models to estimate non-motorized traffic demand.

To model non-motorized traffic demand, in Chapter 4, this thesis showed how socio-economic and land use factors influence on the walking trips, using Pearson's correlation. Two models are estimated considering nine critical factors and consequently showed statistically significant relationship with them; age, driver, education, percentage of the number of vehicles per adults (P_VEH_ADU), non-residential density, road density, and mixed land use (MIX).

Although the sample of home based commuting walk trips was enough to be modeled at the individual level, the sample size was not enough to be aggregated at the tract level. Thus, the model validation for the home based commuting walk trip at the tract level showed the high error in terms of mean absolute forecast error (MAFE). On the other hand, since the actual total walk trips per tract are not provided from Census, the validation for the total walk trip is not actually possible, but may be conducted since the Census factor (CF) can be developed to adjust the NHTS data to Census data. In this case, it should be assumed that the total walk trip has the similar pattern to the home based commuting walk trip in terms of the model estimation results.

The model estimated and validated in this study can be applied to other areas, since the coefficients of land use variables in the model reflect the change of land use between Baltimore metropolitan region and the proposed area. It should be very quick and simple process. However, this transferability might be not precise because the model is estimated from NHTS Baltimore add-on. If other jurisdictions out of Baltimore region need this model and more precise pedestrian demand, it is recommended to consider the change of variables such as road density and non-residential dwelling units.

The researcher recognizes two more limitations. First, the sample size of the home based commuting walk trip should be larger than the number of tracts, in order to validate at the tract level. In this study, having a smaller sample size than the number of tracts caused the several tracts in which nobody walks. Second, it is recommended that the zero walk trips be processed separately when being modeled at the tract level because a lot of zero walk trips cause an increase statistical noise.

To overcome these limitations, the researcher suggests the following analysis methods. "The zero inflated Poisson model" can be applied to describe over-dispersed zero walk trips instead of the general Poisson model. In addition, since transit users who access transit stops by walking are not recorded as walkers in the survey data, they are excluded in the modeling procedures. However, they can be considered as additional walking trips by a simple data manipulation. This approach might increase the sample size for the home based commuting walk trip to overcome the lack of data and validate models at the tract level.

Finally, pedestrian demand models can be applied to a variety of research problems and practical applications. First of all, local government can benefit from the model estimated in this study when they need to predict the amount of construction cost for pedestrian facilities at the tract level to metropolitan level. This model helps designing local transit lines since the stops of the transit line should be selected near the place where pedestrians are generated and attracted. The number of pedestrian should be satisfied with the management cost estimated by transit companies or subsidies estimated by local government. In addition, the results from this study can be used for several research topics such as the impact of land development patterns and urban design on travel behavior, connections between built environment, physical activity and public health outcomes, assessment of potential transit markets, and understanding pedestrian risks. For all applications, the sensitivity analysis can be performed to predict the change in number of walk trips by changing the independent variables.

APPENDICES

APPENDIX A

The estimates of the parameters are found by maximizing the likelihood.

$$L = \prod_{i=1}^{n} P(y_i) = \prod_{i}^{n} \frac{e^{-(\mu_i)}(\mu_i)^{y_i}}{y_i!}$$

However, the log transformation of the likelihood function is used because the function is monotonically increasing. We will maximize the log-likelihood function rather than maximizing *L*: This model can be estimated with the standard maximum likelihood method well organized to yield unbiased estimates for those parameters.

$$\begin{aligned}
LogL &= \sum_{i=1}^{n} P(y_i) = \sum_{i=1}^{n} \left[-\mu_i + y_i \log(\mu_i) - \log(y_i!) \right] \\
&= \sum_{i=1}^{n} \begin{bmatrix} -\exp\left(c + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \cdots + \beta_k \times x_{ki}\right) \\ + y_i \times \log(c + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \cdots + \beta_k \times x_{ki}) - \log(y_i!) \end{bmatrix} \\
&= \sum_{i=1}^{n} \left[-\exp\left(c + \sum_{j=1}^{k} \beta_j \times x_{ji} \right) + y_i \times \left(c + \sum_{j=1}^{k} \beta_j \times x_{ji} \right) - \log(y_i!) \right]
\end{aligned}$$

where $B = (b_1, b_2, \cdots, b_k)$ representing unknown parameters to be estimated.

APPENDIX B

1. Model estimation for the home based commuting walk trip by linear regression

	Coefficient	t - value	P>t
AGE	-0.0005	-1.24	0.214
DRIVER	-0.1067	-4.58	0.000
EDUCATION	-0.0164	-1.39	0.164
INCOME	0.0180	1.46	0.144
WHITE	0.0187	1.23	0.220
P_VHE_ADU	-0.0444	-3.35	0.001
NON-RES DENSITY	0.0001	5.20	0.000
ROAD DENSITY	0.0011	3.05	0.002
MIX	0.0131	4.66	0.000
Constant	0.1623	4.83	0.000

Number of obs	=	3916
F(9, 3906)	=	30.53
Prob > F	=	0
R-squared	=	0.0657
Adj R-squared	=	0.0636
Root MSE	=	0.34344

2. Model estimation for the home based total walk trip by linear regression

	Coefficient	t-value	P>t
AGE	-0.0019	-3.25	0.001
DRIVER	-0.0311	-1.12	0.261
EDUCATION	-0.0816	-3.43	0.001
INCOME	0.0127	0.55	0.579
WHITE	0.1769	6.21	0.000
P_VHE_ADU	-0.1954	-7.92	0.000
NON-RES DENSITY	0.0001	4.86	0.000
ROAD DENSITY	0.0094	12.76	0.000
MIX	0.0085	2.77	0.006
Constant	0.254328	5.57	0

Number of obs	=	7524
F(9, 7,514)	=	76.96
Prob > F	=	0
R-squared	=	0.0844
Adj R-squared	=	0.0833
Root MSE	=	0.90844

BIBLIOGRAPHY

Bates, J. 2000. Chapter 2, Handbook of Transport Modelling.

Batty M. 2001. Agent-based Pedestrian Modeling, Environment and Planning B: Planning and Design 28 321-326

Bauman, A.E., Sallis, J.F., Dzewaltowski, D.A., and Owen, N. 2002. Toward a Better Understanding of the Influences on Physical Activity, American Journal of Preventive Medicine.

Ben-Akiva, M. and Lerman, S.R. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press.

Bradshaw C. 1993. Creating and using a Rating System for Neighborhood Walkability: Towards an Agenda for "Local Heroes. 14th International Pedestrian Conference.

Cervero, R. 1989. Americas' Suburban Centers – The Land Use-Transportation Link. Unwin Hyman, Boston, Mass., pp. 137-142.

Cervero, R. and Radisch, C. Travel choices in pedestrian versus automobile oriented neighborhoods, University of California Transportation Center.

Comsis Corporation. 1983. UTPS Highway Network Development Guide, U.S. Department of Transportation and Federal Highway Administration.

Dixon L. 1996. Bicycle and pedestrian level of service performance measures and standards for congestion management systems, Transportation Research Record 1538: 1-9.

DeMaris, A. 2004. Regression with Social Data: Modeling Continuous and Limited Response Variables. Wiley.

Ewing, R. 2003. Measuring the Health Effects of Sprawl, Smart Growth America.

Ewing, R. and Cervero, R. 2001. Travel and the Built Environment: A Synthesis, Transportation Research Record 1780 : 01-3515.

Federal Highway Administration. 1994. Final Report: The National Bicycling and Walking Study, publication No. FHWA-PD-94-023.

Federal Highway Administration. 1999. Guidebook on Methods to estimate Non-Motorized Travel: Supporting Documentation, United States Department of Transportation, Publication No. FHWA-RD-98-166, Virginia

Federal Highway Administration. 1999. Guidebook on Methods to estimate Non-Motorized Travel: Overview of Methods, United States Department of Transportation, Publication No. FHWA-RD-98-165, Virginia

Hutchinson, B.G. 1974. Principles of Urban Transport System Planning, McGraw-Hill.

Kerridge J, Hine J, Wigen M. 2001. Agent-based modeling of pedestrian movements: the questions that need to be asked and answered, Environment and Planning B: Planning and Design 28 327-341

Kitamura, R., L. Laidet, and P. Mokhtarian. 1997. A Micro-Analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area. Transportation, Vol. 24, pp. 125-158.

Knapp, G.J. and Song, Y. 2002. Measuring Urban Form: Is Portland Winning the War of Sprawl? Journal of American Planning Association, forth coming.

Landis, B, Vatticuti, V, and Ottenberg, R. 2001. Modeling the pedestrian environment: a pedestrian level of service. Transportation Research Record 1678: 38-47.

Matlick, Julie Mercer. 1996. If We Build it, Will They Come? #69 Forecasting Pedestrian Use and Flows, Forecasting the Future, Bicycle Federation of America, Pedestrian Federation of America, Pro Bike/Pro Walk '96, 1996, pp.315-319.

Meyer, M.D. and Miller, E.J. 2001. Urban Transportation Planning 2nd Edition, Mcgraw Hill.

Moudon, A.V, Hess, P.M., Snyder, M.C., and Stanilov, K. 1997. Effects of site design on pedestrian travel in mixed-use, medium-density environments.,Transportation Research Record 1578.

Moudon, A.V. and Lee, C. 2003. Building the methods: Walking and Bicycling: An Evaluation of Environmental Audit Instruments. American Journal of Health Promotion.

Oi, W.Y. and Shuldiner, P.W. 1962. An Analysis of Urban Travel Demands, Northwestern university press.

Porter, C., Suhrbier, J., and Schwartz, W. L. 1999. Forecasting Bicycle and Pedestrian Travel: State of the Practice and Research Needs. Transportation Research Reocord 1674, Paper No. 99-0750.

Pushkarev, B. and Jeffrey, M.Z. 1971. Pedestrian Travel Demand, Highway Research Record 355, Washington, D.C

Raford, N. 2003. Looking both ways: Space Syntax for pedestrian exposure forecasting and collision risk analysis, 4th International Space Syntax Symposium London.

Rajamani, J., Bhat, C.R., Handy, S., Knaap, G., and Song, Y. 2003. Assessing the impact of urban form measures in nonwork trip mode choice after controlling for demographic and level-of-service effects, TRB.

Richard J.J.and Kochtitzky,C. 2002. Creating A Healthy Environment, Sprawl Watch Clearinghouse Monograph Series.

Rothman K. J. 1986. Modern Epidemiology. Little, Brown, Boston.

Rutherford, G.S., E. McCormack, and M. Wilkinson. 1996. Travel Impacts of Urban Form: Implications from an Analysis of Two Seattle Area Travel Diaries. Presented at TMIP Conference on Urban Design, Telecommuting, and Travel Behavior, FHWA, U.S. Department of Transportation,

Targa, F. and Clifton, K. 2005. Bulit Environment and Trip Generation for Non-motorized Travel, Journal of Transportation and Statistics

Teknomo. 2002. Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model, Ph.D Dissertation, Department of Human Social Information Sciences, Graduate School of Information Sciences, Tohoku University, Japan

Turner Fairbank Highway Research Center. 1998. Development of the Bicycle Compatibility Index: A Level of Service Concept, Volume 1: Final Report. Publication FHWA RD-98 XXX. U.S. Department of Transportation.

Urbitran Associates. 2004. Pedestrian Flow Modeling for Prototypical Maryland Cities, Maryland Department of Transportation Division and National Center for Smart Growth, University of Maryland.