Journal of Transport Geography 77 (2019) 39–47

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Developing a pedestrian destination choice model using the stratified importance sampling method

Elmira Berjisian, Meeghat Habibian*

Department of Civil and Environmental Engineering, Amirkabir University of Technology, Hafez St., Tehran 15875-4413, Iran

ARTICLE INFO

Keywords:
Walking
Destination choice model
Choice set
Stratified importance sampling
Rasht

ABSTRACT

Pedestrian destination choice models are among the issues that have not been adequately addressed by transportation researchers. Several issues should be considered in order to develop a pedestrian destination choice model, including the method used to generate destination choice sets as well as the specification of different environmental variables that influence walking. Although some studies use built environment variables to elaborate on pedestrian destination choice models, almost all of them neglect the importance of the choice set generation method and rely merely on the random sampling method. This study illustrates the extent to which a choice set generation method can influence the parameter estimates of a pedestrian destination choice model. Here, in addition to the random sampling method, the method of stratified importance sampling is addressed to develop a pedestrian destination choice model for the city of Rasht, Iran. In addition, an exhaustive array of built environment variables that are known to have a significant effect on walking is employed. The results suggest that using stratified random sampling can improve the destination choice model in both goodness of fit and percent correct of the prediction of pedestrians' destination. The relative importance of environmental variables with respect to altering the pedestrians' destination choices is demonstrated through their elasticities' values. Two scenarios, which aim to improve the walking environment by enhancing network connectivity, are examined. As a result, the pedestrian catchment area is suggested to be the most effective variable of network connectivity. The results of these scenarios can assist policy makers to identify zones that deserve more attention.

1. Introduction

Walking can benefit the transportation system and its users in many ways such as mitigating congestion, providing access to public transit services, trivial cost of travel compared to motorized modes of transportation, and health benefits (Higgins, 2005; Lamíquiz and López-Domínguez, 2015; Lundberg and Weber, 2014; Morency et al., 2007; Schlossberg and Brown, 2004). Furthermore, contrary to the motorized modes of transportation, walking is an environment-friendly mode because it does not consume nonrenewable energy sources, which helps to reduce the amount of greenhouse gas emissions (Higgins, 2005; Morency et al., 2007). Consequently, recognizing the influential factors that make destinations more favorable for walking can be used in policy analysis to enhance walking as an urban mode of transportation in daily trips.

There is a general consensus about the effects of the built environment on walking. Different features of the built environment such as population density, network design, and land use patterns are shown to affect walking frequency or walking as a mode of choice (Dill, 2004; Ewing and Cervero, 2010; Frank et al., 2005; Gori et al., 2014; Schlossberg and Brown, 2004; Hatamzadeh et al., 2017a). On the other hand, few studies have incorporated the features of a built environment into pedestrian destination choice models (Clifton et al., 2016a).

The pedestrian destination choice model is a practical tool to distribute trips, and can thus be used to predict walking trip demand. This approach arms planners with robust tools to decide on different types of pedestrian investments such as pedestrian infrastructures. Its results can also serve as inputs for other pedestrian analyses regarding the health benefits of walking and transportation sustainability, and quantify the potential for mode shift, resulting in the reduction of greenhouse gas emissions (Clifton et al., 2016a). The pedestrian destination choice model that is sensitive to the features of a built environment can help decision makers to formulate policies favoring pedestrian-oriented environments.

There are inherent differences between choice behavior of motorized and non-motorized modes. One of influential factor in destination

* Corresponding author.
E-mail address: Habibian@aut.ac.ir (M. Habibian).

https://doi.org/10.1016/j.jtrangeo.2019.04.009
Received 29 November 2016; Received in revised form 1 September 2018; Accepted 16 April 2019
0966-6923/ © 2019 Elsevier Ltd. All rights reserved.

choice is distance which has different effects on individuals with different modes of travel. Indeed, non-motorized users are more sensitive to distance than motorized users as destination choice studies indicate (Richards and Ben-Akiva, 1974; Yang et al., 2009). The greater disinclination toward distance from non-motorized users is because longer distance is confounded by fatigue and physical fit that may result in greater coefficient of distance in destination choice studies for walking. This effect can be understood through a study by (Heilig et al., 2017) that shows trip length of less than one kilometer has opposite effect on utility of walking compared to public transit and car. Another influential factor in destination choice is travel time. The different weight of in-vehicle and out of vehicle time for transit and car trips shows the different effect of time on motorized and non-motorized part of travel (Richards and Ben-Akiva, 1974). Such innate differences indicate dissimilar choice behavior of users of different travel modes. In case of this study, walking is focused on to elaborate to what extent built-environment variables can influence walking destination choices.

In developing a destination choice model for pedestrians, various issues should be taken into consideration such as the different features of the built environment as well as the choice set generation method for model calibration. Although few studies have used built environment features for destination choice model calibration (Clifton et al., 2016a; Eash, 1999; Khan et al., 2014), to the best of the authors' knowledge, no study has examined the effects of different choice set generation methods on the pedestrian destination choice model. This paper contributes in two ways to enhance the knowledge of pedestrian behavior. It not only develops a pedestrian destination choice model that is sensitive to the built environment features, it also evaluates the effects of two different choice set generation methods on the estimations of a pedestrian destination choice model.

As per suggestions, the generation of choice set is of great importance since it influences the destination choice model estimations (Horni et al., 2010; Pellegrini et al., 1997). The universal destination choice set contains numerous alternatives, which makes model estimation difficult and expensive (Bierlaire and Robin, 2009; Clifton et al., 2016a; Lemp and Kockelman, 2012). Furthermore, an individual realistically considers only a limited portion of the universal choice set (i.e., real choice set) that is conceivable for him/her; therefore, a realistic model should ignore the destinations that are not in an individual's choice set. Additionally, different perceptions of available destinations result in different choice sets for individuals. Therefore, to be more realistic, an analyst should consider different choice sets for different individuals (de Dios Ortúzar and Willumsen, 1994). Although different studies have proposed several methods to generate destination choice sets (Ben-Akiva and Lerman, 1985; de Dios Ortúzar and Willumsen, 1994; Lemp and Kockelman, 2012; Pagliara and Timmermans, 2013), no study has focused on the generation of destination choice set for pedestrians.

It is noteworthy that this paper applies a pedestrian destination choice model to the city of Rasht, Iran. Rasht, as a developing city, is deprived of pedestrian infrastructures and its public transit system does not adequately meet the demands of travelers' transportation. Therefore, the application of a pedestrian destination choice model can give an insight into more walkable destinations for pedestrians and enable city planners to promote a walking environment.

This paper contributes to the literature on walking with its unique evaluation of the methods of choice set generation. Besides, it evaluates an extensive set of built environment variables that are addressed as influential factors regarding walking. Considering the dearth of literature on walking destination choice behavior (Clifton et al., 2016a, 2016b), this study sheds light on a more practical tool for choice set generation and utilizes it to explore the determinants of walking destination choice behavior.

The remainder of this paper consists of the following parts, respectively: First, brief literature on the choice set generation methods and pedestrian destination choice studies are presented. Then, the study area is introduced accompanied by a detailed description of the data. Next, the method and its application in the context of Rasht city are explained followed by its results. The paper concludes with a discussion of the applications of the pedestrian destination choice model developed here.

2. Literature

The methods adopted to generate the choice set affect the estimated parameters of the destination choice model as well as the predicted market share (Pagliara and Timmermans, 2013). To determine the available alternatives in a destination choice set, different methods of choice set generation are employed. These methods use either deterministic rules (de Dios Ortúzar and Willumsen, 1994; Pagliara and Timmermans, 2013) or sampling methods (Ben-Akiva and Lerman, 1985; Lemp and Kockelman, 2012). The deterministic rules allow an analyst to exclude unavailable alternatives from a choice set based on data. The availability of an alternative within a choice set can be determined by using data from the sample or by directly asking respondents (de Dios Ortúzar and Willumsen, 1994; Muhs et al., 2015; Pagliara and Timmermans, 2013).

Sampling methods propose a two-step process to estimate the choice probabilities. The first step is the generation of a choice set by defining a probability distribution function over all the possible choice sets, and in the second step, regarding the probability of each of the specific choice sets determined in the previous step, a probability of choosing each alternative is estimated. Sampling methods reduce the computational difficulty of choice probability estimation, especially in the case of large choice sets (Ben-Akiva and Lerman, 1985). Different approaches are adopted in sampling. While some studies randomly draw a sample because of less complexity, other studies use non-random sampling, which offers a more economical method as well as more precise estimations of parameters (Lemp and Kockelman, 2012; Ben-Akiva and Lerman, 1985; de Dios Ortúzar and Willumsen, 1994). The distinction between these approaches is that in random sampling, the alternatives are selected by uniform probabilities, while in non-random sampling, the alternatives are selected by non-uniform probabilities, which results from assigning different levels of importance to the alternatives based on an analyst's intuition (Ben-Akiva and Lerman, 1985). Ben-Akiva and Lerman developed an importance sampling through the non-random sampling approach. This approach results in more efficient choice sets that contain the alternatives that are likelier to be chosen (Ben-Akiva and Lerman, 1985). Lemp and Kockelman proposed the "strategic sampling scheme" through the concept of importance sampling. This approach confirms better estimations of the parameters on an average compared to random sampling (Lemp and Kockelman, 2012). To prevent the need for calculating the specific selection probability for each alternative, Ben-Akiva and Lerman introduced a method called stratified importance sampling. In this method, the universal set of choices is stratified into mutually exclusive and collectively exhaustive subsets. This importance-based stratification gives different selection probabilities for each strata, while uniform selection probabilities are retained within each strata (Ben-Akiva and Lerman, 1985).

Most pedestrian destination models take advantage of the easiest approach of sampling (i.e., random sampling). In the process of developing a destination choice model for pedestrians, Clifton et al. used choice sets that contained 10 random alternatives within a three-mile (4.8 km) network distance from each origin zone (three miles is equal to or greater than the length of > 99% of the observed walk trips) (Clifton et al., 2016a). In another study on non-motorized modes, Eash drew a sample that included 50 possible destinations within two miles (3.2 km) X-plus-Y distance from each origin zone, in addition to the real destination to form choice sets. It is worth noting that in his study, the majority of the studied trips occurred in < 3.2 km (Eash, 1999). Khan et al. developed a destination choice model for non-motorized trips by

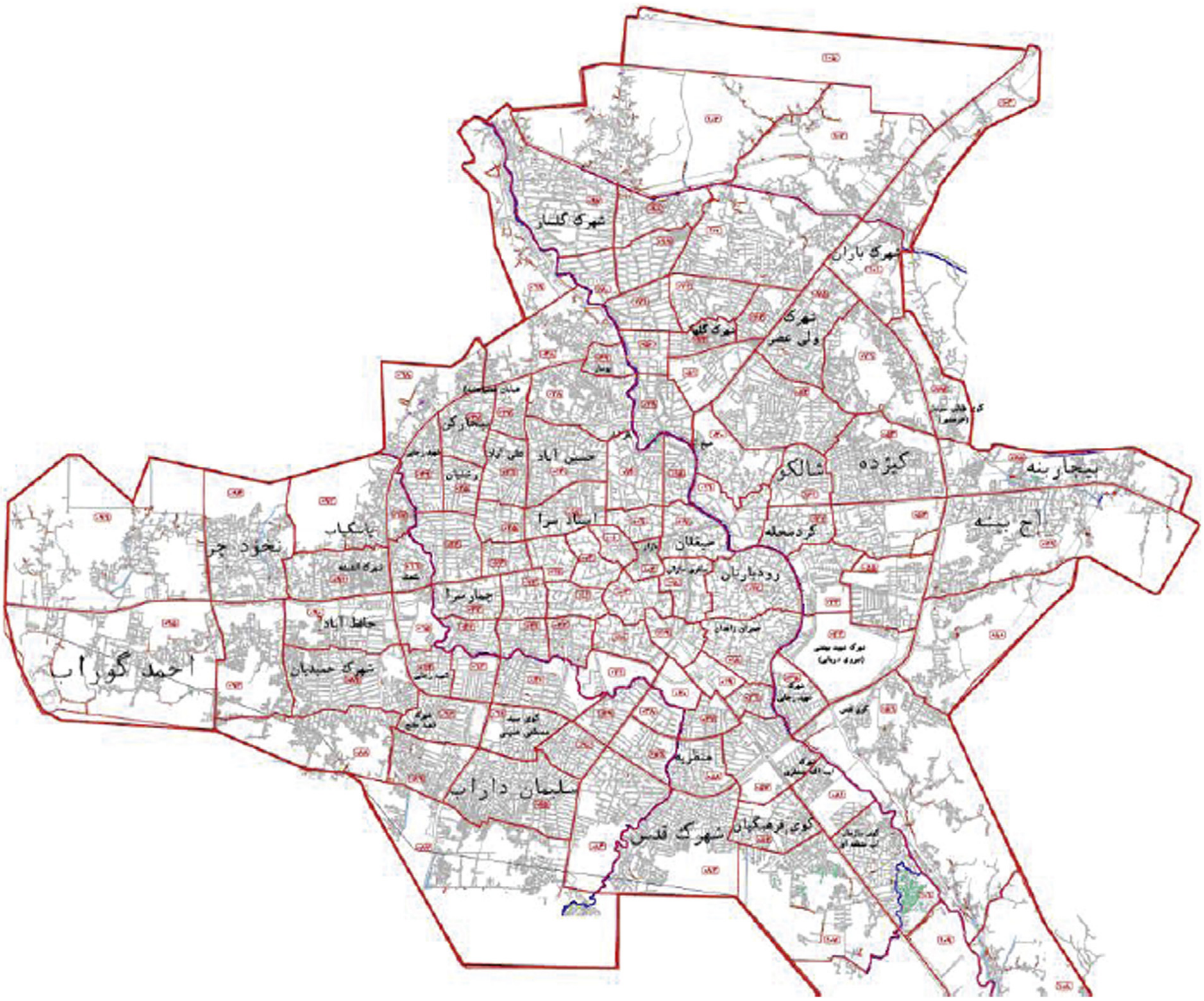

Fig. 1. The study area (Hatamzadeh et al., 2017a).

using random choice sets comprising 40 alternatives, including the real destination (Khan et al., 2014).

The above paragraphs show the different strategies used to generate choice sets in earlier studies. Furthermore, none of the aforementioned studies has delved into the possible effects of the pedestrian destination choice sets on model estimation. Therefore, the present study attempts to address this gap in the literature through the case of Rasht.

3. Study area

Rasht is the largest city on the Caspian Sea coast in Iran (Fig. 1). According to the 2006 census, its population exceeds 550,000 people. The prevailing transportation modes used in daily trips are cars, taxis, motorcycles, buses, and non-motorized modes (i.e., biking and walking) (Hatamzadeh et al., 2017a). It is noteworthy that the existing transit infrastructure cannot adequately meet the citizens' travel demands, and this city does not have any mass transit system. This leads to cars and taxis being the primary modes of transportation for meeting travel demands in the city (Hatamzadeh et al., 2014). The radial pattern of development in Rasht leads to a radial street pattern as observed in the majority of streets. Furthermore, the presence of the traditional bazaar and main retail centers has turned the city's central part into a congested area. Due to this congestion, sparse commercial activity has been emerging out of the city center through the years (Azimi, 2005).

4. Data

In this study, the data on walking trips are derived from a Rasht travel survey carried out in November 2007. In this survey, travel diaries were reported by 5000 resident households of Rasht. Here, the traffic analysis zone (TAZ) is selected as an analysis unit; the city of Rasht is divided into 112 TAZs. The exact locations of the origin and the destination of pedestrians were not available in the data, which led us to take the TAZs' centroid as the respondents' origin/destination. It is worth noting that the average of the TAZs' equivalent radius is 328 m (the equivalent radius is the square root of the quotient that is obtained by dividing the average TAZs' area by pi). This vindicates choosing the TAZ as an analysis unit in this research because people almost walk distances up to 400 m (Mavoa et al., 2012). Out of all the daily trips, roughly 33% comprises walking, which elucidates the significance of walking as one of the important transportation modes in Rasht.

To investigate the potential effects of the built environment on the destination choice behavior of pedestrians, various indices and variables are extracted from the literature. These variables primarily determine land use diversity and transportation network connectivity.

To understand the level of diversity of land uses within origin/destination zones, two indices-entropy (Ewing et al., 2014) and job-population balance (Ewing et al., 2014)-are calculated. Briefly, entropy determines the comparative distribution of different land uses in a zone, and it is calculated with regard to the percentage of different land uses within each TAZ (Ewing et al., 2014). Another land use index is the job-population balance. In short, it indicates the balance between the jobs and the population within a TAZ, and it is calculated with regard to the employment opportunities in a zone and its population (Ewing et al., 2014). Readers interested in further details about the two indices are referred to the mentioned references. As another influential factor on walking, the population density of each TAZ is calculated. Neighborhoods with greater density can generate more walking trips (Ewing et al., 2014).

The transportation network connectivity variables are found to influence walking. Several variables determine the level of network connectivity in the literature, including intersection density (Dill, 2004; Lundberg and Weber, 2014), percentage of four-way intersections (Southworth and Owens, 1993), density of cul-de-sacs (Schlossberg and Brown, 2004), the ratio of intersections per all the nodes in the network (Dill, 2004), links' density (Dill, 2004; Lundberg and Weber, 2014), and the ratio of minor links to major links (Dill, 2004). These variables demonstrate the network connectivity in terms of links and nodes. For the sake of brevity, those who are interested in further details of these variables could look at the mentioned references. To account for multicollinearity among these variables, this paper takes advantage of a connectivity index developed by Hatamzadeh et al. for the city of Rasht by using a principal component analysis (Hatamzadeh et al., 2017a). This index accounts for connectivity in terms of link and node connectivity factors. The link-connectivity factor consists of link density and the ratio of minor roads to major roads, and the node-connectivity factor contains a percentage of four-way intersections and the ratio of intersections to nodes. Readers interested in a more comprehensive explanation of this index are referred to the research of (Hatamzadeh et al., 2017a).

In order to get a sense of the road network circumstances that a pedestrian experiences during his/her travel, the Pedestrian Catchment Area (PCA) is calculated for all TAZs by ARC GIS 10.2. The PCA is derived by dividing the area of a polygon obtained by 400 m of walking, in terms of network distance, from the TAZ's centroid to every direction, by the area of the 400-m-radius Euclidean buffer with the same center (Schlossberg and Brown, 2004). To illustrate this, Fig. 2 shows the PCA for a given zone.

In addition to the previous variables, Southworth and Owens introduced a classification that reflected different network configurations. Other studies use such network configuration introduced in the mentioned study to investigate their effects on walking (Habibian and Hosseinzadeh, 2018; Hatamzadeh et al., 2017b; Sandalack et al., 2013; Schlossberg et al., 2015). These configurations fall into five patterns (Fig. 3). They established their classification of street pattern configuration on the idea that street pattern is an indicative of quality of a neighborhood in terms of devoted infrastructure, route options and street connectivity. Observing residential design over 50 years, Southworth and Owens concluded street patterns are transitioned from interconnected to more disconnected which translates into more loops and cul-de-sacs. Therefore, they used five typologies to delineate changes of street connectivity. They described "the Gridiron" pattern as the most interconnected pattern that suggests a higher choice set of routes. This pattern is accompanied by more intersections and access points and consequently more short trips. This is the pattern that brings about the most walkable neighborhood. Second to Gridiron pattern is “Fragmented Parallels” which is different from former pattern in several aspects. The equal sized grids turn into L-shapes blocks which decrease connectivity by T-intersections and L-shaped corners and, consequently, less route alternatives and access points. The third pattern “Warped Parallels” features with a curvilinear pattern and occasional cul-de-sacs. This pattern encourages automobile use by its decreased connectivity. “Loops and Lollipops” offers limited route choice and creates a user disoriented neighborhood. Frequent loops abstain pedestrians from accessing to anywhere except few choices such as houses and schools. This pattern increases motorized vehicle trips and directs traffic to few arterials which makes neighborhood not suitable from a pedestrian point of view. “Lollipops on Stick” which is on the opposite extreme to the “Gridiron” pattern features with dead-end loops and large blocks. As a result, interconnectivity is limited. Respectively, the level of connectivity decreases from the grid to the lollipops on a stick pattern. Highly connected networks are characterized by a grid pattern, which provides pedestrians with more route alternatives (Ewing et al., 2014; Ewing and Cervero, 2010).

The five different patterns are attributed to each TAZ based on the analyst's intuition of the network within each TAZ. As many as 44 TAZs out of 112 resemble the loops and lollipops pattern, which is the most observed pattern in Rasht. According to the classification by Southworth and Owens (Southworth and Owens, 1993), five dummy variables corresponding to each pattern are assigned to each TAZ to show the network pattern within the TAZ. That is, for each TAZ, one of the dummy variables reflecting the TAZ's network pattern equals 1 and the four remaining dummy variables are set as 0. Table 1 provides the descriptive statistics of the aforementioned variables.

5. Methodology

Discrete choice models are based on the random utility theory and have been used widely in the context of travelers' choice behavior (Ben-Akiva and Lerman, 1985). As per this approach, pedestrian n chooses destination j from the destination alternatives within a choice set C_n that she/he has. Each destination j gives a utility U_{jn} to a specific pedestrian n. In this theory, it is expected that pedestrian n chooses the destination with the highest utility. This utility function consists of a systematic part V_{jn}, and an error term ε_{jn}. Therefore, for each pedestrian:

$$U_{jn} = V_{jn} + \varepsilon_{jn} \tag{1}$$

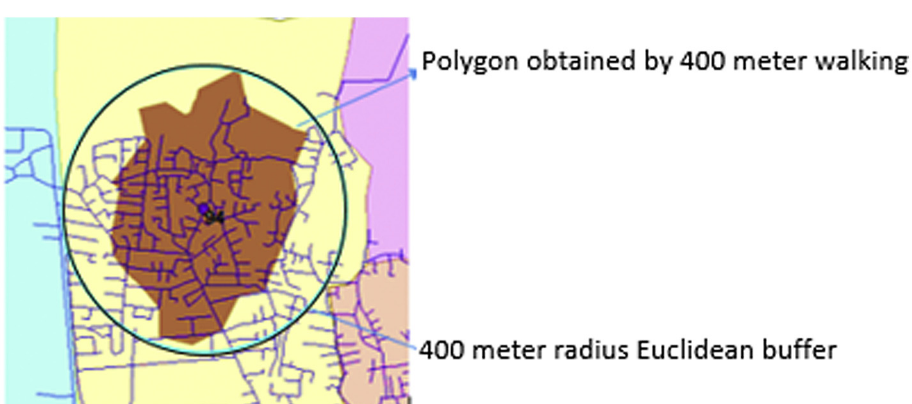

Fig. 2. PCA illustration.

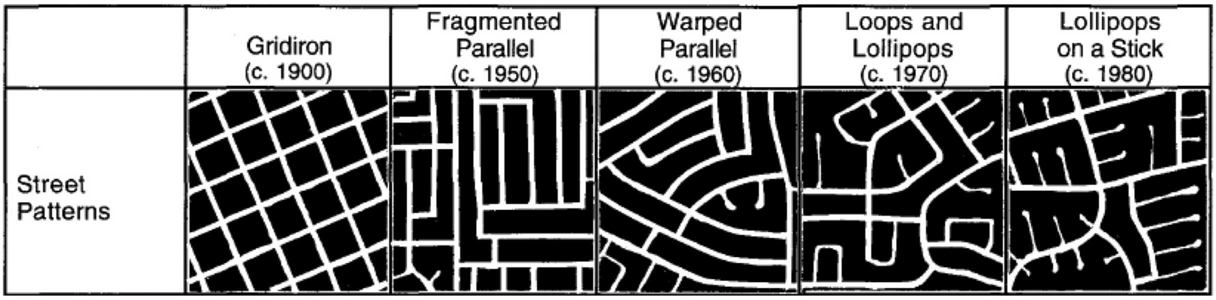

Fig. 3. Different network configuration (Southworth and Owens, 1993).

The systematic part, V_{jn}, explains the observed attributes of destination j for pedestrian n. The most often assumed form for systematic utility is linear function (Eq. (2)).

$$V_{jn} = \beta_{1jn}X_{1jn} + \ldots + \beta_{kjn}X_{kjn} \tag{2}$$

where built environment variables ($X_{1jn}, \ldots, X_{kjn}$) are weighted by parameters ($\beta_{1jn}, \ldots, \beta_{kjn}$) to explain built environment variables' marginal utility.

The error term, ε_{jn} accounts for the unobserved aspects of that specific destination for pedestrian n.

Therefore, the probability that pedestrian n selects destination j is:

$$p_n(j \mid C_n) = \Pr(U_{jn} > U_{in}; \forall\ i \epsilon C_n) \tag{3}$$

The different distributions for ε_{jn} result in various types of choice models. In the context of the destination choice model, the multinomial logit model (MNL) is applied to predict the destination choice behavior (Clifton et al., 2016a). Generally, the MNL model assumes that all the error terms, ε_{jn},are distributed independently and identically, and they follow a Gumble distribution (Ben-Akiva and Lerman, 1985). The MNL model uses the following probability function to predict pedestrian n choosing destination j:

$$p_n(j \mid C_n) = \frac{e^{V_{jn}}}{\sum_{i \in C_n} e^{V_{in}}} \tag{4}$$

Unlike the mode choice models in which the analyst faces small choice sets, the size of destination choice sets can be very large. As a result, the choice-sampling methods can offer more simplicity in estimating the destination choice model. Therefore, it is crucial to adopt a sampling method to develop the destination choice sets instead of using the universal choice set. Due to the bias caused by the sampling method, Eq. (4) should be adjusted. Eq. (5) shows the adjusted probability of choosing destination j by pedestrian n through a sampling method:

$$p_n(j \mid D_n) = \frac{e^{(V_{jn} + \ln(P(D_n \mid j))}}{\sum_{i \in D_n} e^{(V_{in} + \ln(P(D_n \mid i))}} \tag{5}$$

Here $P(D_n|j)$ is the probability of generating a choice set D_n, including the destination j by using a sampling method. If the adopted sampling method meets the positive conditioning property (i.e., $P(D_n|\, i) > 0;\ \forall\, i \epsilon D$), then McFadden shows a sampling method that culminates in consistent parameter estimates for the logit models (McFadden, 1978). It is noteworthy that if random sampling is adopted, the probability of generating any choice set becomes equal and the term $P(D_n|j)$ cancels out from Eq. (5), and thus no adjustment is needed. In this paper, in addition to the random sampling method that randomly draws a sample of 10 alternatives, a stratified importance sampling method is taken into account to enable us to compare the efficiency of the parameter estimates in both sampling methods. In this study, to apply the stratified importance sampling method, the universal choice set for each pedestrian, containing 112 TAZs, is divided into three non-overlapping strata: 1) the origin zone, 2) all zones located within the 90th percentile of cumulative walking trip length distribution (i.e., 1660 m) from the origin zone (but not including the origin zone itself), and 3) the remaining zones in the city of Rasht. Therefore, Srn is the number of choices in stratum r ($r = 1, 2, 3$) for pedestrian n, where $\sum_{r=1}^{3} S_{rn} = 112$. For each pedestrian, the generated choice set consists of all the zones of the first two and none from the third stratum in addition to the observed destination (if it is not included). To explicate this issue, let $\widetilde{S}_{rn}$ be the sample size of stratum r and r(j) be the stratum containing destination j. The stratified importance sampling method draws $\widetilde{S}_{r(j)n} - 1$ samples from the stratum containing destination j (i.e., r(j)) and $\widetilde{S}_{rn}$ samples from the remaining strata. The probability of selecting a choice set D_n including the destination j is then:

$$P(D_n \mid j) = \binom{S_{r(j)n} - 1}{\widetilde{S}_{r(j)n} - 1}^{-1} \prod_{\substack{r=1 \\ r \neq r(j)}}^{3} \binom{S_{rn}}{\widetilde{S}_{rn}}^{-1} \tag{6}$$

By substituting Eq. (6) for Eq. (5), the value of P_{jn} for each pedestrian can be predicted.

After preparing an appropriate dataset for each of the investigated sampling methods, the environmental variables for both endpoints of a trip (i.e., origin zone and potential destination zone) are calculated and used as per the specification of systematic utility.

6. Results

Before model estimation, a correlation test is performed to ensure that no intercorrelation exists between the explanatory variables.

Table 1
Descriptive statistics of built environment variables.

Variable		Average	Standard deviation	Unit
Connectivity	Percentage 4 way intersection	14.10	6.44	-
	Link density	0.22	0.09	km/km^2
	Minor roads to major roads ratio	11.57	25.63	-
	Intersections to total nodes ratio	0.62	0.10	-
	Pedestrian catchment area	0.35	0.13	-
Land use	Entropy	0.33	0.19	-
	Job-Population balance	0.56	0.29	-
Density	Population density	10100	6600	Person/km^2

Table 2
Results of destination choice models.

Variable	Random sampling		Stratified importance sampling	
	Coefficient	*t*-Statistics (p-value)	Coefficient	*t*-Statistics (*p*-value)
Distance	–	–	−0.002	−102.3 (0.0000)
Population density (at destination)	0.006	27.1 (0.0000)	0.003	9.8 (0.0000)
Entropy (at destination)	1.634	19.5 (0.0000)	1.025	11.3 (0.0000)
Job-population balance (at destination)	1.255	27.0 (0.0000)	1.168	21.6 (0.0000)
Link-connectivity factor (at destination)	0.063	5.13 (0.0000)	0.050	3.8 (0.0002)
Loops and lollipops pattern- dummy variable (at destination)	−0.172	−7.5 (0.0000)	−0.244	−10.8 (0.0000)
Pedestrian catchment area (at destination)	1.934	19.2 (0.0000)	1.323	12.5 (0.0000)
Number of observations	9937		9937	
Psuedo-R squared	0.067		0.239	

Therefore, a set of independent explanatory variables containing the environmental features of both the origin and the destination is used to estimate the pedestrian destination choice models based on each sampling method. After model calibration, only the significant variables remain in the final models. The results are consistent with intuition and also confirm previous research on walking. Table 2 shows the results of the pedestrian destination choice models through random and stratified importance sampling methods.

All the variables in Table 2 become significant at the 1% level. The results reveal that the distance deters pedestrians from walking to destinations, which is in line with other studies on walking behavior (Clifton et al., 2016b; Eash, 1999; Khan et al., 2014; Lundberg and Weber, 2014; Millward et al., 2013; Saelens and Handy, 2008). The greater population density at destinations render walking to such destination more likely which is in alignment with another pedestrian destination choice study (Clifton et al., 2016b). Furthermore, both models demonstrate the essential role of diverse land uses in encouraging pedestrians to walk to their intended destinations. This finding is in accordance with other studies that recognize land use diversity as a characteristic of a pedestrian-friendly environment (Clifton et al., 2016b; Khan et al., 2014; Peiravian et al., 2014). The results also show the positive influence of network connectivity on walking to a destination; however, as far as we know, the effect of network connectivity has not been addressed in previous studies on pedestrian destination choice models. It is noteworthy that other studies found a meaningful relationship between connectivity variables used in this study (represented by variables of link connectivity factor, pedestrian catchment area and loops and lollipops pattern) and overall walking behavior (Gori et al., 2014; Hatamzadeh et al., 2017b; Habibian and Hosseinzadeh, 2018).

The model that uses the stratified importance sampling method excels in terms of goodness of fit, which is indicated by the pseudo-R squared value. As pseudo-R squared can be used to compare the different specifications developed on the same data (Ben-Akiva and Lerman, 1985), the results suggest that the stratified importance sampling method may end with better goodness of fit compared to the random sampling method. Furthermore, it is suggested that pseudo-R squared between 0.2 and 0.4 can be compared to values between 0.7 and 0.9 of R-squared with respect to linear regression (Colombo et al., 2005). Besides, in other destination choice studies McFadden R-squared have comparable values (0.04–0.3) (Auld and Mohammadian, 2011; Bernardin Jr. et al., 2009; Huang and Levinson, 2015; Newman and Bernardin, 2010; Yang et al., 2009). Accordingly, the model that employs the stratified importance sampling method can be considered as a good-fit model.

To validate the models, the value of percent correct of each model is determined. To calculate the percent correctness, the model's predicted destination having the highest estimated probability of being selected is compared with the actual destination. If the predicted destination matches the actual destination, the model predicts correctly. The percent of correct predictions of the model using the random sampling method is 25.83%. The stratified importance sampling model correctly assigns destinations to 43.74% of all walking trips, which is twice better than the similar measure of its competitive method. Consequently, this superior estimation justifies the effort put into using the stratified importance sampling method.

To understand the extent to which the variables can change the probability of choosing the actual destination, the relative point elasticity for each variable in the model using stratified importance sampling is calculated. Elasticity gives a unitless measure to compare the importance of variables in terms of predicting the pedestrian's choice. According to its definition (Ben-Akiva and Lerman, 1985), elasticity measures the percentage of change in the probability of choosing an observed destination given 1% of change in a given variable. Fig. 4 shows the variables' elasticity using the method of "probability weighted sample enumeration" by which the elasticity for each pedestrian is calculated and weighted by the pedestrian-associated probability of choosing the observed destination. However, the calculation of elasticity is not meaningful for the dummy variable that takes 0 or 1 (i.e., loops and lollipops pattern).

Fig. 4 demonstrates that distance is the most influential variable that deters pedestrians from walking to their destinations; in other words, a 1% increase in walking distance corresponds to a 0.84% decrease in the probability of choosing the actual destination for walking. On the other hand, diverse land uses at the destination in the form of population density, entropy, and job-population balance surpass network connectivity. It is worth noting that among the network connectivity variables, the PCA index is the most influential. The PCA gives more detailed information about the network circumstances, including information about the relative number of links and their straightness. Therefore, this index gives more robust insights to policy makers, rather than merely using the number and type of links as offered by the link-connectivity factor.

To illustrate the results of the pedestrian destination choice model using stratified importance sampling, the value of the utility function of each TAZ is calculated according to its coefficients in Table 2. In this context, the distance variable for each TAZ is substituted by the average of that TAZ's distance to the other 111 TAZs, while the remaining variables are fixed to their respective values based on environmental data. It is expected that the greater the utility of a TAZ, the greater will be its attraction for pedestrians to choose it for walking. Fig. 5 depicts the normalized values of the 112 TAZs' utility within the city of Rasht.

The greatest extreme of utility values congregates in the center of the city. As previously mentioned, the city center is the location of the traditional bazaar, and provides the majority of potential opportunities to make a trip to. The centered zones are featured with high population density as well as radial pattern of the network, which supports high connectivity in the street network.

7. Application

Policy makers can take advantage of the influential environmental

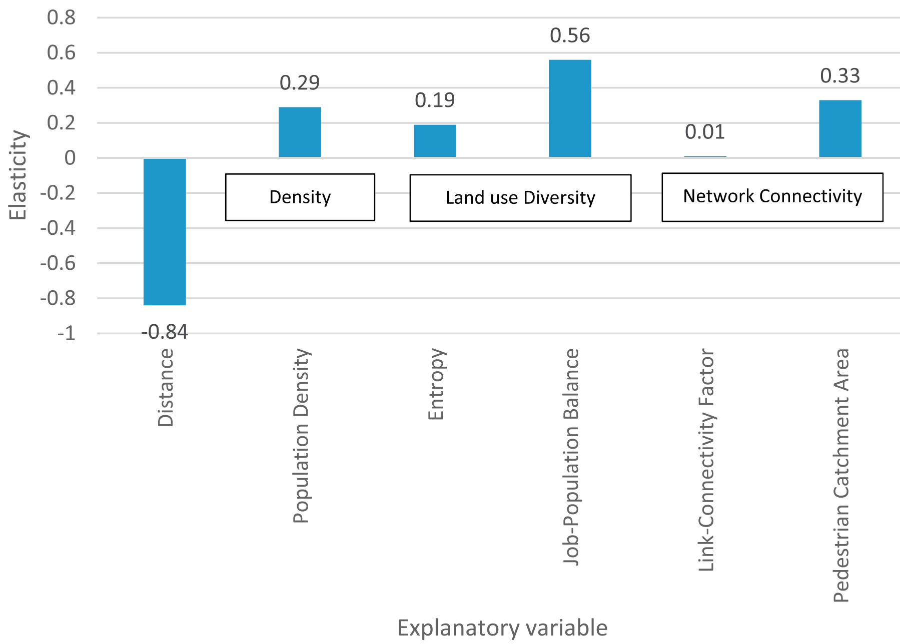

Fig. 4. Explanatory variables' elasticity.

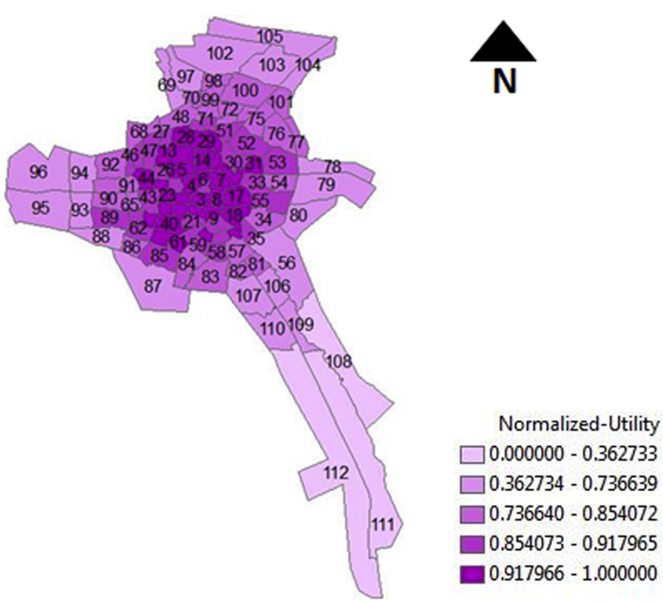

Fig. 5. Normalized utility function values spectrum.

variables in terms of improving a walking environment with the calibrated pedestrian destination choice model. According to the model that uses the stratified importance sampling method, while land use variables (e.g., population density, entropy, and job-population balance) are influential in choosing walking destinations, a transportation planner may be more interested in the distance and network connectivity variables (e.g., the PCA as well as the link-connectivity factor). Therefore, the variables that represent network connectivity can be subject to transportation policies. The PCA demonstrates the resemblance of a network to a well-connected network. A well-connected network allows pedestrians to walk to every direction, resulting in network distances to approach Euclidean distances. Another variable, the link-connectivity factor, represents the link density and the ratio of minor roads to major roads.

To illustrate the effect of policies that aim to improve the walking environment with network connectivity, the PCA and the link-connectivity factor are increased and their consequent effects are investigated, respectively. Each connectivity network variable is subject to a distinct scenario. Therefore, Scenario A corresponds to a 5% increase in the PCA and Scenario B refers to a 5% increase in the link-connectivity factor. Then, the corresponding changes in the destination choices are calculated by using the stratified importance sampling model.

By applying each scenario, the model predicts a new distribution of zones as walking destinations; that is, under each scenario, a pedestrian may change his/her choice of destination, resulting in some changes in the TAZs trip attraction. Fig. 6 demonstrates the results of implementing each scenario in contrast to the present condition. Therefore, the negative sign points to the destinations that lose attraction for some pedestrians, and on the other extreme; the positive sign indicates the destinations that attract more pedestrians.

Fig. 6 shows that Scenario A is more influential than Scenario B, thereby reflecting the stronger effect of PCA than the link-connectivity factor with respect to the 5% increase in their zonal values. This illustration is confirmed by the results of previous elasticities. As mentioned earlier, higher values of PCA correspond to free movement of pedestrians. Therefore, to encourage people to walk, urban planners should avoid designs with barriers to pedestrians. Barriers can be in the form of a road infrastructure, and one should be cautious about its location. To that end, urban planners pursuing new approaches such as pedestrian-oriented development and transit-oriented development should be aware of the PCA's role. It is worth noting that PCA is not just relevant to trips completely made on foot but there is also a similar concept in public transit. That is, PCAs can be measured for public transit stations. Higher PCA for public transit stations can endow the city with more sustainable modes of transportation.

Furthermore, Fig. 6 can assist policy makers to prioritize the zones that deserve more attention with regard to improving the walking environment. Fig. 6 can help to identify the zones that are most likely to be improved by increasing either the PCA or the link-connectivity factor. In other words, these zones are more receptive to walking improvement policies, and thus deserve to be investigated by policy makers. According to Fig. 6, one can observe that the zones located in the central part of the city are more receptive to walking improvement policies due to their higher utilities. These zones, which enjoy diversity of land uses, can be more attractive to pedestrians in the case of

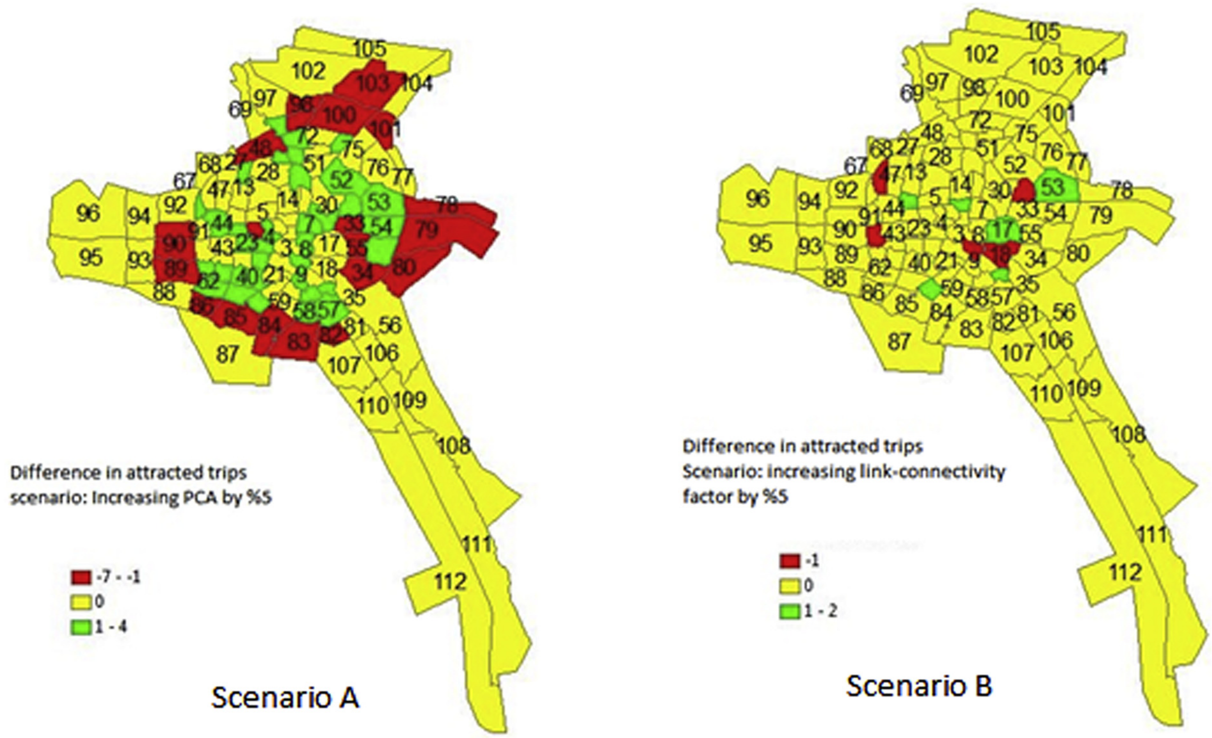

Fig. 6. Differences in attracted trips subject to Scenario A) increase pedestrian catchment area (PCA), and Scenario B) increase link-connectivity factor.

improving network connectivity variables.

In addition, these scenarios show the seemingly appropriate locations for implementing policies aimed at improving walking. As per the results, by enhancing the network connectivity among all zones, a limited number of zones will attract more walking trips. One can expect that the walking infrastructures would be more effective in such zones. Additionally, implementing such scenarios may increase the potential of mode shift in favor of walking, which can occur due to enhanced walking utility through the walking environment; however, this effect is not addressed here.

8. Conclusion

This paper attempts to illustrate the effects of different sampling methods for choice set generation on pedestrian destination choice models, which are not adequately addressed in the literature. This study elaborates on the different methods of sampling for choice set generation, namely random sampling and stratified importance sampling methods. These methods are used in order to simplify the burden of calculations, which arises out of large choice sets with respect to the destination choice models.

Many travel demand models that incorporate disaggregate choice models use a sequential structure that destination choice locates on upper level and mode choice models results affect them in forms of logsums (i.e., inclusive values) (a recent example can be found (Heilig et al., 2017)). This modeling structure makes comparison of destination choice behavior across travel modes difficult as it is assumed that individuals choose their destination regardless of what mode they use. Despite the more traditional model structure, in some cases individuals are more likely to change the destinations based on the mode they use (Newman and Bernardin, 2010).

Although random sampling is a more common method among pedestrian destination choice studies, the results suggest that the stratified importance sampling method leads to more efficient parameter estimates in terms of the pedestrian destination choice model. The percent correct of the model that benefits from the stratified importance sampling method is about twice better than its competitive model that uses the random sampling method. The model that benefits from the stratified importance sampling method can predict nearly half of the pedestrians' destinations correctly.

In addition, this paper takes advantage of a thorough set of built environment variables that are associated with walking. The significance of these variables in predicting pedestrians' destination choices may give a robust insight to transportation planners and policy makers looking for methods to encourage walking as a mode of transportation in daily trips. The results reveal that pedestrians tend to walk to destinations that have a variety of land uses as well as provide them with more route alternatives through a highly connected network. However, literature on pedestrian choice behavior is limited and this topic requires further attention (Clifton et al., 2016b), results of this study are in agreement with other studies in terms of the influence of distance, land use diversity and density on walking destination choice. Although, results of present study reveal that network connectivity can affect likelihood of destination choice for a pedestrian, other walking destination choice studies did not investigate such effect. It is recommended future studies consider network connectivity variables for better understanding of walking destination choice behavior.

The relevant elasticities of the explanatory variables indicate that distance is the most deterring factor to walking, and at the other extreme, diverse land uses at the destination can attract pedestrians. Due to the higher opportunities provided to pedestrians, they can choose the diverse land uses at destinations to fulfill various trip purposes. Additionally, the destinations featured by a highly connected network are found to attract more walking trips. It is worth noting that since network connectivity variables are more related to transportation planning, these variables are more interesting in terms of transportation policies.

Furthermore, the effects of the application of different scenarios aiming to improve the walking environment are also explored. The representative variables of the network connectivity condition in this study, namely the PCA and the link-connectivity factor, show increased support for a more pedestrian-oriented environment with two distinct scenarios, and the consequent changes in model predictions are investigated. The results confirm that the PCA can surpass the link-connectivity factor in providing a more favorable walking environment. Furthermore, a pattern of more favorable walking destinations is identified in the city's central part, which encompasses more diverse land uses. Such zones expectedly enjoy higher values of utility function. Therefore, policy makers can identify the most likely zones in which implementing walking improvement policies can promise an increase in walking. Less dependence on private cars due to the mode shift can be a promising side effect of improving the walking environment through such policies.

It is noteworthy that this study has taken into account only trips made solely on foot due to the available data. However, a majority of urban trips includes some walking known as walking legs of trips (as in access/egress walks to public transit stations, walking to/from parking). While a highly connected area can attract pedestrians, it is also appealing to public transit users who may use walking in the final leg of their trips. Therefore, a highly connected area, which can be promised through enhancing the PCA, can also attract public transit users. However, this issue is open to future research. The more facilitated walking to/from public transit stations can itself encourage people to use public transit, and simultaneously lead to less dependence on private cars, resulting in a beneficial modal shift.

References

Auld, J., Mohammadian, A., 2011. Planning-constrained destination choice in activity-based model: agent-based dynamic activity planning and travel scheduling. Transp. Res. Rec. J. Transp. Res. Board 170–179. https://doi.org/10.3141/2254-18.

Azimi, N., 2005. Restructuring urban morphology, a case study of Rasht in Iran. Geogr. Res. 53, 13–25.

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

Bernardin Jr., V., Koppelman, F., Boyce, D., 2009. Enhanced destination choice models incorporating agglomeration related to trip chaining while controlling for spatial competition. Transp. Res. Rec. J. Transp. Res. Board 143–151. https://doi.org/10.3141/2132-16.

Bierlaire, M., Robin, T., 2009. Pedestrians choices. Pedestr. Behav. 1–26. https://doi.org/10.1108/9781848557512-001.

Clifton, K.J., Singleton, P.A., Muhs, C.D., Schneider, R.J., 2016a. Representing pedestrian activity in travel demand models: framework and application. J. Transp. Geogr. 52, 111–122. https://doi.org/10.1016/j.tra.2016.09.017.

Clifton, K.J., Singleton, P.A., Muhs, C.D., Schneider, R.J., 2016b. Development of destination choice models for pedestrian travel. Transp. Res. Part A Policy Pract. 94, 255–265. https://doi.org/10.1016/j.tra.2016.09.017.

Colombo, S., Hanley, N., Calatrava-Requena, J., 2005. Designing policy for reducing the off-farm effects of soil erosion using choice experiments. J. Agric. Econ. 56, 81–95. https://doi.org/10.1111/j.1477-9552.2005.tb00123.x.

de Dios Ortúzar, J., Willumsen, L.G., 1994. Modeling Transport. Wiley New Jerseyhttps://doi.org/10.1002/9781119993308.

Dill, J., 2004. Measuring network connectivity for bicycling and walking. In: 83rd Annual Meeting of the Transportation Research Board, Washington, DC, pp. 11–15.

Eash, R., 1999. Destination and mode choice models for nonmotorized travel. Transp. Res. Rec. J. Transp. Res. Board 1–8. https://doi.org/10.3141/1674-01.

Ewing, R., Cervero, R., 2010. Travel and the built environment: a meta-analysis. J. Am. Plan. Assoc. 76, 265–294. https://doi.org/10.1080/01944361003766766.

Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., 2014. Varying influences of the built environment on household travel in 15 diverse regions of the United States. Urban Stud., 0042098014560991. https://doi.org/10.1177/0042098014560991.

Frank, L.D., Schmid, T.L., Sallis, J.F., Chapman, J., Saelens, B.E., 2005. Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. Am. J. Prev. Med. 28, 117–125. https://doi.org/10.1016/j.amepre.2004.11.001.

Gori, S., Nigro, M., Petrelli, M., 2014. Walkability indicators for pedestrian-friendly design. Transp. Res. Rec. J. Transp. Res. Board 38–45. https://doi.org/10.3141/2464-05.

Habibian, M., Hosseinzadeh, A., 2018. Walkability index across trip purposes. Sustain. Cities Soc. 42, 216–225. https://doi.org/10.1016/j.scs.2018.07.005.

Hatamzadeh, Y., Habibian, M., Khodaii, A., 2014. Walking behaviors by trip purposes. Transp. Res. Rec. J. Transp. Res. Board 118–125. https://doi.org/10.3141/2464-15.

Hatamzadeh, Y., Habibian, M., Khodaii, A., 2017a. Walking behavior across genders in school trips, a case study of Rasht, Iran. J. Transp. Heal. 5, 42–54. https://doi.org/10.1016/j.jth.2016.08.011.

Hatamzadeh, Y., Habibian, M., Khodaii, A., 2017b. Walking and jobs: a comparative analysis to explore factors influencing flexible and fixed schedule workers, a case study of Rasht, Iran. Sustain. Cities Soc. 31, 74–82. https://doi.org/10.1016/j.scs.2017.02.012.

Heilig, M., Mallig, N., Hilgert, T., Kagerbauer, M., Vortisch, P., 2017. Large-scale application of a combined destination and mode choice model estimated with mixed stated and revealed preference data. Transp. Res. Rec. J. Transp. Res. Board 31–40. https://doi.org/10.3141/2669-04.

Higgins, P.A.T., 2005. Exercise-based transportation reduces oil dependence, carbon emissions and obesity. Environ. Conserv. 32, 197–202.

Horni, A., Charypar, D., Axhausen, K.W., 2010. Empirically approaching destination choice set formation. Transport 25.

Huang, A., Levinson, D., 2015. Axis of travel: modeling non-work destination choice with GPS data. Transp. Res. Part C Emerg. Technol. 58, 208–223. https://doi.org/10.1016/j.trc.2015.03.022.

Khan, M., Kockelman, K.M., Xiong, X., 2014. Models for anticipating non-motorized travel choices, and the role of the built environment. Transp. Policy 35, 117–126. https://doi.org/10.1016/j.tranpol.2014.05.008.

Lamíquiz, P.J., López-Domínguez, J., 2015. Effects of built environment on walking at the neighborhood scale. A new role for street networks by modeling their configurational accessibility? Transp. Res. Part A Policy Pract. 74, 148–163. https://doi.org/10.1016/j.tra.2015.02.003.

Lemp, J.D., Kockelman, K.M., 2012. Strategic sampling for large choice sets in estimation and application. Transp. Res. Part A Policy Pract. 46, 602–613. https://doi.org/10.1016/j.tra.2011.11.004.

Lundberg, B., Weber, J., 2014. Non-motorized transport and university populations: an analysis of connectivity and network perceptions. J. Transp. Geogr. 39, 165–178. https://doi.org/10.1016/j.jtrangeo.2014.07.002.

Mavoa, S., Witten, K., McCreanor, T., O'Sullivan, D., 2012. GIS based destination accessibility via public transit and walking in Auckland, New Zealand. J. Transp. Geogr. 20, 15–22. https://doi.org/10.1016/j.jtrangeo.2011.10.001.

McFadden, D., 1978. Modeling the choice of residential location. Transp. Res. Rec. (673), 72–77.

Millward, H., Spinney, J., Scott, D., 2013. Active-transport walking behavior: destinations, durations, distances. J. Transp. Geogr. 28, 101–110. https://doi.org/10.1016/j.jtrangeo.2012.11.012.

Morency, C., Demers, M., Lapierre, L., 2007. How many steps do you have in reserve?: thoughts and measures about a healthier way to travel. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.3141/2002-01.

Muhs, C.D., Clifton, K., Singleton, P.A., Schneider, R.J., 2015. Development of a Pedestrian Demand Estimation Tool: A Destination Choice Model.

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

Pagliara, F., Timmermans, H., 2013. Choice set generation in spatial contexts: a review. Transp. Lett. https://doi.org/10.3328/TL.2009.01.03.181-196.

Peiravian, F., Derrible, S., Ijaz, F., 2014. Development and application of the Pedestrian Environment Index (PEI). J. Transp. Geogr. 39, 73–84. https://doi.org/10.1016/j.jtrangeo.2014.06.020.

Pellegrini, P.A., Fotheringham, A.S., Lin, G., 1997. An empirical evaluation of parameter sensitivity to choice set definition in shopping destination choice models. Pap. Reg. Sci. 76, 257–284. https://doi.org/10.1111/j.1435-5597.1997.tb00691.x.

Richards, M.G., Ben-Akiva, M., 1974. A simultaneous destination and mode choice model for shopping trips. Transportation (Amst). 3, 343–356. https://doi.org/10.1007/BF00167965.

Saelens, B.E., Handy, S.L., 2008. Built environment correlates of walking: a review. Med. Sci. Sports Exerc. 40, S550. https://doi.org/10.1249/MSS.0b013e31817c67a4.

Sandalack, B.A., Alaniz Uribe, F.G., Eshghzadeh Zanjani, A., Shiell, A., McCormack, G.R., Doyle-Baker, P.K., 2013. Neighborhood type and walkshed size. J. Urban. Int. Res. Placemaking Urban Sustain. 6, 236–255. https://doi.org/10.1080/17549175.2013.771694.

Schlossberg, M., Brown, N., 2004. Comparing transit-oriented development sites by walkability indicators. Transp. Res. Rec. J. Transp. Res. Board 34–42. https://doi.org/10.3141/1887-05.

Schlossberg, M., Johnson-Shelton, D., Evers, C., Moreno-Black, G., 2015. Refining the grain: using resident-based walkability audits to better understand walkable urban form. J. Urban. Int. Res. Placemaking Urban Sustain. 8, 260–278. https://doi.org/10.1080/17549175.2014.990915.

Southworth, M., Owens, P.M., 1993. The evolving metropolis: studies of community, neighborhood, and street form at the urban edge. J. Am. Plan. Assoc. 59, 271–287. https://doi.org/10.1080/01944369308975880.

Yang, M., Wang, W., Chen, X., Wang, W., Xu, R., Gu, T., 2009. Modeling destination choice behavior incorporating spatial factors, individual sociodemographics, and travel mode. J. Transp. Eng. 136, 800–810. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000140.