EXPLORING PEDESTRIAN ROUTE CHOICE PREFERENCES BY DEMOGRAPHIC GROUPS: ANALYSIS OF STREET ATTRIBUTES IN CHICAGO

A Dissertation
Presented to
The Academic Faculty

by

Seung Jae Lieu

In Partial Fulfillment
of the Requirements for the Degree
Master of City and Regional Planning in the
School of City and Regional Planning and
Master of Science in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
May 2023

EXPLORING PEDESTRIAN ROUTE CHOICE PREFERENCES BY DEMOGRAPHIC GROUPS: ANALYSIS OF STREET ATTRIBUTES IN CHICAGO

Approved by:

Dr. Subhrajit Guhathakurta, Advisor
School of City and Regional Planning
Georgia Institute of Technology

Dr. Patricia L. Mokhtarian
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Gulsah Akar
School of City and Regional Planning
Georgia Institute of Technology

Date Approved: April 27, 2023

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisors and committee members for their guidance and invaluable feedback throughout the thesis journey. First and foremost, I would like to thank my advisor, Dr. Subhro Guhathakurta, for his generous supervision during my Research Assistant position. His unwavering guidance and support have been instrumental in helping me achieve dual master's degrees and make significant progress in thesis work. My experience as a member of the Center for Spatial Planning Analytics and Visualization has equipped me with skills and knowledge that have greatly contributed to this study. Additionally, I am grateful to Dr. Patricia Mokhtarian for her patience and detailed guidance. Her expertise in discrete choice model analysis has been crucial in correctly interpreting the results and conducting the modeling. I also wish to sincerely thank Dr. Gulsah Akar for joining my thesis committee and consistently showing interest in my work. I would like to acknowledge Craig Heither from the Chicago Metropolitan Agency for Planning, who granted me access to the survey data, which constitutes a vital part of this study. Finally, I would like to thank my family and friends at Georgia Tech for their unwavering support in my academic pursuits and for making each day enjoyable and fulfilling.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AIC	Akaike Information Criterion
CF	Commonality Factor
CMAP	Chicago Metropolitan Agency for Planning
GPSL	Generalised Path Size Logit
GPS	Global Positioning System
IIA	Independence from Irrelevant Alternatives
IID	Independently and Identically Distributed
MNL	Multinomial Logit
OSRM	Open Source Routing Machine
PS	Path Size
PSL	Path Size Logit
UIC	University of Illinois, Chicago

SUMMARY

Traditional transit accessibility models often overlook travel behavior and fine-grained transit characteristics experienced during first and last-mile walking. Existing models typically assume travelers choose the shortest walking path to minimize travel time, but studies suggest pedestrians do not always follow this pattern.

This study investigates pedestrian route choice preferences in Chicago, Illinois, using a diverse dataset of home-based work walking trajectories collected from a smartphone application. The impact of street attributes on route choice is examined, and a comparison is made of how built environment factors influence preferences among different demographic groups. A path-size logit model with a constrained enumeration approach-based choice set is employed for analysis.

This study also addresses two gaps in pedestrian route choice research. First, unlike most studies that use data constrained to a particular study area or limited participant groups, this research employs a diverse dataset of actual walking trajectories covering a wide range of destinations and participant profiles. Second, this study utilizes GPS data, offering more accurate route choice analysis compared to questionnaires. Such surveys may suffer from recall bias, and they may not capture route choice variability across different times and days.

The findings from this study indicate that factors such as distance, the number of amenities and establishments, sky visibility, greenery, and park accessibility along the route significantly influence route choice. While route distance and the number of

establishments have a negative impact on preference, other factors positively affect route selection. To compare the effect of each variable across gender, age, and income, this study has operationalized the coefficients to use the concept of ‘equivalent walking distance.' This measure quantifies the incremental disutility resulting from various route attributes, represented as an equivalent increase or decrease in walking distance. The analysis shows that male pedestrians are more willing to walk further when there is greater sky visibility. Similarly, individuals aged over 30 years old tend to walk longer distances with increased sky visibility. Notably, we found no significant variables influencing route choice among different income groups.

CHAPTER 1. INTRODUCTION

Travelers seek to minimize travel time and distance to maximize their utility; however, pedestrian behavior tends to deviate from this pattern. Pedestrians make decisions about route selection, walking speed determination, and type and timing of interactions with their surroundings before and during walking. Among these choices, a comprehensive understanding of pedestrian route choice – notably, how street features influence route preferences – can provide valuable insights for designing, planning, and implementing sustainable pedestrian environments and equitable urban mobility systems. For instance, conventional accessibility measurement models often overlook pedestrian behavior and ignore the travel impedance during the first and last-mile trips. Incorporating travel impedance by route choice analysis can therefore enhance the accuracy of accessibility measurements from the traveler's perspective.

Numerous studies have investigated pedestrian route choice behavior using disaggregate or micro-level analyses. While distance and travel time are crucial factors in route choice, pedestrians often opt for safer, more comfortable, or more interesting routes, as long as detours remain within an acceptable range (Sevtsuk & Kalvo, 2021). Various street features along the route have been found to be associated with route choice decisions. Pedestrians are likely to choose streets with wide, continuous sidewalks on both sides of the street for safety and convenience (Rodríguez et al., 2015; Lue & Miller, 2019). Additionally, pedestrian amenities and urban design features such as waterfronts, benches, and greenery along the sidewalk positively influence route choice decisions. (Dessing et al., 2016; Shatu & Yigitcanlar, 2018). Furthermore, a sense of street enclosure, reflecting more detailed visual aspects of streets, has been identified to have impacts on route choice preference (Sevtsuk et al., 2021; Basu & Sevtsuk, 2022).

However, the methods and data used for route choice analysis to date present several limitations in fully understanding general pedestrian behavior. Most studies rely on qualitative data gathered through questionnaires, commonly using intercept surveys (Guo & Loo, 2013; Azegami et al., 2023). These surveys, however, cannot be readily generalized to larger areas as they typically cover small, specific locations and involve limited participant numbers due to the labor-intensive nature of the process. Intercept surveys also depend on self-reported information, which is susceptible to inaccuracies and biases.

Stated preference surveys, which reduce the burden of in-person data collection and allow participants a range of hypothetical alternatives, present their own limitations (Erath et al., 2015; Bellizzi et al., 2021). While the survey results provide data about how much route attributes were valued by participants, responses may not accurately reflect real-life behavior. People might overestimate or underestimate their preferences when not faced with tangible choices. Stated preference surveys also fail to capture real-time walking conditions, choices, and idiosyncrasies. Recently, smartphone applications have been used to collect extensive Global Positioning System (GPS) trajectory data for large-scale route choice modeling. However, a fully anonymized dataset, which aims to protect user privacy, makes it challenging to infer personal attributes such as demographic information and trip purpose.

This study examines pedestrian route choice preferences in Chicago, Illinois by utilizing an extensive dataset of GPS walking trajectories. Since the data includes respondent attributes that supplement traditional travel diary recording, this study explores systematic differences in route choice preferences based on personal characteristics. Therefore, the research aims to determine how specific street features affect pedestrian route choice behavior across demographic groups, including age, gender, and income level. With a focus on home-based work trips, the analysis includes various destinations and geographies. The explanatory street features are based upon previous literature that examined the impact of each such feature on route choice preferences.

The remainder of this paper is structured as follows: Section 2 provides a review of studies focusing on the relationship between route choice decisions, street features, and socio-demographic factors, and route choice analysis modeling frameworks. Section 3 offers a detailed description of the data and methods employed in this paper. In Section 4, the model estimation and results are presented, while the final section, Section 5, concludes the paper by discussing the limitations of this study and suggesting potential directions for future research.

CHAPTER 2. RESEARCH BACKGROUND

This study examines pedestrian route choice preferences by analyzing the impact of street attributes on route choice and comparing how these factors influence pedestrians across different demographic groups. A GPS dataset is employed to investigate actual walking trajectories and personal attributes for each observation, which enables a comprehensive analysis of route choice behavior in diverse urban settings throughout the city of Chicago and among different participant profiles. Furthermore, by estimating the model coefficients of the path size logit model, the concept of an "equivalent walking distance" is introduced to represent the incremental disutility resulting from various street features and compare the effect of each variable across different demographic groups.

2.1 Factors Affecting Pedestrian Route Choice

The route choice for pedestrians can be understood as the process of choosing a route with the highest utility among various options connecting an origin to a destination. While minimizing distance and time is often the primary goal, research has shown that pedestrians do not always follow the shortest route. Instead, they may prefer routes that are safer, more comfortable, or more visually appealing routes than the shortest path.

Many studies have explored the relationship between street features and pedestrian route choice decisions. Broach and Dill (2015) used GPS devices to study the route choices of 283 adults, finding that distance, number of turns, elevation gain, and traffic volumes were negatively or positively related to route selection. Specifically, the study revealed that each additional turn was associated with a decrease in route utility equivalent to approximately 50 meters of distance. Upslopes with a 10 percent gradient were perceived as twice as inconvenient as less steep terrain. Guo (2009) employed a path choice model to analyze the influence of pedestrian environments on walking utility, revealing that factors such as intersection density, sidewalk width, and topography significantly impacted route choice utility. The model showed that an increase of one more intersection per 100m provided a utility equivalent to reducing travel time by 0.3 minutes, while increasing sidewalks by 6 feet offered a utility equivalent to reducing travel time by 0.5 minutes.

Among a wide range of street features, well-designated crosswalks, as a crucial aspect of pedestrian safety, have been shown to play a significant role in route choice behavior. Lue and Miller (2019) found that the presence of sidewalks on both sides of a street was an important variable in the route choice model in Toronto. Their results emphasized the importance of street completeness, with links featuring sidewalks on both sides perceived as 33% shorter in length. Similarly, Sevtsuk et al. (2021) discovered that sidewalk width had a strong positive effect on route choice in San Francisco. They found that a 10-foot increase in sidewalk width could boost the willingness to walk by up to 84 meters in San Francisco while a more modest 13 meters in Boston (Basu & Sevtsuk, 2022). It has also been observed that pedestrians might take slightly longer routes if the sidewalk is wider than the width of 1.5m to 3.0m and separated from the road traffic (Muraleetharan & Hagiwara, 2007). Moreover, sidewalk continuity and condition have been identified as important factors influencing students' route preferences (Shatu & Yigitycanlar, 2018).

Although the impact of sidewalk presence has generally been reported as positive, some studies have found no such effect or even a negative one (Guo & Loo, 2013; Ozbil et al., 2016). Guo & Loo (2013) discovered that sidewalk width had an insignificant impact in New York, while in Hong Kong, it was significant but negatively associated with route choice due to omitted key route attributes like pedestrian crowding, hilly topography, and pedestrian bridges.

Pedestrian amenities and urban design features like open spaces, greenery, and retail frontages also affect route choice decisions. These factors, visually perceivable while walking, can make a route more attractive and, therefore, more likely to be chosen (Ewing & Handy, 2009; Shatu et al., 2019). Borst et al. (2009) explored how environmental street features affect the walking route choices of elderly residents in Dutch urban districts. The study found that older individuals preferred streets with front gardens, dwellings, and shops. Similarly, Erath et al. (2015) discovered that pedestrians in Singapore were most drawn to routes featuring greenery and retail frontages. Although it is difficult to interpret the willingness to walk of each street feature due to the lack of information on the measurement method, Sevtsuk et al. (2021) and Basu and Sevtsuk (2022) supported these findings with a large dataset of walking trajectories from smartphone applications. Basu and Sevtsuk (2022) found that people prefer routes with more vegetation and service amenities. In contrast to San Francisco (Sevtsuk et al., 2021), pedestrians in Boston seem to favor more exposure to the sky (e.g., streets with lower building edges), indicating a preference for openness and less enclosure.

In addition to the street environmental features, personal attributes and trip characteristics influence pedestrians' route choices. Socio-demographic factors, such as age, gender, and income have been found to associate with pedestrian route choice decisions positively or negatively. However, the relationship between age and route choice decision is complex, with few studies investigating its impact on pedestrian decision-making. Liu et al. (2020) examined the relationships between micro-scale built environments and pedestrian preferences using a stated choice experiment in Tianjin, China. The study found that pedestrians generally preferred functional environments and facilities (e.g., greenery, retail shop frontage, lamp density), but the main effects of these features did not significantly differ across age and gender categories. Age and route choice associations were also non-significant in a study in South Korea (Gim & Ko, 2017). However, Brookfield and Tilley (2016) identified that participants aged 65 years and above tended to choose routes with high environmental quality, personal security, and pavement surface quality.

Gender differences in route choice were observed in Wickramasinghe & Dissanayake's (2015) study, which found that male pedestrians typically selected routes with the quickest path, fewer street crossings, and less crowding, while the availability of shops was crucial for female pedestrians. In contrast, Gim and Ko (2017), and Broach and Dill (2015) reported no significant differences in route preferences between male and female pedestrians. Regarding the income level, Shatu et al. (2019) discovered that higher-income students were more likely to choose routes with an oversupply of opportunities, such as shopping and dining options, and less likely to overestimate route directness.

The review demonstrates that pedestrians consider a variety of street characteristics when selecting routes. Consistently, wider sidewalks, the presence of retail shops, sidewalk amenities (e.g., trees), and sky visibility were associated with pedestrian route choice. These findings suggest that improved pedestrian amenities and greener environments encourage more walking through these areas. However, some inconsistencies in the impacts of some environmental features on route choice exist. A plausible explanation is that most of the studies focused on different but specific areas (e.g., schools, downtown) and involved a limited number of participants due to the labor-intensive nature of data collection.

This study aims to infer the common pedestrian route choice preferences at a city-wide scale by encompassing a broad range of geographic locations. The research utilizes actual walking trajectories covering the entire city of Chicago rather than focusing on specific areas. Moreover, personal attributes – age, gender, and income – for each observation will provide further insight into the varying effects of street features on route choices among diverse demographic groups. Previous research employing a substantial number of actual GPS trajectories encountered difficulties in examining individual characteristics or trip purposes due to the anonymized nature of data. While there are studies exploring socio-demographic factors in route choice analysis, they typically concentrate on a particular group of people or rely on small sample sizes.

2.2 Route Choice Models

Assuming that a decision-maker selects the option with the highest utility within a choice set (Ben-Akiva & Bierlaire, 2003), the Multinomial Logit (MNL) model, one of the random utility theory-based models, has been extensively used in discrete choice analysis. In the context of route choice, an individual aims to maximize the utility of a chosen route $k \in R$ from the set of routes R. The utility of route k is expressed as $U_k = V_k + \varepsilon_k$, where V presents the deterministic component of utility, and ε denotes the additive random error term. Consequently, the probability that an individual selects alternative $k \in R$ is:

$$P_k = Pr(V_k + \varepsilon_k \geq V_m + \varepsilon_m, \ \ \forall m \in R, m \neq k) \tag{1}$$

Logit models are characterized by their random variable error terms following a Gumbel distribution (Ben-Akiva & Lerman, 1985). As a result:

$$P_k = \frac{e^{V_k}}{\sum_{m \in R} e^{V_m}} \tag{2}$$

For route choice, the deterministic utility of a route is presumed to be strongly influenced not only by the travel impedance, which includes fixed attributes such as cost and distance, but also by observed route attributes like the presence of trees and sky visibility along the route. Given that each link $a \in k$ has a fixed travel cost t_a, and the travel cost for a route $k \in R$ is the sum of the total cost of its links, the travel cost for route $k \in R$ is computed as: $c_k = \sum_{a \in k} t_a$. Since c_k represents the travel cost of the route, the deterministic utility of route $k \in R$ is given by $V_k = -\theta c_k$, where $\theta > 0$ is the logit scaling parameter, and the negative sign indicates that high cost corresponds to low utility. Thus, if the travel impedance is the only observed component of the utility of a route, then the probability of an individual selecting alternative k is:

$$\mathrm{P_k} = \frac{e^{-\theta c_k}}{\sum_{m \in R} e^{-\theta c_m}} = \frac{1}{\sum_{\mathrm{m \in R}} \mathrm{e}^{-\theta (c_m - c_k)}} \tag{3}$$

In pedestrian route choice analysis, this function calculates the probability of selecting a route based on each alternative's utility, ultimately identifying the optimal route by considering not only travel impedance factors (e.g., cost, distance) but also the attractiveness of the specified attributes (e.g., street features). However, the MNL model's assumption that random error terms are independently and identically distributed (IID) with fixed variances (Sheffi, 1985) results in the Independence from Irrelevant Alternatives (IIA) property. The MNL model assumes route utilities to be independent, but routes with overlapping links share unobserved attributes, violating the assumption that random error terms are IID. In other words, when two routes have overlapping links, they may share unobserved attributes, leading to a correlation between their utilities.

To address the IID assumption violation, various MNL model adaptations have been introduced to account for correlations between routes' unobserved characteristics. These advanced logit models fall into three categories based on their model structures: Generalized Extreme Value-based logit models, Mixed Logit models, and MNL-modification models (Prashker & Bekhor, 2004). The MNL-modification group includes C-Logit (Cascetta et al., 1996), Path Size Logit **(**Ben-Akiva & Bierlaire, 1999), and Generalised Path Size Logit model (Ramming, 2002), which incorporate a correction term to adjust route choice probabilities. Duncan et al. (2020) highlighted that these models feature simple closed-form expressions, which facilitate relatively simple and rapid calculations of route choice probabilities. Additionally, parameter estimation for these models is relatively uncomplicated. Therefore, to ensure efficient computation in route choice analysis, this study conducts a literature review of three models within the MNL-modification group.

2.2.1 C-Logit

Cascetta et al. (1996) developed a model incorporating a deterministic correction by introducing a Commonality Factor (CF). This factor is based on the concept of commonality, which quantifies the degree of similarity among alternatives in the choice set. Consequently, the CF value is less than one or may approach zero for overlapping routes, making these routes appear less attractive. The authors propose four different forms for the CF correction, with one of them defined as:

$$CF_{in} = \beta_{CF}\ ln \sum_{j \in C_n} \left(\frac{L_{ij}}{\sqrt{L_i L_j}} \right)^{\gamma} \qquad \textbf{(4)}$$

where L_{ij} is the length of common links on routes *i* and *j*, while L_i and L_j denote the total lengths of routes *i* and *j*, respectively. β_{CF} is a coefficient to be estimated, and the parameter γ may be estimated or constrained to a convenient value. Although Cascetta et al. (1996) presented various formulations of the CF factor, the lack of theory or guidance on which form of CF should be used is a limitation of the C-Logit methods (Ramming, 2002; Frejinger et al., 2009).

2.2.2 Path Size Logit

Ben-Akiva and Bierlaire (1998) introduced the Path Size Logit (PSL) model, which, like the C-Logit model, incorporates correction terms to account for the correlation arising from overlapping alternatives. A correction for overlapping links is achieved by adding a factor called "Path Size" (PS), which is then applied to the deterministic part of the utility. Therefore, the deterministic component of the utility of route $k \in R$ is $V_k = -\theta c_k + p_k$, where p_k represents the correction term derived from the PS factor. The probability of an individual selecting route $k \in R$ is:

$$\mathrm{P_k} = \frac{e^{-\theta c_k + p_k}}{\sum_{m \in R} e^{-\theta c_m + p_m}} \quad \textbf{(5)}$$

The correction term is $p_k = \beta \, ln(PS_k)$, where $\beta \geq 0$ is the parameter of the PS factor, and PS is the Path Size factor for route $k \in R$. As a result:

$$\mathrm{P_k} = \frac{e^{-\theta c_k + \beta \, ln(PS_k)}}{\sum_{m \in R} e^{-\theta c_m + \beta \, ln(PS_m)}} = \frac{e^{-\theta c_k} \, (PS_k)^{\beta}}{\sum_{m \in R} e^{-\theta c_m} \, (PS_m)^{\beta}} = \frac{1}{\sum_{m \in R} e^{-\theta (c_m - c_k)} \, (\frac{PS_m}{PS_k})^{\beta}} \quad \textbf{(6)}$$

In the context of route choice, Ben-Akiva and Ramming (1998) assumed that decision-makers did not perceive overlapping paths as distinct alternatives. When paths overlap, decision-makers perceive these routes as less distinct, causing the perceived utility of each route to become more similar. This phenomenon occurs because shared links among routes result in a high correlation between them, making it difficult for individuals to differentiate between the alternatives based on their attributes.

The PS factor aims to capture the correlation between similar alternatives and adjust the choice probabilities accordingly. This factor can be interpreted as a measure of the distinctiveness or uniqueness of routes available to an individual when making a route choice decision. A higher PS for a route implies that it has more unique links compared to other routes, increasing the likelihood of being chosen. Conversely, a lower PS indicates greater overlap with other routes, reducing the probability of selecting that route. The value of PS can be greater than 0 but less than or equal to 1, and it is defined by:

$$PS_k = \sum_{a \in \Gamma_k} \frac{l_a}{L_k} \frac{1}{\sum_{m \in C_n} \delta_{am}} \tag{7}$$

where, Γ_k is the set of links in route $k \in R$. l_a and L_k are the length of link a and route k, respectively, with the term $\frac{l_a}{L_k}$ serving as a weight corresponding to the fraction of route impedance coming from a particular link. δ_{aj} is the link-route incidence variable that equals one if link a is on route k and 0 otherwise. $\sum_{m \in C_n} \delta_{am}$ is the number of routes in choice set C_n that share link a.

Figure 1 presents an example of three paths (i.e. routes) for the same origin and destination. Path *1* has no overlapping links, eliminating the need for utility adjustment and yielding a PS factor of one. In contrast, Path *2* and Path *3* share link *b*.

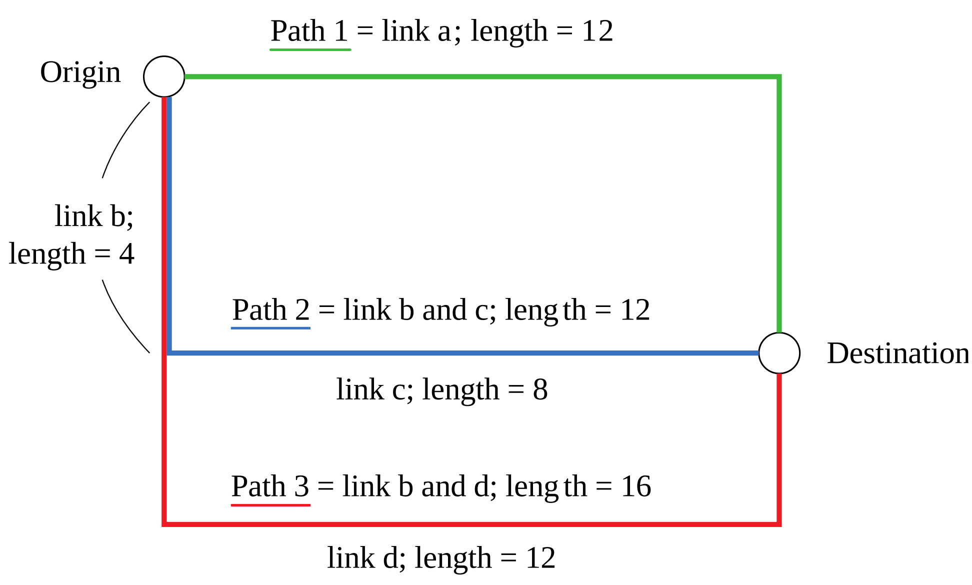

Figure 1 – The overlapping path problem

Table 1 presents the results of calculating the PS for each path and the probability of choosing each path based on the PSL model and the MNL model, which does not account for PS. In this example, the deterministic utility of every path (i.e. d_k) is captured through the travel distance, and the parameter of PS is set to +1.

Table 1 – Path size factor and the probability of choosing each path

Path	Path Size factor accounting for each link				Path Size factor
	link *a*	link *b*	link *c*	link *d*	
1	$\frac{12}{12}\times\frac{1}{(1+0+0)} = 1$	-	-	-	1.000
2	-	$\frac{4}{12}\times\frac{1}{(0+1+1)} = 0.167$	$\frac{8}{12}\times\frac{1}{(0+0+1)} = 0.676$	-	0.843
3	-	$\frac{4}{16}\times\frac{1}{(0+1+1)} = 0.125$	-	$\frac{12}{16}\times\frac{1}{(0+0+1)} = 0.75$	0.875

Path	The Probability of Choosing Each Path Pr [k | C_n]	
	MNL model	PSL model
1	49.5%	53.8%
2	49.5%	45.3%
3	1.0%	0.9%

These results indicate that the PSL model adjusts the probabilities to account for the correlation between overlapping paths, making Path *1* more attractive than Path *2*. Since Path *1* and Path *2* have the same length, the MNL model assigns equal probabilities to both routes. However, the PSL model increases the probability of selecting Path *1* and decreases the probability of selecting Path *2* due to the overlapping link (i.e. link *b*) between Path *2* and Path *3*.

In the PS form, each link *a* in route $k \in R$ is penalized based on the number of routes in the choice set that also use that link (i.e. $\sum_{m\in C_n}\delta_{am}$). The significance of this penalization is weighted according to the prominence of link a in the route (i.e. $\frac{l_a}{L_k}$). A critical issue with the PSL model is that all routes contribute equally to the PS terms since the link-route incidence variable is either 0 or 1. This issue can pose a problem as the choice probabilities of realistic routes are affected by link sharing with unrealistic routes (e.g., paths with excessively long distances) (Duncan et al., 2020).

For example, when the travel distance of Path *3* increases due to a length increase in link *d* (from 12 to 36), as shown in Figure 2, the choice probability of Path *3* decreases (Table 2). As Path *3* becomes an unrealistic alternative due to its extraordinary length, the choice probability of Path *2* should not be penalized for overlapping with Path *3*. However, the PSL model's correction terms dictate that Path *3* contributes equally to the PS of Path *2* regardless of the degree of length increase in link *d*, resulting in the continuous penalization of the choice probability of Path *2*.

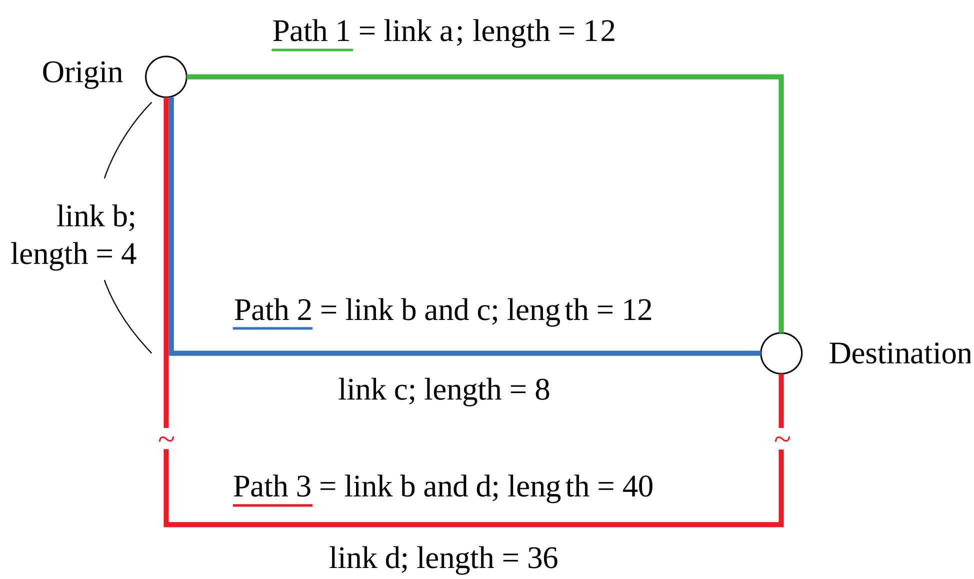

Figure 2 – The overlapping path problem (increase in the length of Path 3)

Table 2 – The probability of choosing each Path (increase in the length of Path 3)

Path	The Probability of Choosing Each Path Pr [k | C_n]	
	The length of Path 3 is 16	The length of Path 3 is 40
1	53.8%	54.5%
2	45.3%	45.5%
3	0.9%	0.0%

2.2.3 Generalised Path Size Logit

Several attempts have been made to reduce the contributions of excessively long paths to the PS of more realistic routes in the choice set. Ramming (2002) introduces a Generalised Path Size Logit (GPSL) model, which modifies the calculation of the Path Size factor. The GPSL model incorporates a weighting function that adjusts the contribution of overlapping routes based on their relative dissimilarity, and the GPSL Path Size for route $k \in R$ is defined as:

$$PS_k^{GPSL} = \sum_{a \in \Gamma_k} \frac{l_a}{L_k} \frac{1}{\sum_{m \in C_n} (\frac{L_k}{L_m})^{\phi} \delta_{am}} \quad \textbf{(8)}$$

noting that when ϕ = 0, the equation is insensitive to length, and the GPSL model is equivalent to the PSL model. The contribution of route *m* to the PS factor of route *k* is weighted according to the ratio between the length of two routes $(\frac{L_k}{L_m})^{\phi}$, thus reducing the contributions of long paths to the Path Size factor of short routes. $\phi \geq 0$ is the Path Size contribution scaling parameter to be estimated.

Hoogendoorn-Lanser et al. (2005) aimed to define overlap in multimodal networks and found that using ϕ value of 14 yielded the best results. However, other studies mentioned that the GPSL model could be problematic for large ϕ values, especially when overlapping routes only have marginally different distances (Ramming, 2002; Frejinger & Bierlaire, 2007). Moreover, the GPSL model has issues with internal inconsistency in assessing the feasibility of routes and scaling parameters (Duncan et al., 2020).

After reviewing the C-Logit, Path Size Logit (PSL), and Generalised Path Size Logit (GPSL) models, this study has opted to use the PSL model for route choice analysis. Despite the noted limitation of the PSL model, where each route contributes equally to the PS factor regardless of its realism, it is deemed suitable for this study. This is because the alternative routes considered in this study have lengths similar to those chosen, suggesting that none of them are excessively long. In addition, the PSL model is often utilized in practice due to its lower computational cost and ease of obtaining parameter estimates. Therefore, the choice of the PSL model is expected to provide accurate route choice predictions while maintaining computational efficiency for this study.

CHAPTER 3. DATA AND METHODS

3.1 Data

The study used a dataset of pedestrian GPS traces in Chicago, Illinois, designed to complement the traditional travel diary managed by the Chicago Metropolitan Agency for Planning (CMAP), known as My Daily Travel Survey 2019. Data collection occurred between September 2018 and May 2019, with 12,391 households participating. These households were distributed across the nine counties that make up the CMAP survey area. Throughout the survey period, app users could record up to 7 days of travel. The smartphone app enabled the collection of detailed information about individuals' stop locations. In total, 5,411 participants from 4,397 households downloaded and initialized the app, completing their travel reports either entirely or partially.

During the trip collection, the app leveraged iOS and Android geolocation features to automatically track participants' travel. After connecting to the user's travel record and obtaining the last known location, the app set a 75-meter auto-start geofence to detect the beginning of a trip and provide a level of privacy by not recording trip origins and destinations with extreme precision. The app activated when the participant left the geofence and collected GPS and accelerometer data at 30-second intervals. Auto-stop geofences were established around known locations, prioritized by their proximity to the current location. A smaller 50-meter auto-restart geofence was set around the last recorded GPS point, with the app collecting 30-second data cycles based on location updates.

The dataset contains identifiable user attributes, trip origin type, destination type, and trip purpose information. The present analysis only includes walking trips related to home-based work that occurred within the city of Chicago **(**Figure 3**)**. The city features an extensive transportation network, including the Chicago Transit Authority (CTA) trains, buses, and Metra commuter rail. Moreover, Chicago was ranked the 4th most walkable city among 2,800 cities in the United States and Canada (Walk Score, 2021) and is located within Cook County (excluding the O'Hare International Airport area).

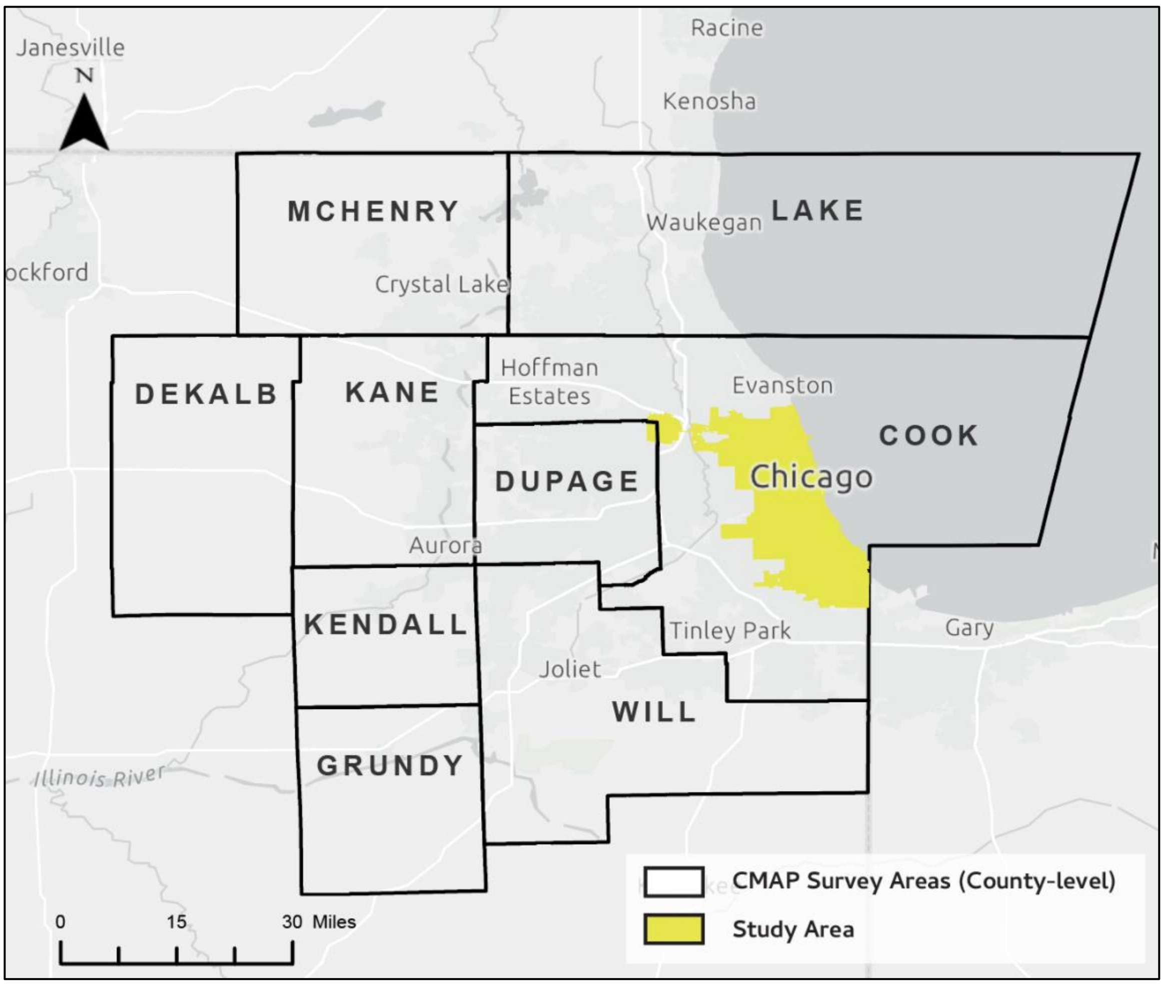

Figure 3 – Study area

3.2 Cleaning GPS Data

Raw GPS traces typically do not align with streets since GPS signals are often noisy and can result in inaccurate positioning data. To address this issue, the actual geolocated dot-data were map-matched to the Open Street Map Road network by the Hidden Markov Map-matching algorithm (Newson & Krumm, 2009), which has previously been applied to GPS data (Raymond et al., 2012). This algorithm identifies the most likely continuous route using a probabilistic approach based on Hidden Markov Models. It first calculates emission probabilities, representing the likelihood of each GPS point being associated with nearby street segments. Next, it computes transition probabilities, estimating the likelihood of moving from one street segment to another based on the street network connectivity. Using such probabilities for the entire GPS trace, the algorithm constructs a probability matrix and identifies the sequence of street segments with the highest overall probability, representing the most likely connected route. This study used the ‘Open Source Routing Machine’ (OSRM) map-matching service to obtain precise representations of pedestrian routes. OSRM is an open-source, high-performance routing engine that calculates optimal routes and provides routing services using OpenStreetMap data. The study employed OSRM as a standalone server and accessed it through its HTTP API. The resulting output was saved as a polyline in a shapefile using QGIS.

Each observation in the data contains a spatial GPS trace, timestamps indicating the start and end of the trip, and speed. To prepare the data for this study, first, observed trajectories that did not appear to be walk trips or which lacked a distance recorded by the phone application were eliminated. Trajectories with travel speed over 4.5 mph were also discarded. Additionally, paths with total travel distances less than 0.3 miles or greater than

2.0 miles were removed, as short paths often limit path choices, and long paths may mistakenly suggest other modes of transportation (e.g., bus, cycling). Second, trips with fewer than three GPS points or large gaps between points were filtered out to create a more accurate walk trip dataset. Trips with significant gaps between points do not offer sufficient accuracy when retracing the path due to the route variability in the gaps. Third, trajectories that were more than 50% longer than the shortest paths were eliminated, as these paths likely included additional stops or motivations other than direct travel to or from work. These paths usually present detours or loop-shaped paths throughout the trip. Fourth, only one-way trips were included, excluding round trips, where the origin and destination were sufficiently distant. Finally, only trajectories where pedestrians walked on streets, not inside buildings or on paths that were not present, were incorporated. Missing walkways included footpaths through parks, spaces, or waterfront areas. To collect street view images, it was necessary to identify trajectories within the Chicago Road network where Google Maps can provide street view images in any direction. Following these constraints, 560 traces out of the initial dataset of 3,981 were selected. It is important to note that 560 trips remained across 380 individuals.

3.3 Alternative Path Generation

In addition to the actual observed path, reasonable alternative paths should be included in route choice analysis. Alternative paths allow a choice model to determine if and how specific street features influence route choice preferences. Pedestrians can have many routes to choose from, which complicates the creation of explicit route choice sets in a model. Furthermore, many alternatives may overlap with the actual route and each other, consequently violating the IIA property in choice decisions. Therefore, researchers have employed various techniques to create plausible choice sets in route choice modeling. Although the concept of stochastic choice sets appears reasonable, Bovy (2009) discovered that most route choice modeling applications in research had applied a deterministic approach. Moreover, a relatively small number of alternatives are typically used in route choice analysis in practice.

Deterministic approaches utilize a predefined set of rules that can be consistently applied without involving randomness when generating path sets. For example, Ben-Akiva et al. (1984) introduced the labeled paths approach, which creates a choice set of optimal paths based on various criteria (e.g., travel time, distance, and congestion), where each route is the optimal one for its specific criterion. This method accommodates a wide range of priorities in seeking routes by applying objective functions such as minimizing travel time. The shortest path method finds the path with the minimal length, assuming pedestrians walk this path without considering street factors. This study implemented the constrained enumeration approach, which identifies all routes between origins and destinations that meet specific constraints, such as maximum detour limits or no segment repetitions (Prato & Bekhor, 2007).

Since Chicago's Street layout closely approximates a perfect grid (Boeing, 2019), it is feasible to generate numerous streets of similar lengths but varying street features. Given this urban context, all potential routes up to 50% longer than the shortest path were identified. Using street network shapefiles containing street centerlines, I eliminated highways and highway ramps from the network shapefile to prevent generating alternative routes on non-walkable roads. The data provided by the City of Chicago was last updated in 2021.

Subsequently, three alternatives are randomly selected from the full set for each path. If paths overlap with the observed route or other alternative paths by more than 30%, the random draw is repeated to maximize the uniqueness of all paths. This iteration is based on the methodology proposed by Basu and Sevtsuk (2022) for generating alternative paths, as they found that a 25% overlap between alternatives and the actual route was an appropriate balance between minimizing overlap and maximizing the diversity of route attributes in the choice set. Consequently, the advantage of this approach lies in the random draw being conducted from a large set of reasonable paths, reducing the likelihood of overlap with the actual path and other alternatives while increasing the diversity of street characteristics in the set (Sevtsuk et al., 2021). The final choice set consists of 560 actual routes and 1,680 non-chosen alternatives.

3.4 Collecting Route Variables

Eight route variables derived from street characteristics are hypothesized to influence pedestrian path choice. As discussed in Section 2, variables such as sky visibility, greenness, and the number of amenities, have been identified as having a significant impact on pedestrian route choice in previous literature. Table 3 provides an overview of these variables, including their descriptions and measurement methods. Variables like greenness, sky view factor, and the number of amenities and establishments were measured at fixed intervals (e.g., 40 meters) along the route, and then averaged to obtain a route-level value. For example, for a total route length of 160 meters, variables are measured at three points, excluding the origin and destination. Other variables were calculated along the entire route.

Table 3 – Route variables

Variable	Description	Measurement Method
Length	Route lengths in meters	Geometry calculation by R script
Turns	Number of turns along the route	R script
Amenities	Number of amenities along the route	Google Place API
Establishments	Number of corporations along the route	Google Place API
Sidewalk	Average area of sidewalks per meter	Geometry calculation by R script, using city of Chicago records
Greenness	Average % of green in Google Street View images along the route	% of green pixels measured by computer vision analysis
Sky View Factor	Average % of sky visibility in Google Street View images along the route	% of sky view pixels measured by computer vision analysis
Park	Average area of accessible parks per meter	Geometry calculation by R script, using city of Chicago records
Path Size Factor	Overlap indicator for alternative paths for the same trip	Python script

3.4.1 Length and Turns

Trip length is the most studied characteristic in route choice analysis since pedestrians generally prefer shorter routes. Additionally, the length variable is used to apply the 'equivalent walking distance' concept to other route variables, providing a consistent way of interpreting the estimated parameters of other variables in comparable length units (Sevtsuk et al., 2021). The ratio of coefficients of the length variable and another route variable can illustrate trade-offs, such as the balance between length and the number of turns, demonstrating how much additional length a pedestrian generally considers walking to avoid each extra turn. The route length, measured in meters (m), is calculated using the 'sf' package from R.

Turns are hypothesized to negatively impact the likelihood of choosing a route. More turns lead to cognitive complexity and frustration while walking, as pedestrians should be aware of their routes to avoid getting lost or making mistakes, such as missing a turn or taking a wrong one. Moreover, turns often involve crossing multiple streets, raising safety concerns and increasing waiting time. In this study, a turn is defined as a change in the direction of more than 45 degrees but less than 120 degrees along the route.

3.4.2 Pedestrian Amenities and Urban Design Features

Greenness and sky view factor represent the extent of visible green space and sky view along the walking route, respectively. A semantic segmentation tool was used to classify each pixel in an image into a specific category (e.g., sidewalk, sky, tree). Using the Street View Static API from Google Maps, Google Street View images were collected at 40-meter intervals along the route, with the heading value (e.g., 0 and 180 indicating

North and South) set as the road direction, considering that most people look ahead while walking. Although there is no strict distance interval for measuring representative street-level greenery or sky visibility, this approach was considered appropriate. To quantify the two variables, the calculation involved determining the percentage of pixels associated with each variable against the total pixels of the image (640 × 640). For example, the resulting greenness value represents the percentage of pixels in the Street View images classified as 'tree,' 'grass,' and 'plant.' Among various semantic segmentation models, the Deeplabv3+ model (Chen et al., 2018) and the PSPNet algorithm (Zhao et al., 2017) were initially used, as they are widely employed in studies evaluating walking environments using street view images. However, after empirically reviewing their performance, the PSPNet algorithm was chosen over the Deeplabv3+. Figure 4 illustrates the processing of original Street View Images to "greenness" and "sky view" categories using the PSPNet algorithm.

Location(lat/lng value): 41.95171, -87.66414
Heading: 179.67

Sky View Factor
Greenness

Figure 4 – The semantic segmentation of a Google Street View image, with (a) the original image and (b) the segmentation results blended onto the image

The number of amenities and establishments was quantified using the Places API from Google Maps. These places within a buffered area along the route (radius = 40m) were captured via the API search. Amenities refer to places that provide convenience or enjoyment to pedestrians, such as recreational facilities (e.g., gyms, amusement parks), services (e.g., laundry, post office), and dining facilities (e.g., café, restaurant). These amenities are not counted as establishments, which count the number of places where a company conducts its business operations or administrative tasks. The amenities and establishments variables are non-overlapping, as the former focuses on commercial spaces for pedestrian interaction, while the latter is more concerned with business operations. Correlations among amenities and establishments will be discussed in Section 4.2. The park index represents park accessibility and is determined by dividing the total area of parks located within a 40-meter buffer of the route by the length of the route.

3.4.3 Sidewalk

A sidewalk can be defined as an independent, safe walking environment, separated from road traffic. While most studies found that pedestrians are likely to choose routes with wide, continuous sidewalks adjacent to the street, the associations of such variables are not consistent in some studies (Basu et al., 2021). Additionally, a few observations in this study revealed that pedestrians walked on trails in parks or along the rivers, which are not captured in the Chicago sidewalk records. Consequently, this study aims to capture the impact of sidewalk quantity on route choice by measuring sidewalk scale in two different ways.

First, the sidewalk variable is quantified using a semantic segmentation tool, similar to the methods employed for quantifying Greenness and Sky View Factor variables. This approach not only detects sidewalks along roads but also identifies other types of walking environments not captured in the GIS shapefile provided by the City of Chicago. For example, Figure 5 illustrates a pedestrian path identified and measured as the 'sidewalk’

Location(lat/lng value): 41.87344, -87.65085
Heading: 87.99

Sidewalk

Figure 5 – Street view image captured inside the campus of UIC, with (a) the original image and (b) the segmentation results blended onto the image

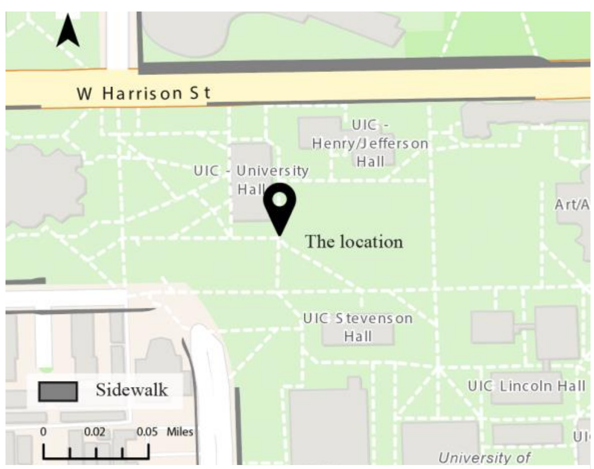

Figure 6 – The location of point where the image is captured

variable through semantic segmentation. However, in Figure 6, the GIS data indicates that no sidewalk exists at this location within the University of Illinois, Chicago (UIC) campus.

While this method of measuring the sidewalk variable offers higher accuracy in detecting sidewalks, it has two critical limitations that affect consistent scale measurement across all images. First, the sidewalk scale may be distorted due to the law of perspective. In the context of street view images, the road width and the lane in which the vehicle is traveling can influence the perceived scale of sidewalks. As the distance between the camera (mounted on the moving vehicle) and the sidewalk increases, the sidewalk appears smaller in the image. Consequently, a wider road can cause the sidewalks on both sides to seem narrower than they actually are.

Additionally, the vehicle’s position on the road can significantly impact the perspective and representation of sidewalks in the images. When the vehicle captures images near the centerline of the street, the camera is positioned at an equal distance from both sides of the road. This position allows for a more symmetrical view of the sidewalks. However, when the vehicle captures images while running closer to the edge of the road, the perspective distortion may be more pronounced, especially for the sidewalk on the opposite side of the road. In this case, the sidewalk closer to the vehicle will appear larger in the image, while the sidewalk on the opposite side will appear smaller due to the increased distance. Figure 7 represents the case when the sidewalk seems smaller when there are more lanes on the road. The road in the left image has about seven lanes, while the right has two lanes. Although sidewalks from both images have similar widths, there is a significant difference in detection (e.g., the value of the sidewalk variable in the left image is 0.011, while for the right image it is 0.091).

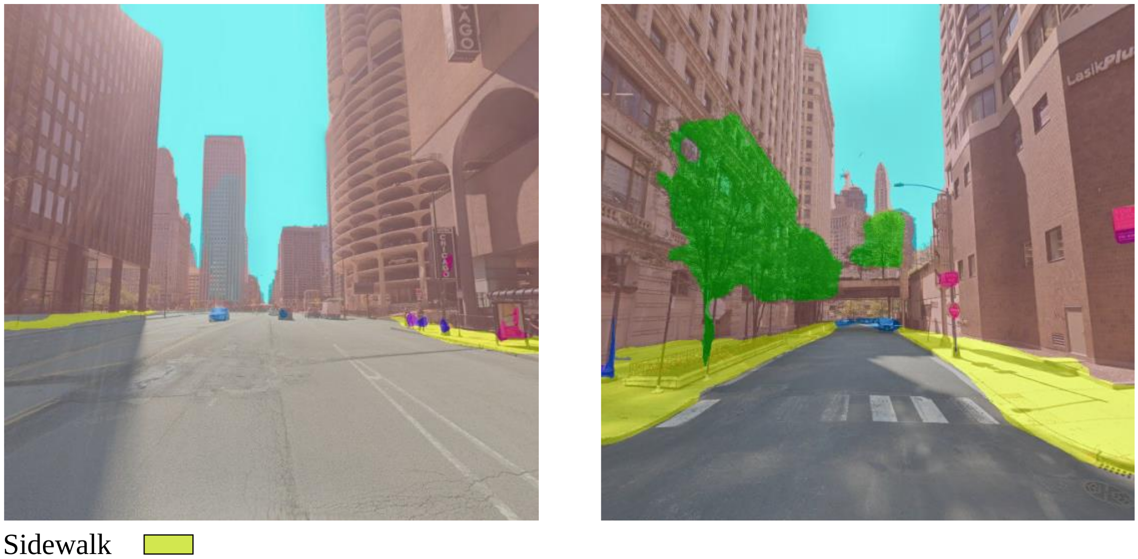

Figure 7– The sidewalk variable captured on (a) a road with seven lanes and (b) a road with two lanes

Second, capturing images on the road has the possibility of obstructions from parked cars, trees, or other stationary obstacles, which could further limit the visibility of sidewalks in the images. Such obstacles in the images may lead to incomplete or inaccurate sidewalk measurements depending on the location and environment. For example, downtown areas typically have higher concentrations of stationary obstacles; conversely, suburban areas tend to have fewer obstacles, resulting in clearer views of sidewalks in street view images.

Instead of using a semantic segmentation tool, this study applied the second approach, which represents the average sidewalk area per meter of the route within a 20-meter buffer when calculating the sidewalk variable. This method involves dividing the sidewalk area, recorded by the city of Chicago, within the buffered area of a route (radius = 20m) by the total length of a trip.

3.5 Path Size Logit Model

As discussed in Chapter 2, the probability $\Pr(k \mid C_n)$ that user n choosing a route k from the choice set *C* is:

$$\Pr(k \mid C_n) = \frac{e^{\mu(V_k)}}{\sum_{m \in C_n} e^{\mu(V_m)}} \tag{8}$$

where C_n represents the choice set, including the actual chosen route, for user *n*. The term μ is the logit scale term, while V_k denotes the deterministic utility function of route *k*, and it is expressed as:

$$\mathrm{V_k} = \sum_{i=1}^{8} \beta_{ki} X_{ki} + \beta_{PS}\, ln(PS_k) \tag{9}$$

where β_{ki} represents a vector of preference coefficients for the eight types of route attributes X_{ki}. By assuming that the marginal utility of each attribute is identical for all alternatives and does not vary with *k*, i.e. that $\beta_{ki} = \beta_i$, these route variables are considered to be generic variables. Although they generally take different values across alternatives, they possess the same coefficient.

PS_k represents the Path Size factor for route *k*, which accounts for the correlation resulting from overlapping alternative routes. Routes with overlapping links become less attractive options due to the decreased distinctiveness and increased similarity with other routes. Therefore, a higher PS value results in negative utility, as the natural logarithm function (i.e. ln(x)) converges to negative infinity when the x variable gets closer to zero. The method for calculating the PS factor can be found in equation (7).

Some reviews have pointed out a limitation of the PSL model: each route contributes equally to the PS factor, suggesting that every route has the same impact on the PS calculation within the travel-impedance variable (with a binary value), regardless of how realistic that route might be. This issue causes the choice probabilities of realistic routes to be negatively affected by link sharing with unrealistic routes, ultimately leading to inaccurate route choice predictions. Nonetheless, the PSL model was employed in this study because the alternative routes have a length similar to the chosen route, indicating that no routes have excessively long distances. Furthermore, this model is commonly applied in practice due to its relatively low computational cost and the ease in obtaining parameter estimates.

CHAPTER 4. RESULTS

This section describes the pedestrian route choice model, which is estimated from GPS data collected in the city of Chicago. The estimated parameters are used to examine the concept of "equivalent walking distance" to compare the impact of route variables on pedestrian route choices among different demographic groups.

4.1 Walk Trip Data

After applying the constraints discussed in Section 3.2, the analysis included 560 tips made by 380 individuals. These trips are recorded across the city of Chicago, as shown in Figure 8, with user demographics presented in Table 4**.**

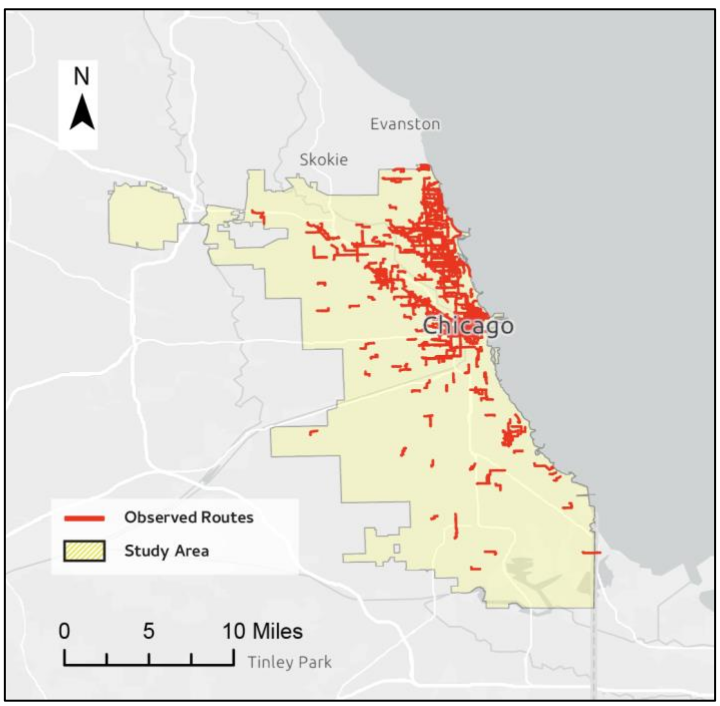

Figure 8 – Observed routes

Table 4 – Demographic characteristics of user sample

	N	%
Gender	380	100%
Male	192	51%
Female	188	49%
Age	380	100%
14 – 18	4	1%
19 – 25	69	18%
26 – 35	165	43%
36 – 45	68	18%
46 – 55	44	12%
56 – 65	20	5%
66 +	10	3%
Household Income	*379	100%
Less than $24,999	36	9%
$25,000 - $49,999	52	14%
$50,000 - $74,999	78	21%
$75,000 - $99,999	46	12%
$100,000 - $149,999	96	25%
$150,000 or more	71	19%
Race	**378	100%
White	299	79%
Others	79	21%

* Prefer not to answer (n=1)
**. Prefer not to answer (n=2)

The dataset was derived from survey respondent attributes collected in the ‘My Daily Travel Survey 2019,” conducted by CMAP. According to 2019 data, there were 2,409,786 individuals aged 10 or older living in the city of Chicago, Illinois (U.S. Census Bureau, 2019). Among them, 1,369,933 were employed and aged over 16 (U.S. Census Bureau, 2019). According to the My Daily Travel Survey 2019 result, 5.18 percent of work trips in

Cook County, Illinois (where Chicago is located) were made by walking mode, which suggests that 70,962 Chicago residents may walk to work. For the sample in this study, 380 individuals reported walking as their primary mode during the survey as they exclusively use walking as their mode of transportation. Therefore, as shown in Figure 9**,** the sample size for this study is 380, representing 0.5% of Chicago's walking commuters. Therefore, this convenience sample is used for exploratory analysis of walking trips related to home-based work trips. Thus, there is no question of extrapolating the results to the entire Chicago population (nor even to the population of commuters), but the analysis can still provide insight into the walk route attributes considered important in a commuting context.

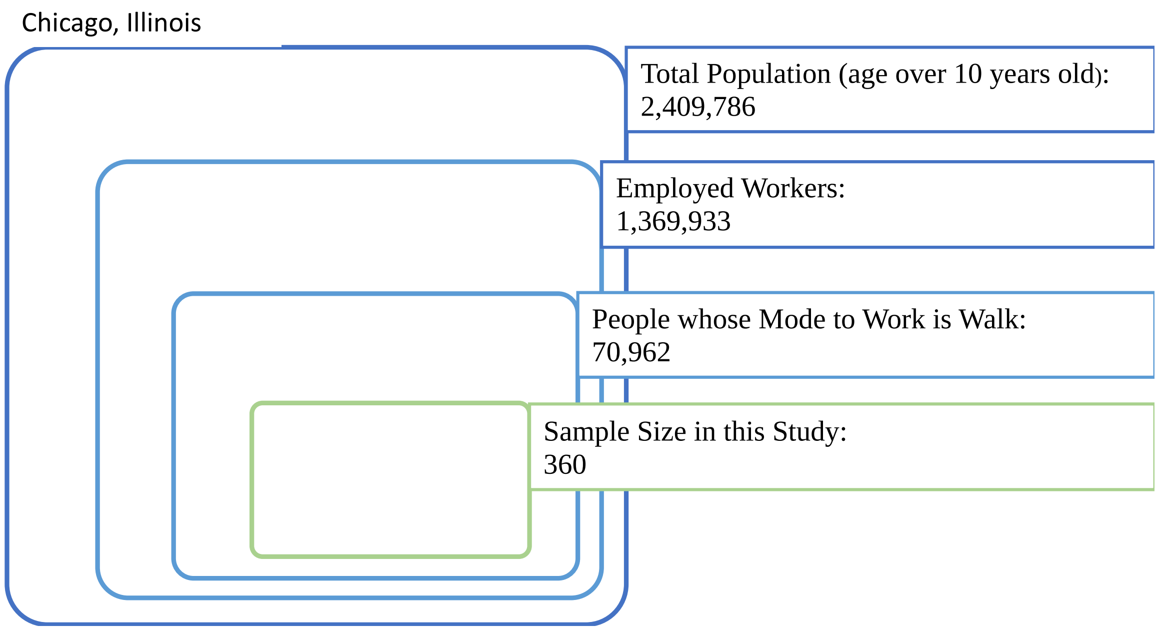

Figure 9 – Sample representation

Notably, some individuals in the study sample recorded multiple trips, while 69% of the total individuals conducted only one walk trip associated with their commute. Table 5 represents the distribution of recorded trips per respondent, ranging from 1 to 8. The problem is that the presence of multiple observations from the same individual means that common unobserved characteristics will likely be shared among those observations, violating the assumption of independence between observations, which produces inconsistent parameter estimates. This is because the observations cluster into groups, with potentially a high degree of dependence within each group. For example, consider a situation where the 380 individuals in the sample are perfectly representative of the population, and each contributes two observations to the sample. Without accounting for the intra-person correlation of error terms, the estimation would incorrectly assume there are 760 (380 x 2) independent observations. In reality, however, these observations cluster into 380 independent groups, each containing two observations from the same individual. This affects the calculation of the standard errors of the parameter estimates and can potentially lead to biased conclusions. It is important to account for this clustering of observations within individuals in order to ensure valid results.

Table 5 – Distribution of recorded walk trips per respondent (N=380)

The number of trips per participant	The number of participants
1	263
2	76
3	28
4	9
5	2
6	0
7	1
8	1

To address this issue, this study uses cluster robust standard errors. This approach is designed to account for the correlated unobserved variables between observations within the same cluster, which is the same individuals in this case, while maintaining the independence assumption between different clusters. By using these errors, more accurate estimates of the standard errors of the coefficient estimators are obtained, resulting in more reliable hypothesis tests and confidence intervals for those estimators.

4.2 Route Variables

Table 6 describes the mean value and descriptions of the route variables for the actual chosen route and three alternatives used in the model. The alternatives for each observation are classified by the value of PS, from low to high.

The actual chosen trips had an average length of 1,190 meters, which is slightly shorter compared to other alternatives. Similarly, the average values for the number of turns and greenness in the actual chosen trips are lower than those of the alternatives. On the other hand, the sky view factor and the park variable values are higher for the chosen routes. The average number of amenities and establishments, and the average sidewalk area are comparable between the chosen trips and alternatives.

Out of all routes (N=2,240), 217 outliers are detected based on the Mahalanobis distance when considering the eight types of route variables. While there are multiple outliers for each route variable when using the interquartile range (IQR) method, this study chose not to exclude them, as each observation has unique and valuable information despite its outstanding value.

Table 6 – Mean attribute values for the chosen route and alternative routes

Variable	Description	Mean			
		Chosen	Alt. 1	Alt. 2	Alt. 3
Length	Route length in meters	1190.3	1215.3	1208.8	1211.0
Turns	Number of turns along the route	2.8	4.0	4.0	3.9
Amenities	Number of amenities along the route	45.3	43.8	45.1	42.5
Establishments	Number of corporations along the route	41.1	41.4	42.6	41.4
Sidewalk	Average area of sidewalks per route-meter	3.1	2.9	3.2	3.1
Greenness	Average % of vegetation in Google Street View images along the route	15.1	15.7	15.4	15.7
Sky View Factor	Average % of sky visibility in Google Street View images along the route	24.6	24.0	24.0	23.8
Park	Average area of accessible parks per route-meter	5.4	5.1	5.3	5.2
Path Size	Overlap indicator for alternative paths for the same trip	0.9449	0.9042	0.9045	0.9082

Figure 8 shows the correlations among route variables for all routes. The average sidewalk area is seen to have a strong positive correlation (0.84) with the park variable. This correlation may be because both variables are likely influenced by the urban design in the area, such as well-planned neighborhoods featuring extensive sidewalks and well-distributed parks. The number of amenities and establishments showed a strong positive correlation (0.81), as expected, because both variables tend to be clustered in commercial or mixed-land use areas. Negative correlations are observed between the greenness/sky view factor and the number of points of interests (i.e. establishments and amenities). This correlation can be attributed to the characteristics of urban environments; densely built-up areas with more points of interests often have taller buildings and less open space, resulting in lower sky visibility and greenness. Despite the presence of multicollinearity among the route variables, this study included all of them in the route choice model, as they have significantly different impacts on pedestrian route choice preferences.

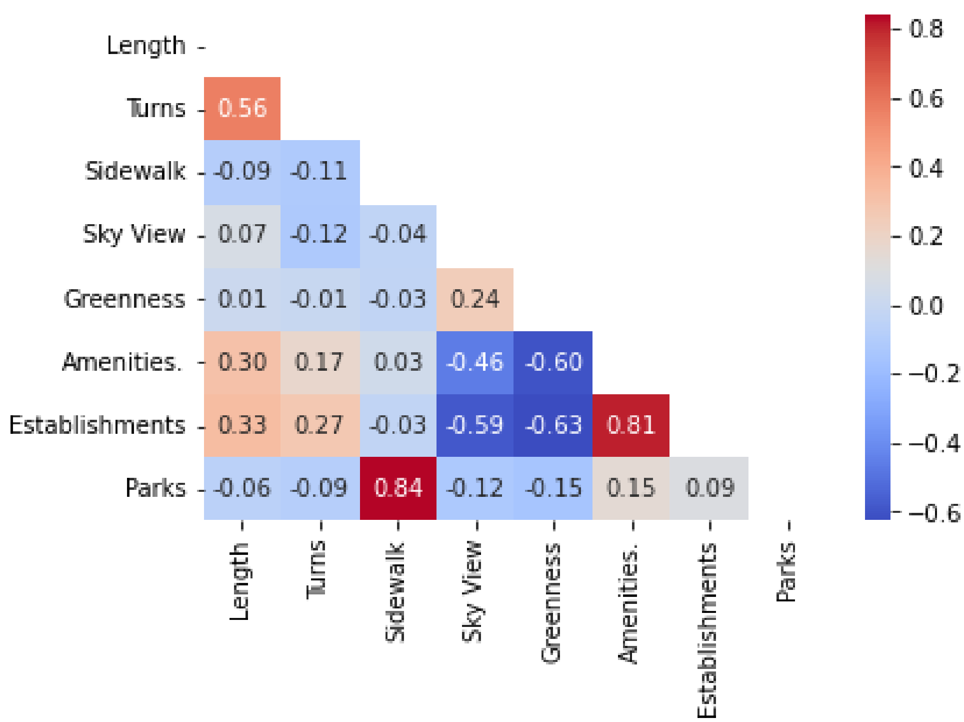

Figure 10 – Route variable correlation matrix

4.3 Model Results

The results of four different models presented in Table 7 include coefficient estimates, statistical significance at various confidence levels, and t-statistics based on the cluster robust standard errors, which account for repeated choices made by the same individual. The model without considering the cluster robust standard errors can be found in Appendix A. Additionally, log-likelihood ratio tests and adjusted rho-squared were calculated to compare different models against one another to identify the best model for evaluating the impact of route variables on route choices.

First, Model one (M1) considers route length as the only explanatory variable using a MNL specification, as the distance is a critical factor in route choice decisions. The negative coefficient for route length in M1 indicates that pedestrians tend to choose shorter routes. Model two (M2) also includes length as the explanatory variable but employs a PSL specification. As the PSL model addresses the overlap issue by incorporating a path size factor, it significantly improves in goodness-of-fit. This improvement can be observed when comparing the adjusted rho-squared values between the two models, with M2 showing a higher value of 0.173 compared to M1's value of 0.005. Although there is relatively little overlap among alternatives on average, this result confirms that the PSL model, which accounts for correlations arising from overlapping alternatives, better explains route choice behavior compared to the MNL model.

Model three (M3) is a PSL model that includes all eight route variables described in Table 7. Most variables are significant at the 10% confidence level, but the 'length' and 'amenities' variables remain insignificant. Despite M3 having the highest goodness-of-fit

Table 7– Parameter estimates for different specifications of MNL and PSL models

Variable	M1			M2			M3			M4		
	Beta		t-stat	Beta		t-stat	Beta		t-stat	Beta		t-stat
Length	- 0.001	~	- 1.8	- 0.002	**	- 2.8	+ 0.000[1]		0.9	- 0.002	**	- 2.7
Turns							- 0.335	***	- 7.5			
Greenness							+ 4.192	~	1.9	+ 5.178	*	2.4
Sky View Factor							+ 6.712	*	2.2	+ 9.663	***	3.4
Sidewalk							- 0.094	~	- 1.8	- 0.085		- 1.6
Park							+ 0.105	*	2.2	+ 0.100	*	2.3
Amenities							+ 0.008		1.6	+ 0.009	~	1.9
Establishment							- 0.017	**	- 3.0	- 0.016	**	- 3.0
ln(Path Size)				+ 16.025	***	14.3	+ 14.726	***	11.7	+ 16.248	***	13.9
Final log-likelihood	- 771.60			- 640.33			- 585.14			- 625.84		
AIC	1545.20			1284.65			1188.28			1267.68		
Adj. rho-squared	0.005			0.173			0.235			0.184		

[1] +0.00051

Significance level: ~p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

In the calculation of Adj. rho-squared, the benchmark is the equally-likely model

(with an adjusted rho-squared of 0.235) and lower log-likelihood, it is challenging to select this model as the final model due to the statistical insignificance of the 'length' variable. To evaluate the impact of route variables on pedestrian route choice using the 'equivalent walking distance' measurement, the 'length' variable is essential, as it helps interpret trade-offs with other route variables and determine how much additional distance an average pedestrian would consider walking to avoid unit increases in corresponding route features.

Consequently, we excluded the 'turns' variable, which significantly reduces the significance of the 'length' variable. We then compared the log-likelihood between the model without 'turns' and 'amenities' variables and the model excluding only the 'turns' variable (the result of the model without 'turns' and 'amenities' variables can be found in Appendix B). As the latter model exhibits a lower Akaike Information Criterion (AIC) and higher goodness-of-fit, this model (i.e. model four (M4)) is chosen as the final model to interpret the impact of each route variable. The exclusion of the ‘turns’ variable, which results in an adjusted rho-squared drop from 0.235 to 0.184, may not seem ideal. However, M4 is chosen to maintain the significance of the ‘length’ variable, allowing for the evaluation of equivalent walking distance and comparison of the impact of route variables across different groups of people. M4 also demonstrates a relatively high goodness-of-fit (with an adjusted rho-squared of 0.184) compared to the model that includes only the 'length' variable, and aligns with typical model expectations. In the context of choice models, McFadden (1980) considered adjusted rho-squared values between 0.2 and 0.4 to be exceptional.

Most statistically significant variables in M4 exhibit expected signs. As anticipated, the route length negatively affects choice, with pedestrians preferring shorter routes. The positive ‘greenness’ coefficient suggests people prefer routes with more visible greenery, which provides benefits such as aesthetic appeal, shade, and reduced noise and air pollution. Similarly, the positive ‘park’ variable sign indicates that pedestrians prefer walking routes with more accessible park areas, which offer relaxation, socializing, and aesthetic improvement opportunities. Moreover, some routes in this study were observed to cross parks along the Chicago River or Lake Michigan instead of streets. The positive ‘amenities’ coefficient reveals that more amenities along a route increase its attractiveness to pedestrians, with routes offering more recreational facilities, services, and dining options more likely to be chosen. Lastly, the correction term derived from the PS factor (i.e. *ln(PS))* also exhibits a positive and significant effect.

Although some studies suggested that people prefer routes more enclosed by built edges and trees rather than exposed to the sky, the sky view factor variable has a strong positive coefficient in this study. One possible reason may be that in the downtown area, none of the chosen routes pass by streets with overhead subway tracks, while some alternative paths do. Elevated railways can completely block sky visibility, making people less likely to prefer walking there. It suggests that pedestrians prefer routes with more sky visibility over those completely out of sight, as they can provide a more open and spacious feeling. Also, the coefficient sign for the 'establishments' variable is negative, while it was expected to be the same as the sign of the 'amenities' variable since pedestrians might find routes with more establishments attractive due to increased opportunities for shopping or socializing. However, this study suggests that pedestrians prefer routes with fewer corporations or businesses, possibly because such routes might be quieter, less congested, or less polluted. Similar to the sky view factor, tall buildings are more likely to block sky visibility, as most corporations are located in tall buildings in downtown Chicago. Moreover, this study raises the possibility that the negative sign for the 'establishments' variable could be attributed to multicollinearity, given the high correlation (0.81) between the 'amenities' and 'establishments' variables.

By calculating the ratio of coefficients for the model parameters, the trade-offs between variables can be determined. Previous literature has utilized trade-offs to understand pedestrian preferences and behaviors, though they may be referred to using different terms. In this study, route length is used as the reference variable so that the effects of other route variables can be expressed through 'equivalent walking distance.' The concept of equivalent walking distance was used in Basu and Sevtsuk’s (2022) study on pedestrian route choice analysis to quantify the perceived effort or disutility associated with various route characteristics, enabling a consistent and comparable interpretation of model results. For instance, the coefficients in Table 7 suggest that the trade-off between one unit of establishment, $\beta_{establishment}$, and walking distance in meters, β_{length} , is:

$$\frac{\beta_{establishment}}{\beta_{length}} = \frac{-\,0.01641044}{-\,0.00173766} \approx 9.4$$

This ratio indicates that the effort of passing one extra establishment is perceived as equivalent to 9.4 meters of walking. Conversely, the distance can be perceived as shorter in the case of the trade-off between amenities and walking distance, computed as:

$$\frac{\beta_{amenities}}{\beta_{length}} = \frac{+\,0.00922258}{-\,0.00173766} \approx -\,5.3$$

This implies that pedestrians are willing to extend their walk by 5.3 meters to pass by one additional amenity. In other words, a route that passes 10 amenities provides the same utility as a route with no amenities that is 53 meters shorter, keeping all other route attributes constant.

The equivalent walking distance for the five route attributes – Greenness, Sky View Factor, Park, Amenities, and Establishment – that are statistically significant in the final model is presented in Table 8. In addition to the variables already discussed, it is found that sky visibility has a considerable impact on the overall utility of a route. A 10-percentage point increase in sky visibility along the route is associated with an increase in route utility equivalent to reducing the actual walking distance by 556.1 meters, on average. While this might seem quite large, accounting for about 50% of the average route length of the sample, one possible explanation for this finding is that pedestrians may perceive routes with higher sky visibility as more open and spacious, providing a more enjoyable and comfortable walking experience. In the context of the study area, unique factors such as the presence of elevated railways or tall buildings in downtown Chicago might significantly affect pedestrians' perception of sky visibility and their route choices. Moreover, routes that cross parks along the lake or river tend to have higher sky visibility, which could make people prefer to walk these routes for the pleasant scenery and natural surroundings. Similar to sky visibility, a 10-percentage point increase in greenery along the route increases the route utility, equivalent to a reduction in actual walking distance of 298 meters. Furthermore, a route with a one m^2 increase per meter of route length, in the size of a park located within the buffered area per meter, is, on average, increases the route utility, equivalent to a reduction in actual walking distance of 57.7 meters.

Table 8 – Equivalent walking distance

Variable	Length equivalent interpretation	in meters
Greenness	An average increase in Greenness by 10-percentage points along the route is associated with a utility increase equivalent to a reduction of:	298.0
Sky View Factor	An average increase in SVF by 10-percentage points along the route is associated with a utility increase equivalent to a reduction of:	556.1
Park	One m^2 of increase per meter of route length in park size along the route is associated with a utility increase equivalent to a reduction of:	57.7
Amenities	Passing one extra amenity is associated with a utility increase equivalent to a reduction of:	5.3
Establishment	Passing one extra establishment is associated with a utility decrease equivalent to an increase of:	9.4

As the equivalent walking distance is based on a utility function calibrated on actual pedestrians' route choices, it can be applied to the preference of specific groups of people based on the different utility functions for each group (Basu & Sevtsuk, 2022). Thus, this study applies this concept to the final model, which includes the length variable, five other route variables (i.e. greenness, sky view factor, park, amenities, and establishments), and the PS factor across different demographic groups.

4.3.1 Gender

Table 9 presents the results of a route choice analysis by gender, revealing differences in preferences for various route variables between males and females. The route length has negatively impacted route choice for both males and females. While most varia-

Table 9 – Parameter estimates for gender group

Variable	Male			Female		
	Beta		t-stat	Beta		t-stat
Length	- 0.001	~	- 1.7	- 0.002	**	-2.6
Greenness	6.667	*	2.2	3.881		1.2
Sky View Factor	11.668	**	2.9	7.770	~	1.9
Sidewalk	- 0.079		-1.5	- 0.091		-0.8
Park	0.154	*	2.1	0.092		0.9
Amenities	0.008		1.2	0.015	~	2.0
Establishment	-0.021	**	-2.6	- 0.011		-1.3
ln(Path Size)	14.748	***	9.3	18.147	***	10.6
Final log-likelihood	- 335.20			-286.04		
AIC	686.41			588.08		
Adj. rho-squared	0.155			0.206		
Num. obs.	293			267		

Significance level: ~p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001
In the calculation of Adj. rho-squared, the benchmark is the equally-likely model

bles show distinct differences in strength and significance, only the ‘sky view factor’ variable is significant for both groups. Consequently, it is possible to compare this impact using the equivalent walking distance concept. Keeping all other route variables constant, males more strongly prefer to walk a route with more sky visibility compared to females, because a 10-percentage point increase in sky visibility is associated with a utility increase equivalent to a reduction of 787.7 meters in walking distance for males, while only 375.3 meters for women. One of the potential reasons for the gender difference in valuing the sky visibility might be that females could be more sensitive to temperature changes or weather conditions, such as sunlight exposure. Another possible reason could be related to a sense of safety, as females might feel more secure with greater enclosure provided by surrounding buildings. As a result, although sky visibility is still a desirable attribute for women as well, they may have a weaker preference for it in view of these drawbacks.

4.3.2 Age

The results of two different models based on the sample of people aged 30 years old or younger and people who are over 30 years old are presented in Table 10. Similar to the gender group, only the ‘length’ and ‘sky view factor’ variables have a significant impact on route choice. As the sign of these coefficients is opposite, the sky view factor helps people reduce the perceived length of the walk. On average, individuals aged over 30 years old place a greater value on routes having more sky visibility than do those aged 30 years old or younger, with the equivalent walking distance for the former group being 869.1 meters, while the latter group shows 287.2 meters). This difference could be attributed to older individuals prioritizing safety and preferring routes with higher sky visibility due to increased visibility and a higher likelihood of open spaces. Additionally, age groups may have different aesthetic preferences for walking environments, with older individuals potentially appreciating more strongly open spaces with greater sky visibility than younger individuals do.

Table 10 – Parameter estimates for age group

Variable	30 years old and younger			31 years and older		
	Beta		t-stat	Beta		t-stat
Length	- 0.003	***	- 3.4	- 0.001	~	-1.7
Greenness	2.185		0.6	7.040	*	2.5
Sky View Factor	7.892	*	2.0	10.743	**	2.7
Sidewalk	- 0.045		-0.5	- 0.105	~	-1.8
Park	0.101		1.1	0.110	~	1.9
Amenities	0.002		0.3	0.015	*	2.2
Establishment	-0.012		-1.2	- 0.021	**	-3.0
ln(Path Size)	14.655	***	8.3	17.786	***	11.0
Final log-likelihood	- 259.74			- 360.99		
AIC	535.48			737.97		
Adj. rho-squared	0.153			0.198		
Num. obs.	228			332		

Significance level: ~p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001
In the calculation of Adj. rho-squared, the benchmark is the equally-likely model

4.3.3 Income

Table 11 presents the results of a route choice analysis across different income groups. As anticipated, route length has a negative impact on route choice decisions for all income groups. However, there are no common variables that are statistically significant across all groups, making it challenging to compare the impact of each route variable among different income levels.

Table 11– Parameter estimates for income group

Variable	~ \$49,999			\$50,000 ~ \$99,999			\$100,000 ~		
	Beta		t-stat	Beta		t-stat	Beta		t-stat
Length	- 0.003	*	- 2.3	- 0.002	**	- 2.9	- 0.001		- 1.2
Greenness	2.461		0.5	5.222		1.4	6.124	~	1.9
Sky View Factor	10.988	~	1.7	12.683	**	2.9	6.523		1.4
Sidewalk	- 0.073		- 0.2	- 0.081		- 1.1	- 0.080		- 0.9
Park	0.178		0.6	0.078		1.1	0.154	*	2.1
Amenities	0.005		0.4	0.010		1.2	0.010		1.4
Establishment	- 0.014		- 0.7	- 0.006		- 0.7	- 0.026	**	- 3.1
ln(Path Size)	15.992	***	8.2	15.253	***	8.3	17.734	***	8.7
Final log-likelihood	- 147.52			- 191.67			- 277.14		
AIC	311.04			399.33			570.27		
Adj. rho-squared	0.163			0.172			0.181		
Num. obs.	134			174			251		

Significance level: ~p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001
In the calculation of Adj. rho-squared, the benchmark is the equally-likely model

CHAPTER 5. CONCLUSION

This study aimed to enhance the understanding of pedestrian route choice behavior by exploring the impact of various street features on route preferences. Utilizing a dataset of GPS walking trajectories in Chicago, Illinois, and incorporating personal attributes, the study focused on how street features affect pedestrian route choice behavior across different demographic groups, such as gender and income level. The findings contribute to the existing body of literature on pedestrian route choice behavior by providing valuable insights for urban planners and policymakers in designing and implementing sustainable pedestrian environments that cater to the needs of specific populations.

The route choice model in this study tested eight different street features, with six variables – route length, greenness, the numbers of amenities and establishments, sky visibility, and park accessibility – demonstrating a significant relationship with pedestrian route choice in the final model. The model revealed that pedestrians generally prefer shorter routes with more greenery, greater sky visibility, increased park accessibility and more amenities. Conversely, a greater number of establishments along the route negatively impacted route choice decisions. The study also utilized the concept of equivalent walking distance to quantify the incremental disutility associated with various street features. For example, on average, a pedestrian is willing to walk an additional 5.3 meters to pass by one more amenity.

The analysis further investigated differences in route choice preferences across demographic groups, hypothesizing that gender, age, and income play a role in shaping pedestrian preferences. The results indicated that males valued sky visibility more than females did. Individuals aged over 30 years old also tended to value sky visibility more highly than younger pedestrians did. However, no common variables were found to be statistically significant across all income groups, making it difficult to compare the impact of route variables among different income levels.

Despite the study's findings, some limitations should be acknowledged. First, the GPS data may not always be accurate, as signal loss or interference in urban environments can result in positional inaccuracies. If an individual is surrounded by tall buildings or inside a building, the obstructed sky view may lead to the inaccurate recording of the actual routes taken. During the process of cleaning GPS data in this study, several paths had significant gaps in distance between GPS points or points are located on the river in downtown Chicago. Second, the study did not include enough street characteristics to explore pedestrian route choice behavior thoroughly. The route variables tested in this study are focused on stationary objects or places along the street. While people walk along the street, they may interact with moving vehicles or other pedestrians. These moving objects can also impact walking behavior, regardless of pedestrians' walking speed. For example, the likelihood of choosing a walking route was reported to increase with a decrease in vehicle traffic volume (Sevtsuk et al.,2021). Furthermore, some studies found that higher pedestrian density along a route leads to a lower likelihood of that route being chosen (Bafatakis et al., 2015; Gim & Ko,2017). Third, future research should include walking trajectories related to non-commuting trips and those occurring in CMAP survey areas, not just the city of Chicago, in order to analyze using larger samples and to better reflect general pedestrian route choice behavior. Lastly, another limitation of this study is the sensitivity of the equivalent walking distance values to the coefficient of route length.

The estimates of equivalent walking distance are heavily dependent on the magnitude of the length coefficient, which, in turn, is influenced by the model specification, particularly the inclusion or exclusion of variables such as the "turns" variable, as discussed in Section 4.3. The choice of the model specification can materially influence the coefficient estimate, potentially affecting the interpretation of trade-offs between route variables and the additional distance an average pedestrian would consider walking to avoid unit increases in corresponding route features. To address this limitation, future research could conduct sensitivity analyses on the equivalent walking distance values across diverse model specifications, assessing the robustness of the findings and providing further insights into how different model configurations might impact the results.

APPENDIX A. PARAMETER ESTIMATES FOR MODELS WITHOUT CLUSTER ROBUST STANDARD ERRORS

This appendix illustrates the four different models which do not account for the cluster robust standard errors. The estimated coefficient of each variable and the performance metrics of each model remains the same as in the Table 7.

Variable	M1			M2			M3			M4		
	Beta		t-stat	Beta		t-stat	Beta		t-stat	Beta		t-stat
Length	- 0.001	~	- 3.0	- 0.002	***	-4.6	+ 0.000[1]		1.1	- 0.002	**	- 4.3
Turns							- 0.335	***	-8.3			
Greenness							+ 4.192	~	2.0	+ 5.178	*	2.6
Sky View Factor							+ 6.712	*	2.5	+ 9.663	***	3.8
Sidewalk							- 0.094	~	- 1.4	- 0.852		- 1.3
Park							+ 0.105	*	1.7	+ 0.100	*	1.7
Amenities							+ 0.008		1.7	+ 0.009	~	2.1
Establishment							- 0.017	**	- 2.8	- 0.016	**	- 2.9
ln(Path Size)				+ 16.025	***	13.5	+ 14.726	***	11.6	+ 16.248	***	13.3
Log-likelihood	- 771.60			- 640.33			- 585.14			- 625.84		
AIC	1545.20			1284.65			1188.28			1267.68		
Adj. rho-squared	0.005			0.173			0.235			0.184		

[1] +0.00051

Significance level: ~p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001

In the calculation of Adj. rho-squared, the benchmark is the equally-likely model

APPENDIX B. PARAMETER ESTIMATES FOR MODLE WITH SIX ROUTE FEATURES (EXCLUDES 'TURNS' AND 'AMENITIES' VARIABLES)

This appendix presents the results of a model that incorporates six route features as explanatory variables. The purpose of this model is to compare it with another model (i.e. M4) that only excludes the 'turns' variable. The comparison reveals that the model with six route features has a slightly lower adjusted rho-squared and higher AIC.

Variable	M_Appendix B		
	Beta		t-stat
Length	- 0.002	**	- 2.7
Turns			
Greenness	+ 4.066	~	1.9
Sky View Factor	+ 9.249	**	3.2
Sidewalk	- 0.083		- 1.6
Park	+ 0.099	*	2.3
Amenities			
Establishment	- 0.009	*	- 2.2
ln(Path Size)	+ 16.322	***	14.1
Final log-likelihood	- 627.97		
AIC	1269.943		
Adj. rho-squared	0.182		

REFERENCES

Azegami, Y., Imanishi, M., Fujiwara, K., & Kusaka, H. (2023). Effects of solar radiation in the streets on pedestrian route choice in a city during the summer season. *Building and Environment*, 110250. https://doi.org/10.1016/j.buildenv.2023.110250

Bafatakis, C., Duives, D., & Daamen, W. (2015, January 11-15). Determining a pedestrian route choice model through a photo survey. In *Transportation Research Board 94th Annual Meeting*. Washington DC, United States.

Basu, R., & Sevtsuk, A. (2022). How do street attributes affect willingness-to-walk? City-wide pedestrian route choice analysis using big data from Boston and San Francisco. *Transportation Research Part A: Policy and Practice*, *163*, 1-19. https://doi.org/10.1016/j.tra.2022.06.007

Bellizzi, M. G., Forciniti, C., & Mazzulla, G. (2021). A stated preference survey for evaluating young pedestrians' preferences on walkways. *Sustainability*, *13*(22), 12434. https://doi.org/10.3390/su132212434

Ben-Akiva, M., Bergman, M. J., Daly, A. J., & Ramaswamy, R. (1984, July 11-13). Modelling inter urban route choice behaviour. In *Papers presented during the Ninth International Symposium on Transportation and Traffic Theory*. Delft, Netherlands.

Ben-Akiva, M., & Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. *Handbook of Transportation Science*, 5-33.

Ben-Akiva, M., & Bierlaire, M. (2003). Discrete choice models with applications to departure time and route choice. *Handbook of Transportation Science*, 7-37.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand (Vol. 9). MIT press.

Bovy, P. H. (2009). On modelling route choice sets in transportation networks: a synthesis. *Transport reviews*, *29*(1), 43-68. https://doi.org/10.1080/01441640802078673

Boeing, G. (2019). Urban spatial order: Street network orientation, configuration, and entropy. *Applied Network Science*, *4*(1), 1-19. https://doi.org/10.1007/s41109-019-0189-1

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

Broach, J., & Dill, J. (2015, January 11-15). Pedestrian route choice model estimated from revealed preference GPS data. In *Transportation Research Board 94th Annual Meeting*. Washington DC, United States.

Brookfield, K., & Tilley, S. (2016). Using virtual street audits to understand the walkability of older adults' route choices by gender and age. *International Journal of Environmental Research and Public Health*, *13*(11), 1061. https://doi.org/10.3390/ijerph13111061

Cascetta, E., Nuzzolo, A., Russo, F., & Vitetta, A. (1996, July 24-26). A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In *Transportation and Traffic Theory. Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Lyon, France.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision.* (pp. 801-818).

Dessing, D., de Vries, S. I., Hegeman, G., Verhagen, E., Van Mechelen, W., & Pierik, F. H. (2016). Children's route choice during active transportation to school: difference between shortest and actual route. *International Journal of Behavioral Nutrition and Physical Activity*, *13*, 1-11.

Duncan, L. C., Watling, D. P., Connors, R. D., Rasmussen, T. K., & Nielsen, O. A. (2020). Path size logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. *Transportation Research Part B: Methodological*, *135*, 1-40. https://doi.org/10.1016/j.trb.2020.02.006

Erath, A. L., Van Eggermond, M. A., Ordóñez Medina, S. A., & Axhausen, K. W. (2015, July). Modelling for Walkability: Understanding pedestrians' preferences in

Singapore. In *14th International Conference on Travel Behavior Research*. ETH Zurich, Switzerland. https://doi.org/10.3929/ethz-b-000106077

Ewing, R., & Handy, S. (2009). Measuring the unmeasurable: Urban design qualities related to walkability. *Journal of Urban Design*, *14*(1), 65-84. https://doi.org/10.1080/13574800802451155

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

Gim, T. H. T., & Ko, J. (2017). Maximum likelihood and firth logistic regression of the pedestrian route choice. *International Regional Science Review*, *40*(6), 616-637. https://doi.org/10.1177/0160017615626214

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

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

Liu, Y., Yang, D., Timmermans, H. J., & de Vries, B. (2020). Analysis of the impact of street-scale built environment design near metro stations on pedestrian and cyclist road segment choice: A stated choice experiment. *Journal of Transport Geography*, *82*, 102570. https://doi.org/10.1016/j.jtrangeo.2019.102570

Lue, G., & Miller, E. J. (2019). Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behaviour and society*, *14*, 34-42. https://doi.org/10.1016/j.tbs.2018.09.008

Muraleetharan, T., & Hagiwara, T. (2007). Overall level of service of urban walking environment and its influence on pedestrian route choice behavior: analysis of pedestrian travel in Sapporo, Japan. *Transportation Research Record*, *2002*(1), 7-17.

Newson, P., & Krumm, J. (2009, November). Hidden Markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International*

Conference on Advances in Geographic Information Systems, 336-343. https://doi.org/10.1145/1653771.1653818

Ozbil, A., Argin, G., & Yesiltepe, D. (2016). Pedestrian route choice by elementary school students: The role of street network configuration and pedestrian quality attributes in walking to school. *International Journal of Design Creativity and Innovation*, *4*(2), 67-84. https://doi.org/10.1080/21650349.2015.1123120

Prashker, J. N., & Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: a review. *Transport Reviews*, *24*(4), 437-463.

Prato, C. G., & Bekhor, S. (2007). Modeling route choice behavior: how relevant is the composition of choice set? *Transportation Research Record*, *2003*(1), 64-73. https://doi.org/10.1080/0144164042000181707

Ramming, M. S. (2002). *Network Knowledge and Route Choice* [Unpublished Doctoral thesis]. Massachusetts Institute of Technology.

Raymond, R., Morimura, T., Osogami, T., & Hirosue, N. (2012, November). Map matching with hidden Markov model on sampled road network. In *Proceedings of the 21st International Conference on Pattern Recognition*, 2242-2245. Tsukuba, Japan.

Rodríguez, D. A., Merlin, L., Prato, C. G., Conway, T. L., Cohen, D., Elder, J. P., Evenson, R. E., McKenzie, T. L., Pickrel, J. L., & Veblen-Mortenson, S. (2015). Influence of the built environment on pedestrian route choices of adolescent girls. *Environment and Behavior, 47*(4), 359-394.

Sevtsuk, A., Basu, R., Li, X., & Kalvo, R. (2021). A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco. *Travel Behaviour and Society*, *25*, 41-51. https://doi.org/10.1016/j.tbs.2021.05.010

Sevtsuk, A., & Kalvo, R. (2021). Predicting pedestrian flow along city streets: A comparison of route choice estimation approaches in downtown San Francisco. *International Journal of Sustainable Transportation*, *16*(3), 222-236. https://doi.org/10.1080/15568318.2020.1858377

Shatu, F., & Yigitcanlar, T. (2018). Development and validity of a virtual street walkability audit tool for pedestrian route choice analysis—SWATCH. *Journal of*

Transport Geography, *70*, 148-160. https://doi.org/10.1016/j.jtrangeo.2018.06.004

Shatu, F., Yigitcanlar, T., & Bunker, J. (2019). Shortest path distance vs. least directional change: Empirical testing of space syntax and geographic theories concerning pedestrian route choice behaviour. *Journal of Transport Geography*, *74*, 37-52. https://doi.org/10.1016/j.jtrangeo.2018.11.005

Sheffi, Y. (1985). *Urban Transportation Networks*. Prentice-Hall, Englewood Cliffs, NJ.

U.S. Census Bureau (2019). *2019 American Community Survey (ACS) 5-year estimates*. Retrieved from https://data.census.gov/table

Walk Score. (2023 April). *2021 City & Neighborhood Ranking*. https://www.walkscore.com/cities-and-neighborhoods/

Wickramasinghe, V., & Dissanayake, S. (2015, January 11-15). Pedestrians' Route Choice Behavior for Intra-CBD Trips in a Developing Country. In *Transportation Research Board 94th Annual Meeting*. Washington DC, United States. Retrieved from https://trid.trb.org/view.aspx?id=1337870

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2881-2890).