Road, Traffic, and Human Factors of Pedestrian Crossing Behavior

Integrated Choice and Latent Variables Models

Eleonora Papadimitriou, Sylvain Lassarre, George Yannis, and Dimitrios I. Tselentis

This study analyzed road, traffic, and human factors of pedestrian crossing behavior through the development of integrated choice and latent variables models. The analysis used recent research as a starting point, in which a two-stage approach was successfully tested, including a separate estimation of human factors and choice models. Data from a dedicated field survey were used: pedestrian field observations of road crossing behavior in different road and traffic scenarios were combined with a questionnaire on pedestrian attitudes, perceptions, motivations, and declared behaviors. The integrated choice and latent variables models were developed for four road types: major urban arterials, main roads, secondary roads, and residential roads. Results suggest that the effect of traffic conditions on pedestrian crossing choices was more important on main and secondary urban roads, whereas on major urban arterials and on residential roads it was nonsignificant. In regard to the effects of human factors, a risk latent variable was found to enhance the explanatory power of most of the models. This variable was estimated on the basis of different indicators in each case, reflecting a clear risk-taking tendency on major and main roads and an optimization tendency on minor roads. Overall, it is indicated that the integration of human factors in pedestrian crossing models provides meaningful and insightful results, and they may be advantageous compared with the two-stage approach.

Modeling pedestrian crossing behavior in urban areas has attracted the interest of many researchers during the past few decades, because it may assist in the better understanding of the interaction between pedestrians and the road and traffic environment, and of the way they balance the need for speed and comfort with the costs of risks and delays (*1–3*). Studying pedestrians' crossing behavior can eventually lead to the better design and management of urban road networks, to improve pedestrians' mobility and safety (*3*).

Signalized junctions provide a protected crossing phase for pedestrians. Nevertheless, it is often observed that pedestrians prefer to use the available traffic gaps for crossing, make midblock and diagonal crossings, and so forth (*4*). Pedestrians generally experience shorter delays than other road users because of their flexibility and adaptability, but the accident risk they are exposed to is higher (*5, 6*).

Road and traffic factors affecting pedestrian crossing decisions have been analyzed by means of gap acceptance models (*7, 8*); level-of-service approaches (*9, 10*); or discrete choice models (*4, 11, 12*). Another part of the related literature is focused on psychological, attitudinal, perceptual, and motivational factors (*13–15*). However, these human factors are rarely incorporated in pedestrian behavior models (*16*).

A first step for the combined analysis of road, traffic, and human factors of pedestrian behavior was presented in Papadimitriou et al. (*16*), where a two-step approach was implemented: first, human factors were calculated by means of principal component analysis on the responses of a questionnaire. Then, these factors were introduced as additional explanatory variables in crossing choice models based on field observations. This approach already provided some interesting results, but it has some known limitations, namely the fact that the error in the estimation of human factors is not taken into account (as these are separately estimated), and this may induce measurement errors in their effects as explanatory variables.

A more pertinent technique for analyzing human factors in discrete choice models is integrated choice and latent variables models (ICLV). ICLV models enhance the understanding of the choice process by merging classic choice models with the structural equation approach for latent variables, and they are a very promising method for capturing attitudes and perceptions of decision makers (*17, 18*). These models have been tested in the fields of transport economics, activity planning, and transport mode choice (*19–21*). However, they have not been used so far for the analysis of pedestrians' choices. This paper therefore presents a more sophisticated and appropriate methodology for the analysis of the data in Papadimitriou et al. (*16*).

More specifically, the objective of this paper is to develop choice models of pedestrian crossing behavior, integrating the effect of human factors (i.e., pedestrian attitudes, perceptions, motivations, and behavior) together with road and traffic factors. Also more specifically, the paper aims to further analyze data from the above mentioned dedicated survey, combining field observations of pedestrian trajectories and a questionnaire on pedestrian human factors, to develop ICLV models of pedestrian crossing behavior (*16*).

Key research parameters are road type, traffic control, traffic volume, and pedestrian demographics, as well as pedestrian risk-taking attitudes and perceptions, walking motivations, opinion on drivers, and so forth. For detailed research hypotheses, please see Papadimitriou et al. (*16*).

E. Papadimitriou, G. Yannis, and D.I. Tselentis, Department of Transportation Planning and Engineering, National Technical University of Athens, 5 Heroon Polytechniou, GR-15773 Athens, Greece. S. Lassarre, Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux, 14-20 Boulevard Newton, Cité Descartes, Champs-sur-Marne, F-77447 Marne Valley Cedex 2, France. Corresponding author: E. Papadimitriou, nopapadi@central.ntua.gr.

Transportation Research Record: Journal of the Transportation Research Board, No. 2586, Transportation Research Board, Washington, D.C., 2016, pp. 28–38.
DOI: 10.3141/2586-04

DATA COLLECTION

In this research, a particular data collection scheme was implemented [Papadimitriou et al. (*16*)]. Crossing behavior at urban pedestrian trips was recorded along with the conditions of the traffic and road environment. Attitudes, perceptions, and behavior with regard to road crossing and accident risk were captured using a questionnaire.

Field Survey Design

The field survey designed and implemented in the present research comprises three walking conditions and eight crossing scenarios. Survey participants were asked to take a trip in the Athens city center, Greece, from Kolonaki square to Evangelismos metro station and back, according to predefined routes presented on the map (Figure 1). The eight survey scenarios were developed so that the choice sets for crossings can be clearly defined; only one crossing of interest will occur for each one of the scenarios, referred to as a primary crossing (*3, 12*).

All types of traffic conditions (free flow to congestion) are encountered during the day for the major urban arterial and the main urban road, while for the minor and residential roads, low to moderate traffic is mainly encountered throughout the day. No major variation of traffic is observed during the day in that area. The survey took place during weekday morning and afternoon hours, with daylight, good weather, fairly constant traffic conditions, and no congestion recorded.

The number of road links for each scenario and the geometric and traffic control characteristics of the roads are summarized in Table 1.

Questionnaire Design

A questionnaire was developed on the basis of several questionnaires from the existing literature (*13, 15, 22, 23*), and it was structured as a list of items to be rated on the basis of 5-point Likert scales (always/never or agree/disagree scales). The questionnaire includes five sections, as shown in Figure 2.

(a)

(b)

FIGURE 1 Crossing scenarios on survey site map.

TABLE 1 Road Type and Geometric and Traffic Control Characteristics of Survey Scenarios

Route	Scenario	Street Name	Road Type	Directions	Lanes	Separation	Traffic Signals	Roadside Parking	Number of Links
From Kolonaki to Evangelismos	1	Patr.Ioakeim	Main	2	2	No	Yes	No	4
	2	Ploutarchou	Secondary	2	2	No	No	Yes	4
	3	Vas.Sofias	Major	2	6	Yes	Yes	No	2
From Evangelismos to Kolonaki	4	Vas.Sofias	Major	2	6	Yes	Yes	No	2
	5	Ploutarchou	Secondary	2	2	No	No	Yes	2
	6	Karneadou	Minor	1	1	No	No	Yes	2
	7	Irodotou	Minor	1	1	No	No	Yes	2
	8	Patr.Ioakeim	Main	2	2	No	Yes	No	2

B	**How many times per week do you travel by each one of the following modes*:**
B1_i	Public transport (metro, bus, trolley bus, tramway)
B1_ii	Pedestrian
B1_iii	Passenger car (driver or passenger)
	Last week, how many kilometers did you travel by each one of the following modes:**
B2_i	Passenger car (driver or passenger)
B2_ii	Pedestrian
B2_iii	Public transport (metro, bus, trolley bus, tramway)
	As a pedestrian, how much would you agree with each one of the following statements*:**
B3_i	I walk for the pleasure of it
B3_ii	I walk because it is healthy
B3_iii	In short trips, I prefer to walk
B3_iv	I prefer taking public transportation (buses, metro, tramway, etc.) rather than my car
B3_v	I walk because I have no other choice
C	**As a pedestrian, how much would you agree with each one of the following statements***:**
C1_i	Crossing roads is difficult
C1_ii	Crossing roads outside designated locations increases the risk of accident
C1_iii	Crossing roads outside designated locations is wrong
C1_iv	Crossing roads outside designated locations saves time
C1_v	Crossing roads outside designated locations is acceptable because other people do it
C2_i	I prefer routes with signalized crosswalks
C2_ii	I try to make as few road crossings as possible
C2_iii	I try to take the most direct route to my destination
C2_iv	I prefer to cross diagonally
C2_v	I try to take the route with least traffic to my destination
C2_vi	I am willing to make a detour to find a protected crossing
C2_vii	I am willing to take any opportunity to cross
C2_viii	I am willing to make dangerous actions as a pedestrian to save time
D	**Compared to other pedestrians, how much do you agree that***:**
D1_i	I am less likely to be involved in a road crash than other pedestrians
D1_ii	I am faster than other pedestrians
D1_iii	I am more careful than other pedestrians
E	**As a pedestrian, how often do you adopt each one of the following behaviors****:**
E1_i	I cross diagonally
E1_ii	I cross at midblock at major urban arterials
E1_iii	I cross at midblock at urban roads
E1_iv	I cross at midblock in residential areas
E1_v	I cross at midblock when I am in a hurry
E1_vi	I cross at midblock when there is no oncoming traffic
E1_vii	I cross at midblock when I see other people do it
E1_viii	I cross at midblock when my company prompts me to do it
E1_ix	I prompt my company to cross at midblock
E1_x	I cross at midblock when there is a shop I like on the other side
E1_xi	I cross even though the pedestrian light is red
E1_xii	I walk on the pavement rather than on the sidewalk
E2_i	I cross between vehicles stopped on the roadway in traffic jams
E2_ii	I cross without paying attention to traffic
E2_iii	I am absent-minded while walking
E2_iv	I cross while talking on my cell phone or listing to music on my headphones
E2_v	I cross even though obstacles (parked vehicles, buildings, trees, etc.) obstruct visibility
E2_vi	I cross even though there are oncoming vehicles

FIGURE 2 Survey questionnaire.

(continued)

F	As a pedestrian, how much would you agree with each one of the following statements***:
F1_i	Drivers are not respectful to pedestrians
F1_ii	Drivers drive too fast
F1_iii	Drivers are aggressive and careless
F1_iv	Drivers should always give way to pedestrians
F1_v	When there is an accident, it is the driver's fault most of the time
F1_vi	I let a car go by, even if I have right-of-way

* (1 = never; 2 = less than once a week; 3 = once a week; 4 = more than once a week; 5 = every day)
** (1 = 1–2 km; 2 = 3–5 km; 3 = 5–20 km; 4 = 20–50 km; 5 = >50 km)
*** (1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree)
**** (1 = never; 2 = rarely; 3 = sometimes; 4 = often; 5 = always)

FIGURE 2 *(continued)* **Survey questionnaire.**

Survey Procedure

The data collection took place in the period July to December 2013, with 75 participants in total. Participants were students of the National Technical University of Athens and young professionals. Fifty-three percent of the survey participants were males, 50% of the participants were 18 to 24 years old, 27% were 25 to 34 years old, 20% were 35 to 45 years old, and 3% were more than 45 years old.

Half of the participants carried out the field experiment after filling in the questionnaire, and half of the participants first carried out the field experiment and then filled in the questionnaire; in this way, the survey was counterbalanced, to minimize the bias of participants possibly adapting their declared behavior to their observed behavior and vice versa. The observed crossing behavior of participants does not appear to be affected by the order of the tasks. Table 2 shows a comparison of observed midblock crossings for the eight scenarios, and in total—between the participants who filled the questionnaire before the walking task and those who filled it after—no significant differences are noticed.

Participants were informed about the purpose of the experiment and the fact that they would be followed along this trip by a researcher who would be unobtrusively recording their behavior. This allows for control over the experiment design (e.g., specific route and scenarios to be examined) and for a larger amount of questionnaire data to be collected. It also complies with privacy protection and informed consent needs. However, limitations exist, for it is possible that participants may alter their behavior if they know that they are being observed. The fact that participants did declare and actually performed risk-taking and noncompliant crossing behaviors within the survey indicates that the degree to which they may have altered their behavior is small.

Once the participant started the trip, a trained researcher followed him or her at a distance of approximately 35 m, to have a sufficient view of the participant and remain unobtrusive, and recorded data on each road link by filling-in a predefined form. For the walking speed data, the researcher recorded the distance walked and the time taken to walk for each road link (from one junction to the other) of the trip. For the traffic volume data, the researcher assessed the traffic conditions on each road link as "empty," low traffic, high traffic or congestion; for the distinction between low and high traffic in particular, an approximate criterion of vehicle headways higher or lower than 3 s was used.

TABLE 2 Comparison of Observed Crossing Behavior for Filling in Questionnaire Before or After Walking Task

	Observed Share of Crossings (%)		
Scenario	Midblock	Junction	Total
All road types	30.0	70.0	100.0
Before	28.7	71.3	100.0
After	31.5	68.5	100.0
Main roads			
Before	23.1	76.9	100.0
After	27.3	72.7	100.0
Secondary roads			
Before	46.8	53.2	100.0
After	52.4	47.6	100.0
Major roads			
Before	5.1	94.9	100.0
After	3.0	97.0	100.0
Minor roads			
Before	40.0	60.0	100.0
After	40.0	60.0	100.0

ANALYSIS METHODOLOGY

A probabilistic discrete choice is involved in determining the location of each primary crossing from the alternatives of the examined scenario (choice set). Previous research has shown that a sequential choice behavior appears to be the optimal assumption for pedestrian crossing choices (*12, 16*). This sequential choice process involves a decision on each road link of the choice set: crossing at midblock, crossing at junction, or no crossing (Figure 3). If no crossing takes place on a given road link, the same choice set is examined on the next road link, and so on, until a primary crossing is made. Therefore, the rest of the choice set (i.e., the subsequent road links for this scenario) is not considered.

INTEGRATED CHOICE AND LATENT VARIABLES MODELS

In an ICLV model, the discrete choice model includes latent variables that capture attitudes and perceptions of the pedestrians. The latent variable model is composed of a group of structural equations describing the latent variables as a function of observable exogenous variables and a group of measurement equations linking the latent variables to the observable indicators. The key feature of the

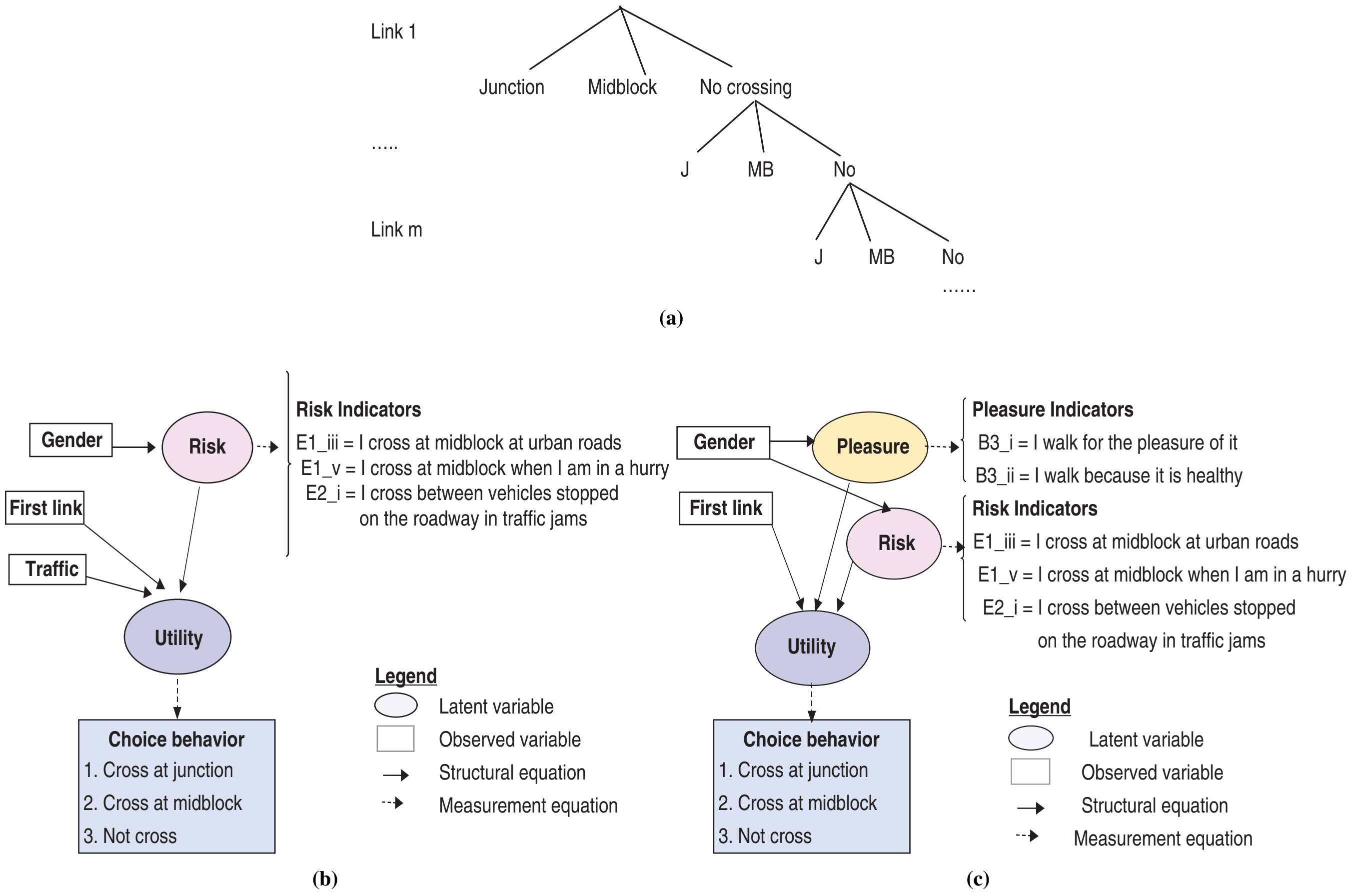

FIGURE 3 Main urban roads: (*a*) sequential logit model of pedestrian crossing behavior, (*b*) ICLV model with latent variable risk, and (*c*) ICLV model with latent variables and risk and pleasure (MB = midblock; J = crossing at junction; no = no crossing).

proposed modeling framework is that the latent variables can be calculated from the observable variables once the model parameters are estimated (integration).

The equations of the ICLV model follow, in the simple case of a binary choice model (choice alternatives *i* and *j*) with two latent variables (Z_1 and Z_2), each one measured by two observed variables (I_1, I_2, I_3, I_4) (*19*):

Structural equations of the ICLV model:

$$U_{in} = b'X_{in} + b_1\tilde{Z}_{1n} + b_2\tilde{Z}_{2n} + \varepsilon_{in}$$

$$U_{jn} = b'X_{jn} + \varepsilon_{jn}$$

Measurement equation of the ICLV model:

$$y_n = \begin{cases} 1 & \text{if } U_{in} > U_{jn} \\ 0 & \text{otherwise} \end{cases}$$

Structural equations of the latent variables model:

$$Z_{1n} = \alpha_1 W_{1n} + \omega_{1n}$$

$$Z_{2n} = \alpha_2 W_{2n} + \omega_{2n}$$

Measurement equations of the latent variables model:

$$I_{1n} = \lambda_1 Z_{1n} + \upsilon_{1n}$$

$$I_{2n} = \lambda_2 Z_{1n} + \upsilon_{2n}$$

$$I_{3n} = \lambda_3 Z_{2n} + \upsilon_{3n}$$

$$I_{4n} = \lambda_4 Z_{2n} + \upsilon_{4n}$$

where

U_{in}, U_{jn} = utility of each alternative, respectively, for individual *n*;
X_{in}, X_{jn} = sets of observed variables;
Z_{1n}, Z_{2n} = latent variables (actually the components accounting for most of the variability of the respective latent variables);
I_{1n}, I_{2n}, I_{3n}, I_{4n} = sets of indicators of the latent variables Z_{1n}, Z_{2n}, respectively;
$\tilde{Z}_{1n}$, $\tilde{Z}_{2n}$ = fitted values of the latent variables, once they are estimated by the structural equations of the latent variable model;

W_{1n}, W_{2n} = sets of observed variables (characteristics of respondent n);
$\varepsilon_{in}, \varepsilon_{jn}$ = extreme value distributed errors;
$\omega_{1n}, \omega_{2n}, \upsilon_{1n}, \upsilon_{2n}, \upsilon_{3n}, \upsilon_{4n}$ = sets of (multivariate normally distributed) errors; and
$b', b_1, b_2, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4$ = sets of unknown parameters to be estimated.

The measurement equations indicators I_{in} in the present research are discrete ordered, as the pedestrians were asked to respond on a 5-point Likert scale, ranging from never to always, or from strongly disagree to strongly agree. The cumulative logit link is used in ordered models. The notation γ_{ij} refers to cumulative probabilities, while π_{ij} designates ordinary probabilities. Formally, cumulative probabilities are defined as

$$\gamma_{ij} = \mathbf{Prob}\left(y_{ij} \geq \mathbf{i}\right) = \sum_{i}^{I} \pi_{ij}$$

where i is the rank of the response category in question. The measurement equation for the latent variable model can therefore be defined as

$$I_i = \log\left(\frac{y_i}{1-y_i}\right) = \log\left(\frac{\Pr\left(y_i \geq i\right)}{\Pr\left(y_i < i\right)}\right) = \lambda_i * Z_{in} + \upsilon_{in}$$

RESULTS

Exploratory Analysis

The model development early indicated that a global model for all scenarios was unfeasible, owing to estimation convergence problems. A review of experiences with such errors in ICLV models revealed that simplifying the model would be the recommended option. Two approaches were found to resolve the estimation problems.

One is that the indicators were recoded in a 3-point Likert scale, by adding up the responses 1 to 2 and 4 to 5, resulting thus in the following coding for the questionnaire variables: 1 = never or rarely; 2 = sometimes; and 3 = often or always. In this way, the number of parameters λ and γ to be estimated, and therefore the degrees of freedom of the models, were substantially reduced; this results in a loss of detail, but may allow for a more robust model.

The other is that the different scenarios were tested separately, because it was indicated that the different road types and sizes of the choice sets made the estimation more difficult.

The results of the two-stage approach were used as a starting point (*16*); a separate estimation of latent variables on this questionnaire indicated two principal components of pedestrian behavior. One related to risk taking (e.g., questionnaire items "I cross diagonally," "I am willing to take any opportunity to cross," "I cross at midblock in urban roads," etc.); and the other related to pleasure from walking (e.g., questionnaire items "I walk for health," "I walk for the pleasure of it," etc.).

The models were developed with the Biogeme v2.3 statistical software (*24*).

Models for Main Urban Roads

This scenario was the first one tested, because pedestrians' choice process seems to present the most variability in this type of road. The best-performing model was the one shown in Figure 3*b*, with indicators E1_iii, E1_v and E2_i forming the latent variable risk, and with pedestrian gender being the risk predictor in the structural equation. More specifically, the model specification is as follows:

$$\text{Structural model of the latent variablerisk} = \text{b\_gender} * \text{gender} + \omega$$

Measurement equations, ordered logit:

$$\text{I\_E1\_iii} = \lambda_1 * \text{risk} + \text{u}_1$$

$$\text{I\_E1\_v} = \lambda_2 * \text{risk} + \text{u}_2$$

$$\text{I\_E2\_i} = \lambda_3 * \text{risk} + \text{u}_3$$

where

$\text{I\_E1\_iii}_1 = 1/[1 + \exp(\lambda_1 * \text{risk} - \gamma_{11})]$;
$\text{I\_E1\_iii}_2 = 1/[1 + \exp(\lambda_1 * \text{risk} - \gamma_{12})] - 1/[1 + \exp(\lambda_1 * \text{risk} - \gamma_{11})]$;
$\text{I\_E1\_iii}_3 = 1 - 1/(1 + \exp(\lambda_1 * \text{risk} - \gamma_{12})]$;
$\text{I\_E1\_v}_1 = 1/[1 + \exp(\lambda_2 * \text{risk} - \gamma_{21})]$;
$\text{I\_E1\_v}_2 = 1/[1 + \exp(\lambda_2 * \text{risk} - \gamma_{22})] - 1/(1 + \exp(\lambda_2 * \text{risk} - \gamma_{21}))]$;
$\text{I\_E1\_v}_3 = 1 - 1/[1 + \exp(\lambda_2 * \text{risk} - \gamma_{22})]$;
$\text{I\_E2\_i}_1 = 1/[1 + \exp(\lambda_3 * \text{risk} - \gamma_{31})]$;
$\text{I\_E2\_i}_2 = 1/[1 + \exp(\lambda_3 * \text{risk} - \gamma_{32})] - 1/(1 + \exp(\lambda_3 * \text{risk} - \gamma_{31})]$; and
$\text{I\_E2\_i}_3 = 1 - 1/[1 + \exp(\lambda_3 * \text{risk} - \gamma_{32})]$.

Choice utility functions:

$$V1 = \text{ASC1} + \text{B\_first} * \text{first} + \text{B\_trafficlow} * \text{trafficlow} + \text{B\_risk} * \text{risk}$$

$$V2 = \text{ASC2} + \text{B\_first} * \text{first}$$

$$V3 = \text{ASC3}$$

where

$V1$ = utility of crossing at midblock,
$V2$ = utility of crossing at junction,
$V3$ = utility of not crossing,
ASC = alternative-specific constants,
trafficlow = low traffic conditions, and
first = first road link of the choice set.

The modeling results are presented in Tables 3 and 4. They can be summarized as follows:

- Parameters λ (λ_1, λ_2, and λ_3) are all statistically significant and positive, indicating that pedestrians with higher scores on these indicators, that is, having reported more frequently the respective risk-taking behavior, have higher risk. More specifically, the latent variable is expressed by the following behaviors:
  - Pedestrians who cross at midblock on urban roads,
  - Pedestrians who cross at midblock when in a hurry, and
  - Pedestrians who cross at midblock between stopped vehicles at congestion.

In this case, the self-reported behavior is matched with the observed behavior. Moreover, crossing of pedestrians between vehicles when

TABLE 3 Estimation Report and Goodness of Fit of ICLV Models for Main Urban Roads

Estimation Estimate	Latent Variable: Risk	Latent Variables: Risk and Pleasure
Number of parameters	15	22
Number of crossings	184	184
Initial log likelihood	−986.670	−3.070.240
Final log likelihood	−607.361	−1.854.286
Likelihood ratio test	758.618	2.431.908
ρ	.384	.396
$\bar{\rho}$ for initial model	.369	.389
Diagnostic	Convergence reached	Convergence reached
Iterations	36	54

the road is congested is a pattern likely to be observed on main urban roads.

- Pedestrian gender is a significant predictor of the latent variable risk (b_gender). Male pedestrians are found to be more risk taking.
- Pedestrians with higher risk taking (B_risk) appear to be more likely to cross at midblock at main urban roads; however, the effect is not statistically significant at 90%.
- The first road link (B_first) was found to have higher probability of being chosen.
- When traffic is low (B_trafficlow), midblock crossing probability increases.

An attempt was made, owing to the previous results of the two-stage model (*16*) for the development of a model with two latent variables: risk and pleasure. The best-performing model of this combination is shown in Figure 3*c*, where indicators B3_i and B3_ii were used for latent variable pleasure and E1_iii, E1_v, and E2_i for risk, with pedestrian gender being the risk predictor for both latent variables in the structural equations.

The presence of the latent variable pleasure seems to improve the significance of the latent variable risk, and the model overall (see Tables 3 and 4). Nevertheless, the latent variable pleasure was not found significant.

Models for Major Urban Arterials

The structure of the ICLV model developed for major roads is summarized in Figure 4*a*. For the risk indicators, a few were marginally significant, such as C2_vii, E1_iv, and E1_v. Modeling results are presented in Tables 5 and 6. They can be summarized as follows:

- Parameters λ (λ_1, λ_2, and λ_3) are all statistically significant and positive, indicating that pedestrians with higher scores on

TABLE 4 Parameter Estimates of ICLV Models for Main Urban Roads

	Latent Variable: Risk			Latent Variable: Risk and Pleasure		
Name	Value	*t*-Test	*p*-Value	Value	*t*-Test	*p*-Value
ASC1	−2.74	−7.40	.00	−2.52	−9.44	.00
ASC2	−1.33	−5.13	.00	−1.31	−7.23	.00
B_first	0.466	1.33	.18	0.427	1.73	.08
B1_trafficlow	1.54	2.56	.01	NA	NA	NA
B_risk	0.342	1.13	.26	−0.410	−1.38	.17
B_pleasure	NA	NA	NA	−0.248	−0.65	.52
b_gender	−0.550	−3.12	.00	0.538	4.88	.00
b2_gender	NA	NA	NA	−0.375	−2.59	.01
λ_1	2.78	3.63	.00	−1.34	−5.58	.00
γ_{11}	−2.04	−3.71	.00	−2.34	−11.31	.00
γ_{12}	1.77	3.52	.00	−0.586	−4.14	.00
λ_2	3.97	2.20	.03	−1.89	−6.17	.00
γ_{21}	−5.92	−2.46	.01	−2.73	−11.24	.00
γ_{22}	−1.97	−1.94	.05	−0.713	−4.42	.00
λ_3	1.38	5.06	.00	−1.57	−5.86	.00
γ_{31}	−2.80	−7.57	.00	−1.00	−6.57	.00
γ_{32}	−0.838	−3.25	.00	1.25	7.54	.00
λ_4	NA	NA	NA	−1.65	−5.12	.00
γ_{41}	NA	NA	NA	−2.46	−10.35	.00
γ_{42}	NA	NA	NA	0.0700	0.43	.67
λ_5	NA	NA	NA	−1.32	−4.18	.00
γ_{51}	NA	NA	NA	−3.03	−10.54	.00
γ_{52}	NA	NA	NA	−1.08	−6.98	.00

NOTE: NA = not available.

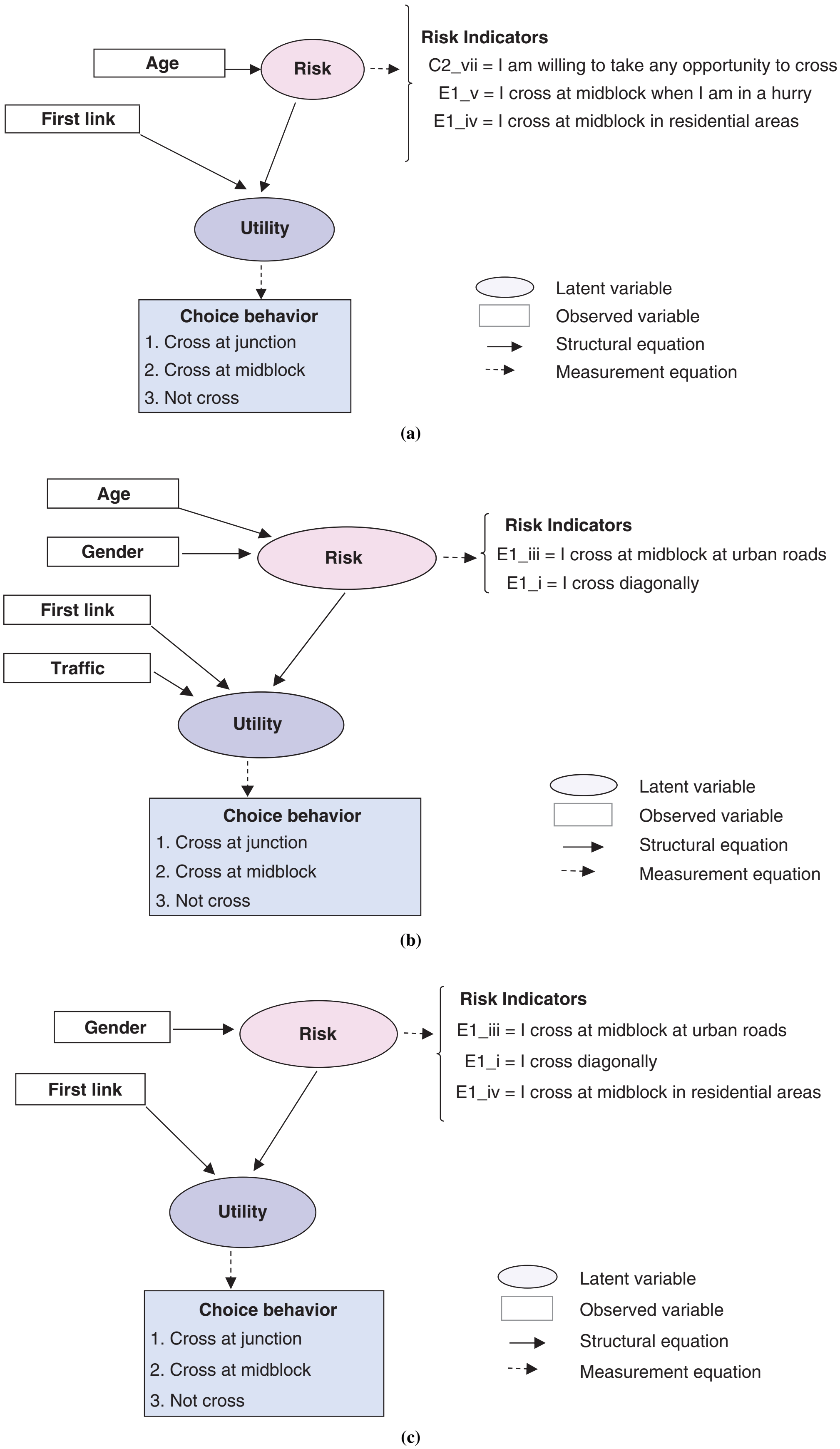

FIGURE 4 Structure of ICLV models with latent variable risk: (*a*) major urban arterials, (*b*) secondary roads, and (*c*) residential roads.

TABLE 5 Estimation Report and Goodness of Fit of ICLV Models for Types of Roads

Model Fit	Major Roads	Secondary Roads	Residential Roads
Number of parameters	8	13	14
Number of crossings	203	263	239
Initial log likelihood	−1.073.024	−965.574	−1.164.738
Final log likelihood	−821.159	−724.802	−894.889
Likelihood ratio test	503.730	481.544	539.700
ρ	.235	.249	.232
$\bar{\rho}$	.227	.236	.220
Diagnostic	Convergence reached	Convergence reached	Convergence reached
Iterations	220	35	32

NOTE: Risk is latent variable for all types of roads.

these indicators have higher risk. Moreover, the latent variable is expressed by the following behaviors:

– Pedestrians who are willing to take any opportunity to cross,
– Pedestrians who cross at midblock when in a hurry, and
– Pedestrians who cross at midblock in residential areas.

It is somewhat surprising that the indicator for crossing midblock in residential areas contributes to the latent variable for major roads; one might expect a stronger role by the crossing at major urban arterials indicator. This may be interpreted in two ways. First, pedestrians are less likely to declare crossing at midblock in major urban arterials, even if they do so under certain conditions. Second, pedestrians who cross frequently at midblock in residential areas may be more likely to cross at midblock in other conditions as well. In this case, the γ parameters were assigned fixed values, for the model to converge.

- Pedestrian age is a significant predictor of the latent variable risk (b_age). Pedestrians older than 25 years (i.e., mostly 25 to 45) are found to be more risk taking compared with pedestrians younger than 25 years.
- Pedestrians with higher risk-taking (B_risk) are more likely to cross at midblock at major roads; although midblock crossing was observed very rarely in this scenario (only 7 out of 203 crossings), it was strongly associated with high-risk reported behavior.
- The first road link (B_first) was found to have lower probability of being chosen in this scenario.
- No effect of traffic was found on this crossing scenario, and this may be attributed to the increased traffic and the high number of lanes of this type of road, leading pedestrians to less variation in their crossing behavior.

Models for Secondary Urban Roads

The structure of the ICLV model developed for secondary roads is summarized in Figure 4*b*. The best performing model was one with indicators E1_iii, E1_i forming the latent variable risk, with pedestrian age and gender being the risk predictors in the structural equation.

Modeling results are also presented in Tables 5 and 6. They can be summarized as follows:

- Parameters λ (λ_1, λ_2) are all statistically significant and positive. The latent variable is expressed by the following behaviors: pedestrians who cross at midblock on urban roads and pedestrians who cross diagonally. In this case, the self-reported behavior is matched with the observed behavior. Moreover, pedestrians crossing diagonally are obviously more likely to cross at midblock, especially on a secondary road.
- Pedestrian gender is not a significant predictor of the latent variable risk (b_gender).
- Pedestrian age is a significant predictor of the latent variable risk, and younger pedestrians (< 25 years) are more likely to exhibit

TABLE 6 Parameter Estimates of ICLV Models for Types of Roads

	Major Roads			Secondary Roads			Residential Roads		
Name	Value	*t*-Test	*p*-Value	Value	*t*-Test	*p*-Value	Value	*t*-Test	*p*-Value
ASC1	2.33	0.10	.92	−1.60	−5.09	.00	−1.91	−5.69	.00
B_first	−6.54	−0.27	.79	0.364	1.28	.20	0.884	2.41	.02
B1_trafficlow	NA	NA	NA	−1.38	−2.88	.00	NA	NA	NA
B_risk	1.10	1.91	.06	0.0189	0.09	.93	0.588	2.77	.01
b_gender	NA	NA	NA	0.0915	0.51	.61	−0.516	−2.80	.01
b_age	1.40	8.09	.00	0.960	4.63	.00	NA	NA	NA
ASC2	7.00	0.29	.77	−1.70	−7.05	.00	−1.29	−5.72	.00
λ_1	0.750	6.25	.00	2.65	1.69	.09	1.40	4.32	.00
γ_{11}	0.5	NA	NA	0.827	1.63	.10	−1.05	−4.20	.00
γ_{12}	1.00	NA	NA	4.04	2.31	.02	1.26	4.94	.00
λ_2	2.15	6.89	.00	0.921	4.25	.00	1.61	3.92	.00
γ_{21}	0.5	NA	NA	−0.153	−0.69	.49	−2.53	−5.23	.00
γ_{22}	1.00	NA	NA	1.83	5.65	.00	−0.757	−2.43	.02
λ_3	1.99	6.53	.00	NA	NA	NA	1.34	4.26	.00
γ_{31}	0.5	NA	NA	NA	NA	NA	−0.581	−2.66	.01
γ_{32}	1.00	NA	NA	NA	NA	NA	1.47	5.56	.00

risk-taking behavior (b_age) than middle-aged pedestrians (25 to 45 years).

- The latent variable risk taking (B_risk) is not statistically significant in the choice model. This is not surprising, because on secondary roads, the road and traffic environment is not very demanding, and conditions for risk-taking behavior may be less present.
- The first road link (B_first) was found to have higher probability of being chosen in this scenario.
- When traffic is low (B_trafficlow), midblock crossing probability decreases—compared, in this case, with the conditions of empty traffic.

Models for Residential Areas

The structure of the ICLV model for residential roads is presented in Figure 4*c*. The various combinations examined resulted in a latent variable with three indicators: E1_i, E1_iii, and E1_iv. Modeling results are presented in Tables 5 and 6. They can be summarized as follows:

- Parameters λ (λ_1, λ_2, and λ_3) are all statistically significant and positive. More specifically, the latent variable is expressed by the following behaviors:
  - Pedestrians who cross at midblock on urban roads,
  - Pedestrians who cross diagonally, and
  - Pedestrians who cross at midblock on residential roads.

In this case as well, the self-reported behavior is matched with the observed behavior. Moreover, pedestrians crossing diagonally are obviously more likely to cross at midblock.

- Pedestrian gender is a significant predictor of the latent variable risk (b_gender).
- The latent variable risk taking (B_risk) is statistically significant in the choice model. On residential roads, the road and traffic environment is not at all demanding, and conditions for optimizing behavior (i.e., midblock crossing, diagonal crossing) may be a common practice.
- The first road link (B_first) was found to have higher probability of being chosen in this scenario.
- No effect of traffic on pedestrian crossing behavior was found in this type of road network, and that was expected.

DISCUSSION OF RESULTS

Overall, the four ICLV models estimated in the present research largely confirm the research hypotheses as per the effects of road, traffic, and human factors of pedestrian crossing behavior (*16*). The effect of traffic volume was nonsignificant on major roads and on minor and residential roads, but it was significant on main and secondary roads. The effect of risk taking was significant on major and minor roads, and marginally significant or nonsignificant on main and secondary roads. Overall, risk taking is a key factor for crossing at midblock when traffic is high, and trip optimization is a key factor for crossing at midblock when traffic is low.

In none of the ICLV models was pleasure a significant latent variable. This finding was somewhat surprising, and it may be partly attributed to the specific trip not being representative of the usual walking motivations of those participants who often walk for pleasure.

Another key finding is that the research hypotheses on the road and traffic factors of pedestrian behavior were largely confirmed, but the research hypotheses on human factors of pedestrian crossing behavior were not fully confirmed. In particular, it was assumed that there were five factors of pedestrian behavior, each one corresponding to one section of the survey questionnaire. However, the survey responses do not confirm this structure, suggesting that the underlying dimensions are in fact few, with the risk-taking dimension being the dominant one.

CONCLUSIONS

Results of the ICLV models indicate that this family of models is very pertinent and useful for addressing the behavioral aspects of pedestrian trips in urban areas. It was clearly indicated that human factors may be important additional predictors of pedestrian behavior.

On the basis of the integrated models tested in this research, as well as the two-stage models tested in previous stages of this research, it appears that both approaches can be meaningful: the measurement error in the two-stage approach appears negligible, since the results of both approaches were similar as per the sign, magnitude, and statistical significance of human factors (*16*). The ICLV approach is theoretically sounder; however, it is a computationally demanding technique, and estimation of a global model was not possible. However, the latent variables estimated by the ICLV models are clearly defined and more easily interpreted.

In general, it would be recommended to implement more pedestrian surveys combining field observations and questionnaires. The present sample is not representative of age groups, and the inclusion of older pedestrians in the sample in future research might reveal additional effects of human factors on crossing behavior. Moreover, the sample size of this field survey is marginally adequate for a structural equation approach for latent variables. Measurements may not be stable and replicable at this sample size, and although the model was simplified to enhance validity, more data would be required to generalize the results to different settings. The present research also has limitations because participants knew that they were being observed, and the role of their usual travel motivations could not be captured. An alternative approach would be to capture crossing behavior of people who are not aware that they are being observed, and then follow up with them to participate in a simplified survey. In this case, however, the researcher would not be able to control for participants' route choices.

The proposed methodology and results need further development, more data, and validation before they can be used for practical applications. The next steps of the research should address in particular the model's validation, internal (e.g., with a small part of the existing data set left out in the model development and used for validation) and external (i.e., by means of new data collected). This analysis allows tackling the question of using such models for prediction.

ACKNOWLEDGMENTS

This paper is based on a research project implemented within the framework of the Action Supporting Postdoctoral Researchers of the Operational Program Education and Lifelong Learning (Action's Beneficiary: General Secretariat for Research and Technology), cofinanced by the European Social Fund (ESF) and the Greek State.

REFERENCES

1. Das, S., C. F. Manski, and M. Manuszak. Walk or Wait? An Empirical Analysis of Street Crossing Decisions. *Journal of Applied Econometrics,* Vol. 20, No. 4, 2005, pp. 529–548.
2. Hamed, M. M. Analysis of Pedestrians' Behavior at Pedestrian Crossings. *Safety Science,* Vol. 38, No. 1, 2001 pp. 63–82.
3. Lassarre, S., E. Papadimitriou, J. Golias, and G. Yannis. Measuring Accident Risk Exposure for Pedestrians in Different Micro-Environments. *Accident Analysis and Prevention,* Vol. 39, No. 6, 2007, pp. 1226–1238.
4. Chu, X., M. Guttenplan, and M. Baltes. Why People Cross Where They Do: The Role of Street Environment. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1878,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 3–10.
5. Grayson, G. B. Pedestrian Risk in Crossing Roads: West London Revisited. *Traffic Engineering and Control,* Vol. 28, 1987, pp. 27–30.
6. King, M. J., D. Soole, and A. Ghafourian. Illegal Pedestrian Crossing at Signalized Intersections: Incidence and Relative Risk. *Accident Analysis and Prevention,* Vol. 41, No. 3, 2009, pp. 485–490.
7. Sun, D., S. V. S. K. Ukkusuri, R. F. Benekohal, and S. T. Waller. Modeling of Motorist-Pedestrian Interaction at Uncontrolled Mid-block Crosswalks. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.
8. Oxley, J., B. Fildes, E. Ihsen, J. Charlton, and R. Days. Crossing Roads Safely: An Experimental Study of Age Differences in Gap Selection by Pedestrians. *Accident Analysis and Prevention,* Vol. 37, 2005, pp. 962–971.
9. Sarkar, S. Evaluation of Safety for Pedestrians at Macro- and Micro-levels in Urban Areas. In *Transportation Research Record 1502,* TRB, National Research Council, Washington, D.C., 1995, pp. 105–118.
10. Baltes, M., and X. Chu. Pedestrian Level of Service for Midblock Street Crossings. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1818,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 125–133.
11. Muraleetharan, T., A. Takeo, H. Toru, S. Kagaya, and S. Kawamura. Method to Determine Overall Level-of-Service of Pedestrians on Sidewalks and Crosswalks Based on Total Utility Value. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
12. Papadimitriou, E. Theory and Models of Pedestrian Crossing Behavior Along Urban Trips. *Transportation Research Part F,* Vol. 15, No. 1, 2012, pp. 75–94.
13. Evans, D., and P. Norman. Understanding Pedestrians' Road Crossing Decisions: An Application of the Theory of Planned Behavior. *Health Education Research,* Vol. 13, No. 4, 1998, pp. 481–489.
14. Diaz, E. M. Theory of Planned Behavior and Pedestrians' Intentions to Violate Traffic Regulations. *Transportation Research Part F,* Vol. 5, 2002, pp. 169–175.
15. Bernhoft, I. M., and G. Carstensen. Preferences and Behavior of Pedestrians and Cyclists by Age and Gender. *Transportation Research Part F,* Vol. 11, 2008, pp. 83–95.
16. Papadimitriou, E., S. Lassarre, and G. Yannis. Introducing Human Factors in Pedestrian Crossing Behavior Models. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
17. Temme, D., M. Paulssen, and T. Dannewald. Incorporating Latent Variables into Discrete Choice Models—A Simultaneous Estimation Approach Using SEM Software. *BuR Business Research,* Vol. 1, No. 2, 2008, pp. 220–237.
18. Ben-Akiva, M., J. Walker, A. T. Bernardino, D. A. Gopinath, T. Morikawa, and A. Polydoropoulou. Integration of Choice and Latent Variable Models. In *Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges* (H. S. Mahmassani, ed.), Elsevier, Oxford, United Kingdom, 2002, pp. 431–470.
19. Kitrinou, E., A. Polydoropoulou, and D. Bolduc. Development of Integrated Choice and Latent Variable (ICLV) Models for the Residential Relocation Decision in Island Areas. In *Proceedings from the Inaugural International Choice Modelling Conference* (S. Hess and A. Daley, eds.), Harrogate, United Kingdom, March 30–April 1, 2009; published 2010, pp. 593–617.
20. Politis, I., P. Papaioannou, and S. Basbas. Integrated Choice and Latent Variable Models for Evaluating Flexible Transport Mode Choice. *Research in Transportation Business and Management,* Vol. 3, 2012, pp. 24–38.
21. Varotto, S. F., A. Glerum, A. Stathopoulos, M. Bierlaire, and G. Longo. Modeling Travel Time Perception in Transport Mode Choices. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
22. Granié, M. A., M. Pannetier, and L. Guého. Developing a Self-Reporting Method to Measure Pedestrian Behaviors at All Ages. *Accident Analysis and Prevention,* Vol. 50, 2013, pp. 830–839.
23. Sisiopiku, V. P., and D. Akin. Pedestrian Behaviors at and Perceptions Towards Various Pedestrian Facilities: An Examination Based on Observation and Survey Data. *Transportation Research Part F,* Vol. 6, 2003, pp. 249–274.
24. Bierlaire, M. BIOGEME: A Free Package for the Estimation of Discrete Choice Models. In *Proceedings of the 3rd Swiss Transportation Research Conference,* Ascona, Switzerland, 2003.

The Standing Committee on Pedestrians peer-reviewed this paper.