Travel Mode Choice Framework Incorporating Realistic Bike and Walk Routes

by

Joseph Paul Broach

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Urban Studies

Dissertation Committee:
Jennifer Dill, Chair
James Strathman
Kelly Clifton
Robert Fountain

Portland State University
2016

Abstract

For a number of reasons—congestion, public health, greenhouse gas emissions, energy use, demographic shifts, and community livability to name a few—the importance of walking and bicycling as transportation options will only continue to increase. Currently, policy interest and infrastructure funding for nonmotorized modes far outstrip our ability to model bike and walk travel. To ensure scarce resources are used most effectively, accurate models sensitive to key policy variables are needed to support long-range planning and project evaluation, and to continue adding to our growing understanding of key factors driving walk and bike behavior. This research attempts to synthesize and advance the state of the art in trip-based, nonmotorized mode choice modeling.

Over the past fifteen years, efforts to model the decision to walk or bike on a given trip have been hampered by the lack of a comprehensive behavioral framework and inconsistency in measurement scales and model specification. This project develops a mode choice behavioral framework that acknowledges the importance of attributes along the specific walk and bike routes that travelers are likely to consider, in addition to more traditional area-based measures of travel environments. The proposed framework is applied to a revealed preference, GPS-based travel dataset collected from 2010-2013 in Portland, Oregon. Measurement of nonmotorized trip distance, built environment, tour-level variables, and attitudinal attributes as well as mode availability are explicitly addressed. Route and mode choice models are specified using discrete choice techniques,

and predicted walking and bicycling routes are tested as inputs to various mode choice models.

Results suggest strong potential for predicted route measures to enhance walk and bicycle mode choice modeling. Findings also support the specific notion that bicycle and pedestrian infrastructure contribute not only to route choice but also to the choice of whether to bike or walk. For decisions to bicycle, availability of low-traffic routes may be particularly important to women. Model results further indicate that land use and built environments around trip ends and a person's home still have important effects on nonmotorized travel when controlling for route quality. Both route and area travel environment impacts are mostly robust to the inclusion of residential self-selection variables, consistent with the idea that built environment differences matter even for households that choose to live in a walkable or bikeable neighborhood. The combination of area and route-based built environment measures alongside trip context, sociodemographic, and attitudinal attributes provides a new perspective on nonmotorized travel behavior relevant to both policy and practice.

iv

pursuing it in the first place. And, with that, I would also thank Portland itself, and all those past and present who have made it the special place it is. I am not sure that any other city could have kindled this research.

Table of Contents

List of Tables

List of Figures

Introduction

Bicycle and Pedestrian Policy, Funding, and Modeling in the US, 1990s-Present

Over the past twenty-five years, bicycle and pedestrian travel have enjoyed increasing policy support. From 1992-2009, federal funding for walking and biking rose from less than 0.25% to 2% of the total transportation budget (FHWA, 2010b). In addition, the 1998 Transportation and Equity Act for the 21st Century (TEA-21) mandated that both pedestrian and bicycle access be considered in long-range regional transportation plans (FHWA, 2008). A nationwide combined bike and walk mode share of 15 percent has been set as a federal policy goal (FHWA, 2008). Some local agencies have set even more ambitious goals. For example, Portland, Oregon's 2030 Bicycle Master Plan sets a bike mode share target of 15 percent by 2020 and 25 percent by 2030 (Portland Bureau of Transportation, 2010). The Federal Highway Administration (FHWA, 2010b) noted that in addition to ongoing traffic congestion and air quality concerns, increased cycling and walking also have the potential to positively affect a raft of 21st century issues: public health, greenhouse gas emissions and energy use, aging populations, and community livability. Given the outcomes at stake, it seems unlikely that the push toward improvements for cycling and walking will subside any time soon.

Predicting nonmotorized travel with travel demand models has lagged behind policy and funding support, although consideration of bike and walk travel has become more common in recent years. Two reviews of the state of the practice conducted in the late 1990s found weak empirical foundations, incomplete theory, few efforts even to begin integration into regional models, and mostly ad hoc methods (Porter, Suhrbier, & Schwartz, 1999; Replogle, 1997). A decade later, the Committee for Determination of the

State of the Practice in Metropolitan Area Travel Forecasting (2007) found that more than half of large Metropolitan Planning Organizations (MPOs) were including walk and/or bike choices in mode choice models ("few" medium MPOs and "almost no" small MPOs were doing so). An even more recent survey by Singleton and Clifton (2013) reported that two-thirds of the 48 largest MPOs included non-motorized modes in their regional models (although less than half differentiated bike and walk). No MPO reported developing a pedestrian route choice or assignment model, however. Since that time, SFCTA (San Francisco) has announced plans to incorporate predicted walk routes in their regional travel model in the future (Bomberg, Zorn, & Sall, 2013). Kuzmyak et al. (2014) provides an updated overview of the state of the practice and notes considerable interest in better integrating nonmotorized travel in travel demand models.

The academic literature has also seen some progress in modeling nonmotorized mode choice; however, most work has not considered regional modeling needs. Instead, models have been constructed to answer specific research questions related to the built environment, children's school travel, or public health.

Models to date have rarely included route-level infrastructure measures, even though a large portion of federal funding for nonmotorized travel goes to improve walking and biking travel networks. For example, about 80 percent of the $100 million Nonmotorized Transportation Pilot Program (NTPP) had been spent to improve or expand on-street and off-street infrastructure (FHWA, 2010a). Even the few models that are sensitive to nonmotorized infrastructure improvements consider the impact only along shortest path routes or in diluted form as part of larger area aggregates. Given a new bike

lane connecting point A and point B, only one existing study out of over twenty reviewed (Chapter 1) could predict a meaningful change in the competitiveness of cycling for a trip from A to B.

Finally, bicycle and pedestrian mode choice has been considered primarily using a trip-based approach. In this framework, choice decisions are made independently from one stop to the next. Meanwhile, the transition from trip-based, four-step models dating to the 1960s toward more advanced tour-based activity models is already underway. The Committee for Determination of the State of the Practice in Metropolitan Area Travel Forecasting (2007) found that 38 percent of large MPOs and 24 percent of all MPOs already were considering moving to the next generation models in the near-term. Promising reduced aggregation and stronger behavioral foundations, activity-based models offer potential advantages for nonmotorized modeling. For that potential to be realized, however, bicycle and pedestrian modeling has some catching up to do.

Toward an Improved Nonmotorized Mode Choice Framework

This research attempted to tie together the threads of existing bicycle and pedestrian mode choice modeling with the emerging area of walk and bike route choice modeling to create a new mode choice framework that better incorporated nonmotorized alternatives. The goal was a choice framework that is behaviorally realistic, policy sensitive, and feasible to implement in regional models. A key modification to existing models was the incorporation of predicted walk and bike routes within the mode choice decision process. In addition, questions of appropriate measurement scale and model

3

specification for nonmotorized modes were addressed. Finally, consideration was given to tour-level variables in the context of nonmotorized travel.

Trip-level mode choice in the context of more complex tours was the focus of this study. Much existing work on built environment and public health policy impacts has focused on more aggregate outcome measures such as total number of walk or bike trips taken by individuals or groups. To develop a behavioral framework, however, the most disaggregate decision units—trips and tours by individuals—seemed most appropriate.

The mode choice modeling framework developed was tested using a revealed preference family activity and travel dataset collected from 2010 to 2013 in Portland, Oregon. The dataset included detailed Global Positioning System (GPS) travel data and a rich set of household and person-level characteristics that aided in developing the proposed framework. Importantly, the levels of bicycling and walking in the sample were great enough to allow meaningful inferences about each of these modes of interest. Although it was acknowledged that the unique sample would not necessarily generalize to other populations, the model framework itself was developed with broader applications in mind.

This research sought to contribute to the existing trip-level nonmotorized travel demand modeling literature primarily in the following ways:

- Use GPS-based travel data to develop a mode choice model without supporting travel diaries.

- Develop a route choice model for utilitarian pedestrian travel using revealed preference GPS data.

- Construct a mode choice modeling framework that incorporates predicted walk and bike routes to represent the relevant travel environment for a trip.

- Compare the use of area versus route-based mode choice factors.

- Compare the impacts of specific bicycle infrastructure on decisions of where to bike versus decisions of whether to bike for specific trips, including testing for interaction effects with gender.

The remainder of the dissertation is organized as follows. Chapter 1 presents a survey of nonmotorized mode choice modeling literature. In Chapter 2, I develop and operationalize a behavioral framework for modeling mode choice by modifying and extending existing trip-based travel demand modeling methods. Chapter 3 then describes the GPS test dataset, how it was processed into trips, routes, and tours, and overviews the supplementary survey data collected on common destinations, socio-demographics, and residential self-selection. Chapter 4 describes the development and adoption of pedestrian and cyclist route choice models. In Chapter 5, I apply the new mode choice framework to the GPS travel data, and report on the estimation of a series of mode choice models including a numerical example applied to a hypothetical policy change. Chapter 6 discusses initial findings, limitations, and promising areas for future research to improve, extend, and apply this work to both regional travel demand modeling and continued academic inquiry.

1 Literature Review

Overview of Nonmotorized Mode Choice Modeling

Most existing efforts to model walk and bike mode choice have focused on one of three primary research questions:

- How do land-use and the built environment affect (usually adult) travel?
- What factors determine children's travel mode to school?
- How can we integrate bike and walk travel into regional mode choice models?

Each question can be further divided by the aggregation of actors and trips in the analysis. Aggregate models consider the total results of many actors taking many trips. The units of analysis in an aggregate study might be neighborhoods, school zones, cities, regions, or nations. Disaggregating the actors into households or individuals leads to what this paper will term "disaggregate actor" models. Reducing the unit of analysis to individual trips results in "disaggregate trip" models. Previous reviews on this topic have not done a good job distinguishing the two disaggregate types, but the distinction is important. Table 1 provides some examples by topic and scale.

This review considers disaggregate trip mode choice studies that have explicitly modeled walk and/or bike travel as alternatives and have included attributes from at least one of the following categories: built environment, attitudes, and tours. Modeling individual decisions allows for more behaviorally realistic models and attenuates risks of ecological fallacy. Using the trip or tour as unit of analysis also allows for more detailed spatial and temporal explanatory variables. The resolution and extent of relevant travel environment attributes is a particular focus of this study, since, as previous research has

noted, nonmotorized trips tend to cover relatively small distances and might be

influenced by smaller scale phenomena (Moudon & Lee, 2003). Modeling bike and walk

options as alternatives to other modes fits with existing regional modeling

Table 1 Examples of Mode Choice Model Questions by Topic and Aggregation Level

| | Scale | | |
Focus	Aggregate	Disaggregate Actor	Disaggregate Trip
Built environment	Does neighborhood density affect the proportion of walk trips by neighborhood residents?	Does density around a residence affect the share of walk trips taken by that household?	Does density at a trip's origin or destination affect the likelihood that a person walks on the trip?
School travel	Does sidewalk coverage in a school zone affect the share of students that walk to school?	Does sidewalk coverage between home and school change the probability that a child walks to school?	Does weather affect whether a child walks to school on a given trip?
Regional modeling	Does the proportion of college students in a region affect bike mode share?	Does student status affect the proportion of trips an individual takes by bike?	Does student status change the odds that a person will bike on a given trip?

practice and the dominant microeconomic theory of travel choice behavior.

With those criteria in mind, peer reviewed articles and technical reports were

screened mainly from reviews covering each of the three major research themes

identified: built environment (Badoe & Miller, 2000; Ewing & Cervero, 2010; Handy,

Boarnet, Ewing, & Killingsworth, 2002), school travel (Davison, Werder, & Lawson,

2008; Pont, Ziviani, Wadley, Bennett, & Abbott, 2009), and nonmotorized regional

modeling (Porter et al., 1999). Online research databases were also consulted to uncover

more recent articles. Table 2 provides a summary of the articles reviewed.

The remainder of this section considers how existing studies have addressed key nonmotorized modeling issues including: measuring time/distance, built environment, trip/tour variables, and attitudes; determining bike and walk availability; and, specifying the choice model structure.
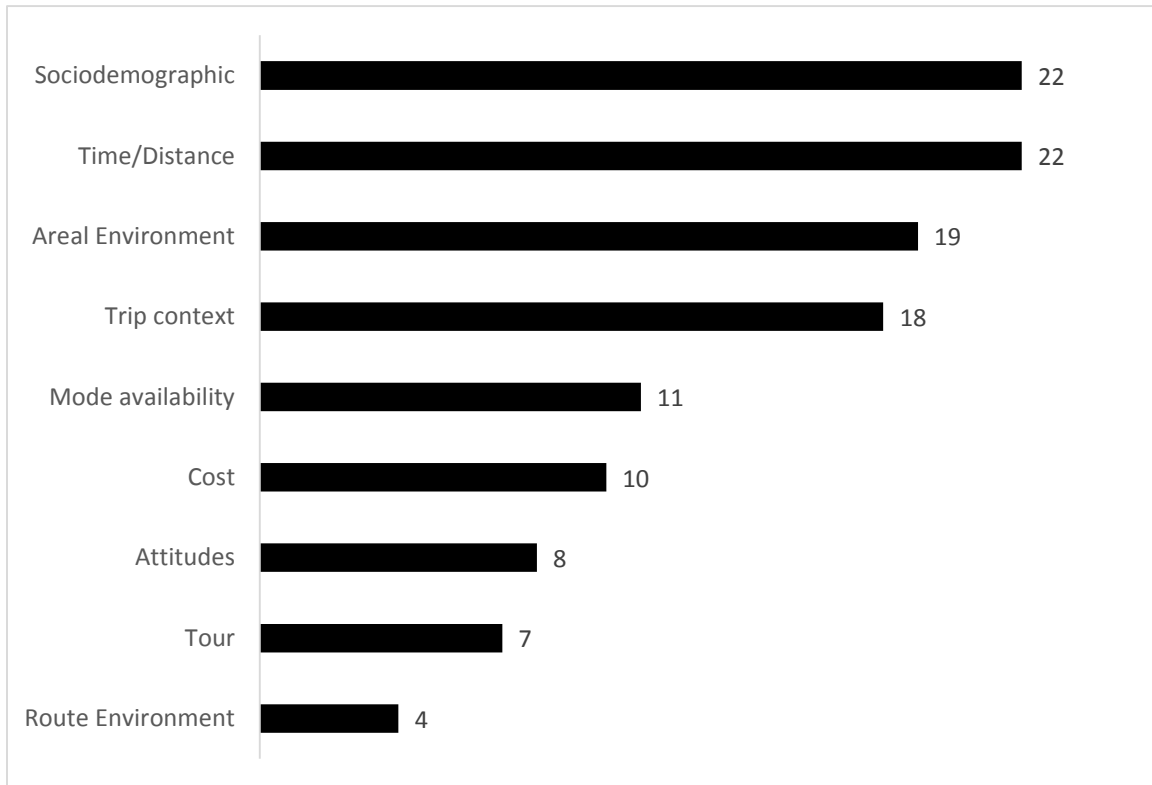
Figure 1 Attribute types found in existing disaggregate trip studies reviewed, 1998-2015, $n = 24$

Table 2 Selected Disaggregate Trip and Tour Nonmotorized Mode Choice Research

Authors	Date	Loc.	Diary	Focus[1]	Age	Modes[2]	*n* bike	*n* walk	Purpose[3]	Models[4]	Time Dist	Cost	Areal Env	Route Env	Soc-Dem	Attitudes	Trip	Tour	Avail
This Research	**2015**	**Portland, OR**	**N**	**RM**	**Adult**	**W,B,A,T**	**1501**	**1419**	**Any**	**MNL**	**X**	[5]	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Akar et al.	2013	Columbus, OH	N	Oth	18+	W,B,A,T	159	233	HBCol	MNL	X		X		X	X	X		
Bergman et al.	2011	Oregon	N	Oth	18+	W,B,A,T	n/a		TAcc	MNL NL	X	X	X	X	X	X	X		X
Black et al.	2001	UK	N	ST	<12	W,A	0	1202	HBSch	BL	X				X	X			
Bradley & Bowman	2005	Sacramento	Y	RM	All	W,B,A,T	196	538	HBW HBS HBN WWOW	NL	X	X	X		X		X	X	X
Cervero & Duncan	2003	San Francisco	Y	BE	All	W,B,A/T	118	980	HBN WNH NW	BL	X		X		X		X		X
DiGuiseppi et al.	1998	London	N	ST	6-10	W,A	0	2294	HBSch	BL	X				X	X	X		
Ermagun & Samimi	2014	Tehran	N	ST	12-17	W,A,SB,T	0	n/a	HBSch	NL	X	X	X		X	X	X		
Ewing et al.	2004	Gainesville FL	Y	ST	<16	W,B,A,SB	24	32	HBSch	MNL NL	X		X		X				X

10

Authors	Date	Loc.	Diary	Focus[1]	Age	Modes[2]	*n* bike	*n* walk	Purpose[3]	Models[4]	Time Dist	Cost	Areal Env	Route Env	Soc-Dem	Attitudes	Trip	Tour	Avail
This Research	**2015**	**Portland, OR**	**N**	**RM**	**Adult**	**W,B,A,T**	**1501**	**1419**	**Any**	**MNL**	**X**	**[5]**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Frank et al.	2008	Seattle	Y	BE	n/a	W,B,A,T	187	706	HBW HBNW WWOW	MNL	X	X	X		X		X	X	X
Kockelman	2007	San Francisco	Y	BE	19+	W/B, A/T	5212		Any	BL	X		X		X				
Lin & Chang	2009	Taipei	N	ST	6-12	W,A,T, Moto	132		HBSch	NL	X		X	X	X		X	X	
McMillan	2007	California	N	ST	8-11	W/B,A	n/a		HBSch	BL	X		X		X	X			
Miller et al.	2005	Toronto	Y	RM	All	W,A,T	0	1202	HBW HBNW WWOW	SIM	X	X					X	X	X
Piatkowski & Marshall	2015	Denver	N	BE	Adult	B	n/a	n/a	HBW	BL	X		X		X	X	X		
Rajamani et al.	2003	Oregon	Y	BE	All	W,B,A,T	28	143	HBNW	MNL	X	X	X		X		X		X
Reilly & Landis	2002	San Francisco	Y	BE	All	W,A,T	n/a		HBNW	MNL			X		X		X		

Authors	Date	Loc.	Diary	Focus[1]	Age	Modes[2]	*n* bike	*n* walk	Purpose[3]	Models[4]	Time Dist	Cost	Areal Env	Route Env	Soc-Dem	Attitudes	Trip	Tour	Avail
This Research	**2015**	**Portland, OR**	**N**	**RM**	**Adult**	**W,B,A,T**	**1501**	**1419**	**Any**	**MNL**	**X**	[5]	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Rodríguez & Joo	2004	Chapel Hill	N	BE	17+	W,B,A,T	59	88	HBW/ HBCol	MNL NL HEV	X	X	X	X	X			X	X
Roorda et al.	2009	Toronto	Y	RM	11+	W,B,A,T,SB,TX*	696	1953	Any	SIM	X	X	X		X		X	X	X
Saneinejad et al.	2012	Toronto	Y	Other	17+	W,B,A,T	612	2087	HBW	MNL	X	X	X		X		X		X
Schlossberg et al.	2006	Oregon	N	ST	11-14	W,B,A/SB	29	86	HBSch	BL	X			X			X		
Schwanen & Mokhtarian	2005	San Francisco	N	BE	18+	W/B,A,T	35		HBW	MNL					X	X			
Soltani & Allan	2006	Adelaide Australia	Y	BE	All	W/B,A,T	295		Any	MNL	X		X		X		X	X	
Yarlagadda & Srinivasan	2008	San Francisco	Y	ST	<18	W,B,A,T,SB	220	1265	HBSch	MNL	X		X		X		X		X
Zhang	2004	Boston	Y	BE	17+	W/B,A,T	168		HBW HBNW	MNL	X	X	X		X		X		

Authors	Date	Loc.	Di-ary	Fo-cus[1]	Age	Modes[2]	*n* bike	*n* walk	Pur-pose[3]	Mod-els[4]	Time Dist	Cost	Are-al Env	Rou-te Env	Soc-Dem	Atti-tudes	Tr-ip	To-ur	Av-ail
This Research	**2015**	**Port-land, OR**	**N**	**RM**	**Ad-ult**	**W,B, A,T**	**1501**	**1419**	**Any**	**MNL**	**X**	**5**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

[1] BE = Built Environment, ST = School Travel, RM = Regional Modeling

[2] W = Walk, B = Bike, A = Auto, T = Transit, SB = School Bus, TX = Taxi, slash indicates combined modes (e.g. W/B = Walk/Bike combined)

[3] TAcc = Transit Access, HBSch = Home-Based School, HBW = Home-Based Work, HBS = Home-Based Social/Recreational, HBNW = Home-Based Non-Work, WOW = Work Other Work, NHNW = Non-Home Non-Work, HBCol = Home-Based College

[4] MNL = Multinomial Logit, NL = Nested Logit, BL = Binary Logit, SIM = Simulation, HEV = Heteroscedastic Extreme Value,

[5] Proxies for parking cost with parking charge area in central city

Variable Measurement

Nonmotorized travel takes place on different transportation networks and at different speeds than motorized travel. Pedestrians and cyclists are also more exposed to the travel environment between origin and destination. Models incorporating these travel modes therefore require different and often more detailed datasets than are typical for auto and transit travel.

Time and distance. Nonmotorized distance has been measured using zonal network distance, geometric approximation, point to point shortest path, and self-reported distance. Zone-based shortest path distances from regional travel model outputs have been used by some studies (Ewing, Schroeer, & Greene, 2004; Rajamani, Bhat, Handy, Knaap, & Song, 2003; Zhang, 2004).[1] Depending on zone size and how intrazonal distances are calculated, aggregation to zones may result in poor distance estimates for short trips. Euclidean (Kockelman, 1997; Miller, Roorda, & Carrasco, 2005) and Manhattan (Roorda, Passmore, & Miller, 2009; Saneinejad, Roorda, & Kennedy, 2012) geometric approximations have also been used. These network-free methods are simple to calculate, but they are insensitive to differences or changes in street and trail layouts. Solving origin to destination distance along the shortest network path solves the latter issue (Akar, Fischer, & Namgung, 2013; Bergman, Gliebe, & Strathman, 2011; Rodríguez & Joo, 2004; Schlossberg, Greene, Phillips, Johnson, & Parker, 2007). The assumption that nonmotorized users will use the shortest route is questionable; for

---

[1] Zone-based measurements assign all potential trip ends within each zone to a single point, termed the zone's centroid. The distance between any two points in different zones is the network distance between the zones' centroids. Trips starting and ending in a single zone would have an implied distance of zero such that rules of thumb have to be applied. For example, Portland Metro has used 0.73 times the distance to the nearest neighboring zone to approximate intrazonal distance.

instance, Broach, Gliebe, and Dill (2012) observed utility cycling trips and reported that half exceeded the shortest path distance by at least 10%. Finally, self-reported bike or walk distance has been used, mostly in school travel studies (Black, Collins, & Snell, 2001; DiGuiseppi, Roberts, Li, & Allen, 1998; Ermagun & Samimi, 2015; Lin & Chang, 2010; McMillan, 2007). Reported distance has the advantage of potentially measuring the actual route considered; however, it is probably not practical to collect as part of regional travel surveys. Respondents would need to supply expected travel times for each mode on every trip taken.

Distance is often converted to walk or bike travel time, most often by applying a constant speed assumption across the sample. Assumed speeds ranged from 2.5 to 3.1 miles per hour (4-5 km/h) for walk and 9 to 12.4 miles per hour (14.5-20 km/h) for bike in the studies surveyed here (Bergman et al., 2011; Ewing et al., 2004; Rajamani et al., 2003; Rodríguez & Joo, 2004; Roorda et al., 2009; Saneinejad et al., 2012; Zhang, 2004). Only two tried adjusting speeds to reflect socio-demographic or environmental factors. Zhang (2004) used an ad hoc age adjustment factor for individuals over and under 30 years old. Rodríguez & Joo (2004) included time adjustment variables calculated from the slope along the shortest path route. Many delay factors were also embedded in the utility-weighted distance used by Bergman et al. (2011) to estimate bike travel time.

The relatively short distances covered by many nonmotorized trips make aggregate, zonal distance measures less reliable. At the same time, distances need to be consistent across travel modes. The relative importance of these conflicting goals has not been tested in existing work.

14

Cost. Nonmotorized modes are typically assumed to have no monetary marginal cost. Surprisingly, over half of reviewed studies either extended this assumption to auto and transit travel or found that motorized mode costs were not significant model variables. A number of studies included transit fares (Bradley & Bowman, 2006; Frank, Bradley, Kavage, Chapman, & Lawton, 2008; Miller et al., 2005; Rajamani et al., 2003; Rodríguez & Joo, 2004; Roorda et al., 2009; Saneinejad et al., 2012; Zhang, 2004). Fuel, parking, and/or toll costs were specified for auto trips (Frank et al., 2008; Rajamani et al., 2003; Rodríguez & Joo, 2004; Roorda et al., 2009; Saneinejad et al., 2012; Zhang, 2004).

Built Environment. This paper follows Handy, Boarnet, Ewing, and Killingsworth (2002) and includes under the built environment umbrella urban design, land-use, and transportation infrastructure.[2] Disaggregate trip mode choice models have with few exceptions focused on built environments at one or both trip ends. This seems a reasonable approach for enclosed, motorized modes for which the intermediate, "line haul" portion of the trip is accounted for more or less completely by time and money costs. For walking or biking, however, ignoring the immediate environment along the travel route itself is more troubling. Relative to drivers and passengers, pedestrians and cyclists are much more exposed to the built environment experienced between a given origin and destination pair. Figure 2 presents the methods that have been used to capture trip-level built environment attributes, graphically highlighting some potential shortcomings.

---

[2] Handy et al. (2002) also include dynamic transportation features in the built environment (e.g. traffic volume, transit frequency), but these are instead considered as "trip" variables here.

Zone-based: Aggregate built environment into irregular zones around trip origin and destination (may not cover entire trip).	Buffer: Aggregate built environment into circular or network-based polygon buffers around trip origin and destination (may not cover entire trip).	Route: Measure built environment around or along shortest path or actual (reported) path (shortest path may not correspond to actual path; reported path may not correspond to actual path for all modes).

Actual route ········· **Measurement area** ----- **Origin/destination** ◉

Figure 2 Different approaches to measuring trip-level built environment

Table 3 lists built environment variables from existing mode choice studies by measurement scale. Reflecting a lack of consensus, only one variable has been measured at the same scale in different studies (intersection density in combined origin and destination zones). Land-use and urban design attributes have been measured most commonly within quarter mile to one mile straight line buffers around origins and destinations. The implied theoretical link is not particularly intuitive between trip mode choice and land-use/design patterns some distance from the traveled route once trip distance, cost, and socioeconomic characteristics are controlled for. Measurement of transportation system attributes has been fairly evenly split between origin and destination buffers and measures along shortest or reported routes. Measuring linear infrastructure such as bike lanes with areal measures (miles of bike lane per unit area)

16

leads to some unusual interpretations. For example, it would suggest that each bike lane within a given area equally affects the decision to bike, regardless of whether or not a bike lane is useful for a given trip.

Five existing studies have considered route-level built environment factors for walk or bike travel. Routes have been defined in three ways: 1) shortest path, 2) self-reported route by usual travel mode, and 3) reported route child would walk/bike to school. Rodríguez and Joo (Rodríguez & Joo, 2004) calculated the percentage of sidewalks, slope time cost, and off-street path time savings along shortest path routes from respondents' homes to the University of North Carolina-Chapel Hill. This method ignores attributes along alternative routes, relying heavily on the assumption that walk and bike trips always will use the shortest path. Schlossberg et al. (2007) measured several variables within a quarter-mile corridor around the shortest route from home to school: intersection density, dead-end density, major road crossing, and railroad crossing. The authors termed this corridor a "walking zone," although it was used for bike trips as well. The shortest path corridor method has the advantage of potentially capturing features of alternate routes near the shortest path at the expense of potential aggregation errors. Lin and Chang (2010) considered built environment variables along the reported travel route by a child's usual school travel mode. For each direction along the reported route, the authors calculated sidewalk coverage, shade tree density, number of intersections, slope, and building density. Walk was the only nonmotorized option considered in the model. A potential shortcoming of this technique is that a single route

17

Table 3 Built Environment Measures Used in Nonmotorized Mode Choice Models

Land-use	Measures
accessibility	network distance to nearest commercial (O[1])
	30 minute walk network distance buffer (O)
	TAZ[1] (D[1])
detached housing	1/4 mi buffer (O)
	1 mi buffer (O)
job density	TAZ (D)
	TAZ (O)
	1 mi buffer (D)
	1 mi buffer (O)
	5 mi buffer (O)
land-use mix	1/4 mi buffer (Home)
	1/4 mi buffer (O)
	1/4 mi buffer (School)
	1/2 mi buffer (O)
	1 mi buffer (D)
	1 mi buffer (O)
	1 km network buffer (O)
	census block split at arterial (Home)
	Li^1 (Home)
	mean over 0.8 km buffers for each developed hectare in census tract (D)
	mean over 0.8 km buffers for each developed hectare in census tract (O)
	TAZ (D)
	TAZ (O)
park area per housing unit	census block split at arterial (Home)
parking spaces	transit station (D)
population density	1 mi buffer (D)
	1 mi buffer (O)
	TAZ (D)
	TAZ (O)
	TAZ (O+D)
	school zone

Table 3 (continued)

retail FAR	1 km network buffer (D)
	1 km network buffer (O)

Transportation system	*Measures*
intersection density	1/8 mi shortest path buffer
	1/4 mi buffer (O)
	1 km network buffer (O)
	TAZ (O+D) [2 studies]
intersections	reported route
cul-de-sacs/dead-ends	census block split at arterial (Home)
	TAZ (O)
railroad crossing	1/8 mi shortest path buffer
major street crossing	1/8 mi shortest path buffer
off-street paths	shortest network path
	½ mi (Home)
slope	point elevation (O,D)
	reported route
	shortest network path
local street density	reported route
bike lanes	1/4 mi buffer (Home)
	TAZ (O+D)
sidewalks	1/4 mi buffer (School)
	reported route
	shortest network path
	TAZ (O+D)
arterial street proportion	TAZ (O+D)
perceived traffic speed	reported route
composite utility	least-cost route

Urban design	*Measures*
block size	Li^1 (Home)
building density	reported route

Table 3 (continued)

cyclist environment factor	neighborhood (Home)
housing windows facing street	1/4 mi buffer (School)
housing age	1/4 mi buffer (O)
	1/2 mi buffer (O)
pedestrian environment factor	neighborhood (Home)
ped/bike factor	1 mi buffer (D)
	1 mi buffer (O)
street trees	reported route
	TAZ (O+D)

[1]O = origin, D = destination, TAZ = Traffic Analysis Zone, *Li* = basic administrative unit of Taiwan

was provided for all travel modes. This assumes that respondents would not consider

mode-specific routes. McMillan (2007) gathered data on parents' perceived traffic speeds

along the route their child would walk and bike to school. This improves on reported

route by usual mode, but it still leaves the question of whether walk and bike routes

would be identical. Furthermore, parents whose children do not walk or bike to school

might be less knowledgeable about the best routes by those modes. Finally, Bergman et

al. (2011) calculated a utility-weighted distance along a predicted least-cost path for bike

trips based on an existing bike route choice model from Broach et al. (2009). The

composite utility included route-level measures of slope, bike facilities, traffic, and

intersection types.

To date, little consensus has formed around the correct scale to measure built

environment variables for nonmotorized travel. The majority of studies have used zonal

and origin/destination buffers. These measures might be too coarse to capture the

reduced-scale travel environments experienced by cyclists and pedestrians. A handful of studies have calculated more precise route-level measures of built environment—especially urban design and transportation infrastructure components. Choosing the appropriate routes to measure is a key issue.

Trip context and tour attributes. Trip context variables are defined as attributes that can change value for different instances of a trip, or those that place the trip into context beyond the built environment or personal characteristics of the traveler. For example, the same trip may be taken on a weekend or weekday, and trip purpose establishes a different choice context for otherwise similar trips. Tours define a series of connected trips. A tour variable is defined here as any attribute derived from related sequences of trips. For instance, the distance from a child's school to a parent's workplace would be a tour-level variable.

Among trip variables, purpose has been the most commonly measured attribute. Some studies have chosen to estimate separate models by purpose (Bradley & Bowman, 2006; Frank et al., 2008; Zhang, 2004). Others have included trip purpose as an additional mode-specific constant (Cervero & Duncan, 2003; Miller et al., 2005; Rajamani et al., 2003; Reilly & Landis, 2002; Roorda et al., 2009). School trips are single purpose by definition; however, differentiating between from-school and to-school travel appears to be important for predicting walk trips, demonstrating the power of context in decision making (DiGuiseppi et al., 1998; Lin & Chang, 2010; Schlossberg et al., 2007; Yarlagadda & Srinivasan, 2008). Other trip variables have included weekend/weekday (Cervero & Duncan, 2003; Reilly & Landis, 2002; Soltani & Allan, 2006), peak/off-peak

(Roorda et al., 2009; Saneinejad et al., 2012), day/night (Cervero & Duncan, 2003), and weather variation (Saneinejad et al., 2012).

True tour-based models consider joint mode choice across related trips. Four studies have explicitly considered nonmotorized travel in tour models (Bradley & Bowman, 2006; Frank et al., 2008; Miller et al., 2005; Roorda et al., 2009). In addition, Yarlagadda and Srinivasan (2008) considered the joint choice of to-school and from-school travel mode and found significant differences from the naïve assumption that each direction could be viewed as an independent trip.

As an intermediate step toward complex tour-based models, some authors have included tour-level variables in otherwise trip-based models. Rodríguez and Joo (2004) modeled trips to a college campus. For the bike alternative, they included the maximum slope delay in either direction such that having to climb a steep hill on the return trip would affect the choice of biking to campus. Lin & Chang (2010), in a model of school travel, included a variable capturing whether the child's school was near a parent's commute route. Soltani & Allen (2006) included an estimate of the potential number of stops each mode would permit between a given origin and destination.

Attitudes and self-selection. Traveler attitudes rarely have been captured by travel surveys; however, there has been some concern that apparent effects of built environments on travel behavior actually reflect underlying attitudes about travel (Committee on Physical Activity, Health, Transportation, 2005). Self-selection refers to the concept that households might choose to locate where travel options best fit pre-existing household attitudes and preferences. Thus, neighborhood built environment

22

might simply reflect who lives there instead of actually impacting travel behavior. Since controlled experiments are difficult in an urban environment, measuring and controlling for residential self-selection, travel, and related attitudes might be the best we can do.

A handful of nonmotorized mode choice studies have included some attitudinal variables. DiGuiseppi et al. (1998) asked caregivers' opinions on perceived danger from traffic and crime, relating these attitudes to the child's school mode choice. McMillan (2007) asked parents how important it was that their children interact with others and whether they considered driving generally more convenient than other modes. As expected, importance of interaction increased the odds of nonmotorized travel to school, while pro-driving views decreased them. Schwanen and Mokhtarian (2005) asked a series of attitudinal items and developed summary factors such as pro-density, adventure seeker, travel freedom, pro-environment, frustration factor, and status seeker. In addition, they classified respondents by how well their attitudes about density matched their neighborhood. They found some evidence of self-selection effects for commute mode choice. Piatkowski and Marshall (2015) used a two-stage approach, first segmenting cyclists by willingness to cycle regularly, and then modeling the effects of attitudes and other variables separately for the two groups. Akar et al. (2013) found significant effects on bicycling to a college campus of attitudes toward gas prices, the sense of having travel options, and feeling safe biking on campus after dark. As that study demonstrates, often the line between perceptions and attitudes is a blurred one.

To the extent self-selection exists, it seems likely that it would relate most strongly to areal measures of built environment near respondents' homes. Route level

measures for specific trips, and areal measures at destination ends would likely be less correlated with household travel attitudes. To date, only McMillan's (2007) school travel study has combined attitudes with route and destination built environment attributes. Included were route-level perceived traffic speed along with sidewalk coverage, percent of residential windows facing the street, and land-use mix within a quarter mile of school. The built environment variables were significant predictors of mode choice even when controlling for parent attitudes and social/cultural norms. Measuring built environment at the smaller scales of route and destination might help to disentangle neighborhood choice from built environment effects on travel. For example, a pro-bike household may locate in a generally bike friendly neighborhood; however, specific trips will still vary in terms of the environment along the way. Measuring the built environment only at the neighborhood scale may be too coarse to pick up potentially important differences.

Sociodemographic variables. Excluding studies of children's school travel, which are less relevant to the adult travel models developed here, a handful of sociodemographic variables have been consistently identified as significant predictors of bike and walk travel. Gender, car ownership, and age were found to be significant factors in a majority of existing models, while income, age, and race were less often reported to have significant effects. Most of the time sociodemographic attributes were specified as having the same impact on bike and walk travel, though when tested independently, gender was more often correlated with decisions to bike. One study specified a full gender segmentation model but did not report whether differences were significant (Akar et al., 2013).

Travel Mode Availability

Discrete choice models commonly used in mode choice studies require the joint specification of a choice set.[3] The choice set contains the alternatives actually considered by the decision maker. Mis-specified choice sets can lead to inconsistent model estimation and faulty conclusions about relationships between variables (Ben-Akiva & Boccara, 1995). Getting the choice set right is particularly important when walk and bike alternatives are included in models because each mode's availability is closely related to a key model variable: distance/time. For example, assuming that walk and bike are available for all trips, regardless of distance, would likely bias distance coefficients for these modes.

Table 4 Bike (B) and Walk (W) Availability Distance Thresholds

Type	Distance
Universal choice set	Always available (B, W)
Rule of thumb	1.8 mi (W), 7.2 mi (B)
	3 mi (W), 12 mi (B)
	4 mi (W)
	5 mi (W,B)
	6 mi (W,B)
	10 mi (W), 30mi (B)
Sample-based	3 mi (W), 8 mi (B) [85th & 95th percentiles]
	4.8-8.1 mi (B) [95th, 97th, 98th percentiles by purpose]
	maximum sampled distances
Tour-based	bike on tour

---

[3] Or, more accurately, a set of possible choice sets and each particular set's probability (Manski, 1977). In practice, a single choice set is usually chosen deterministically in some *ad hoc* fashion for each case, or else the universal choice set is assumed the relevant one for all cases.

As summarized in Table 4, existing work has used four techniques to specify availability: universal choice set (all modes always available), rule of thumb distance thresholds, sample-based time/distance thresholds, and tour-based availability. A surprising number of studies have assumed that nonmotorized modes are always available (Akar et al., 2013; Black et al., 2001; DiGuiseppi et al., 1998; Ermagun & Samimi, 2015; Kockelman, 1997; Lin & Chang, 2010; McMillan, 2007; Reilly & Landis, 2002; Schlossberg et al., 2007; Schwanen & Mokhtarian, 2005; Soltani & Allan, 2006). In school travel studies of public elementary schools, this assumption may be reasonable (DiGuiseppi et al., 1998; McMillan, 2007). In other cases, however, assuming the universal set is likely to bias parameter estimates significantly.

Several studies used rules of thumb to hopefully reduce bias (Bradley & Bowman, 2006; Cervero & Duncan, 2003; Ewing et al., 2004; Rodríguez & Joo, 2004; Saneinejad et al., 2012; Yarlagadda & Srinivasan, 2008). Some rule of thumb distances seem too long to be of much use, however; for instance, Bradley and Bowman (2006) consider walk and bike to be available as long as distance is less than ten and thirty miles, respectively.

More behaviorally defensible are sample-based thresholds. Bergman et al. (2011) used the 85th percentile observed walk distance (3 mi) and the 95th percentile observed bike distance (8 mi) to eliminate nonmotorized modes from consideration. Roorda et al. (2009) incorporated maximum sampled thresholds for walk availability that varied by trip purpose: 13 km for work trips, 12 km for school trips, and 8km for other trips. Rajamani

et al. (2003) specified the maximum sampled distance for both walk and bike trips as availability thresholds. Finally, tour-level variables can inform trip-level bike mode availability since bikes, like cars, usually need to be brought from home (Roorda et al., 2009). This represents a key advantage for tour models in estimating bike mode choice, although public bicycle sharing systems are eroding the logic behind deterministic rules. A person could leave home without a bike only to pick one up at a bike share station along the way.

Bike ownership data rarely has been collected as part of a modeling effort, but like auto ownership for driving, it is likely a key factor in whether cycling is considered an option. Yarlagadda & Srinivasan (2008) used bike ownership as an availability criterion along with a rule of thumb distance threshold. Cervero & Duncan (2003) collected bike ownership data but curiously chose not to use the information to limit choice sets.

Overall, in most existing work, bike and walk availability either has been ignored or handled with questionable rules of thumb. Availability thresholds have ranged from 1.8 to 10 miles for walk and from 5 to 30 miles for bike. Bike ownership data and tour-level variables seem like potential ways to improve bike availability specification. Lacking additional data, observed distance distributions seem like the next best option.

Mode Choice Model Structure

All of the reviewed studies modeled mode choice as a random utility maximization problem and used maximum likelihood techniques to estimate parameters (Ben-Akiva & Boccara, 1995; McFadden, 1973). The majority of studies estimated

binomial logit (BL) or multinomial logit (MNL) models. MNL is the more appropriate

model when more than two alternatives are available. Logit models are popular primarily

because their properties are well understood, and they are relatively easy to estimate;

however, they depend on the assumption that errors are independent and identically

distributed (IID) across alternatives and cases (Ben-Akiva & Lerman, 1985). If

alternatives or groups of observations share some unobserved attributes, then model

estimates will be biased. A number of studies tested more flexible model structures to

account for non-IID errors.

The nested logit (NL) model allows for groups of alternatives to share unobserved

variation and therefore to be closer competitors (Ben-Akiva & Lerman, 1985). Among

the seven studies that tested NL model structures, three rejected them in favor of the

simpler MNL specifications (Ewing et al., 2004; Saneinejad et al., 2012; Schwanen &

Mokhtarian, 2005). This suggests, among other things, that bike and walk modes are not

necessarily closer substitutes for one another than for motorized modes. Bergman et al.

(2011) specified a nesting structure for transit access mode choice (bike/walk and

auto/transit) that outperformed the MNL model. Bradley and Bowman (2006) specified

NL models with nonmotorized nests for tour mode choice. They reported success with

this structure for all purposes except work-based trips. Rodríguez and Joo (2004)

specified a nesting structure that grouped bike with auto in a "personal vehicle" nest. The

NL model did not improve model fit, but the correlation structure was significant and

valid. Lin and Chang (2010) took a unique approach, estimating an NL model of school

mode choice with nests for independent (without parent) and dependent travel. They did

not report an MNL specification, but the nesting structure appeared to be valid and significant. Ermagun and Samimi (2015) specified a three-level nesting structure and reported relatively small but significant differences in some cross-elasticities.

Nested logit models relax the IID assumption between nests but still rely on IID errors within nests. Rodríguez and Joo (2004) tested an unusual heteroskedastic extreme value (HEV) specification that constrained alternative specific standard deviations to be equal within a nesting structure. The HEV model as usually applied allows each alternative's error variance to differ, but as applied the result was just a restricted version of the NL model. Yarlagadda and Srinivasan (2008) specified an interesting joint MNL model of from-school/to-school mode choice. Since assuming IID errors across from/to mode pairs was questionable, they reported testing various mixed logit (MMNL) structures. In the end, the MMNL models were rejected and the more restrictive joint MNL form was retained. Miller et al. (2005) and Roorda et al. (2009) estimated linked tour and trip mode choice models using simulated log-likelihoods. The authors argued that complexity of tour-based models would not have fit well within a traditional discrete choice framework. Frank et al. (2008) circumvented this challenge by modeling only the main mode of each tour and ignoring individual trip modes.

In addition to unobserved correlation between different modes, error correlation among individuals or groups of individuals is also possible. Existing nonmotorized mode choice models have relied on socioeconomic characteristics to capture heterogeneity among travelers. Furthermore, models to date have considered heterogeneity across groups in terms of preferences for specific modes. Another possibility is that different

groups or individuals vary in how they value specific attributes of modes. For example, women may not have a bias against biking, as is often reported, but could instead be less willing to trade off time or have a lower tolerance for biking in heavy traffic. Models that allow for unobserved heterogeneity such as the random parameter logit (RPL) and latent class model (LCM) are natural extensions to existing work.

In terms of model specification, existing work suggests that more flexible error structures than MNL might not be necessary. There is, however, considerable scope for exploring more nuanced testing of group or individual heterogeneity in nonmotorized choice models. Tour-based models could require simulation-based modeling to model trip-level mode choices.

2 Conceptual Model

Theoretical Framework

   Vernez Moudon and Lee (2003) divided determinants of walking and cycling into

three parts: intra- and inter-personal factors, environmental factors, and trip

characteristics. The authors further conceptualized measures of walking and bicycling

environments into three components: the specific trip origin and destination,

characteristics of the chosen route, and characteristics of the area where the trip occurs.

Not explicitly addressed was how to measure route quality when a chosen route is either

not observed (e.g. in common travel diary data) or not available (e.g. the walking or

cycling route that would have been used by a person who drove). The framework

developed here incorporates the three determinant factors and adopts the three

components of walking and cycling environments while making the route component

explicit. What were referred to as origin and destination variables in the existing

framework (Moudon & Lee, 2003) are here labeled trip context attributes to further

distinguish them from area measures around origins and destinations.

   Figure 3 outlines the proposed behavioral framework for modeling nonmotorized

mode choice. The decision structure synthesizes existing research and extends it by

incorporating predicted walk and bike routes as inputs to mode choice. It is hypothesized

that in addition to other factors, cyclists and pedestrians will be sensitive to conditions

along specific, considered routes for each mode. Some existing studies have included

variables along or around a route, but those studies have assumed shortest paths or else

have relied on reported routes. Shortest paths are behaviorally questionable,
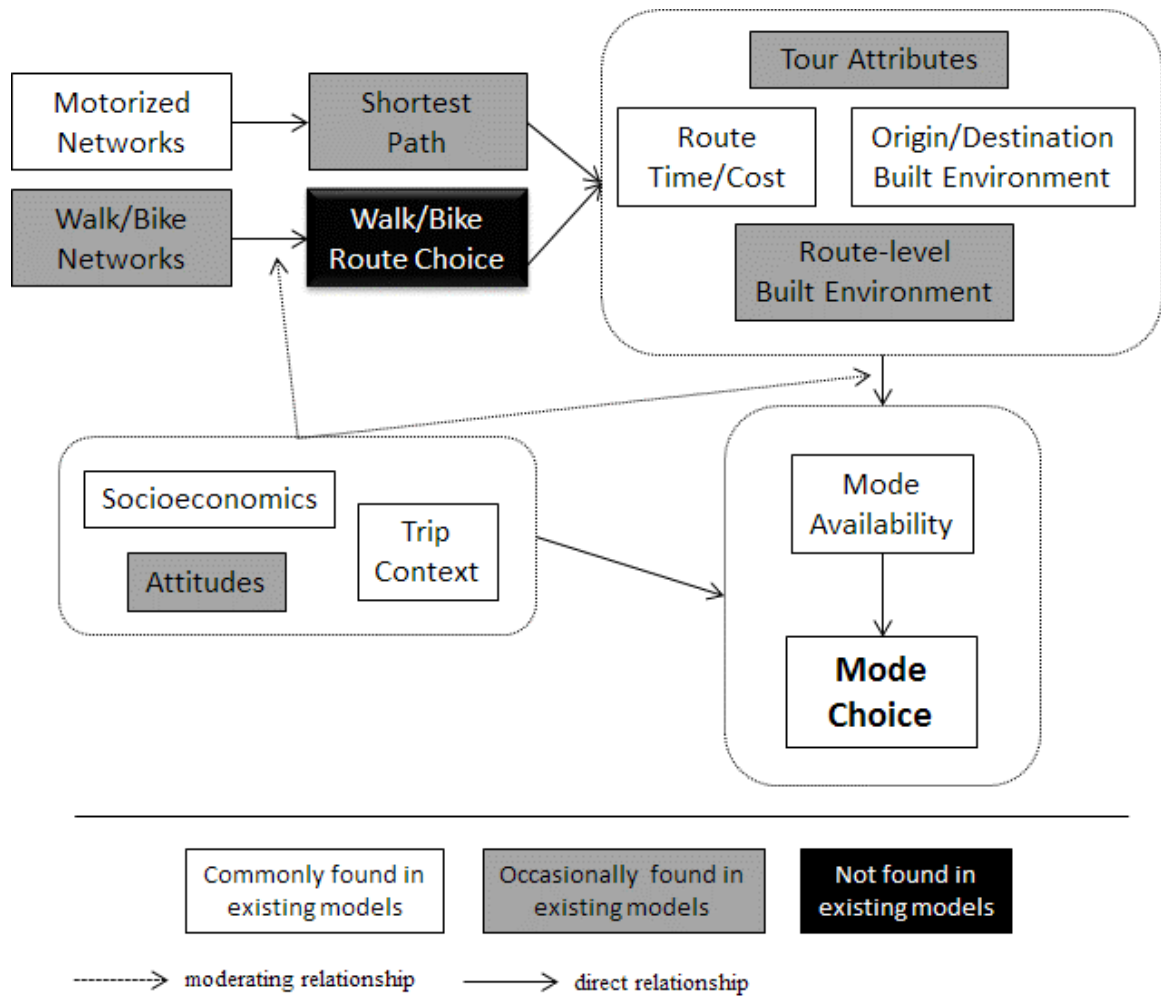
Figure 3 Proposed mode choice framework

especially for bike trips, and reported routes are not usually available in large-scale travel

surveys. Only Bergman et al. (2011) have used predicted paths and then only for bike

travel. They applied a route choice model developed by Broach et al. (2009).

Based on random utility maximization from economic consumer choice theory,

the framework assumes that travelers of various types weigh travel options based on the

attributes of each mode and the context of the choice situation (Ben-Akiva & Lerman,

1985; McFadden, 1973). Socioeconomic status and traveler attitudes, trip context, tour

attributes, time and cost, and the built environment around trip ends and routes all are posited to affect mode availability and mode choice directly. In addition, attributes of the household, traveler, and trip context may interact with (i.e. moderate) tour, origin-destination, and route-level variables as well. The same variables can moderate network attributes' impact on route choice for walk and bike trips. Shortest paths and predicted route choices determine which portions of travel networks are considered for each mode.

The proposed framework was developed as a first step and as a testing environment and therefore relies on some additional simplifying assumptions to ensure tractability. First, mode choice decisions are assumed to be made independently from given destination choice decisions. This is the standard assumption in traditional four-step mode choice models, but joint models and destination conditional on mode specifications have shown promise (Newman & Bernardin, 2010). Where tours are considered, it is further assumed that the entire sequence of trips is determined exogenously. Mode choice decisions at one stage only affect mode availability at later stages.

Second, the framework assumes that travelers have complete information about available routes and route attributes for walk and bike trips and use that information to mentally calculate an optimal route. In reality, information access may vary across decision-makers, and even given information, they may choose a simpler heuristic to consider routes for a mode choice decision.

Third, estimation of the route and mode choice models is sequential, implying that choosing considered routes precedes and is independent of choosing a mode.

Finally, it is assumed that all framework variables and choice situations are independent of one another. In reality, built environment might affect attitudes and even socioeconomics (e.g. car/bike ownership) to some degree. Mode choice habits also might affect attitudes. Travelers might fall into a pattern of choosing a certain mode or exhibit variety seeking behavior. Each of these possibilities would be interesting to explore in future research but would substantially complicate the modeling framework.

As indicated in Figure 3, route choice is treated as the outcome of an exogenous choice process determined by network attributes, traveler characteristics, and trip context. This permits an additional level of preference heterogeneity. The optimal bike or walk route for a given origin-destination pair could differ across travelers or trip context, resulting in different routes entering the mode choice model for different groups or in different contexts.

This two-level heterogeneity property could have some interesting consequences for policy analysis. As an example, imagine group "A" cyclists have a strong *route choice* preference for low-traffic streets, while group "B" cyclists are less sensitive to traffic levels. Both groups have a strong *mode choice* preference for bike lanes. Group "A" cyclists would be less likely to consider a high-traffic route with a new bike lane, and thus the new facility might impact mode choice only among group "B" cyclists. A similar policy response could occur based on trip context (e.g. purpose, time of day, child on trip) as well. In general, the proposed framework suggests that new infrastructure projects would only impact nonmotorized mode choice if the projects result in routes that

are attractive relative to existing alternatives. Thus, the choice structure implicitly recognizes diminishing marginal utility for projects near competing routes.

Operational Framework

The theoretical framework was operationalized as shown in Figure 4 to fit the scope of this initial research and the data readily available for model specification. All components were retained, though with different degrees of complexity. In particular, tradeoffs had to be made between non-motorized and motorized components and also built environment versus attitudes and sociodemographics. Generating point to point auto and transit travel times and costs at the resolution required was judged beyond the scope of this research. Access and egress distance, for transit, travel to and within the central city analysis district, for autos, and shortest path distances for both modes were used as proxies for parking, travel time, and other costs.

Travel attitudes likely have complex relationships with route and area-based built environments, the focus of this research. I decided to focus only on reported residential self-selection as an attitudinal predictor.

Sociodemographic variability was limited in the test data available, since participants were all families with children in Portland's residential North and East sides. Variables were limited to household car ownership and gender.

Finally, although data were generated by a two-cohort panel, travel data were pooled for analysis with the exception of a time period effect for mode choice to control for underlying time trends or seasonal variations.

Figure 4 Operationalizing the proposed mode choice framework

3 Data Description and General Methodology

To test the mode choice framework, several secondary datasets were obtained and processed for analysis. Data sources are summarized in Table 5.

Table 5 Data Sources

Source	Years	Description
City of Portland	2005	Interpolated traffic volumes for all local streets
City of Portland	Nov 2010	GIS[1] sidewalks, crosswalk markings, medians, and curbs
City of Portland	2010-2013	GIS Bicycle facilities by installation date
Family Activity Study	2010-2013 (Jul-Nov, 2010/12, May-Jul, 2011/13)	Two 5-day, person-based, 4-second interval GPS travel surveys and hip-mounted accelerometer data for 499 adults in 333 Portland, Oregon households
Family Activity Study	2010-2011	Survey questionnaires including household and person sociodemographic, household common destination addresses, and household residential self-selection items
Portland Metro	2012	GIS travel network of all local streets, multi-use paths, and trails, including bicycle facilities
Portland Regional Land Information System (RLIS)	May 2012	GIS parcel-based land use classifications and building footprints, public transit network
Portland State University Department of Geography	n/a	Digital Elevation Model (DEM) based on 1m LIDAR data
US Census (Decennial)	2010	Population and housing units by Census Block

[1] Geographic Information System

GPS Travel and Traveler Data

The primary travel data used in the study were collected for 499 adults in 333 households in Portland, Oregon. Nineteen neighborhood street corridors were identified in the largely flat, gridded, residential areas east of downtown Portland. Eight of the corridors were scheduled to receive bicycle boulevard treatments about one year after the site was recruited for the study, but the households were unaware that the study was

focused on bicycling behavior. Bike boulevards, sometimes called neighborhood greenways, are mostly quiet, residential streets that prioritize walking and, especially, bicycling using traffic calming, diversion of motor vehicles, pavement markings, and signage. Further details about the study's design and primary motivations are available elsewhere (Dill, McNeil, Broach, & Ma, 2014).

Recruitment. All non-vacant households within 1000 feet (305m) of the corridor street received a flyer on their doorstep or by mail, and eligible families self-selected into the study. Eligibility required at least one adult and one child aged 5-17 willing to participate, but for this research the child data was used only to identify when a child was on a trip with an adult. Each participating member had to have a working bicycle available for use and no physical limitations preventing active travel. Participants would agree to answer two rounds of survey questionnaires and collect two rounds of 5-day GPS and accelerometer data over a 2-year period. Families were offered $75 as an incentive for each round of successful data collection.

Using Census (ACS 5-year, 2011) counts of families with children 6-17 within each study site, we estimated a recruitment rate of only 3.2%. While low, the estimate likely understates the true rate somewhat, since some households would have had only children aged 0-4. Still, the rate was lower than a broadly comparable study in the Puget Sound Region, which reported a 6% initial recruitment rate (Vernez Moudon et al., 2009). We attributed the low rate to a combination of strict eligibility requirements, relatively low incentive, demanding data collection, and a two-year commitment from families with busy schedules.

38

Initially, each cohort was scheduled to collect travel and activity data one year apart. Due to delays in completing the bike boulevard treatments, cohorts collected two years apart. Cohort 1 collected travel data for five consecutive days from July-November, 2010, and then again during the same months in 2012. Cohort 2 collected during April-July, 2011, and then again over the same months in 2013. Each collection were scheduled to include at least one weekend day.

Representativeness. Table 6 provides comparisons with both the broader Portland population and the population within block groups overlapping the study site boundaries. Multnomah County's boundary is nearly identical to the city boundary. The study sites generally reflected the characteristics of families in the city as a whole. Compared with other family households with children, study participants had similar incomes and gender splits, were more likely to identify as non-Hispanic white, and were much less likely to rent their residence. A larger share of participants had college degrees than the population 25 years and older. Relative to all Portland households, participating households had a slightly higher number of vehicles, but compared with other multiple person households, the rate of zero-car households was lower. Some of the differences may reflect a bias toward participating in the specific study or research in general.

Table 7 compares trips within the city of Portland from the full GPS based travel dataset to the same subset of trips in the most recent traditional regional travel survey, intended to be representative of travel in the Portland region. Relative to the regional survey, the GPS sample included smaller shares of walk and transit trips and larger shares of bicycle and auto trips. Restricting the regional survey sample to households with

Table 6 Study Participants and Broader Population

Measure	Participating Households	Multnomah County[1,2]	Study Area Block Groups[2,3]
Median Annual Family Income	$50,000-$75,000	$57,143	$58,458[7]
% Female Adults	62.4%	58.7%	52.2%
% Female Children	47.9%	49.0%	48.1%
% Married Couple Families	58.5%	64.6%	68.9%
% White, Non-Hispanic Adults	84.9%	68.7%	71.4%
% Renters	19.2%	37.8%	40.4%
% Four-year college degree or higher	59.8%	37.5%[4]	34.1%[4]
Vehicles per household	1.7	1.5[5]	1.5[5]
% Zero car households	4.2%	6.3%[6]	-
Median adult age	41.0	35.4	-

[1] Census ACS 5-year, 2010

[2] households with children under 18, unless noted

[3] Census ACS 5-year, 2011

[4] population 25 years or older, all households

[5] all households

[6] households with more than one person

[7] family households (two or more related individuals)

children, GPS sample walk and auto mode shares were comparable to the representative survey (within 10%). Transit share was still considerably lower (-48%), and bike share was higher (+84%) in the GPS survey. Only those able to bike and with access to a bicycle were included in the GPS sample, so the higher rate of cycling was expected. Since the transit mode imputation success rate was relatively high (89%, Appendix A), it appears that the GPS sample might have underrepresented transit users. Shortest path

Table 7 Comparison of GPS and Regional Survey Travel Data

Item	Family Activity Study (2010-2013)	Regional Travel Survey (2011)[1]	Regional Travel Survey- Households w/ children
GPS points (4s interval)	18,620,392	-	
GPS trip stages (single-mode trip segments)	55,741	-	
GPS Trips	38,402	-	
Trips (within city of Portland)	30,885	17,304	8,165
walk			
…number	7,515	5,095	2,150
…share	24.3%	29.5%	26.4%
…GPS distance (mi)	0.40	-	-
…shortest path distance (mi)	0.32	0.33	0.37
bike			
…number	3,125	982	451
…share	10.1%	5.7%	5.5%
…GPS distance	1.60	-	-
…shortest path distance	1.31	2.10	1.93
auto			
…number	19,329	9,698	4,943
…share	62.6%	57.0%	60.8%
…avg. GPS distance	3.29	-	-
…avg. shortest path distance	2.56	3.00	2.87
transit[2]			
…number	811	1,299	405
…share	2.6%	7.5%	5.0%
…GPS distance	2.97	-	-
…shortest path distance	2.49	3.03	3.16
other			
…number	105	230	180
…share	0.3%	1.3%	2.2%

[1] Oregon Travel and Activity Survey; for comparability, only trips by city of Portland residents within the city boundaries were included. Applying regional sampling weights had only small effects on the results, and since the weights were not intended for subsamples, unweighted statistics are presented.

[2] Bus, light rail (MAX), streetcar, and aerial tram

distances were lower in the GPS sample across all modes. Without further information, it was difficult to distinguish to what extent this represented true sample differences versus differences in trip definitions or omissions. Shorter average differences are consistent with previous findings that traditional travel diaries tend to miss shorter distance trips (Stopher, FitzGerald, & Xu, 2007).

Trip and tour data. GPS derived travel data is mapped by mode in Figure 5 and summarized in Figure 5. The 18.6 million raw GPS points, consisting of time and location stamps at four-second intervals, were transformed into trip stages (single-mode trip segments) and trips by adapting existing GPS processing algorithms (Schuessler & Axhausen, 2009b). Trip stages were then assigned a most likely travel mode by choosing the highest probability mode after applying a mode choice prediction model, described in Appendix A, to the processed GPS and accelerometer data. A subsample of GPS trips was used for mode choice modeling (Chapter 5).

Low probability modes (maximum < 0.4) were randomly checked by hand in a GIS to catch any systematic errors in the method. One problem identified was boat trips in the local rivers falsely identified as bicycling due to similar speed and activity levels. These water trip stages were eliminated by flagging all trips with more than 80% of GPS points over water, and then manually flagging the 70 trip stages that were actually on water.
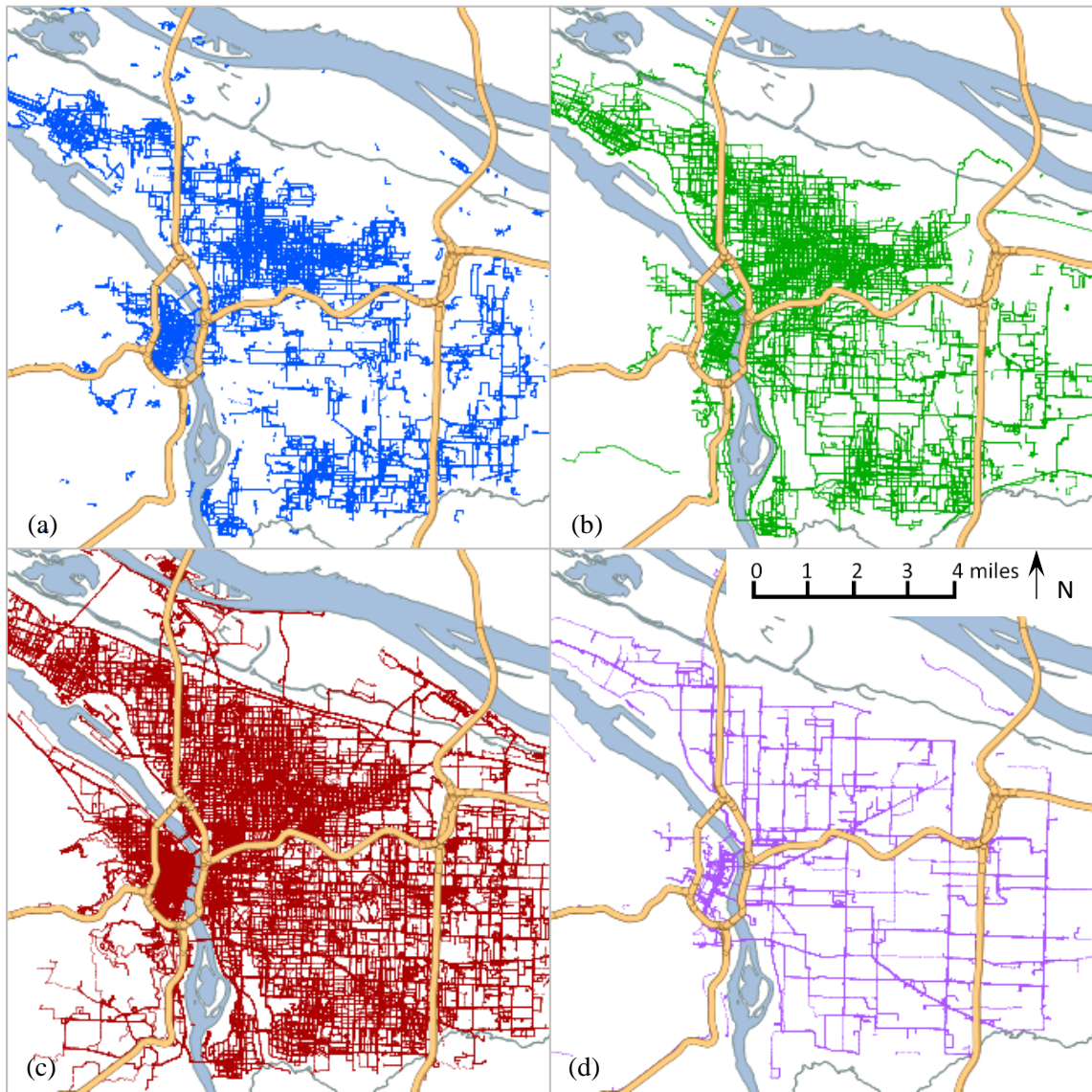
Figure 5 GPS data geographic extent for (a) walk, (b) bike, (c) auto, and (d) transit

Once modes were assigned to trip stages, a multiple hypothesis map matching technique was adapted from existing work and applied to all bicycle and pedestrian trips (Schuessler & Axhausen, 2009a). Map matching is the process of assigning a series of

GPS points to network links (street segments between intersection nodes). The multiple hypothesis method makes use of network topology to ensure that only feasible routes are chosen, in contrast to proximity-based algorithms that may match to nonsense routes (e.g. jumping back and forth between the lower and upper decks of a bridge). Modifications included handling short stretches of off-network travel, since walking and even cycling trips often made use of informal cut through paths in parks, plazas, shopping centers, campuses, and other locations despite an extremely detailed travel network. The algorithm was amended so that it could find its way around obstacles and pick up the trace on the other side. I found through experimentation that u-turns in the GPS data were much more common for walking trips than I had previously dealt with in bicycle travel, and procedures had to be added to ignore them; otherwise such movements can easily derail the map matching logic, which relies on sequential, least cost paths through the network.

Trips were assigned a primary mode based on the longest distance trip leg. Tours were constructed by splitting trip chains each time a trip ended within 200 meters (656 ft) of home or the travel day ended (2:00 A.M. the following morning, to catch late night returns). All analysis was restricted to travel within the city of Portland, although the data extended farther, and different analyses used different subsamples of the data, described here in the appropriate sections.

I also imputed joint household travel. Not all households had full participation in the study; for others, only a subset of household members enrolled. Children under 5 or over 16 when the study began were ineligible, as were adults with physical limitations

44

restricting walking and cycling and those without access to a bicycle. Others simply chose not to participate. Overall, 41% of households included all members, 39% were missing one or more adults, 36% were missing one or more children, and 6% of households did not provide data on missing participants. Since each household had at least two participants, there was always a possibility of joint travel despite the incomplete data. Joint travel was assumed at the trip stage level when more than half of paired GPS points were within 4 seconds and 50 meters of each other.

While traditional travel diaries have their own demonstrated biases, it should be noted that the deterministic trip, trip end, and route identification, as well as the probabilistic mode prediction methods, are each subject to unknown error rates. It was hoped that by following existing methodological standards and validating new methods as much as possible, results would remain consistent with actual travel despite the data's imperfections.

Where travel data processing errors did occur, I could at least speculate on the likely modeling impacts. Raw GPS data were divided into discrete trips, travel between an activity origin and destination, based on dwell time and GPS point density calculations adapted from Schuessler and Axhausen (2009b). In some cases, trips might have been split before a destination was reached (e.g. an exceptionally long delay in traffic or stopping to answer a mobile phone call). In others, trips might have failed to split when an activity was too brief to meet the time or density thresholds (e.g. dropping off library books or picking up a waiting passenger). Trip spitting errors were expected to have only minor impacts on model results. In mode choice modeling, trip definition errors were

45

further mitigated by including tour (linked trip) distances as well as trip distance, so that even if a trip were split into two segments, the total distance still would factor in mode choice. Trip purpose was imputed based on a provided list of common destinations, resulting in potential confounding between a destination being commonly visited and the specific trip purpose. In modeling mode choice, I partially addressed this issue by testing control variables reflecting "any common destination," regardless of purpose, but they were in all cases insignificant. Nevertheless, the purpose results are best interpreted as trips to specific classes of common locations rather than true purpose measures. Map matching of GPS points to GIS travel networks was subject to data limitations. In particular, informal shortcuts through parks and other spaces were not included in modeling, though they may well impact pedestrian and cycling route or mode choice. Future work might consider ways to include at least the portion of these route alternatives through public spaces. Finally, the imputation of travel mode itself was subject to errors (Appendix A). External validation suggested that bicycling detection errors were most likely to be of Type II (i.e. false negative, an actual bicycle trip stage is coded as another mode, primarily walking), while assigned walk errors were more likely to be of Type I (i.e. false positive, a non-walk trip stage is treated as walking, primarily bicycle and auto travel). The result of mode imputation errors would be to assign preferences for attributes to the wrong mode. Assigning (mostly very low speed) bicycle trip preferences to walking did not seem particularly troubling, while applying actual auto preferences to walk travel did seem problematic. Mitigating the trip stage errors, each trip was assigned only a primary mode, so that an error on any one stage did not necessarily result in a trip

46

mode assignment error. Efforts should continue to improve the imputation of travel mode from GPS data.

Household and traveler data. Both household and individual survey questionnaires were distributed three times as part of the study. Except to correct missing data that was assumed to be mostly static (e.g. gender) and in rare cases to update residential self-selection responses after a move, only the initial year's data were used for each cohort. Participants were instructed to answer household-related questions jointly, and I assumed that responses applied equally to all family members.

Missing data for household vehicles per licensed driver were replaced with the sample mean, since I had no reason to expect non-response bias by vehicle ownership (there was instruction to note if zero vehicles were owned). I was less confident in imputing missing data for residential self-selection importance ratings, since it seems reasonable that those who had not considered walking and biking when choosing a home, or participants confused by the concept, might also be more likely to skip those items. These cases were treated as missing in all analysis requiring the self-selection responses.

GIS Built Environment Data

In addition to the survey data, a database of Geographic Information System (GIS) measures was populated using data provided mainly by city and regional government agencies as shown in Table 5.

Time consistency of data. Because travel took place over a 4-year period, ideally built environment data likely to change over that time interval would be updated at least annually, but that was not always feasible. Data availability and resource constraints

47

forced decisions on which data could reasonably be treated as dynamic and which to treat as fixed. On-street bicycle facility and multi-use path variables were known to have changed considerably during the study's timeframe and were also a key focus of the research and framework. Therefore, the base 2012 travel network provided by Portland Metro was updated with annual construction data from the City of Portland for 2010-2013 to construct year by year GIS layers of bike facility locations. Stop signs, traffic signals, and pedestrian signals related to the eight project sites near participants were updated by hand based on project drawings.

Household buffer variables that included bike facilities were calculated for each of the four years of data, and the correct year's value was applied based on the year of the trip in calculating trip end buffer attributes. Other attributes were measured in a year as close as possible to the midpoint of data collection, and generally at least within the data span. An exception was traffic volume, which is only interpolated to local streets irregularly, the last time in 2005. A colleague performed a statistical comparison between the 2005 data and updated data from 2008-2012 at 51 locations and reported a correlation of 0.99 with mean error of just +200 vehicles per day, well below the level of precision specified in my models (A. Bigazzi, personal communication, December 19, 2013).

Roadway slope and block face variable calculations. Special techniques were developed to calculate roadway slope from a digital elevation model (DEM) based on 1 meter (3.28 ft) resolution LIDAR[4] data and to calculate linear block face (versus areal) built environment and land use measures. Figure 6 provides a graphical representation of

---

[4] LIDAR is a technology used to scan and map terrain at very high resolution.

the slope and block face calculations. For each street segment, elevation was measured

from each endpoint to the other and in each direction at 10 meter (32.8 ft) intervals or

less. Point elevations were then used to calculate gross elevation gain and loss in each

direction. Finally, the gain measured in one direction was averaged with the measured

loss in the opposite direction and vice versa, to account for the different measurement

points in each direction. The method was adapted slightly from one developed at Portland

Metro. Bridge roadway elevations were constructed manually by overlaying the DEM

and aerial photos of the bridge structures, and applying a similar technique.

(a)

(b)
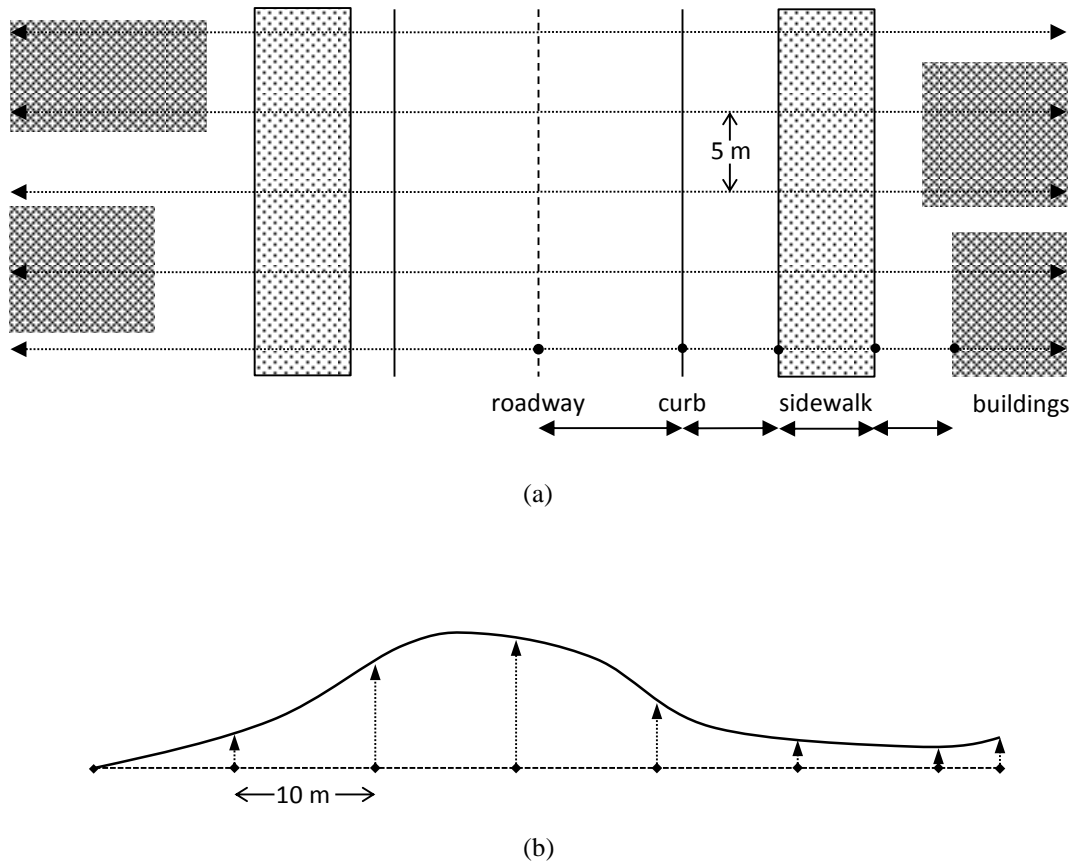
Figure 6 (a) Automated GIS street audit at 5m intervals and (b) Calculating gross
elevation gain and loss along segment at 10m intervals (shown in one direction only)

An automated, GIS street audit method was developed for measuring feature

distributions and other attributes (e.g. proportion commercial frontage or sidewalk

widths) linearly along street segments. The concept derived from a manual audit

methodology developed by Park (2008). For each street segment, scan lines were

generated at 5 meter (16 ft) intervals extending perpendicularly 30 meters (100 ft) to

either side of the street centerline. The intersections of the scan lines with curbs,

sidewalks, buildings, and land parcels were used to sample the block features over each 5

meter interval on either side of the street. By measuring at street level rather than from

the air, I hoped to better represent the environments experienced by people walking and

biking. While many of the more detailed measures (street width, sidewalk width, building

setbacks, enclosure, skyline height) were not ultimately used in this initial research, I

have plans to use them in future extensions to provide region-wide bicycle and,

especially, pedestrian micro-scale built environment measures.

Random utility maximization and the logit model

The modeling framework presented in Chapter 2 assumes that individuals make

route and mode choice selections that maximize the overall utility (satisfaction) derived

from each choice. Utility is specified as a function of attributes of the alternatives. Some

attributes are either unobservable or not fully measureable by the analyst, and so the

utility of each alternative in a given choice situation is random:

$$U_i = f(V_i + e_i) \qquad (1)$$

where U_I is the utility of alternative i, V is the observed component, and e is the unobserved or error component. Observed utility is most commonly specified as a sum of products of unobserved linear parameters and observed attributes:

$$V_{it} = \sum \mathbf{BX_{it}} \qquad (2)$$

where $\mathbf{B}$ is a vector of linear parameters and $\mathbf{X}_{it}$ is a vector of attribute values for alternative i and choice situation t. Since individuals are assumed to choose the alternative with highest utility, the probability of choosing an alternative is just the joint probability that the alternative's utility exceeds that of all other alternatives. Assuming that e_i is independently and identically (IID) Gumbel (type I extreme value) distributed leads to the multinomial logit (MNL) with choice probabilities:

$$P_{it} = \frac{e^{V_{it}}}{\sum_i e^{V_{it}}} \qquad (3)$$

The IID error assumption poses two potential problems for the current project. First, alternatives are assumed to have no correlation among unobserved attributes. In route choice sets, some alternatives are likely to physically overlap and thus would almost certainly share unobserved characteristics along those portions. In mode choice models, some alternatives may share unobserved attributes (e.g. both bike and walk require physical exertion and expose travelers to the elements).

Second, in the MNL specification, the parameter vector **B** is fixed across individuals and choice situations. If sensitivity to certain attributes actually varies among the population, more flexible model forms may be required. Significant heterogeneity due to unobserved individual attributes might also suggest more flexible forms. The panel nature of the data also poses challenges for the simple error structure.

On the other hand, the MNL model is simpler to specify and interpret. It has also been shown to be surprisingly robust to violations in assumptions. Given that the analysis here was an exploratory test of a new mode choice framework, and that the specific results were not intended to be applied directly to modeling generalized choice problems, I chose to retain the basic MNL structure, addressing the most obvious problems of route overlap and, to some degree, panel effects, and otherwise specifying sociodemographic and trip context interactions to allow for basic preference heterogeneity within the systematic portion of utility. Future applications will relax the IID assumptions and compare to this preliminary work.

4 Route Choice Models

In order to predict most likely walk and bicycle routes for each trip, route choice models are needed to specify the utility of alternative paths through a travel network from each trip origin to destination, even when walk, bike or neither are the chosen alternative. This chapter describes the development of a pedestrian route choice model using walk trips from the Family Activity Study GPS dataset, and the re-estimation of an existing bicycle route choice model with bike trips from the Family Activity Study.

Pedestrian Route Choice Model

While several recent cyclist route choice models have been estimated from GPS, there has been less activity on modeling route choice for pedestrian utility travel. Most pedestrian studies have instead focused on simulating aggregate flows and dynamic pedestrian movements in response to crowding and evacuation scenarios. Those studies were not considered relevant to the current problem, but a review may be found in Antonini et al. (2006). Only two studies were found that had developed models suitable for generating predicted routes, but while these studies provided useful information, the relatively small samples of adolescent girls (Rodríguez et al., 2015) and walking trips in very dense cities (Guo & Loo, 2013) seemed unlikely to be readily transferable to the sample of adults in Portland. There have, however, been a number of studies over the years that provided me with ideas about what pedestrians value, and whether their decision making process was reducible to a discrete choice problem at all.

Relevant existing work. Verlander and Heydecker (1997) compared a sample of walk trips in an urban area of the UK to shortest paths. They reported that 75% of walk trips followed the shortest available route, suggesting that distance is a potentially near-

deterministic factor for many walk trips. They also suggested two potential choice set generation techniques for walk trips: K-shortest routes and K-dissimilar routes. No choice model was estimated.

Seneviratne and Morrall (1985) examined walk trips in downtown Calgary and also found distance to be the primary route choice factor, but crossings, crowding, attractions, weather protection, pollution, and safety were also found to be important. Importance of factors was found to vary by trip purpose.

A survey of Jerusalem pedestrians found distance, "convenience," scenery, and habit to be the most important stated factors in walk route choice (Bovy & Stern, 1990).

Agrawal, Schlossberg, and Irvin (2008) collected stated route preferences and recalled routes from morning commuters at five rail stations in San Francisco and Portland, Oregon. Minimizing distance was the dominant factor, while traffic safety, crossing delays, sidewalk condition, route attractiveness, and presence of other pedestrians were also rated important factors by a majority of respondents. Women were considerably more likely than men to rate safety as important.

Borst et al. (2009) considered walk route choice among elderly residents of three Dutch urban districts. Unlike previous studies, they found that their elderly sample chose the shortest path only 20% of the time. They used an ad hoc iterative fitting method to match link counts by adjusting stochastic link friction factors. Audited, link-level built environment variables were then regressed on the best fitting friction factors. A number of factors were found to impact link choice: pavement (+), slopes/stairs (-), green strips (-

54

), front gardens (+), blind walls (-), litter (-), first floor dwellings (+), shops (+), parks (-), traffic volume (+), and segment length (+).

Rodríguez et al. (2015) developed a discrete choice-based model based on GPS and travel diary data collected by 303 adolescent girls in San Diego, California and Minneapolis, Minnesota. The branch and bound heuristic technique developed by Prato and Bekhor (2006) was used to identify all alternatives to chosen routes that met a set of logic conditions specified by the authors. Manual audits were performed to gather built environment information on all street segments along both chosen and alternative routes. This method was feasible due to the relatively small number of trips recorded (n=112) and a low average number of alternative routes, about three per trip. Separate route models were estimated for each sample location using a Path Size Logit (PSL) model form that accounts for overlap between route alternatives. Distance (-), proportion greenway (multi-use path, +), presence of traffic lights (+), proportion of route with medians (-), abandoned buildings (+), parks (+), food establishments (+), destinations index (+), and a safety index (+) were each found to be significant at the 5% level in at least one location. The authors noted the potential confounding effects of pedestrian amenities such as sidewalks, public transit stops, pedestrian signals and destinations often being located near detracting features like high traffic roads and temporarily abandoned buildings in busy districts.

Guo et al. (2013) developed pedestrian route choice models for two sub-samples of pedestrians in specific neighborhoods of Hong Kong and New York City. A total of 321 pedestrians were interviewed during an intercept survey, and each marked their

current walking route on a paper map, which was later digitized. Both utilitarian and recreational walking trips were included in the model. A combination of in-person audits, GIS data, and Google Street View audits were used to construct route attributes. In New York City, 5-minute manual vehicle traffic counts were performed due to the lack of volume data available for local streets. To generate alternative routes, the authors used a modified labeling approach similar to one used in Broach et al. (2011) to generate bicycle route alternatives, though without the calibration step. In the combined model, route distance (-), percentage retail frontage (+), and percentage open space (+) are all found to be significant route choice factors at 5-percent significance level.

Data selection. Walking trip stages from the Family Activity Study GPS Travel dataset were the units of analysis, where a trip stage was a single-mode segment of a trip (e.g. walking to access transit, or simply walking from origin to destination). Only trip stages where walking or transit was the primary mode were considered. For the remainder of the description, trip will be used interchangeably with stage for readability.

Walking trips were screened by distance, directness, and use of the travel network to fit the theory of utility maximization chosen for modeling. Only trips of at least 400m (0.25 mi, or about 2-5 blocks in the study area) were retained. Trips traveling more than three times the straight-line distance were considered likely to be recreational or loop trips and were not used in this analysis. Statistics for sampled routes are provided in Table 8.

Next, despite a detailed pedestrian network that included off-street paths, some walking trips still included sections of off-network travel. For those cases, if the off-

network segment occurred at the start or end of a trip (e.g. leaving a campus or park), the off-network segments were simply clipped; otherwise, the trip was rejected. Finally, where u-turns occurred in the data, I clipped the u-turn if the remaining route was still logical (e.g. a short jog up and back on a side street).

Table 8 Walk Trip Stage Sample and Alternative Route Characteristics

Attribute	Chosen	Alternative
number of walk trip stages	1167	-
number of individuals	283	-
mean distance	876 m	1651 m
	(0.54 mi)	(1.03 mi)
walk as primary mode	95.5 %	-
transit as primary mode	4.5 %	-
trips by females	72.4 %	-
travel on streets with complete sidewalks (both sides)	80.9 %	78.4 %
travel on off-street paths	4.2 %	3.4 %
travel with another household member (imputed)	10.2 %	-
mean distinct alternatives	-	16.8

The data comprised a panel with two time periods and multiple trips for most individuals. The average number of trips per person was low (4.1), however, and the maximum number of eligible walk trips for an individual was just 25 (2% of the total). Given the relatively even distribution of trips and the fact that data were collected on all travel covering both weekdays and weekends, I did not expect panel effects to be particularly worrisome.

Challenges peculiar to walking data. Determining side of street and street crossing behavior was a challenge. Even after filtering, the GPS data were not always precise

Figure 7 Pedestrian crossing rules: GPS points could not reliably determine side of street. The walk trip beginning at lower left above would have been assumed to use the crosswalk at SE Umatilla St. (by rule 1, see text) and a crosswalk at SE 15th Ave. and SE Tacoma St (by rule 2).

enough to accurately determine on which side of the street someone is traveling. In addition, when street side in unknown, pedestrians have at least two crossing options available at each intersection. Tree canopy, urban canyons, and variation in how a

person-based GPS device is worn can all degrade signals such that they wobble back and forth across centerlines. Figure 7 provides an example of how the following crossing rules were applied:

- Rule 1: Pedestrians passing straight through an intersection were assumed to have made use of the qualitatively most protected crossing to either side of the street or any midblock crossings on streets entering the intersection.

- Rule 2: If a turn sequence occurred that necessitated a crossing (e.g. left then right, or right then left), pedestrians were assumed to have crossed only once and to have used the qualitatively most protected crossing available between the two turns.

- The crossing protection hierarchy assigned was: traffic signal > crosswalk with median > crosswalk > crossing with none of these treatments.

Generating alternative routes. I used a stochastic, random walk-based method developed by Frejinger et al. (2009) to generate up to twenty alternatives to the chosen path. Choice set generation is a heavily-studied area within route choice modeling (Prato, 2009). Choice sets can be specified deterministically or using a random process. My colleagues and I had developed a deterministic technique based on optimizing certain attributes along each alternative to estimate a cyclist route choice model (Broach et al., 2011), and a similar method has been applied to pedestrian route choice (Guo & Loo, 2013). However, in that case we had a better idea—including a survey of stated preferences—of the factors important to cyclists along routes. The pedestrian route choice literature is sparse and uncertain by comparison. The random walk method assumes only that distance is important, and unlike competing stochastic methods, selection probabilities are known and can be corrected for in modeling (Frejinger et al., 2009). A more recent method avoids the choice set generation problem entirely, and may

59

be preferable, but it was not ready to share at the time of this analysis (Fosgerau,

Frejinger, & Karlstrom, 2013). The random walk around a shortest path was applied as

follows:

- Define the origin (O) and destination (D) as the nearest points on network links.

- Define the universal choice set as all possible paths between O and D such that:

    o No node is traversed twice (acyclic).

    o No u-turns are needed.

    o The path does not exceed three times the shortest network path
      between O and D.

    o The path does not pass by the destination link.

- For each of twenty iterations:

    o Including the origin, randomly select the next outgoing link i (from
      node v to w) from the set of all valid outgoing links M using the
      probability formula in Equation 1, where SP is calculated as the
      shortest valid path given links already traversed.

    o If a dead end is encountered, restart up to ten times before giving up
      on that iteration.

    o Through trial and error, we set b1 equal to 5 and b2 equal to 1. This
      provided a reasonable balance between generating a variety of
      alternatives and oversampling relatively direct routes.

$$\Pr(i) = \frac{1 - (1 - \frac{SP(v,D)}{cost(i) + SP(w,D)}^{b1})^{b2}}{\sum_{i \in M} 1 - (1 - \frac{SP(v,D)}{cost(i) + SP(w,D)}^{b1})^{b2}} \qquad (4)$$

Table 8 describes characteristics of the observed and sampled alternative routes. The randomly generated paths were nearly twice as long on average as the chosen path; however, the model weights shorter, more likely paths more heavily, so that the coefficient estimates should have remained unbiased. Only ten trips had to be discarded because no alternative was generated by the random walk.

Pedestrian environment and trip-related variables. I hypothesized that pedestrians would be sensitive to attributes both along (e.g. sidewalks) and adjacent to (e.g. land use) potential walking routes. GIS network, elevation, and land use data were used to calculate variables at the link level that were then aggregated to form route measures.

Relevant adjacent environment variables were calculated for both sides of each street segment and then averaged. An automated GIS-based audit method, described in Chapter 3, was used to capture attributes including commercial and other land-use frontages, sidewalk coverage, and enclosure to test in the model. In addition, a few trip context attributes such as weekend/weekday, joint travel with other participating family members, and walk as transit access mode were assembled. It should be noted that the joint travel measures were known to be limited, since we only had information about other family members who were participating in the study. Table 9 provides a list of variables considered in model development.

Table 9 Pedestrian Route and Trip Variables

Name	Description
distance (m)	route distance in meters
turns	intersection angle at least 30 degrees onto street with different name
steep upslope (m)	distance along links with average upslope of at least 10 percent
substandard street (m)	distance along links classified as unpaved or alleys
busy street (m)	distance along streets with collector or higher classification (approximately 10,000+ vehicles per day)
traveling together	another family member on trip (child or adult)
neighborhood commercial (m)	proportion commercial frontage (average of both sides) along collector and minor arterial streets (average 10,000-13,000 vehicles per day) times street distance
unsignalized arterial crossings	number of arterial (average 13,000-23,000 vehicles per day) crossings with no signal (auto or pedestrian)
unmarked collector crossings	number of collector (average 10,000 cars per day) crossings with no marked crosswalk or signal
additional variables tested	segmentation by gender, transit access trip, weekend, parks, paths, residential streets, downslope, missing sidewalk, median refuges, enclosure, pre-1946 buildings

Behavioral theory and model form. One may argue that, when walking, people choose their path at random, solely due to habit, or even dynamically from decision point to decision point. I chose instead to consider that pedestrians might not differ so drastically from travelers using other modes. The model form adopted assumes that pedestrians choose a route before traveling by selecting the bundle of route attributes that maximizes their utility given the options available. I argued that pedestrians, like other travelers, were likely willing to make tradeoffs among the bundles of attributes. For example, they might go out of their way to avoid a difficult crossing, or to walk along a more interesting street at the cost of slightly more distance walked.

Based on the assumption of utility maximizing behavior, I used a discrete choice modeling framework, in which individuals choose from among a discrete number of

alternatives, in this case a set of possible paths from origin to destination (Ben-Akiva &

Lerman, 1985). Following Frejinger et al. (2009), I used an expanded path-size logit

model. This model modifies the common multinomial logit model to account for overlap

among path alternatives and corrects for the importance sampling alternative generation

technique described in the previous section. The model form is shown in Equation 5 and

the expanded path size formulation in Equation 6.

$$\Pr(i|C_n) = \frac{e^{\mu(V_{in}+\ln(EPS_{in}))+\ln(\frac{k_{in}}{q(i)})}}{\sum_{j \in C_n} e^{\mu(V_{jn}+\ln(EPS_{in}))+\ln(\frac{k_{jn}}{q(j)})}} \qquad (5)$$

where C_n is the choice set for choice situation n
μ is the usual logit scale term
V is the systematic portion of utility
EPS is the expanded path size factor (see Equation 3)
k_{in} is the number of times alternative i is randomly drawn (+1
if chosen
   alternative, which is always included)
$q(i)$ is the selection probability, calculated as the product of
each link choice
   probability in the random draws

$$EPS_{in} = \sum_{a \in \Gamma_i} \frac{L_a}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{aj} \Phi_{jn}} \qquad (6)$$

Γ_i is the set of links in path i
L_a is the length of link a
L_i is the length of path i
δ_{aj} is 1 if path a contains link j, 0 otherwise
Φ_{jn} is 1 if path j is the chosen path or is expected to be
   drawn at least once

Results. Results of the final model explaining the choice of the observed routes

over plausible alternatives are shown in Table 10. Many intermediate models were

estimated, and likelihood ratio tests were conducted to determine whether coefficients

should be retained, combined, or dropped. In general, I was pleased with the model's

ability to explain the data. Overall model fit was strongly significant, signs were as expected, and the specification did not seem to be overly sensitive to small specification changes. The expanded path size parameter was positive and significant and appeared to slightly improve estimation efficiency when compared with the normal path size term.

Coefficients in logit models cannot be compared directly beyond sign and significance. One useful tool to interpret model results is the relative willingness to trade off against other attributes, defined by the marginal rate of substitution. Elasticities provide the expected change in probability for a small change in each variable, given the variable values in each choice situation. Marginal rates of substitution provide a better idea of the potential effect size of each attribute relative to another. Table 11 presents marginal rates of substitution relative to distance for each significant attribute.

Distance, steep upslope, and substandard streets all had significant effects on route choice. Steep upslopes of 10% are perceived as twice as costly as travel on less steep ground. Lower grade thresholds were tested along with various downslope variables. None were found to be significant. It is interesting if not unexpected that pedestrians are not nearly as sensitive to slopes as cyclists, who previous work has found to avoid upslopes as low as 2% (Broach et al., 2012).Due to data limitations, it was difficult to separate alleyways from unpaved streets, and so the two were combined. These were prevalent in the study areas. By definition, neither of these substandard street types has sidewalks, and this fact likely masked any independent correlation between sidewalks and route choice. The model suggests an alley or unpaved street added about 50% to perceived distance.

In contrast to findings from cycling studies, I did not find a benefit to off street paths separate from possible distance savings. It should be noted that I selected walk trips to exclude likely recreational walking travel, and paths may well be attractive options for such trips.

Table 10 Path-size Logit Walk Route Choice Results

Attribute	coeff	t-stat[1]
distance (m)	-0.015	-28.4
turns	-0.809	-20.2
steep upslope (m)	-0.015	-2.43
substandard street (m)	-0.008	-6.01
busy street (m)	-0.002	-5.12
distance * traveling together (m)	-0.013	-3.19
neighborhood commercial (m)	0.004	3.70
unsignalized arterial crossings	-1.090	-4.35
unmarked collector crossings	-0.419	-2.71
ln(EPS)	0.128	5.53
Log-likelihood (Null)	-2,919	
Log-likelihood (Model)	-1,047	
Adjusted Pseudo R^2	0.638	
Num. Walk Trip Stages	1,167	

[1]all coefficient estimates significant at the 5% level, standard t-tests

Crossing busy streets without aids significantly reduced the utility of a route. To avoid an additional unsignalized arterial crossing, a pedestrian would be willing to go over 70 meters (230 ft) farther via an alternate path. Though significant, this only represents about one Portland block. According to the model, signalized crossings would need to be quite frequent to be useful to the average pedestrian. Walkers were also willing to deviate somewhat to avoid crossing collector streets (average AADT around

10,000) without at least a painted crosswalk. To what extent these intersection crossing attributes reflect disutility due to delay, perceived safety, or some other factor could not be determined from the analysis here but would be an interesting topic for future work.

Table 11 Distance equivalent (Marginal Rate of Substitution) for attributes

Attribute	Distance Equivalent
Per additional…	
turn	+54 m
unsignalized arterial crossing	+73 m
collector crossing w/o marked crosswalk	+28 m
Change in perceived distance along…	
ten percent upslope	+99 %
unpaved or alleyway	+51 %
busy street (collector or larger)	+14 %
neighborhood commercial	−28 %
Increase in detour cost…	
traveling with another family member	+85 %

Even after controlling for the intersection crossing effects just described, each additional turn along a route was equivalent to about 50 meters (164ft) distance, or about one-half to three-quarters of a typical Portland block. Turns are not likely to be a delay factor for pedestrians as they are for cyclists and motorized travelers. I propose that the cost of turns might be in terms of cognitive wayfinding cost; a route with fewer turns is simpler to navigate. Another possibility suggested to me is that people walking might choose a preferred "line haul" street and prefer to stay on it until they need to turn toward their destination. It is also feasible that the turn finding is at least in part an artifact of the alternative route generation technique. Future work could compare models estimated

using different methods to generate alternative walking routes and check the sensitivity of the turn finding.

Only two adjacent environment factors were found to be significant in pedestrian route choice. Neighborhood commercial corridors, defined as the proportion of commercial frontage along minor arterial and collector streets, were attractive to people walking, perhaps because of increased street activity or possibly because of the opportunity to make an unplanned stop. Walking distance on those streets was perceived as up to 28% less costly than a similar non-commercial street. Commercial districts on major arterials did not attract pedestrians passing through. Routes adjacent to busy streets (collector or greater) added the equivalent of 14% more distance. I found no significant difference among the classes of heavily trafficked adjacent streets. I was unable to obtain suitable measures of tree canopy and other street amenities but hope to test these in the future.

The only trip-level attribute significant in the final model was travel with another household member. Imputation and limitations of my joint travel measure are described in Chapter 3. Compared to traveling without them, willingness to detour declined sharply. This was equally true whether the companion was a child or another adult. Trip purpose had not yet been imputed when this part of the analysis was completed, and the might have reflected a difference in common trip purposes when walking with someone else; for instance, walking with a child to school or walking with a partner to get lunch. There may also be other reasons that joint trips place a higher importance on time regardless of purpose. While joint travel behavior is usually not represented in trip-based models at this

level of detail, newer activity-based travel demand models such as Portland Metro's

DASH often contain joint travel sub-models (Gliebe & Kim, 2010).

Discussion. The successful estimation of a walk route choice model for a large

sample of varied adult walk trips builds on recent related work (Guo & Loo, 2013;

Rodríguez et al., 2015) with smaller samples and children in support of treating walk

travel as utility maximizing behavior, at least for utilitarian travel. Pedestrians were

sensitive to attributes of the pedestrian network, intersection crossing aids, and elements

of the street and block face environment. They were willing to go out of their way to use

more attractive facilities, but their tolerance for detours is limited, perhaps more so even

than for cyclists. Crossing aids must be densely placed along major streets to be useful.

Neighborhood-scale commercial streets might serve as both attractive destinations and

walking routes. Alleyways and unpaved streets did not seem to be working well for

pedestrians, who only were predicted to use them if the distance savings was quite large.

Only extremely steep terrain was perceived as a barrier, and then only in the uphill

direction, although this result may not hold for all populations. Finally, I found that those

traveling jointly with another household member may prefer more direct routes than solo

travelers, a potentially useful finding to investigate further within detailed, activity-based

models.

A number of factors hypothesized to be influential in pedestrian route choice were

insignificant in the analyzed sample of trips. I tested for differences in female pedestrian

preferences in terms of distance, adjacent street types, crossing aids, and joint travel but

found no evidence of gender-based preference heterogeneity. Counter to expectations,

68

there was no significant increase in sensitivity to distance when walking as part of a transit trip ($n = 53$) or on weekends ($n = 386$). Street width was too closely correlated with traffic volumes to include in the model separately, and other measures of local design along routes, such as measures of enclosure and proportion of pre-1946 buildings were not significant predictors. Local design measures may need to be aggregated in some way other than simply averaging them over a route, but this is left for future research. Finally, it was surprising that tested interactions between various route attributes and walking with one or more children were insignificant. I had expected those walking with a child to be more sensitive to traffic levels, difficult crossings, and perhaps adjacent commercial land uses that might make the pedestrian environment feel less safe or predictable. Because trip purpose information was not available, I would speculate that time pressure might be more important than safety perception on some common trips with children, such as going to school or other activities, while more flexible purposes might have shown the hypothesized additional willingness to detour.

Bicyclist Route Choice

Ideally, both route choice models could have been developed from travel data separate from trips used for mode choice analysis to avoid potential endogeneity issues. That the participant households in the Family Activity Study were chosen specifically because they were located more than 1/4 mile from existing bicycle boulevards—a potentially important factor in route choice—created an additional concern.

For walking route choice, no suitable existing models were available, nor was a suitable exogenous dataset available. For cyclist route choice, a candidate route choice

model did exist. With colleagues, I had previously developed a model of revealed preference route choice from a sample of GPS-based bicycling trips also recorded in Portland, Oregon (Broach et al., 2012). I was concerned, though, that the difference in samples might make transferring the model to the current sample unrealistic.

The existing route choice model was estimated from data gathered in 2007 as part of the Bicycle GPS Study (Dill, 2009). Some participants were drawn from a broader random sample of the region, but most were recruited via non-random sampling strategies including media promotion. Female cyclists were purposefully oversampled, and since it was important to gather sufficient data, frequent cyclists were targeted. Among cyclists in the Bike GPS study, 84% reported riding 7 days or more per month during the summer, and 68% reported riding 7 or more days even during Portland's rainy winter months. Among Family Activity Study cyclists and non-cyclists, in contrast, 46% reported riding 7 or more days a month in summer, and just 19% reported riding that often in winter. Thirty-two percent identified themselves as "interested, but concerned" about cycling in Portland. There had also been substantial changes in the bicycle network since 2007.

In order to get a sense of the transferability of route choice behavior between the two samples, I estimated a route choice model using the earlier methodology and the new data. Full model results are provided in Appendix B alongside the original model estimates. Figure 8 compares predicted marginal rates of substitution (rates at which cyclists were willing to trade off more or less distance for other attributes) from route

choice models estimated with the different datasets. Missing bars indicate non-significant parameter estimates in the relevant model ($\alpha = 0.05$).

The consistency between model results was encouraging. Almost all of the parameters agree both in significance and direction, and the effect sizes of most factors were reasonably similar. Compared with the earlier results, the sample of cyclists used in this research were sensitive to lower levels of mixed vehicle traffic (~5-10,000 vehicles per day), less sensitive to bridge facilities, less sensitive to certain turning movements across moderate traffic, and showed a small but significant residual preference for on-street bike lanes after controlling for traffic volumes. They also revealed a small but significant counterintuitive preference for increased stop sign density along routes, possibly reflecting a preference for low-speed, quiet residential streets not adequately captured by other variables. Overall, I felt the existing route choice model would produce paths that reasonably represented the current sample's preferences, thus avoiding potential endogeneity issues. Testing the sensitivity of mode choice model estimates to the route choice model specification would be an interesting topic for future research.
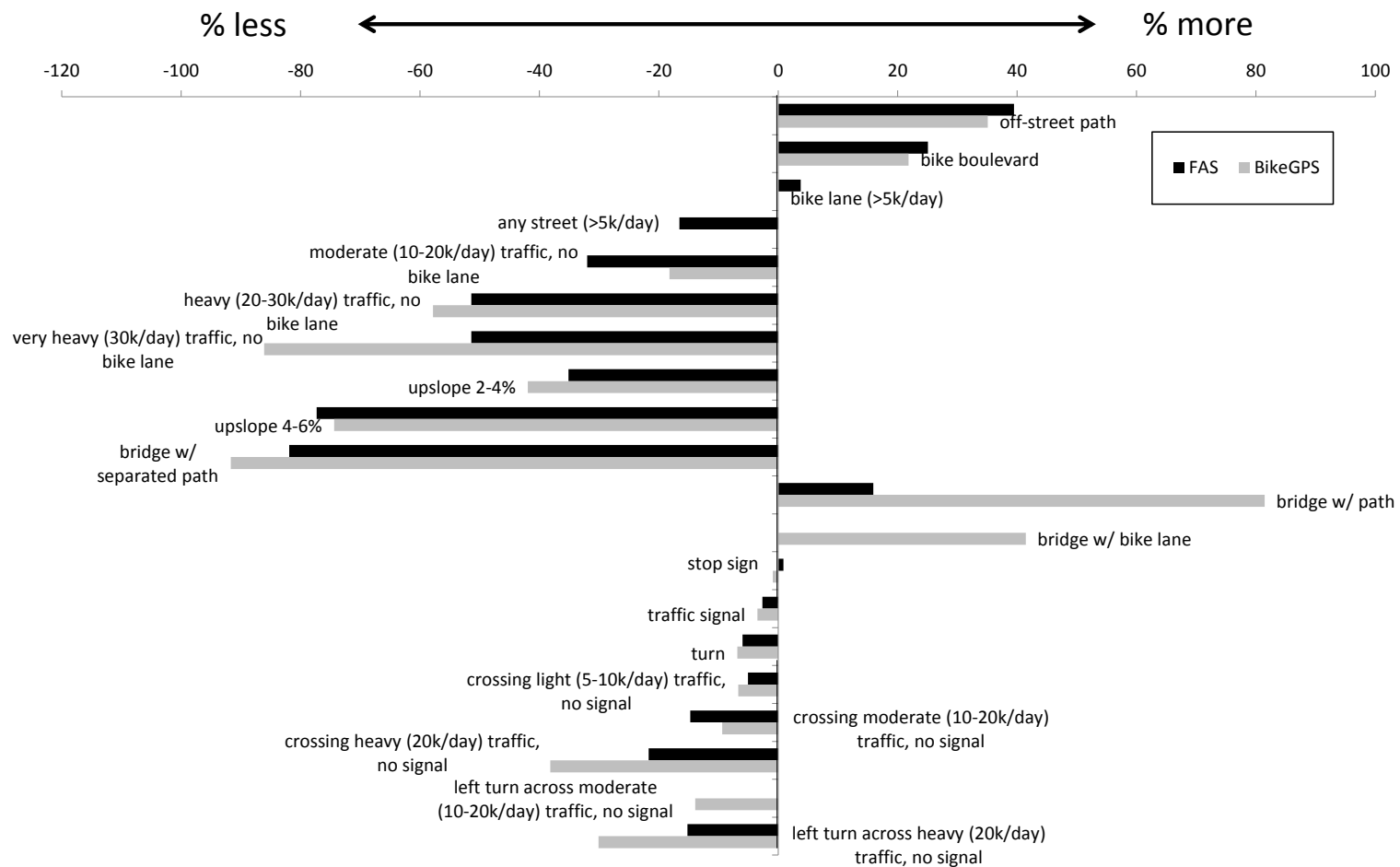
Figure 8 Calculated compensating variations in terms of distance for various bicyclist route choice attributes

5 Mode Choice Models

The walking and bicycling route choice models outlined in Chapter 4 were
combined with the Family Activity Study GPS travel and questionnaire data along with a
set of generated area-based measures to assemble a dataset capable of applying the mode
choice framework developed in Chapter 2. A series of multinomial logit (MNL) mode
choice models were developed to test the framework, comparing route and area-based
travel and built environment measures and examining what the data sample and
framework could reveal about nonmotorized mode choice.

Data Processing Methods

Mode choice dataset and case selection. Mode choice data came from the
Family Activity Study GPS travel data described in Chapter 3. Only primary trip mode,
defined as the mode used on the longest distance leg of a trip, was considered. Although
both child and adult travel were recorded, only adult travel is examined in this study, with
the exception of variables relating to having another household member on the trip.

Simple home-based trip chains had been constructed in order to incorporate some
tour-level variables, such as total distance, number of stops, and bi-directional slope.
Each tour started either at home or at the start of the travel day (2:00 A.M.), and ended
either with a return trip home, or at the end of the travel day (2:00 A.M. the following
day). For trip-based work, simple home to home tours seemed appropriate to capture the
most relevant information for mode choice decisions.

The full dataset included 22,445 adult trips. To match the primary research
motivation, nonmotorized travel behavior, I limited trips to distances over which those

modes would be competitive by dropping any trip with an all network shortest path distance greater than the 99th percentile shortest path for observed cycling trips, 7 miles (11.3 km). I also chose to exclude trips flagged as loop trips (thought likely to be recreational), those that were part of known cycling or walking events (e.g. Sunday Parkways or Providence Bridge Pedal), those with a mode other than walk, bike, auto, and transit (e.g. run, watercraft), those extending beyond the city of Portland GIS database, and very short trips with traveled GPS distance less than one quarter mile (402 m). The final analysis set consisted of 13,261 trips, about 59% of the full sample of adult trips.

Having a working bicycle and being able to ride were study recruitment criteria, and cycling was considered available for all analyzed trips, as were transit and auto, while walking was available based on shortest path trip distance. All households lived within one-half mile of a public transit line, although of course some origins and destinations were farther away, up to 3.9 miles from the nearest transit stop. Since biking or even driving were available access modes, I elected to consider transit an available mode for all trips. Walking was considered available for any trip with a shortest path less than the 99th percentile observed shortest path for walk trips, 2.35 miles (3.8 km, 64% of trips). Shortest path distances were based on the full network, including streets and paths open only to bicycle and walk travel. The use of universal shortest paths allowed for separating simple origin-destination proximity effects from detour effects related to mode-specific network quality. Table 12 describes attributes appearing in the final model specifications.

Table 13 compares the mode choice GPS data subset to comparable regional survey data (for comparison with the full datasets see Table 7). The GPS data included smaller shares of walking and transit trips, and larger shares of bicycling and auto trips. Shortest path trip distances were similar across surveys for walking and driving, while bicycle and transit trips were shorter on average in the GPS sample.

Table 12 Mode Choice Variable Descriptions

Name	Description	Mean[1]	Min	Max
cohort 1, year 2	First study cohort, second year (2012)	0.27	0	1
constant	Mode-specific constants	1.00	1	1
Trip Context				
commute	Direct trip home to work or work to home	0.04	0	1
grocery	Either trip end within 200m of self-identified grocery store	0.07	0	1
school	Either trip end within 200m of self-identified child's school	0.11	0	1
work	Either trip end within 200m of self-identified household work location	0.05	0	1
weekend	Saturday or Sunday	0.32	0	1
central city, one end	One trip end in central city analysis district	0.13	0	1
central city, both ends	Both trip ends in central city analysis district	0.02	0	1
rivcross * male	Major river crossing * Male	0.04	0	1
ln(miles_sp)	Log of shortest path miles along full travel network	0.44	-1.39	1.95
ln(tour_miles_sp)	Log of tour shortest path miles along full travel network	1.56	-1.38	3.68
tour stops	Number of stops between leaving or returning home[2]	4.00	1	26

Table 12 (continued)

Name	Description	Mean[1]	Min	Max
ln(transit stop to dest.)	Log of shortest path meters along bike/ped network from nearest transit stop to destination	4.84	0.00	8.74
ln(transit stop to orig.)	Log of shortest path meters along bike/ped network from trip origin to nearest transit stop	4.78	0.00	8.73

Sociodemographics

Name	Description	Mean[1]	Min	Max
cars per driver[3]	Household cars per licensed driver	0.93	0.00	3.00
female	Self-identified female	0.67	0	1
zero car household	Household with no motor vehicles	0	1	0.03

Areal transportation infrastructure[4]

Name	Description	Mean[1]	Min	Max
bike blvd. miles 1/2mi_od * female[5]	Miles of bike boulevard * Female (trip ends)	0.48	0.00	3.20
bike blvd. miles 1/2mi_h * female[5]	Miles of bike boulevard * Female (home)	0.42	0.00	2.70
com. miles 1/2mi_od	Miles of secondary arterial (~10-20,000 vehicles per day) and higher order streets (~20,000-30,000) vehicles per day) weighted by proportion commercial block faces	3.00	0.01	17.00
bike lane miles 1/2mi_od * male	Miles of on-street, striped bike lanes * Male	0.30	0.00	2.93
local street miles 1/2mi_od	Miles of local streets, lower order than collector (~0-5,000 vehicles per day)	16.60	0.02	22.40
local street miles 1/2mi_od * with other	Local street miles * Other household member on trip (adult or child)	2.86	0.00	21.80
nbh. com. miles 1/2mi_od	Miles of collector class streets (~5-10,000 vehicles per day) weighted by proportion commercial block faces (neighborhood commercial); e.g. a 1/4 mile block that is half commercial frontage would equal 1/8 mile nbhcom	0.66	0.00	4.10
path miles 1/2mi_od	Miles of off-street local and regional multi-use paths	0.48	0.00	4.56

Table 12 (continued)

Name	Description	Mean[1]	Min	Max
reg. path miles 1/2mi_h * female	Miles of regionally significant off-street multi-use paths (e.g. rails to trails, commuter paths) * Female	0.11	0.00	2.13
reg. path miles 1/2mi_od	Miles of regionally significant off-street multi-use paths (e.g. rails to trails, commuter paths)	0.23	0.00	3.45
prop. swlk. 1/2mi_od	Proportion of streets with complete sidewalks (not counting off-street paths or alleys)	0.77	0.14	0.98
prop. vol. 5k 1/2mi_od * female	Proportion of streets with traffic volumes between from 5,000-10,000 vehicles per day * Female	0.05	0.00	0.31
prop. vol. 10k 1/2mi_od * female	Proportion of streets with traffic volumes between from 10,000-20,000 vehicles per day * Female	0.05	0.00	0.95
prop. vol. 20k 1/2mi_od	Proportion of streets with traffic volumes 20,000 vehicles per day or higher	0.04	0.00	0.95
tour prop. slp. 2% 1/2mi_od[6]	Proportion streets and paths along tour with average slope (up or down) from 2-4%	0.17	0.00	0.53
tour prop. slp. 4% 1/2mi_od[6]	Proportion streets and paths along tour with average slope (up or down) 4% or more	0.07	0.00	0.61
xwalks per mi. 1/2mi_od	Marked crosswalks (at traffic signals or elsewhere) per street mile	2.77	0.23	20.50

Predicted routes

Name	Description	Mean[1]	Min	Max
com. miles	Miles of commercial street frontage along route, any type of street	0.29	0.00	1.84
left turn unsg. 10k per mi.	Unsignalized left turns per mile from street with 10,000 plus vehicles per day	0.16	0.00	4.00
nbh. com. miles	Miles of collector class streets along route (~5-10,000 vehicles per day) weighted by proportion commercial block faces (neighborhood commercial)	0.08	0.00	1.18
path miles	Miles of local and regional multi-use paths along predicted walk route	0.02	0	2.04
prop. blvd. * female[5]	Proportion of route with bike boulevard * Female	0.05	0.00	0.99

Table 12 (continued)

Name	Description	Mean[1]	Min	Max
prop. reg. mu. path	Proportion of route with off-street regional multi-use path	0.03	0.00	1.00
prop. vol. 5k * female	Proportion of route along streets with 5,000-10,000 vehicles per day, with or without on-street bike lanes * Female	0.11	0.00	1.00
prop. vol. 10k * female	Proportion of route along streets with 10,000-20,000 vehicles per day, with or without on-street bike lanes * Female	0.11	0.00	1.00
prop. vol. 20k * female	Proportion of route along streets with 20,000 or more vehicles per day, with or without on-street bike lanes	0.07	0.00	1.00
ratio to SP	ratio of predicted route length to shortest path	1.09	1.00	9.04
ratio to SP * with other	ratio of predicted route length to shortest path * Other household member on any stage of trip (adult or child)	0.20	0.00	1.23
swlk. miss. miles[7]	miles of missing sidewalk along route	0.16	0.00	1.87
tour prop. upslp. 2%[6]	proportion of segment lengths along entire tour route with average upslope from 2-4 percent	0.07	0.00	0.81
tour prop. upslp. 4%[6]	proportion of segment lengths along entire tour route with average upslope 4 percent or more	0.03	0.00	0.47
unsg. art. xing	number of secondary or primary arterial crossings along route without traffic or pedestrian signal control	0.30	0	6
Land Use[4]				
entropy 1/4mi_h	Land use entropy (0-1) using three categories: residential, commercial, and all other (home); calculated as $-\text{sum}(K*\ln(K))/\ln(\text{length}(K))$, where K is vector of land use proportions	0.26	0.06	0.66
entropy 1/4mi_od	Land use entropy (0-1) using three categories: residential, commercial, and all other (trip ends); calculated as $-\text{sum}(K*\ln(K))/\ln(\text{length}(K))$, where K is vector of land use proportions	0.37	0.00	0.98
FAR 1/2mi_od	Mean floor area ratio calculated as building area / parcel area	0.04	0.00	0.37

Table 12 (continued)

Name	Description	Mean[1]	Min	Max
isect. 4-way ratio 1/2mi_h	Ratio of four-way intersections to all intersections, excluding alleyways (home)	0.44	0.05	0.77
isect. 4-way ratio 1/2mi_od	Ratio of four-way intersections to all intersections, excluding alleyways (trip ends)	0.42	0.00	0.80

Self-selection[8]

Name	Description	Mean	Min	Max
biking nbh. self-select	Household importance of good biking neighborhood when looking for current home, from "not at all important" to "extremely important" (1-5)	3.79	1	5
walking nbh. self-select	Household importance of good walking neighborhood when looking for current home, from "not at all important" to "extremely important" (1-5)	4.17	1	5

[1] Statistics calculated over all trips included in the analysis. Statistics provided are accurate for the sample in Models 1-4. Model 5 had a slightly smaller valid sample due to missing self-selection data.

[2] Travel days assumed to span 2:00 A.M. to 2:00 A.M. the following day; tours that have not returned home by 2:00 AM are clipped at that point and re-start the following day.

[3] Data for driver's licenses and household vehicles were missing for 7 households out of 333. Mean values for cars per license were substituted for these cases.

[4] _od indicates the averaged value within circular buffers of given radii at each trip end, origin and destination; for tour variables, it is the average of all trip ends on the tour. _h indicates the value within a circular buffer of given radius around the traveler's home, regardless of the origin and destination of the trip and tour.

[5] Bike boulevards, sometimes called neighborhood greenways, are mostly quiet, residential streets that prioritize walking and, especially bicycling with traffic calming, diversion of motor vehicles, pavement markings, and signage.

[6] Average upslope and downslope for each segment are calculated separately as the sum of elevation gain and loss divided by segment length, where elevation change is measured as the net change over each 10 meter (32.8 ft) interval along a segment.

[7] Sidewalks were measured using a GIS algorithm and compared with curb length along both sides of each street, so that any proportion of sidewalk completeness from 0-1 is possible for each street segment, and missing sidewalk can be up to twice the street segment length, accounting for both sides.

[8] Due to missing data, statistics for self-selection variables reflect reduced sample used for Mode Choice Model 5.

Table 13 GPS Mode Choice Dataset Compared with Regional Travel Survey

Item	Family Activity Study (2010-2013)	Regional Travel Survey (2011)[1]	Regional Travel Survey- Households w/ children
Trips (within city of Portland)	13,261	10,834	4,279
walk			
…number	1,419	1,783	655
…share	10.7%	16.5%	15.3%
…shortest path distance (mi)	0.53	0.56	0.56
bike			
…number	1,501	799	319
…share	11.3%	7.4%	7.5%
…shortest path distance	1.72	2.31	2.27
auto			
…number	9,957	7,145	3,035
…share	75.1%	65.9%	70.9%
…avg. shortest path distance	2.44	2.54	2.47
transit[2]			
…number	384	1,107	270
…share	2.9%	10.2%	6.3%
…shortest path distance	2.27	2.96	3.33

[1] Oregon Travel and Activity Survey; distance thresholds applied as in GPS data. Applying regional sampling weights had only small effects on the results, and since the weights were not intended for subsamples, unweighted statistics are presented.

[2] Bus, light rail (MAX), streetcar, and aerial tram

Trip Context Measures. Each trip end was checked against a geocoded list of participant supplied common destinations to impute purpose to some degree.[5] A trip end within 200 meters (656 ft) of a common location was assigned a purpose based on that origin or destination. The distance threshold was taken from a study of common destinations based on a sub-sample of the same dataset used in this research (Dill &

---

[5] Common destination categories besides home were: work, child's school, and grocery store.

Broach, 2014). A hierarchy was used to handle cases where multiple common destinations fall within 200 m of a trip end: Home > Work > School > Grocery.

A special category was created for direct trips in either direction between home. and work and termed a commute trip. This follows from finding such trips to have unique impacts on travel behavior in previous work on bicyclist route choice (Broach et al., 2012). Such trips are more likely to be taken at fixed times and repeated frequently.

Returning home was not considered as a common destination, since it is assumed that the place where an activity has just occurred (now the trip origin) has a greater impact on mode choice than the fact that one is going home. It is possible that returning home does have an impact within a multimodal travel day; for instance, one might need to get a private car or bike back home at the end of a trip chain.

Two central city variables capture trips that start, end, or stay within the central city analysis district as defined by the local Metropolitan Planning Organization (MPO). I included these variables

Figure 9 Central city analysis district, metered parking, and bike accessible river crossings in Portland

primarily to proxy for parking cost and travel speed, which were not readily available at

the small scale required. As shown in Figure 9, the central city district includes most of the metered parking in Portland. The area also covers the Downtown Transit Mall where many bus and all light rail lines converge and have additional roadway and signal priority. Most of the central district is a tight grid of one-way, often congested streets, with traffic lights timed to keep speeds low and protect pedestrians. One versus two downtown trip ends were considered separately based on the hypothesis that moving a car within downtown would require an additional parking search, possibly incur an additional charge, and likely not be significantly faster than competing modes.

Figure 9 also displays six of the eight bike accessible Willamette River crossings in relation to the central city area. Two additional crossings existed, one to the north and one to the south, at the time of the study. A new transit and nonmotorized bridge, Tilikum Crossing, was not yet constructed during the data collection period.

My treatment of distance differed somewhat from common practice and highlights one advantage of using predicted routes rather than shortest paths. I hypothesized that distance would have two effects on walk and bike mode choice decisions: 1) a proximity effect captured by the shortest network path between origin and destination, and 2) an excess distance effect captured by the ratio of the least cost, predicted path to the shortest path (e.g. a ratio of 1.5 would indicate a typical cyclist might have to go 50% out of their way if they cycled on this trip). The first effect is commonly captured in mode choice models and represents the fact that human powered transportation modes are slower and require effort in proportion to distance and so compete less well as distance increases. The second effect is not generally captured and

82

represents a sort of summary measure of network quality. An ideal connection would provide a cyclist or pedestrian with their preferred route along the shortest path. As they are forced to detour in search of better routes, the less competitive a mode becomes, all else equal. Shortest path distance is fixed across modes to capture just the proximity effect of mode competition.

Panel Effects. The travel data used were generated by two unbalanced panels with each recording data across two time periods. Some participants naturally recorded more trips in each time period such that their preferences are overrepresented in the model. There might also be differences between time periods in each cohort or even underlying trends in mode preferences progressing across all four time periods. Controlling for person-specific effects in non-linear discrete choice models is more difficult than in linear modeling frameworks. Fixed effects cannot be used with the multinomial logit (MNL) model alongside other person-specific variables such as sociodemographics or home-based built environment measures and leads to identification issues. Random effects are possible but require more complex model forms such as Mixed Multinomial Logit (MMNL) with the associated specification, estimation, and interpretation complexities that come with them.

An alternative, if the main concern is representativeness, is applying weights to the individuals based on number of trips taken. This was done by Hood et al. (2011) to address a highly unbalanced volunteer smartphone sample for route choice modeling. The panel used here showed reasonable balance, with the most frequent traveler having taken just 0.8% of trips (while comprising 0.2% of the sample). Mean trip frequency was

within two trips of the median. In other words, there did not appear to be a person or group of people whose preferences would seriously bias the model.

It was initially tempting to at least exclude trips repeated between the same origin and destination, since these would seem to add no new information while potentially biasing sample preferences. However, identifying a repeated trip requires specifying some threshold distance, since GPS trip ends are not perfectly precise like geocoded addresses. And, what about the same trip repeated in a different context, say on a weekend instead of a weekday, or with another person along instead of alone? These do add new information, even if mode choice does not change. Finally, if I deleted repeated trips only when mode choice stayed fixed, I would only be shifting the preference bias to multimodal travelers. In the end, I decided to acknowledge the limitation of the individual-specific panel effects but estimate the model with all trips having equal weight.

Potential time trends were more troubling, since Portland was actively pursuing policies to increase walk and bike travel over the collection period, and the sample selection had been related to areas specifically slated to receive improved street designs. Furthermore, since each cohort collected within about a three month window, there were potential period effects due to weather differences and construction projects on key facilities, particularly for cycling. Non-random attrition and survey fatigue might also influence both cohorts in the second year. For these reasons, I decided to control for panel period effects by testing for an overall second period trend and individual cohort second year trends.

Area-Based Measures. Area-based built environment and land use attributes were calculated using simple circular buffers around each trip origin and destination as well as each geocoded home location. Based on the lack of consensus for optimal buffer size, a range of distances were specified for testing: 1/8 mile (200 m), 1/4 mile (400 m), 1/2 mile (800 m), and 1 mile (1600 m). For both walking and cycling, the 1/2 mile (800 m) buffers consistently outperformed the others regardless of the attribute measured.

Network-based shortest path buffers were considered, but the close correlation of network buffer size to street connectivity and barriers resulted in some odd effects that would have made interpretation more difficult. Areas with high street connectivity tend to be overrepresented in network buffers, inflating values for features correlated with connectivity such as density. For example, a neighborhood with a small section of dense streets and the rest mostly inaccessible (e.g. water, park, golf course) will have most of its buffer in the dense, gridded area. Conversely, an area with many private or semi-private ways missing from the travel network, such as a commercial or university campus, may not be included in a network buffer, but may actually be a dense area and add to land use diversity. For these reasons, and because the issue of connectivity is already controlled for with network-based shortest and predicted paths, network buffers were rejected in favor of simple circular buffers.

Sidewalks and marked crosswalk data were only available within Portland city limits. Some buffered trip ends, however, extended beyond the boundary. For these variables, a second set of buffers was created following the boundary. Since counts and

sums would be deflated in the smaller areas, sidewalk and crosswalk attributes were normalized by city street miles for use in modeling.

In total, over 114,000 trip end buffers and over 5,000 household buffers were intersected with 33 spatial built environment variables. Because of the large size of the intersecting data features, the technique used for each calculation had to weigh the importance of precision against the processing time, which for some attributes could stretch to multiple days on a desktop workstation. Linear travel infrastructure attributes were judged to benefit the most from precision, so they were intersected exactly with buffers, rather than allowed to dangle over the edge of a buffer area. Street, path, and block face attributes were weighted by the length of the clipped feature. Census blocks and tax lot parcels, on the other hand, were reduced to centroids and tallied as counts, instead of intersecting areas. This greatly reduced processing time, and the idea that each parcel counts once regardless of size matches the concept of access to opportunities for nonmotorized travel, rather than access to square footage.

Two attributes, sidewalk coverage and commercial land use share by street type, were available from the block face street profiles developed as part of the pedestrian route choice models and described more completely in Chapters 3 and 4. The block face scans used an automated GIS method to take a cross section of the street and adjacent infrastructure, buildings, and land use every 5 meters (32.8ft) and summarized to street segment level. In each case, the block face measures outperformed the areal measures of the same attributes, suggesting some promise for the method going forward. Many other variables were calculated by the block face scans (e.g. sidewalk width, skyline height,

building setbacks, building age, enclosure etc.), but the high degree of correlation suggested factor analysis or some other data reduction technique that was beyond the scope of this research.

I tested a number of variants of land use intensity, diversity, and overall design, and model results consistently found some versions superior to others, so that only three appear in the final models. For intensity, reliable employment data was not easily available at an appropriately small scale, and among alternatives floor-area ratio (FAR) consistently outperformed variants of residential population density. For diversity, a three-class land use entropy measure (residential, commercial, and all other) outperformed a four-class version (splitting residential into single and multi-family), and proportion of commercial block faces by street type explained more than either for walk choice.

Following the literature, I initially tested origin and destination buffers separately; however, there is a logical problem when attempting to differentiate trip ends for all daily travel. Each destination becomes the origin for the next trip in sequence, and each original origin, usually home, becomes the final destination for a tour or travel day. Since each location's buffer measurements are nearly identical between trips, there is too much correlation across cases to reliably identify a model with origin and destination measures. This paradox also makes interpretation difficult, given the linkages between trips. A more complex, tour-based model would be needed to examine differences between origin and destination factors across sequences of related trips. For a trip-based solution to the problem, I simply averaged origin and destination trip ends for all trips. Since only

differences in attributes matter in discrete choice models, the moderation effect on

extreme values should not matter.

Bike route generation

      To generate potential bike routes, I adapted a route choice model developed by

Broach et al. (Broach et al., 2012), after first replicating the model with the sample of

bike trips used in this research to check for transferability between the different samples.

The replication effort is described in Chapter 4 and Appendix B. The model distinguished

between commute (defined by as any direct trip between home and work in either

direction) and non-commute travel, but no other trip or person characteristics were found

to be significant predictors.

      The original path-size logit route choice model included a non-linear distance

term, and most variables depend on the total length of the route. This made it impossible

to use the utility function directly to generate a link-based cost-minimizing path between

an origin-destination pair for which the final route distance is unknown. Broach et al.

(Broach et al., 2012) also provided marginal rates of substitution (MRS, i.e. the rate at

which one is willing to trade off changes in one attribute for changes in another), but

these still require that the path distance be known *a priori*.

      In order to circumvent the problem of undefined route length, I first calculated the

shortest path, and then used that distance in place of the actual route distance. This

preserved the substitution rates between all distance-normalized variables in the model

(e.g. if a cyclist were willing to go up to 20% farther to avoid a busy street, the method

would select an alternate route on quiet streets as long as the detour were 20% or less, and the same for point-based intersection variables, etc.).

Two bridge facility attributes were included in the route choice model that were not distance based, and the MRS for those attributes will only be preserved exactly in the adapted method if the alternative is the shortest path. As the competing route alternative increases in length from the shortest path, the MRS for bicycle-accommodating bridges will be biased downward (i.e. the method will underestimate willingness to detour). The downward bias is reasonably small over most likely ranges of route detours, at most -4% when an alternative is 10% longer than the shortest path and -15% for an alternative 50% longer.

Table 14 provides the cost function components for the least cost path selection, all based on the marginal rates of substitution provided in Broach et al. (2012) and included in Appendix B. Distance was set to 1.0 divided by the shortest path distance, for convenience. Dijkstra's algorithm was applied to the network to generate a single least-cost path between origin and destination. Where trips did not start or end at an intersection node, they were joined to the nearest location on a network link, and the distance and cost of that link was pro-rated based on the starting or ending position along the segment.

Table 14 Adapted Cost Function Parameters for Generating Least Cost Bike Paths

Attribute	Non-commute	Commute
Divided by shortest path distance		
Distance	1.000	1.000
Turns	0.074	0.042
Distance with upslope 2-4 %	0.723	0.371
Distance with upslope 4-6 %	2.904	1.230
Distance with upslope >= 6 %	11.066	3.239
Traffic signal exc. right turns	0.036	0.021
Stop sign (/mi)	0.009	0.005
Left turn, unsig., AADT 10-20k[1]	0.162	0.091
Left turn, unsig., AADT 20k+	0.431	0.231
Unsig. cross AADT >= 10k right turn	0.067	0.038
Unsig. cross AADT 5-10k exc. right turn	0.072	0.041
Unsig. cross AADT 10-20k exc. right turn	0.104	0.059
Unsig. cross AADT 20k+ exc. right turn	0.617	0.322
Prop. bike boulevard	-0.179	-0.108
Prop. bike path	-0.260	-0.160
Prop. AADT 10-20k w/o bike lane	0.223	0.368
Prop. AADT 20-30k w/o bike lane	1.373	1.400
Prop. AADT 30k+ w/o bike lane	6.194	7.157
Not divided by shortest path distance		
Bridge w/ bike lane	0.282	0.156
Bridge w/ sep. bike facility	0.814	0.414

[1] AADT=Average Annual Daily Traffic

Walk Route Generation

I used the pedestrian route choice model described in Chapter 4 to generate least-cost pedestrian paths. Since the parameters of the model were linear and did not depend on route length, the inverse of the parameters could be applied directly to generate a single maximum utility (least cost) path using Dijkstra's algorithm.

The route choice model was developed using the same dataset used in the mode choice component of this project. While not ideal, I do not believe this represents a substantial problem for my objectives here. The worry would be that it introduces some endogeneity into our modeling, since the route choices generated were developed from the same trips included in the mode choice model. An offsetting advantage is the lack of need to worry about the route choice model generalizing to our sample. Future work will apply the method to an external mode choice sample.

Mode Choice Model Development

Based on the assumption of utility maximizing behavior, I specified the familiar multinomial logit (MNL) model form to predict the probability of choosing each of four modes (walk, bike, auto, or transit). I followed standard practice in developing each model, using hypothesis tests to choose between competing model specifications, dropping insignificant variables (with certain noted exceptions for clarity in comparing models), and constraining related parameters that did not meet the specified significance level ($\alpha = 0.05$).

The five mode choice specifications are summarized in Table 15. Model 1 includes only trip context variables. Model 2 adds areal measures of transportation infrastructure. Model 3 replaces areal measures with attributes calculated along the predicted maximum

91

utility routes for walking and cycling. In Models 2 and 3, I was careful to provide

comparable measures of each attribute in the area and route specifications so that the

model results could be meaningfully compared. Model 4 adds density, diversity, and

design land-use and built environment variables, as well as measures of transportation

infrastructure around a traveler's home. Model 5 adds measures of residential self-

selection. Table 16-Table 20 present full results of each model, with the added sets of

attributes listed first.

Table 15 Mode Choice Models Summary

	Constants Only	Model 1 Trip Context	Model 2 Areal Infra-structure	Model 3 Predicted Routes	Model 4 Land Use + Routes	Model 5[1] + self-selection
N param. estimated	3	27	40	38	48	51
n trips	13,261	13,261	13,261	13,261	13,261	12,981
LL(Null)	-17007.4	-17007.4	-17007.4	-17007.4	-17007.4	-16652.9
LL(Model)	-9972.6	-7854.0	-7695.7	-7624.7	-7511.1	-7187.2
pseudo-R^2 adjusted	0.413	0.537	0.547	0.549	0.556	0.565
likelihood ratio test chi-square stat from left[2]		4237.2	388.6	-	227.2	-
non-nested hypothesis test z-stat from left[2]		-	-	-8.5	-	-
mean predicted probability of chosen mode as percent…[3]						
…walk	16.7 %[4]	41.0 %	43.5 %	43.9 %	44.1 %	44.6 %
…bike	11.1 %	18.1 %	19.4 %	19.8 %	20.9 %	23.4 %
…auto	76.0 %	81.6 %	81.9 %	81.9 %	82.2 %	82.6 %
…transit	2.9 %	17.0 %	16.9 %	16.5 %	17.8 %	17.7 %
…overall	60.2 %	68.2 %	68.9 %	68.9 %	69.3 %	70.0 %

[1] Self-selection responses were missing for some households; statistical tests cannot be performed since samples differ.

[2] For each test performed, the earlier model can be rejected as the true model at the $\alpha=0.01$ significance level.

[3] Mean predicted probability represents the expected portion of mode choices correctly identified (in-sample) by applying the modeled probabilities case by case. Due to the non-linear nature of logit models, order matters, so relative contributions to predictive power cannot be compared meaningfully as variables are added.

[4] equal to mode shares for constants only model

Table 16 Mode Choice Model 1 Trip Context (MNL coefficients with t-stats in parentheses)[1]

Variable	Walk	Bike	Auto	Transit
cohort 1, year 2	-0.20 (-2.44)	-0.40 (-6.23)[a]		-0.40 (-6.23)[a]
constant	-0.98 (-9.32)	0.19 (1.86) [‡]		2.46 (8.89)
Trip Context				
commute		1.36 (12.4)[b]		1.36 (12.4)[b]
grocery	-0.58 (-5.84)[c]	-0.58 (-5.84)[c]		
school	0.76 (7.75)	0.25 (2.74)		
work	0.44 (3.76)[d]	0.44 (3.76)[d]		
weekend	0.23 (2.87)	-0.17 (-2.85)[e]		-0.17 (-2.85)[e]
central city, one end			-0.70 (-9.14)	
central city, both ends		-1.56 (-11.0)[f]	-1.56 (-11.0)[f]	
rivcross * male		1.09 (8.44)		
ln(miles_sp)	-2.13 (-29.0)	-0.75 (-15.3)		
ln(tour_miles_sp)	-0.68 (-17.3)	-0.27 (-6.23)[g]		-0.27 (-6.23)[g]
tour stops		-0.13 (-7.19)		0.09 (4.65)
ln(transit stop to dest.)				-0.49 (-11.7)
ln(transit stop to orig.)				-0.73 (-18.9)
Sociodemographics				
cars per driver			0.76 (10.4)	
female		-0.51 (-8.37)		
zero car household	0.71 (4.44)			1.45 (8.30)
% cases mode available	63.9%	100%	100%	100%
Log-likelihood (Null)	-17007.4			
Log-likelihood (Model)	-7854.0			
Pseudo-R^2 Adjusted	0.535			
Num. parameters est.	31			
Number of trips	13261			

Table 16 (continued)

Table 17 Mode Choice Model 2 Areal Bicycle and Pedestrian Infrastructure (MNL coefficients with t-stats in parentheses)[1]

Variable	Walk	Bike	Auto	Transit
cohort 1, year 2	-0.24 (-2.77)	-0.47 (-7.05)[a]		-0.47 (-7.05)[a]
constant	-1.53 (-7.12)	-1.49 (-4.56)		2.55 (9.21)
Areal transportation infrastructure				
blvd. miles 1/2mi_od * female		0.30 (4.38)		
com. miles 1/2mi_od	-0.27 (-7.09)[b]			
com. miles 1/2mi_od	-0.27 (-7.09)[b]			
local street miles 1/2mi_od		0.10 (5.77)		
local street miles 1/2mi_od * with other	-0.06 (-10.2)			
nbh. com. miles 1/2mi_ od	0.46 (3.22)[c]			
nbh. com. miles 1/2mi_ od	0.46 (3.22)[c]			
path miles 1/2mi_od	0.35 (4.66)			
reg. path miles 1/2mi_od		0.65 (8.03)		
prop. swlk. 1/2mi_od	0.75 (3.15)			
prop. swlk. 1/2mi_od				
prop. vol. 5k 1/2mi_od * female		-0.04 (-0.04)†		
prop. vol. 10k 1/2mi_od * female		-4.55 (-4.34)[d]		
prop. vol. 20k 1/2mi_od		-4.55 (-4.34)[d]		
tour prop. slp. 2% 1/2mi_od		1.04 (2.62)		
tour prop. slp. 4% 1/2mi_od		-1.20 (-1.90) ‡		

Table 17 (continued)

Variable	Walk	Bike	Auto	Transit
path miles 1/2mi_od	0.34 (4.58)			
prop. swlk_1/2mi_od	0.49 (2.15)			
prop. vol. 5k 1/2mi_od * female		-0.13 (-0.12)[†]		
prop. vol. 10k 1/2mi_od * female		-4.67 (-3.13)		
prop. vol. 20k 1/2mi_od		-5.16 (-3.68)		
reg. path miles 1/2mi_od		0.63 (7.32)		
tour prop. slp. 2% 1/2mi_od[6]		1.12 (2.81)		
tour prop. slp. 4% 1/2mi_od[6]		-0.90 (-1.42)[†]		
xwalkspermile_1/2mi_od	0.18 (4.11)			
Trip Context				
commute		1.35 (12.1)[e]		1.35 (12.1)[e]
grocery	-0.58 (-5.82)[f]	-0.58 (-5.82)[f]		
school	0.80 (8.02)	0.17 (1.82)[d]		
work	0.36 (2.93)[g]	0.36 (2.93)[g]		
weekend	0.33 (3.98)	-0.20 (-3.18)[h]		-0.20 (-3.18)[h]
central city, one end			-0.63 (-7.62)	
central city, both ends		-1.24 (-6.47)[i]	-1.24 (-6.47)[i]	
rivcross * male		0.83 (5.92)		
ln(miles_sp)	-2.16 (-28.6)	-0.70 (-14.0)		
ln(tour_miles_sp)	-0.70 (-16.4)	-0.28 (-6.31)[j]		-0.28 (-6.31)[j]
tour stops		-0.12 (-6.57)		0.10 (4.86)
ln(transit stop to dest.)				-0.49 (-11.9)
Sociodemographics				
cars per driver			0.76 (10.1)	
female		-0.43 (-3.63)		
zero car household	0.95 (5.61)			1.39 (7.92)
% cases mode available	63.9%	100%	100%	100%

96

Table 17 (continued)

Log-likelihood (Null)	-17007.4
Log-likelihood (Model)	-7659.7
Pseudo-R^2 adjusted	0.537
Number of parameters	40
Number of trips	13261

[1] Parameter estimates significant at the α=0.05 level unless noted otherwise

[‡] Significant at the α=0.10 level

[†] Not significant at the α=0.10 level

[a-j] Equality constrained parameters, no significant difference (α=0.05)

Table 18 Mode Choice Model 3 Predicted Bicyclist and Pedestrian Routes (MNL coefficients with t-stats in parentheses)[1]

	Walk	Bike	Auto	Transit
cohort 1, year 2	-0.26 (-3.05)	-0.48 (-7.27)[c]		-0.48 (-7.27)[c]
constant	-0.40 (-3.06)	1.74 (11.5)		2.50 (9.06)
Predicted routes				
com. miles	-1.56 (-5.64)			
nbh. com. miles	1.80 (4.42)			
L. turn unsg. 10k per mi.		-0.27 (-3.52)		
path miles	2.72 (7.63)			
prop. blvd. * female		0.78 (3.58)		
prop. reg. mu. path		1.29 (4.12)		
prop. vol. 5k * female		-0.75 (-5.09)[a]		
prop. vol. 10k * female		-0.75 (-5.09)[a]		
prop. vol. 20k * female		-2.26 (-7.62)		
ratio to SP		-1.10 (-11.2)[b]		
ratio to SP * with other	-1.10 (-11.2)[b]			
swlk. miss. miles	-1.40 (-4.18)			

97

Table 18 (continued)

Variable	Walk	Bike	Auto	Transit
tour prop. upslp. 2%		-1.48 (-2.83)		
tour prop. upslp. 4%		-4.69 (-4.39)		
unsg. art. xing	-0.42 (-4.28)			
Trip Context				
commute		1.36 (12.3)[d]		1.36 (12.3)[d]
grocery	-0.52 (-5.22)[e]	-0.52 (-5.22)[e]		
school	0.68 (6.95)			
work				
weekend	0.29 (3.52)	-0.20 (-3.26)[f]		-0.20 (-3.26)[f]
central city, one end			-0.67 (-8.65)	
central city, both ends		-1.60 (-10.8)[g]	-1.60 (-10.8)[g]	
rivcross * male				
ln(miles_sp)	-1.91 (-21.0)	-0.73 (-14.5)		
ln(tour_miles_sp)	-0.66 (-16.2)	-0.26 (-5.83)[h]		-0.26 (-5.83)[h]
tour stops		-0.12 (-6.78)		0.09 (4.59)
ln(transit stop to dest.)				-0.49 (-11.7)
ln(transit stop to orig.)				-0.73 (-18.9)
Socio-demographics				
cars per driver			0.78 (10.5)	
female		-0.36 (-4.68)		
zero car household	1.18 (8.69)[i]			1.18 (8.69)[i]
% cases mode available	63.9%	100%	100%	100%
Log-likelihood (Null)	-17007.4			
Log-likelihood (Model)	-7624.7			
Pseudo-R^2 adjusted	0.549			
Num. parameters est.	38			
Number of trips	13,261			

[1] Parameter estimates significant at the α=0.05 level unless noted otherwise

[a-h] Equality constrained parameters, no significant difference (α=0.05)

Table 19 Mode Choice Model 4 Areal Infrastructure, Land Use, and Route Attributes (MNL coefficients with t-stats in parentheses)[1]

Variable	Walk	Bike	Auto	Transit
cohort 1, year 2	-0.43 (-7.55)[a]	-0.43 (-7.55)[a]		-0.43 (-7.55)[a]
constant	-0.58 (-4.19)	-0.00 (-0.01)[†]		1.23 (3.61)
Area Infrastructure				
blvd. miles 1/2mi_h * female		0.13 (3.31)[b]		
bike lane miles 1/2mi_od * male		0.13 (3.31)[b]		
reg. path miles 1/2mi_h * female		0.29 (2.83)		
Area Land Use				
entropy 3-class 1/4mi_h				2.27 (5.00)
entropy 3-class 1/4mi_od				1.18 (2.89)
floor area ratio 1/2mi_od	5.88 (4.89)		-2.68 (-2.97)	
isect. 4-way ratio 2_h		1.12 (3.66)		
isect. 4-way ratio 1/2mi_od		1.79 (4.94)		
Predicted Routes				
com. miles	-1.78 (-6.12)			
nbh. com. miles	1.82 (4.39)			
L. turn unsg. 10k per mi.		-0.23 (-2.99)		
path_miles	2.58 (7.08)			
prop. blvd. * female		0.66 (2.98)		
prop. reg. mu. path		1.64 (4.94)		
prop. vol. 5k * female		-0.76 (-5.06)[c]		
prop. vol. 10k * female		-0.76 (-5.06)[c]		
prop. vol. 20k		-1.19 (-3.81)		
ratio to SP		-1.08 (-10.9)[d]		
ratio to SP * with other	-1.08 (-10.9)[d]			
swlk. miss. miles	-1.54 (-4.44)[e]			

99

Table 19 (continued)

Variable	Walk	Bike	Auto	Transit
tour prop. upslp. 2%		-1.54 (-2.86)[e]		
tour prop. upslp. 4%		-5.79 (-5.28)		
unsg. art. xing	-0.40 (-4.00)			
Trip Context				
commute		1.41 (12.0)		0.98 (3.66)
grocery	-0.50 (-5.03)[f]	-0.50 (-5.03)[f]		
school	0.72 (7.00)	0.22 (2.33)		
work	0.39 (3.18)[g]	0.39 (3.18)[g]		
weekend	0.30 (3.65)	-0.18 (-2.82)[h]		-0.18 (-2.82)[h]
central city, one end			-0.39 (-4.27)	
central city, both ends		-0.66 (-3.19)[i]	-0.66 (-3.19)[i]	
rivcross * male		0.92 (6.30)		
ln(miles_sp)	-1.82 (-19.8)	-0.67 (-12.9)		
ln(tour_miles_sp)	-0.71 (-16.8)	-0.28 (-6.20)[j]		-0.28 (-6.20)[j]
tour stops		-0.10 (-5.75)		0.09 (4.34)
ln(transit stop to dest.)				-0.48 (-11.5)
ln(transit stop to orig.)				-0.71 (-18.1)
Sociodemographics				
cars per driver			0.69 (8.29)	
female		-0.39 (-4.85)		
zero car household	1.24 (8.06)[k]	0.57 (3.14)		1.24 (8.06)[k]
% cases mode available	63.9%	100%	100%	100%
Log-likelihood (Null)	-17007.4			
Log-likelihood (Model)	-7511.1			
Pseudo-R^2 adjusted	0.558			
Num. parameters est.	48			
Number of trips	13261			

Table 19 (continued)

[1] Parameter estimates significant at the α=0.05 level unless noted otherwise

[†] Not significant at the 10 percent level

[a-k] Equality constrained parameters, no significant difference (α=0.05)

Table 20 Mode Choice Model 5 Areal Infrastructure, Land Use, Route Attributes, and Self-Selection (MNL coefficients with t-stats in parentheses)[1]

Variable	Walk	Bike	Auto	Transit
cohort 1, year 2	-0.42 (-7.30)[a]	-0.42 (-7.30)[a]		-0.42 (-7.30)[a]
constant	-1.99 (-8.15)	-1.88 (-7.49)		1.93 (4.63)
Residential Self-Selection				
biking nbh. self-select		0.54 (15.3)		
walking nbh. self-select	0.30 (6.92)			-0.16 (-3.04)
Area Infrastructure				
blvd. miles 1/2mi_h * female		0.14 (3.31)[b]		
bike lane miles 1/2mi_od * male		0.14 (3.31)[b]		
reg. path miles 1/2mi_h * female		0.26 (2.48)		
Area Land Use				
entropy 3-class 1/4mi_h				2.10 (4.50)
entropy 3-class 1/4mi_od				1.04 (2.52)
floor area ratio 1/2mi_od	5.58 (4.56)		-1.99 (-2.17)	
isect. 4-way ratio 2_h		0.19 (0.59) [†]		
isect. 4-way ratio 1/2mi_od		1.79 (4.81)		
Predicted Routes				
com. miles	-1.65 (-5.60)			
nbh. com. miles	1.84 (4.40)			
L. turn unsg. 10k per mi.		-0.24 (-3.03)		
path miles	2.49 (6.73)			

Table 20 (continued)

Variable	Walk	Bike	Auto	Transit
prop. blvd. * female		0.63 (2.74)		
prop. reg. mu. path		1.58 (4.61)		
prop. vol. 5k * female		-0.71 (-4.62)[c]		
prop. vol. 10k * female		-0.71 (-4.62)[c]		
prop. vol. 20k		-0.98 (-3.05)		
ratio to shortest path		-1.08 (-10.7)[d]		
ratio to SP * with other	-1.08 (-10.7)[d]			
swlk. miss. miles	-1.22 (-3.46)			
tour prop. upslp. 2%		-1.42 (-2.56)		
tour prop. upslp. 4%		-5.25 (-4.78)		
unsg. art. xing	-0.38 (-3.81)			
Trip Context				
commute		1.52 (12.1)		0.99 (3.71)
grocery	-0.51 (-5.05)[e]	-0.51 (-5.05)[e]		
school	0.73 (7.09)	0.24 (2.47)		
work	0.49 (3.90)[f]	0.49 (3.90)[f]		
weekend	0.28 (3.28)	-0.19 (-3.01)[g]		-0.19 (-3.01)[g]
central city, one end			-0.45 (-4.89)	
central city, both ends		-0.76 (-3.63)[h]	-0.76 (-3.63)[h]	
rivcross * male		0.93 (6.18)		
ln(miles_sp)	-1.88 (-20.0)	-0.68 (-12.7)		
ln(tour_miles_sp)	-0.71 (-16.6)	-0.26 (-5.59)[i]		-0.26 (-5.59)[i]
tour stops		-0.11 (-5.98)		0.09 (4.26)
ln(transit stop to dest.)				-0.48 (-11.3)
ln(transit stop to orig.)				-0.72 (-17.9)
Sociodemographics				
cars per driver			0.62 (7.44)	
female		-0.37 (-4.49)		

102

Table 20 (continued)

Variable	Walk	Bike	Auto	Transit
zero car household	1.28 (8.26)[j]	0.64 (3.36)		1.28 (8.26)[j]
% cases mode available	64.0%	100%	100%	100%
Log-likelihood (Null)	-17007.4			
Log-likelihood (Model)	-7187.2			
Pseudo-R^2 adjusted	0.565			
Num. parameters est.	51			
Number of trips	12981			

[1] Parameter estimates significant at the α=0.05 level unless noted otherwise

[†] Not significant at the α=0.10 level

[a-j] Equality constrained parameters, no significant difference (α=0.05)

As shown in Table 15, each subsequent model except Model 5 was used to test the hypothesis that the previous specification represents the true model. The preferred statistical test for nested model hypotheses (i.e. one model represents a restricted version of another) is the likelihood ratio test (LRT). The LRT generates an approximately chi-square distributed statistic for the null hypothesis of equivalence between restricted and unrestricted versions of the same model. The test statistic is calculated as:

$$D = -2(LL(R) - LL(U)) \sim \chi^2 \qquad (7)$$

where LL is the likelihood value at the estimated parameters, R is the restricted model, and U the unrestricted model (Ben-Akiva & Lerman, 1985). When neither model can be cast as a restricted version of the other, as in the case of Models 2 and 3, an adjusted rho-square test statistic or Akaike Likelihood Ratio Index (LRI) can be used to calculate a z-statistic as:

$$z = -[-2((\bar{\rho}_2^2 - \bar{\rho}_1^2) * LL(0)) + (K_2 - K_1)]^{1/2}) \qquad (8)$$

where z is a standard normal variate, $\bar{\rho}^2 = 1 - \frac{LL(\mathbf{B})-K}{LL(0)}$, $LL(0)$ is the log likelihood with

no coefficients, $LL(\mathbf{B})$ is the log likelihood with estimated coefficients, and K is the

number of model parameters estimated (Ben-Akiva & Lerman, 1985; Ben-Akiva &

Swait, 1986; Koppelman & Bhat, 2006). The test statistic indicates the (asymptotic)

probability that the observed difference $\bar{\rho}_2^2 - \bar{\rho}_1^2$ would occur under the null hypothesis

that model 1 was the true model.

Using either the LRT, or the LRI as appropriate, Models 2-4 each reject the

preceding specification. That Model 3 with route-based measures rejects Model 2 with

area-based measures is consistent with the hypothesis that measuring walking and biking

travel environments along likely routes provides better information than measuring

around trips ends. That Model 4 rejects Model 3 indicates that there is additional,

independent information contained in areal measures of land use around trip origins and

destinations, and also contained in bike infrastructure around a person's home, regardless

of where a specific trip takes place. While Model 5's addition of self-selection variables

appears to improve model fit based on the adjusted fit statistic, the improvement cannot

be meaningfully compared to prior models, since missing data resulted in a different

sample.

Overall, there is considerable consistency across models in terms of trip context

variables as additional information is added. Trip purpose, proximity, cost and time

proxies, weekday/weekend, and sociodemographics contribute predictably to mode

choice, particularly walking and biking, as information about the specific trip built

environment is added. The remainder of this section describes results by variable group.

Because MNL coefficients are not readily interpreted beyond sign and significance, much

of the discussion refers to various predicted effects calculations provided in Table 21.

Table 21 Expected Attribute Effects on Walk and Bike for Selected Attributes Based on Model 4

	Point Elasticity	Point Elasticity (when present)	%Chg. Prob. For Unit Change	MRS Detour (equiv. chg. detour ratio)[1]
Bike				
cohort 1, year 2			-35.8%	
Area Infrastructure				
blvd. miles 1/2mi_h * female	0.04	0.11		-0.12
bike lane miles 1/2mi_od * male	0.04	0.12		-0.12
reg. path miles 1/2mi_h * female	0.02	0.23		-0.27
Area Land Use				
isect. 4-way ratio 2_h	0.44			-1.04
isect. 4-way ratio 1/2mi_od	0.67			-1.66
Predicted Routes				
L. turn unsg. 10k per mi.			-19.6%	0.21
prop. blvd. * female	0.03	0.15		-0.61
prop. reg. mu. path	0.04	0.22		-1.52
prop. vol. 5k * female	-0.05	-0.16		0.70
prop. vol. 10k * female	-0.04	-0.14		0.70
prop. vol. 20k	-0.04	-0.11		1.10
ratio to SP	-0.92			1.00
tour prop. upslp. 2%	-0.08	-0.10		1.43
tour prop. upslp. 4%	-0.10	-0.13		5.36

Table 21 (continued)

	Point Elasticity	Point Elasticity (when present)	%Chg. Prob. For Unit Change	MRS Detour (equiv. chg. detour ratio)[1]
Trip Context				
commute			214.0%	
grocery			-37.9%	
school			20.8%	
work			40.0%	
weekend			-15.2%	
central city, one end[2]			32.1%	
central city, both ends			-45.7%	
rivcross * male			116.4%	
ln(miles_sp)	-0.53			
ln(tour_miles_sp)	-0.22			
tour stops			-9.7%	
Sociodemographics				
cars per driver[2]			-64.1%	
female			-37.5%	
zero car household			63.3%	
Walk				
cohort 1, year 2			-33.8%	
Predicted Routes				
com. miles	-0.14	-0.18		
nbh. com. miles	0.05	0.14		
path miles	0.04	0.23		
ratio to SP * with other	-0.08	-0.75		
swlk. miss. miles	-0.08	-0.09		
unsg. art. xing			-31.1%	

Table 21 (continued)

	Point Elasticity	Point Elasticity (when present)	%Chg. Prob. For Unit Change	MRS Detour (equiv. chg. detour ratio)[1]
Area Land Use				
floor area ratio 1/2mi_od	0.12			
Trip Context				
Grocery			-36.2%	
school			76.2%	
work			37.5%	
weekend			26.5%	
central city, both ends			52.1%	
ln(miles_sp)	-1.00			
ln(tour_miles_sp)	-0.39			
Sociodemographics				
cars per driver[2]			-55.1%	
zero car household			168.5%	

[1] Marginal rate of substitution, detour for attribute (e.g. a unit change in miles of bike path within 1/2 mile of home for women is equivalent to a -0.27 change in detour ratio, or a change from 1.27 to 1.0)

[2] Cross elasticity, auto

Elasticities provide the expected percent change in probability of choosing a mode for a percent change in a given attribute. They are a useful measure for continuous variables of how much a given attribute impacted the odds of selecting a mode given the choice context in the sample. I follow Hensher et al.'s (2005) recommendation and calculate the probability-weighted sample enumeration elasticities, which weights each case-level elasticity by the predicted choice probability of a given alternative. Rare attributes (e.g. bike paths or steep slopes) will be deflated in elasticity calculations

107

because where an attribute is zero, the elasticity for that case will also be zero. For that reason, I also provide an average probability-weighted elasticity calculated only for cases where the attribute is present (e.g. only for cases where a bike path is available for at least part of a trip).

For non-continuous variables, point elasticities are difficult to interpret, since a 1% change in a variable like "weekend day" is not very meaningful. For those variables, I calculated the average probability change for a unit increase in the attribute. To ensure calculations remained within the range of the sample, for an attribute greater than the minimum value, I subtracted one and measured the probability change from the new value to the original value. For cases where the attribute was already at the minimum sample value, I increased the value by one, calculating the probability change between the original value and the increased value.

Also included are marginal rates of substitution (MRS) for select variables with regard to the detour ratio between the predicted and shortest path route. The MRS provides the ratio at which a decision maker would be indifferent to trading off one attribute for another. MRS provides a better approximation of the relative value of different features in a hypothetical choice situation or policy scenario versus elasticities, which focus on the features that contributed most to a specific set of observed choices. The next section discusses model results by variable group.

Mode Choice Results by Attribute Group

Distance. For cycling trips, there were clear effects of both proximity and excess distance. A 1% increase in shortest path distance or the detour ratio decreased cycling

probability by about 0.5% and 0.9%, respectively. Having to go out of their way was about 1.8x as costly as having to pedal farther in deciding whether to ride. For walking, as might be expected, the proximity effect dominates, with a 1% increase in shortest route distance decreasing the odds of walking by an equal amount. Mirroring findings in the pedestrian route choice model, there was a significant excess distance effect when walking with another household member but not when walking alone.

Transit access and egress distances are likewise important. Each 1% increase in distance at the origin end yields about a -0.6% change in transit probability, and a 1% increase in distance from stop to destination results in a -0.4% change. The higher weight on access to stops from an origin was significant and consistent across models. Perhaps the extra distance on the origin end adds to the uncertainty of catching a bus or train, while the destination has no schedule pressures.

Purpose. For the most part, trip purpose has similar predicted effects on cycling and walking. School and work-based trips increase the odds of choosing a nonmotorized mode, while it is less attractive to grocery shop on foot or by bike, all else equal. Even controlling for related factors such as traveling to the central city, direct commutes attracted cycling trips more than any other purpose. For a given trip, a person was more than three times as likely to bike if the trip ends were home and work. Direct commutes were also attractive for transit travel.

Other trip context attributes. Trips into or out of the central city increased the chance of using any non-auto mode. Since the central city in the models captures both a parking charge area and congestion on downtown streets, the two likely impacts cannot

be disentangled. Travel within the central city made walking and transit even more attractive; however, cycling was surprisingly no more competitive than driving for travel between downtown locations. The finding is consistent with bicycle commuting behavior, since many downtown commuters have off-street storage for their bikes. Getting the bike out for a midday errand would require the extra hassle of first retrieving it from a bike room or similar. There may also be theft concerns locking bikes on the street in the central district. Another possibility is that while the routes into and out of downtown Portland have received much policy attention over the years, the downtown core itself has received noticeably less. While the infrastructure variables should control for these differences to some degree, the relatively wide, high volume streets in the one-way grid downtown may be a barrier to less confident cyclists. A more positive spin might be that the ease of getting around the core by transit and walking simply erodes any advantages of cycling.

Men were more than twice as likely to bike on trips that cross one of the Willamette River bridges that connect downtown to the east side of Portland (compared to all trips not crossing one of the bridges?). Women, on the other hand, are no more likely to bike on such trips. While many of the bridges have bicycle accommodations—multi-use paths or bike lanes—the result suggests that they may not go far enough to make all cyclists feel safe on river crossings. This is something that warrants further investigation, and in light of this finding, it will be particularly interesting to see how Portland's newest bridge, designed exclusively for nonmotorized and transit use, fares in attracting female cyclists.

The data sample studied showed pronounced weekend mode shifts. Taking the same trip on a weekend versus a weekday increases the chance of walking by about 27% but decreases the probability of biking or riding transit by about 15% and 17%, respectively. One possibility for this is that people would have preferred to walk for more of their travel, but time constraints, likely more prevalent on weekdays, made walking impractical for many weekday trips. Trip purposes not captured here, such as social/recreational, dining out, and non-grocery shopping trips, might have been more common on weekends, and perhaps better served in some cases by walking or driving instead of transit and cycling. Finally, transit typically runs less frequently on weekends, reducing its attractiveness.

Tour variables. While the mode choice models were all trip-based, some tour-level variables were significant predictors of trip mode. Tour distance had about 40% as large an estimated impact on walking and biking probabilities as trip distance. A 1% increase in tour distance decreases the chance of walking or biking by about 0.4% and 0.2%, respectively. The odds of transit travel also diminishes with tour distance, while trip distance has no significant effect, likely reflecting the added costs of transfers and trip planning on more complex transit trip chains.

The number of stops on a tour has been used as a measure of potential travel complexity (Soltani & Allan, 2006). As expected, as tour complexity increased, cycling became less and less attractive. Each additional stop decreased probability by about 10%. Unlike walking, each stop generally requires locking and unlocking a bike, and as the trip chain grows, so does the burden of finding suitable cycling routes and carrying things

111

picked up along the way. The finding that the number of tour stops would increase the odds of a transit trip was less expected. One possibility is that Portland's public transit agency, TriMet, uses time-based fares, offering unlimited transfers until the fare expires. Another possibility is noise in the GPS data processing. If a transfer took long enough and did not involve much walking, it could have been incorrectly flagged as an intervening activity.

Tour slope variables were borrowed from bicycle mode choice specifications in Rodríguez and Joo (2004), and they outperformed trip-level slope attributes, which only capture the current trip direction. Many downhill bike trips will include an uphill return. Public bike share or improved transit-bike integration could potentially reduce this effect. Grades of 2% and greater, as they are for route choice, were powerful deterrents to cycling when present. Based on the marginal rate of substitution, just 10% of a tour climbing up such hills reduced the odds of cycling by the same amount as a 14% detour from the shortest path.

Trip end infrastructure. Model 2 specifies a range of area-based infrastructure attributes meant to capture the quality of walking and cycling conditions. Since those area-based measures were mostly displaced by better fitting route-level attributes, I will not describe them in detail here, instead referring interested readers to Table 17. It is worth highlighting that of the trip end buffers tested, from 1/8 to 1 mile in radius, 1/2 mile versions consistently outperformed other sizes for predicting walk and bike travel. Since Model 2 was mainly intended for comparison to Model 3, the "best" insignificant versions of various route measures were left in the model for presentation purposes.

112

Removing those parameters did not change the results of the model rejection test. In a few cases, even significant parameter estimates were unintuitive in Model 2. As an example, moderate slopes at the tour-level positively affected the probability of cycling. This likely reflects the inclusion of irrelevant sloped areas in the buffer calculations. For example, there are steep slopes all along the Alameda Ridge in northeast Portland, but cyclists rarely need to traverse it. Such results aside, the area-based parameters generally agreed fairly closely with the route-based measures, suggesting that it is still useful to measure specific infrastructure as well as land-use even when network analysis is infeasible. The numeric example later in this chapter provides a stronger reason not to rely on area measures alone: poor sensitivity to network changes.

Model 4 included area based infrastructure and land use measures alongside predicted route variables. The results were intriguing and are taken up at greater length in the discussion in Chapter 6. Especially for decisions to bike, having infrastructure around one's home appeared to be an important factor, but which specific infrastructure matters is split across genders. For women, a 10% increase in miles of bicycle boulevards or off-street paths within a half-mile of home increases the chance of biking on any trip by 1.1% and 2.3%, respectively. This is an example of when the elasticity when present is more appropriate, since the majority of households had neither boulevards nor bike paths nearby. For men, bike lanes near home had a similar impact on cycling, with a 10% increase in bike lane miles expected to shift biking odds on any trip by 1.2%, when bike lanes were present. For walking, land-use intensity, as measured by floor area ratio, at trip ends encouraged walk travel.

Well-connected street grids, as measured by the ratio of all intersections that are 4-way within either a 1/2 mile of home or of trip ends, were strong predictors of cycling in the model. The effect of a more connected grid equaled or exceeded that of even proximity or route directness.

The fact that land use intensity impacted only walking relative to other modes, instead of, for example, simply reducing the utility of driving, suggests that there was something specifically about walking between denser places that offered an advantage over even cycling and transit, even after controlling for pedestrian amenities like sidewalks, crossing aids, and proximity.

Finally, land-use mix, measured by the entropy of residential, commercial, and other uses within 1/4 mile of home or of trip ends, were significant factors in transit use. The result matched earlier findings by Cervero (1996), who surmised that mixed uses near stops level the playing field for transit riders that otherwise find it difficult to compete with driving when wanting to combine other errands with commutes. Smaller, 1/4 mile buffers fit better in this case, further supporting the idea that mixed uses need to be in close proximity to stops to be useful.

Predicted route attributes. Route measures along the single, predicted highest utility path included infrastructure attributes and interactions between infrastructure and adjacent land-use. A gender split was again evident for cycling infrastructure variables, as it was for area-based measures. Among women, traffic volumes from 5,000-20,000 vehicles per day reduced the odds of cycling, regardless of whether striped bike lanes were present. Bicycle boulevards, on the other hand, increased the probability that

114

women would bike considerably, the equivalent of reducing route detour by 61%. A 10%

increase in the portion of a route that uses a boulevard raised the probability of a woman

choosing to bike on that trip by 1.5%, when bike boulevards were present along the way.

For both men and women, routes along off-street bike paths were strongly

attractive, equivalent to a 2.5x reduction in route detour, or 2.2% for each 10% rise in the

proportion following a path, when paths were present. At high levels of traffic, men and

women's preferences converged, and they were equally discouraged from cycling when

facing streets with 20,000 vehicles a day or more, with or without a bike lane. A host of

intersection variables was found to be important for cyclist route choice, but only the

most difficult maneuvers factored in mode choice. Left turns across moderate to heavy

traffic and without a traffic signal reduced the chance of cycling on a trip by about 20%

for each additional such crossing encountered per mile.

No gender differences were evident in route factors predicting walking mode

choice. For every 10% increase in missing sidewalk along route, walking probability fell

by about 1%. Similar to results for cycling, only the most onerous crossings seemed to

factor in pedestrian mode choice. Each arterial crossing without a traffic or pedestrian

signal reduced walk choice probability by 31%.

Modeling suggested that the nuanced relationship between walking and

commercial streets extended to mode choice as well as route choice for pedestrians.

Commercial land use along a route in general had a negative impact on the odds of

walking. This could be linked to higher motor vehicle activity levels, especially in and

out of commercial driveways, or it could simply be that people walking prefer quieter

streets until they arrive at their ultimate destination. Commercial development on smaller streets, which I labeled as neighborhood commercial here, exactly offset the negative general impact of commercial land use along a route, essentially leaving it not preferred, but equally attractive compared with a quiet residential street. This is an area of investigation that might benefit from a specification that allows for heterogeneous preferences. There may well be a group that enjoys commercial, and another that prefers walking through areas not quite as bustling. Finally, off-street paths encouraged people to walk to destinations even though recreational trips were excluded, though interestingly paths did not factor in route choice for pedestrians.

Sociodemographics. As expected from existing work, even after controlling for perhaps the largest set of factors to date, women were considerably less likely to bike than men in otherwise identical circumstances. Increases in cars per licensed driver left all non-auto modes less likely alternatives. Zero car households (n=14) in the study were most likely to fill their transportation gap with walking and transit, and to a lesser extent cycling.

Attitudes. Model 5 introduced two residential self-selection variables into the model structure. Due to missing data, Model 5 was not strictly comparable to the other models; however, the importance of attitudes was clear from the results. All else equal, including the area measures around a household's residence, those who chose a neighborhood for its good biking qualities were more likely to bike (around 39% more likely for a unit increase on the 5-point scale) on any given trip. Those choosing their neighborhood for walkability were more likely to walk (about 21% for a one-point

116

increase) on given trip but less likely to use transit, perhaps because they have more of what they need within walking distance. Counter to expectations, only one home area environment variable, network connectivity, was crowded out by the addition of self-selection attributes.

Numeric example. A simple example demonstrates how using predicted routes instead of areal measures improves model sensitivity when policy is focused, as it often is, on specific corridors. Consider the case of an origin currently connected to a destination two miles away by an on-street bike lane along a busy arterial (20,000 or more vehicles per day). Planners are considering improving the connection by developing a bike boulevard along quieter streets parallel to the arterial. The new boulevard would allow cyclists to avoid the higher traffic street, but where the bike lane follows the shortest path, the new boulevard route would be 2.2 miles long, requiring cyclists to detour about 10% out of their way.

The mode choice models developed in this chapter can provide predictions for scenarios like the given one. For Model 2, before the change, 0.5 miles of bike lane fall within each trip end half-mile buffer, there are no existing bike boulevards, all other areal measures are set to their average values, the terrain is flat, the trip is outside the central city, and it will not cross the river. For a simple tour from home to destination and back again (so that the tour is 4 miles long with 1 stop), and for none of the special purposes included in the model (e.g. perhaps a non-grocery shopping trip or social-recreational trip), Model 2 predicts that the odds of cycling for each leg of the two mile trip are 1.7% for men and 1.3% for women. The gender gap reflects both the positive impact of area

117

bike lanes for men and the generally lower probability for women choosing to bike, all else equal. Adding the hypothetical bike boulevard affects the predicted outcome only due to the new 0.5 miles of boulevard that would fall within each trip end buffer. This change results in no predicted increase in the propensity for men to cycle on the given trip, since they are insensitive to area boulevards, and a small rise from 1.3% to 1.5% for women.

With similar assumptions as used for Model 2, and also assuming that the bike lane is currently the predicted route, Model 3 predicts the chance of cycling as 3.1% for men and 2.1% for women. In this case the gender gap reflects only the general downward bias for women choosing to cycle, since men and women were equally averse to riding in a bike lane alongside heavy traffic in the route-based mode choice model. Adding the bike boulevard option would shift the predicted route from the bike lane to the boulevard, reflecting the fact that cyclists value the new facility more than the 10% additional distance (Broach, Gliebe, & Dill 2012, Appendix B). The predicted route change would shift the probability of cycling via changes in three attributes: the shift away from a heavy traffic route (positive for men and women), the shift onto a bike boulevard (positive for women), and the increased detour (negative for both). The net result in this case is strongly positive, with the chance of cycling for the trip increasing from 3.1% to 21.5% for men, and from 2.1% to 30.1% for women. In this case, the predicted effect of moving bike traffic off of the high traffic street easily outweighs the required detour, and the presence of a bike boulevard for the entire route reverses the gender gap for the hypothetical trip.

118

6 Discussion

The primary motivation for this research was to develop and a new method that uses predicted pedestrian and cyclist routes, and the features along them, to predict the choice of whether to walk or bike on a given trip. I argued that measuring large areas around trip ends, as is most commonly done, seemed like an inefficient way to measure travel environments, that these aggregate measures likely measure much that is irrelevant at the trip level, while potentially diluting or missing what is actually important. Statistical comparisons of matched area and route-based models supported those hypotheses. In the tested sample of data, a pair of representative routes averaging 1.1 miles (1.8 km) for pedestrians and 2.4 miles (3.8 km) for cyclists provided significantly more information about mode choice decisions than areal summaries of, on average, 46 miles (74 km) of streets and paths within trip end buffers. Particular routes captured walking and bicycling environments with surprising efficiency.

Furthermore, I argued that there was a fundamental mismatch between areal measures and common nonmotorized policies aimed at making strategic investments along specific routes and corridors, or between specific origins and destinations. Portland recently constructed a new transit, bicycle, and pedestrian bridge, Tilikum Crossing, that links predominately residential areas east of the Willamette River with a large regional employment, health care, and education cluster on the west side. Before the new bridge, any cyclist traveling north faced a mile-long detour to reach a bridge with reasonable accommodation (the bridges farther south are high traffic with narrow sidewalks on just one side). Using a mode choice model based on areal measures, the new bridge would have either no predicted effect on choice outcomes, if the bridge fell outside of the trip

end buffers, or a diluted effect from the marginal addition of multiuse paths on the bridge itself to the area as a whole. The area-based framework has no way to properly weight the usefulness of a specific route connection. Using a route-based framework, the predicted cycling and pedestrian routes would shift to the new bridge for many predicted trips, and route attributes and detour measures would reflect the new, improved path. Increased sensitivity to marginal route improvements is a key feature of the predicted route-based mode choice framework developed in this research.

Other advantages of the proposed approach include more compelling tests of facility-level impacts on walk and bike decisions and an allowance for tradeoffs between various attributes impacting mode choice. When route-level attributes are measured within larger areas (e.g. total miles of bike lane), the connection between trip-level decisions and the built environment is less clear. How is it that miles of bike lane would impact a decision to bike when the facilities do not connect an origin and destination? In that case, the measurement of bike lanes per area might just be serving as a proxy for other features that increase the propensity to bike. In contrast, adding the condition that a bike lane must connect an origin and destination within a reasonable detour makes the test of a specific facility's impact on behavior more convincing. In addition, the embedded random utility framework allows for walking and bicycling suitability measures that allow tradeoffs among network attributes. Unlike all or nothing approaches to measuring network connectivity such as bicycle level of stress, which rates route quality based on the worst single segment (Mekuria, Furth, & Nixon, 2012), the method developed here weights positive and negative factors along the entire route. For example,

a bicycle route mainly along attractive off-street paths but with a brief, high-stress (high disutility) stretch along a busy roadway. The level of stress route quality would be determined entirely by the short, high traffic stretch, while the predicted route method introduced in this research would weigh the positive impact of off-street paths against the negative impact of the busy stretch, resulting in a network quality measure somewhere between the two. The predicted route method could be extended from route to network-level quality scores in future research and its ability to predict walking and bicycling behavior compared with level of stress and other existing approaches.

While route-based measures have certain advantages, the mode choice modeling exercise also indicated that trip end infrastructure measures and land use in a person's home neighborhood provide complementary information about decisions to bike and walk. An unforeseen result of specifying route-based travel infrastructure attributes was their greater independence from traditional density, diversity, and design land use measures. That allowed for an identifiable model incorporating both specific route attributes, trip end land use measures, and home neighborhood areal infrastructure and land use variables. That model outperformed models using only areal or only route measures, controlling for trip context. This result is at least consistent with the possibility of land use and infrastructure effects on the formation of walk and bike habits and perhaps also lower entry barriers. Traveling or living in more walkable or bikeable areas might increase the odds of walking and biking for other trips, forming habits that increase the chance of walking or biking for a specific trip regardless of its route quality. For bicycling, which arguably has a higher cost of entry than walking for many due to safety

perceptions, having lower traffic-stress options near home, including off-street paths, bike boulevards, and a well-connected local street grid, might reduce this entrance barrier and increase the probability of cycling on other trips. Each of these interpretations of the findings is speculative, and there are other possible explanations, including simply living near other cyclists (Wang, Akar, & Guldmann, 2015), but the implications are intriguing enough to warrant further efforts to identify the source of the residual area-based correlations.

A second major aim of the research was to compare the impact of route-level attributes on choices of where to walk and bike versus decisions of whether to walk and bike. Generally speaking, findings showed that most of the important factors influencing route choice also impacted mode choice, particularly in the case of bicycling.

For bicycling, low traffic-stress bike infrastructure, including bicycle boulevards and off-street paths, increases both the attractiveness of a route and the odds that a person will bicycle on a given trip in the first place. The same is true for more direct routes, routes with fewer moderate and steep hills, and routes with fewer difficult turning movements at major street intersections. There were also key differences between factors in route choice—for those that have already chosen to ride—and the choice of whether to bike in the first place. Most striking was the gender gap in preferences for lower traffic routes and specific bicycle infrastructure.

Route choice models developed both in this research and earlier projects with colleagues had found no significant interactions between gender and route preferences. This was surprising, given the United States' well-known gender gap in cycling. Mode

choice models, however, revealed significant gender gaps in decisions to cycle for specific trips. Women were deterred by even moderate levels of traffic, as low as 5,000 vehicles per day; while for men, the choice to cycle was not affected until traffic levels reached relatively high volumes of 20,000 vehicles per day or more. And, while both men and women are willing to go out of their way to use low-stress bicycle boulevards when available, only women were more likely to cycle when boulevards connected origins and destinations.

Evidence also suggested a gender difference in cycling across the Willamette River, which separates Portland's downtown core from the largely residential east side. While men were more than twice as likely to cycle on cross-river trips, perhaps because of time and cost advantages of cycling into downtown, women were no more likely to ride on such trips, suggesting that some factor might be offsetting the advantages for them. Possibilities include the perceived inadequacy of cycling infrastructure downtown, on the bridge crossings themselves, or some other factor beyond those controlled for in the models. Even after accounting for these specific preference differences between genders and trip context, route, and built environments, women were around 38% less likely to cycle. Lack of access to low traffic-stress routes explains part of the gender gap in cycling, and for specific trips, bicycle boulevards can even close or eliminate the gap, but there is still a statistical bias against women cycling that we do not completely understand. Possible avenues for exploration include gender-specific attitudes toward cycling, and household role differences that may impose time or other constraints that make cycling less attractive.

The value of simple, striped on-street bike lanes adjacent to parking, road shoulders, or curbs was another aspect on which route and mode choice results differed. In route choice models, bike lanes had the effect of offsetting adjacent vehicle traffic volumes. Cyclists were indifferent toward using a high traffic street with bike lane or an otherwise similar, parallel low traffic street. In decisions of whether to bike, however, bike lanes were found to have no significant direct impact. They do not offset the negative impact of low to moderate adjacent traffic volumes, for women, or the negative impact of high traffic streets for both genders (men were insensitive to low and moderate traffic volumes). Results are consistent with that notion that bike lanes are useful to cyclists but unlikely to induce people to ride. In this context, it will be fascinating to study the newer generation of protected bike lanes now appearing around the US.

For walking, comparisons were more limited due in part to pedestrian route choice modeling being in an earlier stage of development, both in this research and in the literature, but some consistent factors were identified. Subpar walking facilities showed slightly different but consistently negative correlation with the choice of route and whether to walk. Having to cross a busy arterial street without a traffic or pedestrian signal was a significant deterrent in both route and mode choice. Unpaved streets and alleyways, in route choice modeling, were mirrored by a slightly different attribute, missing sidewalks, in mode choice, but I surmise that both reflect similar incomplete street environments that discourage walking.

There is an interesting, complex relationship between walking choice and commercial streets. In decisions of where to walk, commercial districts along smaller

124

streets (minor arterials or neighborhood collectors) were attractive, and pedestrians were willing to walk out of their way to use them, while similar non-commercial streets were avoided. For the choice of whether to walk, only arterial commercial districts reduced the odds of walking, while neighborhood commercial districts and busy non-commercial streets were neutral. More nuanced measures may be needed to more fully understand the relationship between walking, traffic, and commercial design, and for now I can only say that neighborhood-scale commercial on lower order streets appeals to those already walking but is unlikely to draw additional walking trips unless it brings attractive destinations closer to travelers, reducing trip distance.

The pedestrian route choice model did not find that people were willing to walk out of their way to use an off-street path, given a low-traffic street alternative. The presence of paths along predicted routes, however, did have a positive influence on choosing to walk. It appears that although not willing to go much out of their way to use them, people are more likely to walk if a path happens to be on the way.

While attitudinal measures were limited to residential self-selection into walkable and bikeable neighborhoods, this research is one of the first to test self-selection effects on mode choice at the trip level, with controls for the route and area-level factors that comprise such neighborhoods. Participants were asked to rate the importance of choosing "a good biking neighborhood" or a "good walking neighborhood" when looking for their current home. The self-selection effect was found to hold up under such controls, increasing the probability of cycling by 39% and the chance of walking by 21% for a one-point increase on the five-point scale. While these impacts are substantial, they do

125

fall within the effect range of many other attributes in the mode choice model, consistent with the idea that attitudes are important but far from the only thing that matters.

Including self-selection in the mode choice model crowded out some of the explanatory power of the area-based built environment measures, but route-level features were largely unaffected. This suggests that route quality has an impact on walking and biking rates that is independent of residential self-selection effects. Those that placed importance on living in a good walking neighborhood were less likely to use transit, a somewhat counterintuitive finding, given the importance of walking as a transit access mode. One possibility is that it reflects a tradeoff between better transit access—in Portland often along arterials or near freeways—and more walkable neighborhoods. While only suggestive, it is worth exploring further to understand if some opportunities are being missed to provide better transit service or better connections to existing service in walkable places outside of the downtown core.

A final aim was to test the practicality of incorporating predicted walk and bike routes into a model structure similar to existing, trip based regional travel demand models. With one important caveat, the model form developed and tested in this research could be implemented within many regional modeling frameworks immediately. In fact, an early version of the bicycling portion has been incorporated Portland Metro's trip-based regional modeling system. A key remaining hurdle is at what spatial scale to apply the predicted routes. In this research, a predicted walk or bike route only represented the travel environment between a single origin and destination pair for a single prospective trip. A regional model typically operates on a zone-based system, and having a single

route represent all potential walk or bike trips by all travelers between two zones is problematic, to say nothing of how to handle trips that start and end in a single zone. It would also undercut one of the modeling framework's notable advantages, sensitivity to route improvements, since a route-level improvement would only affect mode choice to the extent it improved the specific connection between zone centroids. In the long-run, improving spatial resolution for all modes may be the solution. In the interim, I would suggest fixing shortest path distance to the traditional zone-based measure but then sampling a number of origin and destination pairs between and within zones, averaging over predicted routes to estimate expected route quality for a given zone pair or intrazonal area.

There are other limitations both to the conceptual framework and the specific application presented here to bear in mind. Throughout, I have mentioned various shortcomings and workarounds for coercing what are really sequences of trips—trip chains or tours—into a trip by trip choice process. Partly, this was a necessary simplification to develop and test a new modeling technique; after all, models of tour-based walk and bike route choice to my knowledge do not yet exist, so I am not sure how one would go about predicting such routes other than falling back to individual trips. If I included a variable for previous mode in the mode choice models estimated here, I expect I would find that it explained a large portion of mode choice likelihood. Much of the mode choice decision is made when leaving the house in the morning, and this phenomenon is commonly captured via mode availability in tour-based models. There is a sense that this kind of travel behavior thinking may already be losing relevance as the

next mode choice becomes less fixed to the last. In many cities, car sharing, bike sharing, and ride sharing mean that I could drive to work (in a shared car), bike to lunch (on a shared bike), take transit to a restaurant, and walk home—a nightmare for a tour-based model, but no problem in a trip-based framework.[6]

In addition to assuming independence of individual trip mode choice, another simplification in the conceptual framework is treating trip generation and distribution (the destination) as exogenous to the mode choice process. It is more likely that, to some degree, people make decisions to travel at all, and if so to where, with a chosen mode already in mind. This is the difference between the hypothetical base of models in this research, "I'm going to this particular place, what mode should I use, considering the route?" and, "I'd like to walk to some place, where shall I go?" Parallel work by others, including some of my own colleagues, is working to better understand how specific types of places produce and attract nonmotorized trips. It will be useful at some point to better integrate these modeling steps for walk and bike travel.

A final pair of limitations to discuss, not with the framework but the application presented here, is the limited treatment of motorized modes in the mode choice models, and the specific travel sample used. To the extent important variables of auto and transit utility were omitted in the models, estimates of the remaining parameters would be inconsistent, and the interpretations provided suspect. That the majority of estimated effects were consistent with existing travel behavior theory and research, and that many

---

[6] Although, my use of tour-level variables would become problematic, since things like slope are less relevant when one can coast downhill on a bike share bike and take some other mode home. There is also the problem of forming expected tours to calculate the tour level variables—here, they were taken as given from the observed travel data.

mode choice effects were consistent with route choice effects were taken as signs that the omitted variable bias was not too great. It will be interesting in future work to examine the sensitivity of these estimates to more complete transit and auto specifications.

The sample used for estimation in this research is not necessarily generalizable to other populations and geographies. The sample of families was self-selected into a study confined to specific street corridors in predominately single family, residential neighborhoods with well-connected street grids, mostly complete sidewalks, and generally at least basic bicycle accommodation. While the route and mode preferences of families with children under 16 might not reflect precisely those of the larger population, they are an important travel demographic, accounting for nearly half of all trips nationally (NHTS, 2009). I look forward to testing the framework in other settings and with different populations. There is nothing inherent in the developed framework that would limit broader application.

Implicit in much of travel demand theory is Tobler's so-called first law of geography, "[E]verything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). For people walking and biking, nothing is nearer than the particular routes that they use. In thinking about decisions to walk and bike, then, it only makes sense to incorporate our best approximation of those routes. This research presented a means of doing so, using walk and bike route choice models to generate predicted routes traversing the specific travel environment someone might consider, along with other factors, in choosing how to travel. Calculated sensitivities to various route-level factors were shown to be useful predictors of mode choice, adding

129

significant information even after controlling for trip context, sociodemographics, area-based built environment, and residential self-selection. Tools, methods, and frameworks were developed to aid in replicating the research elsewhere using increasingly common GPS travel data. I hope that this contribution might, in some small way, help to uncover what motivates bicycling and walking, and how policy might better encourage sustainable travel choices to the benefit of urban places.

References (includes Appendix references)

Agrawal, A. W., Schlossberg, M., & Irvin, K. (2008). How Far, by Which Route and Why? A Spatial Analysis of Pedestrian Preference. *Journal of Urban Design*, *13*(1), 81–98. http://doi.org/10.1080/13574800701804074

Akar, G., Fischer, N., & Namgung, M. (2013). Bicycling Choice and Gender Case Study: The Ohio State University. *International Journal of Sustainable Transportation*, *7*(5), pp 347–365. Retrieved from http://dx.doi.org/10.1080/15568318.2012.673694

Antonini, G., Bierlaire, M., & Weber, M. (2006). Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, *40*(8), 667–687. http://doi.org/10.1016/j.trb.2005.09.006

Badoe, D. A., & Miller, E. J. (2000). Transportation - Land Use Interaction: Empirical Findings in North America and their Implications for Modeling. *Transportation Research - D*, *5*(4), 235–263. http://doi.org/10.1016/S1361-9209(99)00036-X

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, *12*(1), 9–24. http://doi.org/10.1016/0167-8116(95)00002-J

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge, MA: The MIT Press.

Ben-Akiva, M., & Swait, J. D. (1986). The Akaike Likelihood Ratio Index. *Transportation Science*, *20*(2), 133–136. http://doi.org/10.1287/trsc.20.2.133

Bergman, A., Gliebe, J., & Strathman, J. G. (2011). Modeling Access Mode Choice for Inter-Suburban Commuter Rail. *Journal of Public Transportation*, *14*, 23–42.

Black, C., Collins, A., & Snell, M. (2001). Encouraging Walking: The Case of Journey-to-school Trips in Compact Urban Areas. *Urban Studies*, *38*(7), 1121–1141. http://doi.org/10.1080/00420980124102

Bomberg, M., Zorn, L., & Sall, E. (2013). lncorporating user based perspective of livabitity proj ects in SF-CHAMP mode choice modets. *Transportation Letters*, *5*(2), 83–95.

Borst, H. C., de Vries, S. I., Graham, J. M. A., van Dongen, J. E. F., Bakker, I., & Miedema, H. M. E. (2009). Influence of environmental street characteristics on walking route choice of elderly people. *Journal of Environmental Psychology*, *29*(4), 477–484. http://doi.org/10.1016/j.jenvp.2009.08.002

Bovy, P. H. L., & Stern, E. (1990). *Route choice: wayfinding in travel networks*. Boston: Kluwer Academic Publishers.

Bradley, M., & Bowman, J. L. (2006). *SACSIM/05 Activity-based travel forecasting model for SACOG featuring DAYSIM- the person day activity and travel simulator* (Technical memo No. number 4). Retrieved from http://jbowman.net/ProjectDocuments/SacSim/SACOG tech memo 4--Mode

Choice.20060802.pdf

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. http://doi.org/10.1016/j.tra.2012.07.005

Broach, J., Gliebe, J., & Dill, J. (2009). Development of a Multi-class Bicyclist Route Choice Model Using Revealed Preference Data. In *IATBR*. Jaipur, India.

Broach, J., Gliebe, J., & Dill, J. (2011). Calibrated Labeling Method for Generating Bicyclist Route Choice Sets Incorporating Unbiased Attribute Variation. *Transportation Research Record: Journal of the Transportation Research Board*, *2197*, 89–97. http://doi.org/10.3141/2197-11

Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American housing survey. *Transportation Research Part A: Policy and Practice*, *30*(5 PART A), 361–377. http://doi.org/10.1016/0965-8564(95)00033-X

Cervero, R., & Duncan, M. (2003). Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area. *American Journal of Public Health*, *93*(9), 1478–1483. http://doi.org/10.2105/AJPH.93.9.1478

Committee for Determination of the State of the Practice, & Forecasting, in M. A. T. (2007). *Metropolitan Area Travel Forecasting Current Practice and Future Direction. Transportation Research Board Special Report 288.* Retrieved from http://onlinepubs.trb.org/onlinepubs/sr/sr288.pdf

Committee on Physical Activity, Health, Transportation, and L. U. (2005). *TRB Special Report 282: Does the Built Environment Influence Physical Activity* (TRB Special Report No. 282). Washington, D.C.: Transportation Research Board Institute of Medicine. Retrieved from http://www.trb.org/publications/sr/sr282.pdf

Davison, K. K., Werder, J. L., & Lawson, C. T. (2008). Children ' s Active Commuting to School : Current Knowledge and Future Directions. *Preventing Chronic Disease*, *5*(3), 1–11. http://doi.org/A100 [pii]

DiGuiseppi, C., Roberts, I., Li, L., & Allen, D. (1998). Determinants of Car Travel on Daily Journeys to School: Cross Sectional Survey of Primary School Children. *British Medical Journal*, *316*(7142), 1426–1428.

Dill, J. (2009). Bicycling for transportation and health: the role of infrastructure. *Journal of Public Health Policy*, *30 Suppl 1*(2009), S95–S110. http://doi.org/10.1057/jphp.2008.56

Dill, J., & Broach, J. (2014). Travel to Common Destinations. *Transportation Research Record: Journal of the Transportation Research Board*, *2413*, 84–91. http://doi.org/10.3141/2413-09

Dill, J., McNeil, N., Broach, J., & Ma, L. (2014). Bicycle boulevards and changes in physical activity and active transportation: findings from a natural experiment.

Preventive Medicine, *69 Suppl 1*, S74–8.
http://doi.org/10.1016/j.ypmed.2014.10.006

Ermagun, A., & Samimi, A. (2015). Promoting Active Transportation Modes in School Trips. *Transport Policy*, *37*, pp 203–211. Retrieved from http://dx.doi.org/10.1016/j.tranpol.2014.10.013

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. http://doi.org/10.1080/01944361003766766

Ewing, R., Schroeer, W., & Greene, W. H. (2004). School Location and Student Travel Analysis of Factors Affecting Mode Choice. *Transportation Research Record*, *1895*(1), 55–63. http://doi.org/10.3141/1895-08

FHWA. (2008). Bicycle and pedestrian provisions of federal transportation legislation. Retrieved April 18, 2012, from http://www.fhwa.dot.gov/environment/bikeped/bp-guid.htm

FHWA. (2010a). Nonmotorized Transportation Pilot Program investments and project types. Retrieved from http://www.fhwa.dot.gov/environment/bikeped/ntpp/funding_type.pdf

FHWA. (2010b). *The National Bicycling and Walking Study: 15–Year Status Report*. United States Department of Transportation. Retrieved from http://trid.trb.org/view.aspx?id=920129

Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, *56*, 70–80. http://doi.org/10.1016/j.trb.2013.07.012

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. http://doi.org/10.1007/s11116-007-9136-6

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

Gliebe, J., & Kim, K. (2010). Time-Dependent Utility in Activity and Travel Choice Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *2156*, 9–16. http://doi.org/10.3141/2156-02

Gonzalez, P. A., Weinstein, J. S., Barbeau, S. J., Labrador, M. A., Winters, P. L., Georggi, N. L., & Perez, R. (2010). Automating mode detection for travel behaviour analysis by using global positioning systems-enabled mobile phones and neural networks. *IET Intelligent Transport Systems*, *4*(1), 37. http://doi.org/10.1049/iet-its.2009.0029

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. http://doi.org/10.1016/j.jtrangeo.2012.11.013

Handy, S. L., 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. 1), 64–73. http://doi.org/10.1016/S0749-3797(02)00475-0

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied Choice Analysis: A Primer*. Cambridge: Cambridge UP.

Kockelman, K. M. (1997). Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area. *Transportation Research Record*, *1607*(1), 116–125. http://doi.org/10.3141/1607-16

Koppelman, F. S., & Bhat, C. R. (2006). *A Self Instructing Course in Mode Choice Modeling : Multinomial and Nested Logit Models*. *USDOT; FTA*. Retrieved from http://www.ce.utexas.edu/prof/bhat/COURSES/LM_Draft_060131Final-060630.pdf

Kuzmyak, J. R., Walters, J., Bradley, M., & Kockelman, K. M. (2014). *Estimating Bicycling and Walking for Planning and Project Development: A Guidebook*. Retrieved from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_770.pdf

Lawson, C. T., Chen, C., & Gong, H. (2010). *Advanced Applications of Person-based GPS in an Urban Environment*.

Lin, J.-J., & Chang, H.-T. (2010). Built Environment Effects on Children's School Travel in Taipai: Independence and Travel Mode. *Urban Studies*, *47*(4), 867–889. http://doi.org/10.1177/0042098009351938

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3), 229–254. http://doi.org/10.1007/BF00133443

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (pp. 105–142). http://doi.org/10.1108/eb028592

McMillan, T. E. (2007). The relative influence of urban form on a child's travel mode to school. *Transportation Research Part A: Policy and Practice*, *41*(1), 69–79. http://doi.org/10.1016/j.tra.2006.05.011

Mekuria, M. C., Furth, P. G., & Nixon, H. (2012). *Low-Stress Bicycling and Network Connectivity*.

Miller, E. J., Roorda, M. J., & Carrasco, J. A. (2005). A tour-based model of travel mode choice. *Transportation*, *32*(4), 399–422. http://doi.org/10.1007/s11116-004-7962-3

Moiseeva, A., Jessurun, J., & Timmermans, H. J. P. (2010). Semiautomatic Imputation of Activity Travel Diaries. *Transportation Research Record: Journal of the Transportation Research Board*, *2183*(-1), 60–68. http://doi.org/10.3141/2183-07

Moudon, A. V., & Lee, C. (2003). Walking and bicycling: An evaluation of

environmental audit instruments. *American Journal of Health Promotion*, *18*(1), 21–37. http://doi.org/10.4278/0890-1171-18.1.21

Newman, J. P., & Bernardin, V. L. (2010). Hierarchical ordering of nests in a joint mode and destination choice model. *Transportation*, *37*(4), 677–688. http://doi.org/10.1007/s11116-010-9277-x

Park, S. (2008). Defining, Measuring, and Evaluating Path Walkability, and Testing Its Impacts on Transit Users' Mode Choice and Walking Distance to the Station. *University of California Transportation Center*, *UCTC Disse*(150), 239. Retrieved from http://www.uctc.net/research/diss150.pdf

Piatkowski P, D., & Marshall E, W. (2015). Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting. *Travel Behaviour and Society*, *2*(3), pp 166–173. Retrieved from http://dx.doi.org/10.1016/j.tbs.2015.02.001

Pont, K., Ziviani, J., Wadley, D., Bennett, S., & Abbott, R. (2009). Environmental correlates of children's active transportation: A systematic literature review. *Health and Place*, *15*(3), 827–840. http://doi.org/10.1016/j.healthplace.2009.02.002

Porter, C., Suhrbier, J., & Schwartz, W. (1999). Forecasting Bicycle and Pedestrian Travel: State of the Practice and Research Needs. *Transportation Research Record*, *1674*(1), 94–101. http://doi.org/10.3141/1674-13

Portland Bureau of Transportation. (2010). *Portland Bicycle Plan for 2030*. Portland, OR: City of Portland. Retrieved from ftp://ftp02.portlandoregon.gov/PBOT/Bicycle_Plan_for_2030/Plan_Documents/Complete_Plan/Portland_Bicycle_Plan_for_2030_as-adopted.pdf

Prato, C. G. (2009). Route choice modeling: Past, present and future research directions. *Journal of Choice Modelling*, *2*(1), 65–100. http://doi.org/10.1016/S1755-5345(13)70005-8

Prato, C. G., & Bekhor, S. (2006). Applying Branch-and-Bound Technique to Route Choice Set Generation. *Transportation Research Record: Journal of the Transportation Research Board*, *1985*(1985), 19–28. http://doi.org/10.3141/1985-03

Rajamani, J., Bhat, C. R., Handy, S. L., Knaap, G., & Song, Y. (2003). Assessing Impact of Urban Form Measures on Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects. *Transportation Research Record*, *1831*(1), 158–165. http://doi.org/10.3141/1831-18

Reilly, M., & Landis, J. (2002). The Influence of Built-Form and Land Use on Mode Choice. *Development*, *4*(1), 1–51.

Replogle, M. (1997). *Intergrating Pedestrian and Bicycle Factors into Regional Transportation Planning Methods: Summary of the State-of-the Art And Suggested Steps Forward*.

Rodríguez, D. A., & Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D: Transport and Environment*, *9*(2), 151–173. http://doi.org/10.1016/j.trd.2003.11.001

Rodríguez, D. A., Merlin, L., Prato, C. G., Conway, T. L., Cohen, D., Elder, J. P., … Veblen-Mortenson, S. (2015). *Influence of the Built Environment on Pedestrian Route Choices of Adolescent Girls*. *Environment and Behavior* (Vol. 47). http://doi.org/10.1177/0013916513520004

Roorda, M. J., Passmore, D., & Miller, E. J. (2009). Including Minor Modes of Transport in a Tour-Based Mode Choice Model with Household Interactions. *Journal of Transportation Engineering*, *135*(12), 935–945. http://doi.org/10.1061/(ASCE)TE.1943-5436.0000072

Saneinejad, S., Roorda, M. J., & Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport and Environment*, *17*(2), 129–137. http://doi.org/10.1016/j.trd.2011.09.005

Schlossberg, M., Greene, J., Phillips, P. P., Johnson, B., & Parker, B. (2007). School Trips: Effects of Urban Form and Distance on Travel Mode. *Journal of the American Planning Association*, *72*(3), 337–346. http://doi.org/10.1080/01944360701653235

Schuessler, N., & Axhausen, K. W. (2009a). *Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT). Working Paper: Transport and Spatial Planning*. Zurich. Retrieved from http://www.baug.ethz.ch/ivt/ivt/vpl/publications/reports/ab568.pdf

Schuessler, N., & Axhausen, K. W. (2009b). Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, *2105*(-1), 28–36. http://doi.org/10.3141/2105-04

Schwanen, T., & Mokhtarian, P. L. (2005). What affects commute mode choice: Neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*, *13*(1 SPEC. ISS.), 83–99. http://doi.org/10.1016/j.jtrangeo.2004.11.001

Seneviratne, P. N., & Morrall, J. F. (1985). Analysis of factors affecting the choice of route of pedestrians. *Transportation Planning and …*, *10*(2), 147–159. http://doi.org/10.1080/03081068508717309

Singleton, P. A., & Clifton, K. J. (2013). Pedestrians in Regional Travel Demand Forecasting Models: State-of-the-Practice. In *93nd Annual Meeting of the Transportation Research Board*.

Soltani, A., & Allan, A. (2006). Analyzing the Impacts of Microscale Urban Attributes on Travel: Evidence from Suburban Adelaide, Australia. *Journal of Urban Planning*

and Development, *132*(3), 132–137. http://doi.org/10.1061/(ASCE)0733-9488(2006)132:3(132)

Stopher, P., FitzGerald, C., & Xu, M. (2007). Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, *34*(6), 723–741. http://doi.org/10.1007/s11116-007-9126-8

Tobler, W. R. (1970). Clark University. *Economic Geogrpahy*, *46*(Supplement: Proceedings. International GeographicalUnion. Commission on Quantitative Methods), 234–240. http://doi.org/10.1126/science.11.277.620

Train, K. (2009). *Discrete choice models with simulation* (2nd ed.). Cambridge: Cambridge UP. Retrieved from http://elsa.berkeley.edu/books/choice2.html

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 & Science in Sports & Exercise*, *40*(5), 972–978. http://doi.org/10.1249/MSS.0b013e318164c407

Tsui, S., & Shalaby, A. (2006). Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems. *Transportation Research Record*, *1972*, 38–45. http://doi.org/10.3141/1972-07

Verlander, N. Q., & Heydecker, B. G. (1997). Pedestrian route choice: An empirical study. In *Transportation Planning Methods: Proceedings of Seminar F held at the European Transport Forum Annual Meeting, Volume P415* (pp. 39–49). Brunel University, England. Retrieved from http://discovery.ucl.ac.uk/55507/

Wang, C.-H., Akar, G., & Guldmann, J.-M. (2015). Do your neighbors affect your bicycling choice? A spatial probit model for bicycling to The Ohio State University. *Journal of Transport Geography*, *42*, pp 122–130. Retrieved from http://dx.doi.org/10.1016/j.jtrangeo.2014.12.003

Yarlagadda, A. K., & Srinivasan, S. (2008). Modeling children's school travel mode and parental escort decisions. *Transportation*, *35*(2), 201–218. http://doi.org/10.1007/s11116-007-9144-6

Zhang, M. (2004). The Role of Land Use in Travel Mode Choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association*, *70*(3), 344–360.

Appendix A   GPS Travel Data Mode Imputation Model

For a number of reasons, interest in augmenting or even replacing traditional diary-based household travel surveys with Global Positioning System (GPS) data has been growing over the past decade. Promises of reduced cost and respondent burden, greater detail, and feasibility of longer-term data collection are certainly attractive. This optimism about the future of travel data is tempered mainly by the burden of processing and questions about accuracy. Unlike a travel diary, a GPS data file is only a record of movements in space and time. Trip ends, purposes, and travel modes must be imputed in some way, and, while much progress and innovation has occurred, there are still no standards for processing the raw data. When comparing research results across studies, it is important to know whether apparent differences reflect underlying behavioral distinctions or simply different methods of data processing.

This paper proposes a new method for imputing travel mode from raw data, one important aspect of GPS data processing. While examining data from a GPS-based travel study our research team was conducting, we became concerned that existing methods were not detecting mode as accurately as we would like. Particularly concerning was the fact that bicycling, a focus of the study, appeared to be especially poorly predicted.

Using a unique dataset collected as part of a larger study in Portland, Oregon, I develop and test a multinomial logit (MNL) mode imputation model using GPS and accelerometer data. The MNL model is already well-known in travel demand modeling circles, and it has some appealing advantages in this application in terms of transferability and integration with other models. Initial results show that the MNL models developed here can predict urban travel modes—including bicycling—with a fairly high degree of accuracy, although as with any new method, there remains room for improvement. It is my hope that the relatively simple and familiar methodology will encourage replication and refinement within the larger research and practice community.

Existing Work

A number of other researchers have grappled with the problem of GPS mode imputation. Lawson et al. provide a recent review (2010). The majority of studies have used very small convenience samples and exploratory, heuristic classification techniques. These methods generally establish a hierarchy of rules based on test data and expert judgment. Trips are then deterministically binned by mode with implied 100 percent certainty. Without diminishing the importance of all work in the area at this stage of development, this review focuses on studies with larger, more representative samples and/or application of systematic, statistical models.

Fuzzy logic. One step up in complexity from hierarchical rule-based methods are fuzzy logic techniques. Instead of hard cutoffs, boundaries (e.g. between walking and biking speed) are represented as overlapping. A series of such rules can be applied, and then aggregated based on a membership criterion such as the maximum value of the lowest-scoring rule. The result is a mode assignment and a membership score. This score does not have a natural interpretation beyond higher values representing more certainty.

Another potential drawback is that the rules and fuzzy ranges must all be specified exactly beforehand. There is no model estimation beyond trial and error, and thus the results are largely tied to the skill of the fuzzy rules' architects. Tsui and Shalaby applied fuzzy rules to 109 GPS trips in Toronto, Canada (2006). Average and 95th percentile speeds, median acceleration, and data quality were used to form the rules. When bus or bike membership values were at least 0.4, a binary variable capturing whether the GPS track followed a GIS transit route was applied in a hierarchical manner. They reported an overall success rate of 91 to 94 percent, and cycling was predicted correctly from 72 to 86 percent of the time.

Schuessler and Axhausen applied a similar method minus the transit network component to a sample of nearly 5,000 Swiss participants (2009b). An additional check was made based on the reasonableness of predicted mode chains. No actual trip data was available to test prediction success, but aggregate shares were compared to Swiss national travel survey data. Auto and, to a lesser extent, bike travel appeared to be overestimated at the expense of public transit trips. Lawson et al. (2010) adapted (Schuessler & Axhausen, 2009b) to a 101 trip database in New York City. They reported a relatively poor success rate of 67 percent, attributed to the difficulty of distinguishing bus and auto travel in dense traffic. Cycling was not included as a travel mode.

Neural and Bayesian networks. Neural networks and Bayesian belief networks have also been proposed and tested for mode imputation. Gonzalez et al. developed a neural network application to impute travel mode from mobile phone GPS tracks (2010). They included average and maximum speeds and acceleration, as well as measures of how frequently speed and heading changed over time and distance. The algorithm was applied to a 114 trip dataset in Tampa, Florida and recorded an overall success rate of 91 percent. Bike travel was not included. Lawson et al. (2010) adapted (Gonzalez et al., 2010) to their small New York City dataset, and reported an 84 percent success rate, lowest for auto and bus modes.

Moiseeva et al. presented the first application of a Bayesian network to the mode imputation problem and applied their TraceAnnotater system to a sample of 1,554 trips from residents of Eindhoven, The Netherlands (2010). An extensive list of variables was specified: average, maximum and standard deviation of speeds; average and maximum acceleration; distance per time period; GPS accuracy measures; GIS railway proximity; and car and bike ownership. Initial results were promising with an overall success rate of 92 percent. Cycling was least accurately predicted at a still high 85.5 percent. The mode imputation algorithm itself is just one component of the TraceAnnotater system.

Both neural and Bayesian networks are "learning" models that can adapt both their parameters and structure to fit new data without researcher input. In this way they may be more transferable, since the algorithm is never fixed. Neural networks, however, lack easily interpretable prediction measures for individual observations. Bayesian networks are built on conditional probabilities, but it is not clear whether overall probabilities can reasonably be calculated. Both methods require considerable time and processing power to set up.

Discriminant analysis & accelerometer data. Perhaps closest to the present work in method, Troped et al. used discriminant analysis to impute mode from a small, 29 trip test dataset in Boston (2008). Since the focus was public health instead of travel, walking, jogging, and biking were done on a recreational path, while only driving was done in an on-street environment. The study was the first to combine GPS data with accelerometers for mode imputation. Median activity counts, median recorded steps, and median GPS speed were included in the final model. Results suggested that the two accelerometer variables alone were sufficient, with the GPS speed improving results only marginally.

It is unclear how these results might transfer to on-street travel. As a classification method, discriminant analysis is similar to the multinomial logit (MNL) model and posterior mode probabilities can be calculated.

Additions to the literature. The remainder of this paper develops the first MNL model of mode imputation, adding a well-established statistical technique to the literature. In addition, I extend work on combining GPS and accelerometer data, more commonly used in health research, by considering its application to urban, on-street travel. Models are developed from the largest North American dataset used for mode imputation in the literature, and one which includes both children and adults.

Objectives

My primary objective was to test the feasibility of a multinomial logit (MNL) model for completely automated mode imputation. A special focus was given to correctly identifying urban bicycle trips, a task our research team have found to be difficult in ongoing research. Many existing mode imputation datasets derive from Europe, where cycling tends to be slower and more separated from general traffic. This would tend to make imputing cycling trips easier. To our knowledge, these models have not been tested on US cycling trips.

A secondary objective is to test the relative importance of GPS, GIS transit network, accelerometer, and socio-demographic variables for mode imputation. Each additional layer adds cost and complexity, and it is important to know what the relative benefits are for mode prediction.

Finally, I try to provide enough information that our methods can be easily replicated elsewhere. An advantage of the MNL model is that software and expertise is already common within the travel demand modeling community. It is our hope that this low entry barrier will encourage others to test and refine the model, pushing the field toward some agreement in this critical area of GPS data processing.

The remaining sections describe the data and methodology employed, discuss the model results, and conclude with considerations for the course of future development and testing.

Data and Methodology

Data drawn from larger study. The data used in this study were collected in 2012 in Portland, Oregon as part of the second phase of an ongoing longitudinal study on family travel and physical activity. In the larger study, 975 participants from 323 households are collecting 5 consecutive days of person-based GPS (GlobalSat DG-100) and accelerometer (Actigraph GT3X) data during two different study phases. GPS points are recorded once every 4 seconds as long as a speed greater than 1 mile per hour (1.6 km/h) is detected. Hip-based acceleration forces in 3 dimensions are recorded constantly and aggregated to 15 second intervals, or epochs. In addition, survey questionnaires capture demographic data for the household and each participant. No travel diaries are kept by participants.

Validation data collection. In order to test our data processing accuracy, during the second phase of data collection, participants in a subsample of households were invited by email to provide recalled mode and purpose for up to 20 recently recorded trips. The initial household response rate was 40 percent for the supplemental survey. A total of 926 trips by 80 participants were available for initial model development.

A website allowed participants to view up to 20 trips one at a time using a Google Maps interface as shown in Figure 10 Web-based survey for mode and purpose recall. In addition to the trip segment start/end points and GPS track, date, weekday, and start/end times are provided to aid recall. Related trip segments were also displayed (e.g. the walking segment after a transit trip), but potentially leading details such as average speed or activity level were left out to avoid bias. The map was "live" and could be panned, zoomed, and switched among road map, satellite, and street view (Google's street level imagery). No complaints were received, and a number of participants left comments that they enjoyed the activity.

Figure 10 Web-based survey for mode and purpose recall

In order to maximize useful variation, we stratified the sample of trips for each person by speed, acceleration, transit network adjacency, and recorded activity level. Using thresholds from existing literature, previously collected Portland cycling GPS data, and the research team's judgment, we oversampled trips with data profiles that we thought fit walking, cycling, and transit travel. We also oversampled trips with profiles that had conflicting patterns or fell at the edge of cutoffs used in other studies. In order to reduce the effect of data noise at trip ends, we drew first from trips of at least 5 minutes, drawing shorter trips only when fewer than 20 long trips were available.

142

The recall sample used for model development included 398 trip segments made by children aged 5-17 and 528 by adults. While the sample was self-selected, we do not feel in general that differences between the sample and larger population are likely to be important for mode imputation. There are, however, certain demographics such as young college students and the elderly who are missing from the larger sample, and it is possible that their GPS and accelerometer patterns could differ for certain modes (e.g. by cycling or walking speed).

Data processing. In addition to recalled mode, the trip-level statistics shown in Table 1 were calculated from GPS, accelerometer and GIS transit network data. Raw GPS data were first partitioned into single-mode trip stages using existing methods that assume a mode change must include a walking stage (Schuessler & Axhausen, 2009b; Tsui & Shalaby, 2006). Trip segments were then joined to the best matching single transit route using a map matching technique developed by Schuessler and Axhausen (2009a).

Table 22 Mode Imputation Variable Descriptions

Name	Source	Description	Mean	Std. Dev.	Min	Max
95th acceleration	GPS	95th percentile acceleration (m/s^2)	0.98	0.50	0.05	2.05
95th acceleration^2	GPS	95th percentile acceleration squared	1.22	0.88	0.00	4.18
95th speed	GPS	95th percentile speed (m/s)	13.9	8.39	0.94	34.5
95th speed^2	GPS	95th percentile speed squared	263	251	0.88	1191
CV speed	GPS	speed coefficient of variation (m/s)	0.47	0.16	0.05	1.31
mean distance to transit route match	GIS	average distance from gps points to closest single transit route match (1000m)	3.47	1.30	0.00	4.00
median speed	GPS	median speed (m/s)	8.68	6.09	0.49	30.4
pcount	ACC	median 1-minute count[1] total (axis 3: perpendicular to travel) (/1000)	0.42	0.70	0.00	4.76
pcount^2	ACC	median 1-minute count[1] total squared (axis 3: perpendicular to travel)	0.66	1.86	0.00	22.7
vcount	ACC	median 1-minute count[1] total (axis1: vertical) (/1000)	0.57	1.13	0.00	7.26

[1] a count is a measure of directional force over time; one count = 0.01664 g/sec/count where g is the gravitational constant

Although it has not been used in this application before, I felt the multinomial logit (MNL) model was well-suited to the application of mode imputation. First, its statistical properties are well established (Ben-Akiva & Lerman, 1985; Train, 2009). Second, it requires only that the model structure be specified. Threshold and fuzzy logic models, by contrast, also require the researcher to specify value ranges and relative weights for each variable, adding to the potential for bias and limiting transferability. Third, an MNL model's mode predictions have an easily interpreted measurement unit: the probability that a predicted mode is the true one. The predicted probabilities could be useful for reporting standard cutoffs for observations (e.g. "cases for which the mode probability was >=60% were retained", etc.) or for incorporating as observation weights in more sophisticated models. Finally, MNL models have a long tradition in travel demand modeling, and estimation software and expertise already exists widely within the field.

In this context the MNL model maximizes the likelihood that the reported modes would have been predicted by the model. The modeled probability that a trip segment s was by mode m is given by:

$$-\Pr(m)_s = \frac{\exp(V_{m,s})}{\sum_m^M \exp(V_{m,s})} \tag{9}$$

where the observed utility $V_{m,s}$ of each mode is a linear-in-parameters function of trip segment attributes (Ben-Akiva & Lerman, 1985). Since the sampling strategy was exogenous (i.e. without knowledge of the actual travel modes), no adjustment was necessary in MNL model estimation (Ben-Akiva & Lerman, 1985).

Model Development and Findings

Three MNL models were developed, representing incremental increases in data collection and processing cost. One of the objectives was to determine the marginal benefit of additional data and processing for mode imputation. Model 1 uses only GPS data. A model form like this could be used with minimal post-processing and does not require any GIS, survey, or accelerometer data. Model 2 adds a transit network correspondence variable that requires matching GPS data to a transit network. Model 3 adds accelerometer data, requiring an additional device and processing software. I also tested for effects by age but found no significant differences.

Recognizing that the GPS, accelerometer, map matching, and recalled mode were each subject to error, I attempted to identify questionable observations. Standard practice was followed to examine outliers based on low predicted probabilities in initial model runs (Ben-Akiva & Lerman, 1985). I inspected each observation with a predicted probability below 0.05 and made a judgment call based on the data. Seventeen observations were removed as outliers that likely resulted from data errors (e.g. recalled walk but average speed was 20 mi/h [32km/h], recalled transit but trip did not follow

transit network, or recalled drive but trip was partly on a separated bike path). Models were then re-estimated with the spurious observations removed.

For each model, utility functions were estimated relative to walk as the base mode. Variables are defined in Table 1. Model results are shown in Table 2, prediction success is presented in Table 3, and predicted probability distributions are displayed in Figure 2.

Model 1 (GPS data only). Model 1 included only speed-related variables from GPS data. Median speed and 95th percentile speed were found to fit the reported data better than average and maximum speed. Acceleration and the coefficient of variation in speed also proved to be useful predictors for some modes. It should be noted that speeds below 1 mile per hour (1.6 km/h) were not recorded such that time spent stopped during an active trip stage is effectively ignored in our data.

Overall model fit was encouraging, but there was considerable variability in mean probability of choice by mode, a measure of the odds of selecting the chosen mode averaged over all cases. Auto (93.5%) and walk (84%) trip stages were predicted well. Bicycle trips had an expected prediction success rate of just 68.5%, with misclassifications split between walk and auto. On the high end, bike speeds overlap with auto speeds in urban conditions, and, on the low end, very slow biking—as one might do with a young child—overlaps with walking speed. Transit trips could not be distinguished from auto trips using only speed data. The sample of transit trips was small (n=22), and I was unable to consider bus and rail separately in this initial effort.

Model 2 (GPS + GIS transit network). Model 2 added a measure of correspondence with the transit network. GPS points were matched to single transit lines, and the minimum average distance to the line was included as a transit predictor. Overall model fit improved significantly, and transit prediction average probability improved sharply (4.5% to 54.5%). If it were sufficient to distinguish among walk, bike, and motorized modes, GPS data alone might be adequate, but if transit and auto travel needed to be treated separately joining data to a GIS transit network appears to be essential.

Model 3 (GPS + GIS transit network + accelerometer data). Model 3 incorporated data from accelerometers worn on the hip. Two of the three axes were found to be useful for mode detection. Vertical accelerometer counts are commonly associated

Table 23 MNL Model Results

Base = Walk	Model 1 GPS only		Model 2 +GIS transit network		Model 3 +Accelerometer	
	coeff.	t-stat	coeff.	t-stat	coeff.	t-stat
Bicycle						
constant	-6.1	-9.9	-5.9	-10.0	-6.4	-6.3
median speed	4.2	8.6	4.1	8.7	3.7	7.2
median speed^2	-0.3	-5.6	-0.3	-6.2	-0.2	-5.3
95th speed	-0.5	-3.1	-0.4	-3.2	-0.5	-2.7
vertical intensity					-2.3	-3.1
vector magnitude					3.6	4.6
vector magnitude^2					-0.5	-3.0
Auto						
constant	-4.9	-9.2	-4.3	-10.0	-2.3	-4.2
median speed	1.3	4.5	0.9	4.9	0.7	4.3
95th acceleration	3.6	4.3	3.8	5.5	2.6	3.6
vertical intensity					-0.9	-3.5
*Transit**						
constant	-7.0	-4.7	-3.1	-2.5	-1.5	-1.2*
median speed	1.4	4.6				
95th acceleration	2.4	2.1				
95th speed			1.0	5.9	0.8	4.9
Coefficient of variation in speed	3.5	1.8*				
Mean distance to transit route match			-0.2	-4.2	-0.2	-3.9
vertical intensity					-0.9	-3.5
LogL (constants)	-741		-741		-741	
LogL (model)	-207		-187		-158	
McFadden R^2	0.72		0.75		0.79	
N	926		926		926	

* All coefficient estimates significant at p<0.05 unless noted with asterisk.

** Transit considered only when trip start and end within 1km (0.6mi) of a single transit line.

Table 24 MNL Model Mean Probability of Recalled Travel Mode

(percent)	Model 1 GPS only	Model 2 +Transit network	Model 3 +Accelerometer	n
Walk	84.0	84.0	86.0	150
Bicycle	68.5	68.5	74.2	89
Auto	93.5	94.9	95.3	665
Transit	4.5	54.5	54.5	22
All modes	*87.5*	*89.6*	*90.8*	*926*

with ambulatory activity; thus, the negative values for bike and auto relative to walk are to be expected. No relationship was found for transit, perhaps because transit passengers may be either seated or standing. Accelerometer values perpendicular to travel direction enter the bike model as an initially positive factor that gradually diminishes, becoming negative at high levels. I investigated this effect, and found that, particularly riding at low speeds, the accelerometers tend to record very low vertical values but increasing perpendicular values. Presumably this is due to balance "wobble" at low speeds. In any case, while the accelerometer data showed only modest effects on overall model fit, prediction of bike travel improved considerably (68.5 to 74.2%). Accelerometers show promise for distinguishing urban bike trips in conjunction with GPS data.

Age differences. I hypothesized that children's travel behavior might differ from adults. Specifically, I thought that children might bicycle at lower speeds and might move around more energetically while using any modes, perhaps manifesting as higher accelerometer readings. The sample was partitioned into three age groups: 5-10, 11-16, and adult. Model specifications were tested with interaction terms between bike speed and age and between accelerometer counts and age. Surprisingly, segmenting by age had a negligible impact on parameter estimates, suggesting that child and adult speed and activity patterns were practically indistinguishable in our sample.

External Validation

After collecting an additional 534 trips via the web survey, I applied Model 3 as an out of sample prediction test. The highest probability choice was selected deterministically, as would be done for the data used in the main study. Results are shown in Table 25 Out of Sample Validation (Model 3). Mode identification success rates ranged from 79% (bike) to 97% for driving, with an overall success rate of 94%. Errors for bike identification were all Type II (false negative), while errors for walk identification were mostly Type I (false positive). Manual inspection of errors suggested that many of the bicycle trips incorrectly identified as walking were cases of walking a bike or else riding very slowly, perhaps with children or on a sidewalk.

147

Table 25 Out of Sample Validation (Model 3)

	Walk	Bike	Auto	Transit	Total	Detection
Walk	82	0	3	1	86	95 %
Bike	11	66	7	0	84	79 %
Auto	9	0	337	0	346	97 %
Transit	0	0	2	16	18	89 %
Total	102	66	349	17	534	
Precision	80 %	100 %	97 %	94 %		94 %

Conclusion and Future Directions

In order for GPS data to useful for travel demand modeling without supplemental travel diaries, we must be able to consistently identify characteristics of the recorded travel such as mode, destination, and purpose. In order to identify travel mode for the mode choice framework used in this research, I developed multinomial logit (MNL) models for imputing travel mode from GPS, GIS, and accelerometer data. Three models were developed, each reflecting an incremental increase in data and processing cost. Using only GPS data, motorized and non-motorized travel could be distinguished with fairly high accuracy, walk from bicycle with reasonable accuracy, and transit from auto travel not at all. With the addition of transit network data, transit mode could be predicted in more than half of cases. Adding accelerometer data to proxy for activity intensity improved overall prediction modestly but considerably enhanced identification of bike trips. Segmenting by age did not significantly affect the results.

This exercise marks the first use of logit modeling for travel mode imputation. It is argued that logit models have a number of potential advantages over competing methods such as threshold cutoffs, fuzzy logic, and neural and Bayesian networks. In particular, prediction values have an easy to understand scale and meaning—probability—which lends itself more easily to setting reporting standards. In addition, the predicted probabilities may be useful as observation weights when estimating models with imputed data, such that more certain observations carry greater estimation weight. This will be an interesting direction for future research.

Though our unique sample and local conditions may limit the direct transferability of the MNL models developed, they are ready to be replicated with any combination of GPS and accelerometer data in other locations. The extended data set allowed testing of the model's out of sample prediction performance, and results were encouraging.

148

Appendix B Bicyclist Route Choice Re-estimation Results

Results from Broach et al.'s (2012) bicyclist route choice model were compared with the Family Activity Study (FAS) data used in this research. The aim was to decide whether the existing model could be used to represent likely routes for the current sample and updated network. Primary concerns were that the original sample collected in 2007 as part of the Bike GPS Study was made up mainly of avid, year round cyclists, most of whom signed up for the study specifically because they were frequent riders. The FAS sample, on the other hand, was made up of families with children, and while all had a bicycle to use, self-reported cycling frequencies were much lower, and they did not need to be a current rider or even interested in cycling to participate.

Model Estimation and Comparison

Following the earlier study, I used the same calibrated labeling method to generate alternatives (Broach et al., 2011). After choice sets were established, I used the same Path-size Logit form to estimate a route choice model. Starting from the earlier specification, I proceeded to test variations, using the usual method of t-tests of parameter significance and Likelihood Ratio Tests to accept or reject more parsimonious model versions.

The final model is shown in Table 26, side-by-side with the original model. In general, the similarity of results is striking, though there are a few differences. Most notably, there was no clear commute preference segment as there had been in the earlier model. This was somewhat expected, since commute trips (defined as direct trips between home and work or vice versa, were considerably rarer in the FAS cycling data (4% of trips vs. nearly 30% in the original dataset). Other differences included stop signs (negative in original, positive in FAS), sensitivity to lower traffic volumes, and a weak but significant residual value of striped, on-street bike lanes, after subtracting the traffic effect. There was also a relative lack of bridge crossing data, and since no cyclist used a bridge with bike lane, that parameter could not be estimated. Alleyways had been added to the travel network since the original work, and a parameter had to be added to account for the strong aversion cyclists have toward using them. A similar result was found in the Pedestrian route choice model developed in Chapter 4. Alleyways do not seem to be useful to people walking and biking in Portland.

Table 26 Route Choice Model Estimation Comparison

Variable	(Broach et al., 2012)		FAS	
	coeff.	t-stat	coeff.	t-stat
Ln(distance)	-5.22	*-10.9*	-5.68	*-14.6*
Ln(distance) * commute	-3.76	*-5.14*		
Turns (/mi)	-0.37	*-15.4*	-0.35	*-18.2*
Prop. upslope 2-4 %	-2.85	*-4.57*	-2.46	*-3.52*
Prop. upslope 4-6 %	-7.11	*-6.11*	-8.43	*-5.61*
Prop. upslope >= 6 %	-13.0	*-8.57*	-9.71	*-3.96*
Traffic signal exc. right turns (/mi)	-0.19	*-5.73*	-0.15	*-3.06*
Stop sign (/mi)	-0.05	*-2.10*	0.05	*1.88‡*
Left turn, unsig., AADT 10-20k (/mi)	-0.78	*-4.19*		
Left turn, unsig., AADT 20k+ (/mi)	-1.87	*-4.70*	-0.94	*-3.63*
Unsig. cross AADT >= 10k right turn (/mi)	-0.34	*-2.32*	-0.77	*-6.03*
Unsig. cross AADT 5-10k exc. right turn (/mi)	-0.36	*-5.39*	-0.30	*-4.08*
Unsig. cross AADT 10-20k exc. right turn (/mi)	-0.52	*-5.39*	-0.91	*-8.92*
Unsig. cross AADT 20k+ exc. right turn (/mi)	-2.51	*-11.5*	-1.39	*-8.07*
BETA_prop_bike_lane, AADT 5k+			1.24	*3.83*
Prop. bike boulevard	1.03	*5.17*	1.27	*5.30*
Prop. bike path	1.57	*4.64*	1.89	*2.90*
Prop. AADT 5k+			-1.03	*-3.53*
Prop. AADT 10-20k w/o bike lane	-1.05	*-3.02*	-2.19	*-4.30*
Prop. AADT 10-20k w/o bike lane * commute	-1.77	*-2.28*		
Prop. AADT 20-30k w/o bike lane	-4.51	*-6.04*	-4.10	*-6.05*
Prop. AADT 20-30k w/o bike lane * commute	-3.37	*-2.24*		
Prop. AADT 30k+ w/o bike lane	-10.3	*-4.67*		
Prop. AADT 30k+ w/o bike lane * commute	-8.59	*-1.96*		
Bridge w/ bike lane	1.81	*-4.71*		
Bridge w/ sep. bike facility	3.11	*-4.96*	0.84	*3.21*
BETA_prop_alley			-21.0	*-11.4*
Ln(path size)	1.81	*20.8*	1.12	*17.6*
Number of observations		1,449		1388
Null log-likelihood		-4058.7		-3548.9
Final log-likelihood		-3020.0		-2670.3

Table 26 (continued)

Rho-square	0.256	0.248

[1] Parameter estimates significant at the α=0.05 level unless noted otherwise

[‡] Significant at the α=0.10 level

[†] Not significant at the α=0.10 level

Another way to compare is via relative effect sizes of the two parameter sets. Table 27 allows for comparison. Given that the sampling strategies were different, the travel network is constantly being updated, both digitally and in reality, and that the data were collected 3-6 years apart, I found the consistency of most parameters to be a pleasant surprise. Based on this work, I chose to adopt the original model to generate routes for the FAS sample used in this research.

Table 27 Relative attribute values (unit change)

	Distance value (equiv. % dist)		
	(Broach et al., 2012)		FAS
Attribute	Non-commute	Commute	All trips
Turns (/mi)	7.4	4.2	6.4
Prop. upslope 2-4 %	72.3	37.1	54.2
Prop. upslope 4-6 %	290.4	120.3	341.1
Prop. upslope >= 6 %	1106.6	323.9	452.6
Traffic signal exc. right turns (/mi)	3.6	2.1	2.7
Stop sign (/mi)	0.9	0.5	-0.01
Left turn, unsig., AADT 10-20k (/mi)	16.2	9.1	0
Left turn, unsig., AADT 20k+ (/mi)	43.1	23.1	18.0
Unsig. cross AADT >= 10k right turn (/mi)	6.7	3.8	0
Unsig. cross AADT 5-10k exc. right turn (/mi)	7.2	4.1	0
Unsig. cross AADT 10-20k exc. right turn (/mi)	10.4	5.9	17.3
Unsig. cross AADT 20k+ exc. right turn (/mi)	61.7	32.2	27.7
Prop. bike lane, AADT 5k+			-3.6
Prop. bike boulevard	-17.9	-10.8	-20.0
Prop. bike path	-26.0	-16.0	-28.3
Prop. AADT 5k+			19.9

Table 27 (continued)

| | Distance value (equiv. % dist) | | |
| | (Broach et al., 2012) | | FAS |
Attribute	Non-commute	Commute	All trips
Bridge w/ bike lane	-29.3	-18.2	0
Bridge w/ sep. bike facility	-44.9	-29.2	-13.7