ROAD, TRAFFIC AND HUMAN FACTORS OF PEDESTRIAN CROSSING BEHAVIOUR: INTEGRATED CHOICE AND LATENT VARIABLES MODELS

Eleonora Papadimitriou, PhD (Corresponding author)
Research Associate
National Technical University of Athens
Department of Transportation Planning and Engineering
5 Heroon Polytechniou st., GR-15773 Athens
Tel: +302107721380, Fax: +302107721454
E-mail: nopapadi@central.ntua.gr

Sylvain Lassarre, PhD
IFSTTAR - 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 la Vallée Cedex 2
Tel: +33145925764
E-mail: sylvain.lassarre@ifsttar.fr

George Yannis
Professor
National Technical University of Athens
Department of Transportation Planning and Engineering
5 Heroon Polytechniou st., GR-15773 Athens
Tel: +302107721326, Fax: +302107721454
E-mail: geyannis@central.ntua.gr

Dimitrios I. Tselentis
Researcher
National Technical University of Athens
Department of Transportation Planning and Engineering
5 Heroon Polytechniou st., GR-15773 Athens
Tel: +302107721575, Fax: +302107721454
E-mail: dtsel@central.ntua.gr

ABSTRACT

This paper aims to jointly analyze road, traffic and human factors of pedestrian crossing behavior, through the development of Integrated Choice and Latent Variables (ICLV) models. The analysis uses 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 are used, in which 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. ICLV models were developed for four different road types, namely major urban arterials, main roads, secondary roads and residential roads. The results suggest that the effect of traffic conditions on pedestrian crossing choices was more important on main and secondary urban roads, while on major urban arterials and on residential roads it was non-significant. As regards 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 may be advantageous compared to the two-stage approach.

Key-words: pedestrian behavior; human factors; integrated choice and latent variables models.

BACKGROUND AND OBJECTIVES

Modelling pedestrian crossing behavior in urban areas has attracted the interest of many researchers during the last few decades, as 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, 2, 3*). Studying pedestrians' crossing behavior can eventually lead to the better design and management of urban road networks in order 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 mid-block and diagonal crossings, etc. (*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, 14, 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 (*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 and 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 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 are 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 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, 20, 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 (*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. 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 (*16*), in order to develop ICLV models of pedestrian crossing behavior.

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

DATA COLLECTION

In this research, a particular data collection scheme was implemented [see also (*16*)]. Crossing behavior at urban pedestrian trips was recorded along with the conditions of the traffic and road environment. Attitudes, perceptions and behavior with regards 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 (see 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 '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/ 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 on weekdays' 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.

TABLE 1 Road type and geometric / traffic control characteristics of the 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

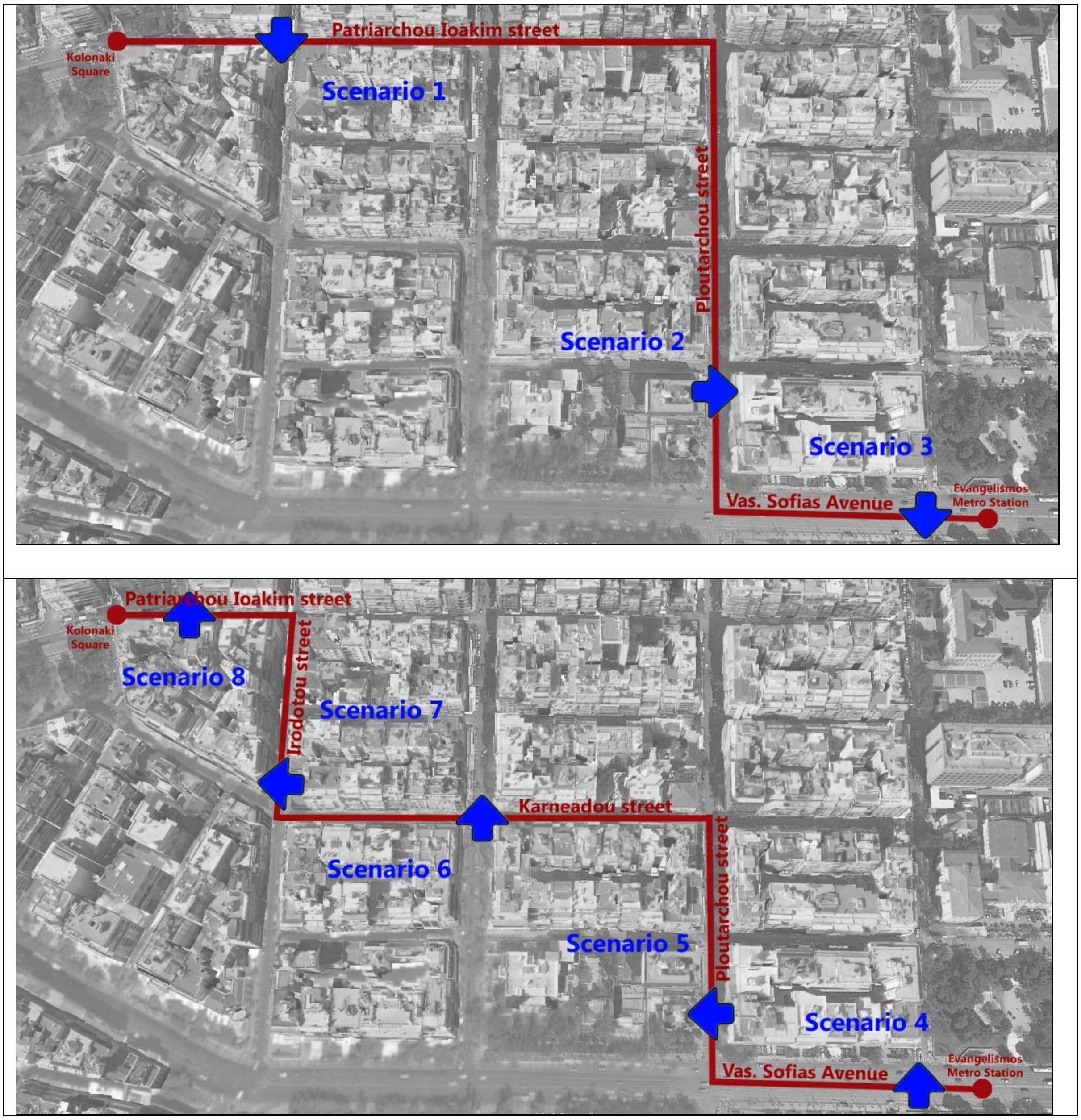

FIGURE 1 Presentation of the crossing scenarios on the survey site map

Questionnaire design

A questionnaire was developed on the basis of several questionnaires from the existing literature (*13*, *15*, *22*, *23*), and 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 5 sections, as shown in Table 2:

TABLE 2 Survey questionnaire

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.) than my car
B3_iv	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***:**
D_i	I am less likely to be involved in a road crash than other pedestrians
D_ii	I am faster than other pedestrians
D_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
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 times
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)

Survey procedure

The data collection took place in the period July - December 2013 with 75 participants in total. Participants were students of the National Technical University of Athens (NTUA) and young professionals. Fifty three percent of the survey participants were males, 50% of the participants were 18-24 years old, 27% were 25-34, 20% were 35-45 and 3% were >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 in order 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 3 shows a comparison of observed mid-block crossing 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.

TABLE 3 Comparison of observed crossing behavior for filling in the questionnaire before or after the walking task

		Observed share of crossings		
		Mid-block	**Junction**	**Total**
All participants		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%

The 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, and also complies with privacy protection and informed consent needs. However, there are limitations, as 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 non-compliant 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 meters, in order 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 that 3 seconds was used.

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 (*12, 16*) has shown that a sequential choice behavior appears to be the optimal assumption for pedestrian crossing choices. This sequential choice process involves a decision on each road link of the choice set: crossing at mid-block (MB), crossing at junction (J) or no crossing (No) (see Figure 2). 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, and therefore the rest of the choice set (i.e. the subsequent road links for this scenario) are not considered.

Integrated Choice and Latent Variables Models (ICLV)

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 proposed modelling 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

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

Structural equations of the latent variables model

$$Z_{1n} = \alpha_1 W_{in} + \omega_{1n}$$
$$Z_{2n} = \alpha_2 W_{in} + \omega_{2n}$$

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} denote the utility of each alternative respectively, for individual *n*; X_{in}, X_{jn} are sets of observed variables; Z_{1n}, Z_{2n} are the latent variables (actually the "components" accounting for most of the variability of the respective latent variables); $I_{1n}, I_{2n}, I_{3n}, I_{4n}$ are sets of the indicators of the latent variables Z_{1n}, Z_{2n} respectively; $\tilde{Z}_{1n}, \tilde{Z}_{2n}$ are the fitted values of the latent variables, once they are estimated by the structural equations of the latent variable model; W_{1n}, W_{2n} are sets of observed variables (characteristics of respondent *n*); $\varepsilon_{in}, \varepsilon_{jn}$ are extreme value distributed errors; $\omega_{1n}, \omega_{2n}, \upsilon_{1n}, \upsilon_{2n}, \upsilon_{3n}, \upsilon_{4n}$ are sets of (multivariate normally distributed) errors; $b', b_1, b_2, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4$ are 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:

$$\gamma_{ij} = \text{Prob}(y_{ij} \geq i) = \sum_{i}^{I} \pi_{ij} \text{ ,}$$

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

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

RESULTS

Exploratory analysis

The models development early indicated that a global model for all scenarios was unfeasible, due 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:

- The indicators were re-coded in a 3-point Likert scale, by adding up the responses 1-2 and 4-5, resulting thus in the following coding for the questionnaire variables: "1 - never or rarely", "2 - sometimes", "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 different scenarios were tested separately, as 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 that there are 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 mid-block in urban roads', etc.) and one 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, as pedestrians' choice process seems to present the most variability in this type of road. The best performing model was the one shown in Figure 2 (bottom left panel), with indicators E1_iii, E1_v and E2_i forming the latent variable "risk", with pedestrian gender being the risk predictor in the structural equation. More specifically, the model specification is as follows:

Structural model of the latent variable

$risk = b\_gender * gender + \omega$

Measurement equations: ordered logit

$I\_E1\_iii = \lambda_1 * risk + u_1$
$I\_E1\_v = \lambda_2 * risk + u_2$
$I\_E2\_i = \lambda_3 * risk + u_3$

Where:

$I\_E1\_iii_1 = 1 / [1 + \exp(\lambda_1 * risk - \gamma_{11})]$
$I\_E1\_iii_2 = 1 / [1 + \exp(\lambda_1 * risk - \gamma_{12})] - 1 / [1 + \exp(\lambda_1 * risk - \gamma_{11})]$
$I\_E1\_iii_3 = 1 - 1 / (1 + \exp(\lambda_1 * risk - \gamma_{12})]$
$I\_E1\_v_1 = 1 / [1 + \exp(\lambda_2 * risk - \gamma_{21})]$,
$I\_E1\_v_2 = 1 / [1 + \exp(\lambda_2 * risk - \gamma_{22})] - 1 / (1 + \exp(\lambda_2 * risk - \gamma_{21}))]$
$I\_E1\_v_3 = 1 - 1 / [1 + \exp(\lambda_2 * risk - \gamma_{22})]$
$I\_E2\_i_1 = 1 / [1 + \exp(\lambda_3 * risk – \gamma_{31})]$,
$I\_E2\_i_2 = 1 / [1 + \exp(\lambda_3 * risk - \gamma_{32})] - 1 / (1 + \exp(\lambda_3 * risk - \gamma_{31})]$
$I\_E2\_i_3 = 1 - 1 / [1 + \exp(\lambda_3 * risk - \gamma_{32})]$

Choice utility functions

V1 = ASC1 + B_first * first + B_trafficlow * trafficlow + B_risk * risk
V2 = ASC2 + B_first * first
V3 = ASC3

Where V1 is the utility of crossing at mid-block, V2 the utility of crossing at junction and V3 the utility of not crossing, ASC are alternative-specific constants, 'trafficlow' refers to low traffic conditions, and 'first' refers to the first road link of the choice set.

The modelling results are presented in Table 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, i.e. 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 mid-block on urban roads
    - Pedestrians who cross at mid-block when in a hurry
    - Pedestrians who cross at mid-block between stopped vehicles at congestion.

  In this case, the self-reported behavior is matched with the observed behavior Moreover, pedestrians crossing between vehicles when the road is congested is a likely pattern 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 mid-block 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), mid-block crossing probability increases.

An attempt was made, due 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 2 (bottom right panel), where indicators B3_i and B3_ii where used for latent variable "pleasure" and E1_iii, E1_v and E2_i for "risk", and 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 Table 4). Nevertheless, the latent variable 'pleasure' was not found significant.

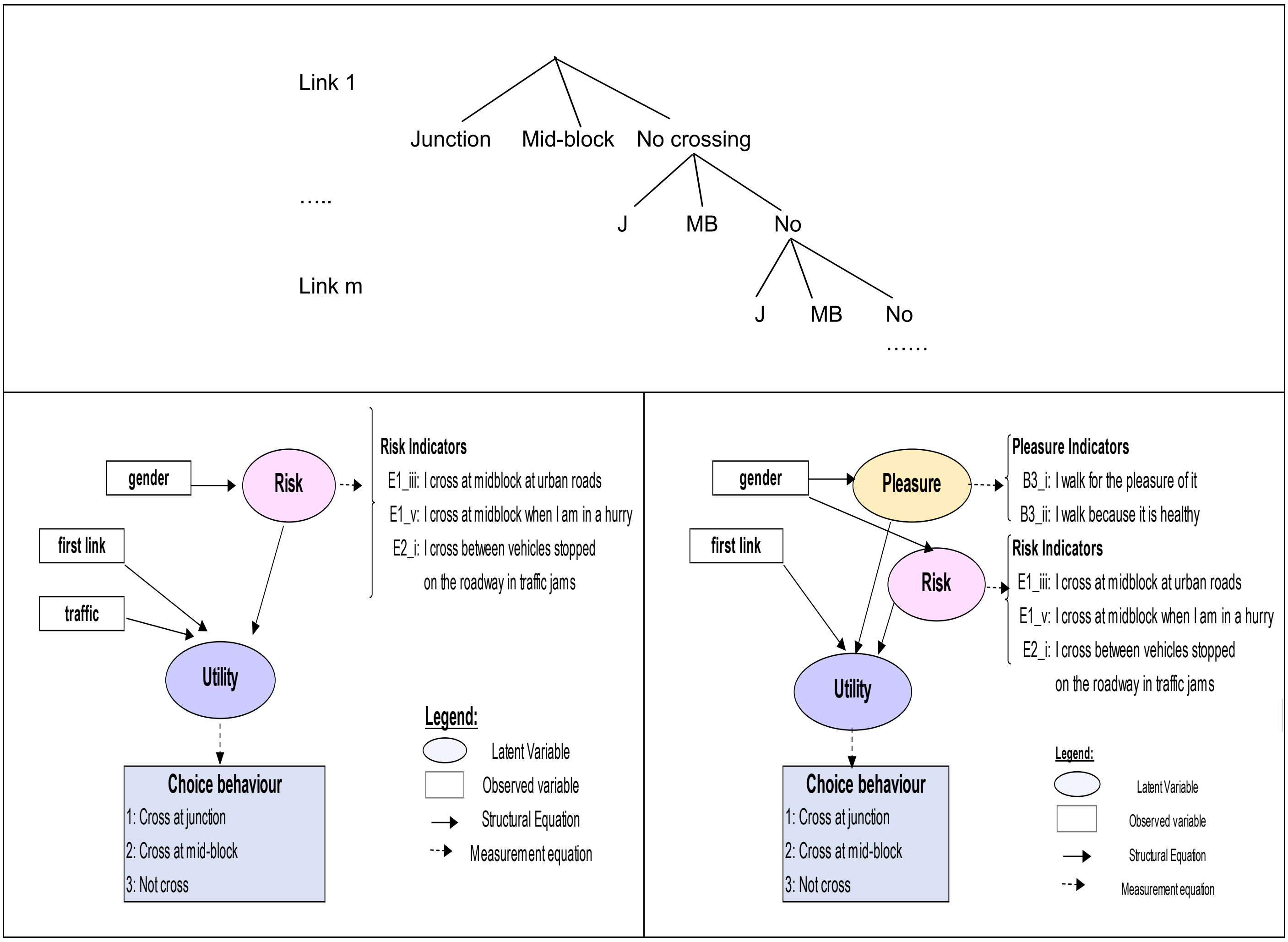

FIGURE 2 Sequential logit model of pedestrian crossing behavior (top panel) - Structure of the ICLV model for main urban roads: Latent variables: risk (bottom left panel) - risk & pleasure (bottom right panel)

TABLE 4. Parameter estimates of ICLV models for main urban roads - Latent variables: risk (left panel), risk & pleasure (right panel)

Estimation report	***Latent variable: risk***			***Latent variables: risk & plesure***		
Number of parameters:	15			22		
Number o crossings:	184			184		
Init log-likelihood:	-986.670			-3.070.240		
Final log-likelihood:	-607.361			-1.854.286		
Likelihood ratio test:	758.618			2.431.908		
Rho	0.384			0.396		
Rho bar for the init. model:	0.369			0.389		
Diagnostic:	Convergence reached...			Convergence reached...		
Iterations:	36			54		
Estimated parameters						
Name	**Value**	**t-test**	**p-value**	**Value**	**t-test**	**p-value**
ASC1	-2.74	-7.40	0.00	-2.52	-9.44	0.00
ASC2	-1.33	-5.13	0.00	-1.31	-7.23	0.00
B_first	0.466	1.33	0.18	0.427	1.73	0.08
B1_trafficlow	1.54	2.56	0.01	-	-	-
B_risk	0.342	1.13	0.26	-0.410	-1.38	0.17
B_pleasure	-	-	-	-0.248	-0.65	0.52
b_gender	-0.550	-3.12	0.00	0.538	4.88	0.00
b2_gender	-	-	-	-0.375	-2.59	0.01
lamda1	2.78	3.63	0.00	-1.34	-5.58	0.00
gamma11	-2.04	-3.71	0.00	-2.34	-11.31	0.00
gamma12	1.77	3.52	0.00	-0.586	-4.14	0.00
lamda2	3.97	2.20	0.03	-1.89	-6.17	0.00
gamma21	-5.92	-2.46	0.01	-2.73	-11.24	0.00
gamma22	-1.97	-1.94	0.05	-0.713	-4.42	0.00
lamda3	1.38	5.06	0.00	-1.57	-5.86	0.00
gamma31	-2.80	-7.57	0.00	-1.00	-6.57	0.00
gamma32	-0.838	-3.25	0.00	1.25	7.54	0.00
lamda4	-	-	-	-1.65	-5.12	0.00
gamma41	-	-	-	-2.46	-10.35	0.00
gamma42	-	-	-	0.0700	0.43	0.67
lamda5	-	-	-	-1.32	-4.18	0.00
gamma51	-	-	-	-3.03	-10.54	0.00
gamma52	-	-	-	-1.08	-6.98	0.00

Models for major urban arterials

The structure of the ICLV model developed for major roads is summarized in Figure 3 (top panel). For the “risk” indicators, there were a few that were marginally significant, such as C2_vii, E1_iv and E1_v. The modelling results are presented in Table 5. 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 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 mid-block when in a hurry
    - Pedestrians who cross at mid-block in residential areas

  It is somewhat surprising that the indicator for crossing mid-block in residential areas contributes to the latent variable for major roads; one might expect a stronger role of the "crossing at major urban arterials" indicator. This may be interpreted in two ways: first, pedestrians are less likely to declare crossing at mid-block in major urban arterials, even if they do so under certain conditions; and second, pedestrians who cross frequently at mid-block in residential areas may be more likely to cross at mid-block in other conditions as well.

  In this case, the γ parameters were assigned fixed values in order for the model to converge.
- Pedestrian age is a significant predictor of the latent variable "risk" (b_age). Pedestrians older than 25 (i.e. mostly 25-45) years are found to be more risk-taking compared to pedestrians younger than 25 years.
- Pedestrians with higher risk-taking (B_risk) are more likely to cross at mid-block at major roads; although mid-block 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 3 (middle panel). 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.

The modelling results are also presented in Table 5. 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 mid-block on urban roads
    - 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 mid-block, 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 risk-taking behavior (b_age) than middle aged pedestrians (25-45 years).
- The latent variable risk-taking (B_risk) is not statistically significant in the choice model. This is not surprising, as 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, to the conditions of empty traffic.

Models for residential areas

The structure of the ICLV model for residential roads is presented in Figure 3 (bottom panel). The various combinations examined resulted in a latent variable with three indicators, namely E1_i, E1_iii and E2_i. The modelling results are presented in Table 5:

- 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 mid-block on urban roads
    - Pedestrians who cross diagonally
    - Pedestrians who cross at mid-block 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 mid-block.
- 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. mid-block 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, which was expected.

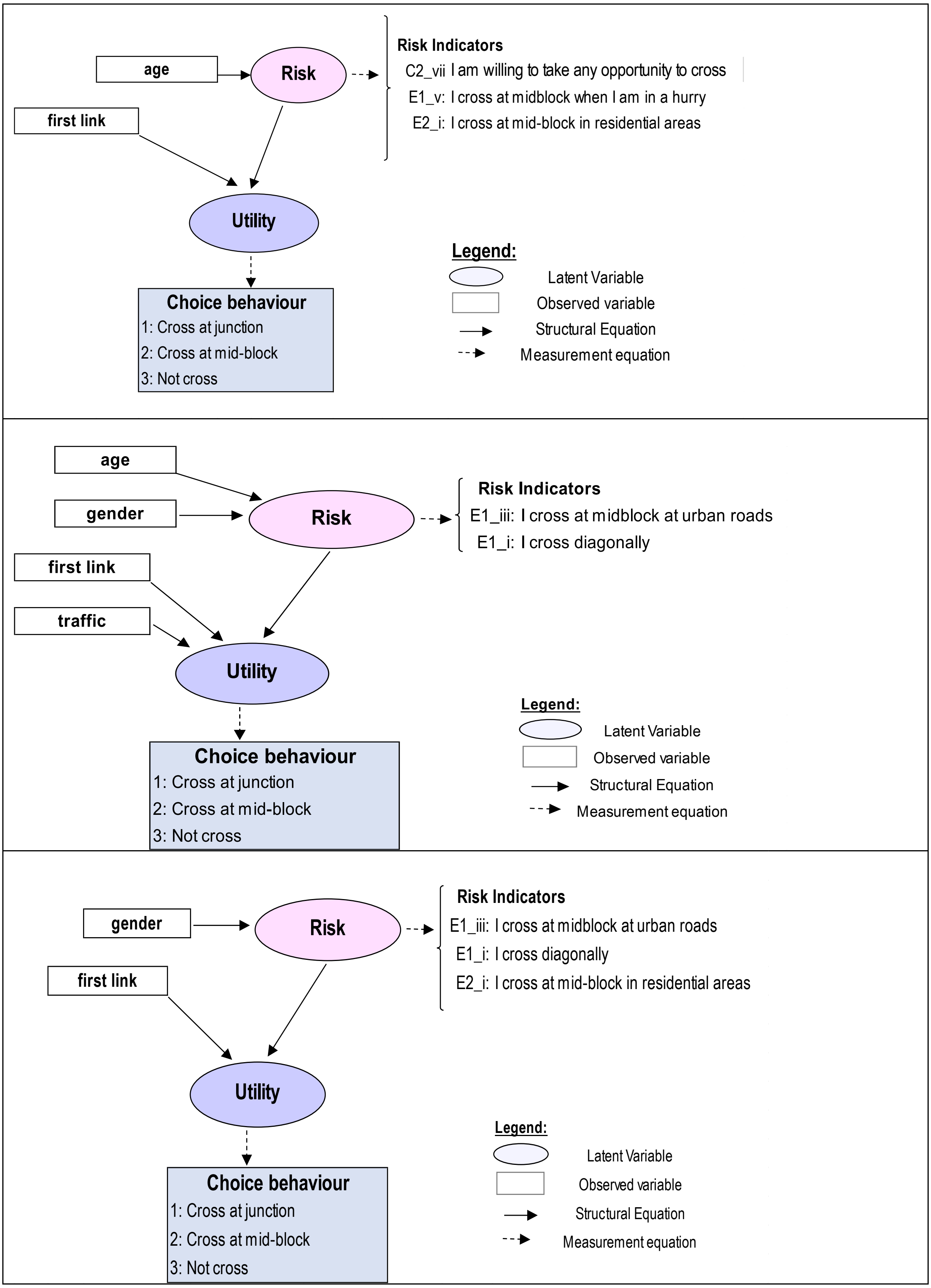

FIGURE 3 Structure of the ICLV models for: a) major urban arterials - Latent variable: risk (top panel), b) secondary roads - Latent variable: risk (middle panel), c) residential roads - Latent variable: risk (bottom panel)

TABLE 5 Results of ICLV model for a) major urban arterials - Latent variable: risk (left panel), b) secondary roads - Latent variable: risk (middle panel), c) residential roads - Latent variable: risk (right panel)

Estimation report	***Major roads***			***Secondary roads***			***Residential roads***		
Number of parameters:	8			13			14		
Number of crossings:	203			263			239		
Init 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		
Rho	0.235			0.249			0.232		
Rho bar	0.227			0.236			0.220		
Diagnostic:	Convergence reached...			Convergence reached...			Convergence reached...		
Iterations:	220			35			32		
Estimated parameters									
Name	**Value**	**t-test**	**p-value**	**Value**	**t-test**	**p-value**	**Value**	**t-test**	**p-value**
ASC1	2.33	0.10	0.92	-1.60	-5.09	0.00	-1.91	-5.69	0.00
B_first	-6.54	-0.27	0.79	0.364	1.28	0.20	0.884	2.41	0.02
B1_trafficlow	-	-	-	-1.38	-2.88	0.00	-	-	-
B_risk	1.10	1.91	0.06	0.0189	0.09	0.93	0.588	2.77	0.01
b_gender	-	-	-	0.0915	0.51	0.61	-0.516	-2.80	0.01
b_age	1.40	8.09	0.00	0.960	4.63	0.00	-	-	-
ASC2	7.00	0.29	0.77	-1.70	-7.05	0.00	-1.29	-5.72	0.00
lamda1	0.750	6.25	0.00	2.65	1.69	0.09	1.40	4.32	0.00
gamma11	0.5	-	-	0.827	1.63	0.10	-1.05	-4.20	0.00
gamma12	1,00	-	-	4.04	2.31	0.02	1.26	4.94	0.00
lamda2	2.15	6.89	0.00	0.921	4.25	0.00	1.61	3.92	0.00
gamma21	0.5	-	-	-0.153	-0.69	0.49	-2.53	-5.23	0.00
gamma22	1,00	-	-	1.83	5.65	0.00	-0.757	-2.43	0.02
lamda3	1.99	6.53	0.00	-	-	-	1.34	4.26	0.00
gamma31	0.5	-	-	-	-	-	-0.581	-2.66	0.01
gamma32	1,00	-	-	-	-	-	1.47	5.56	0.00

DISCUSSION

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

In none of the ICLV models was 'pleasure' a significant latent variable. This finding was somewhat surprising, and 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.

CONCLUSION

The 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 (*16*), it appears that both approaches can be meaningful: the measurement error in the two-stage approach appears negligible, as the results of both approaches were similar as per the sign, the magnitude and the statistical significance of human factors. The ICLV approach is theoretically sounder; however, it is a computationally demanding technique and the estimation of a global model was not possible. On the other hand, 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 a 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 due to the fact that 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 their route choice.

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 dataset left out in the model development and used for validation) and external (i.e. by means of new data collected). This analysis allow tackling the question of using such models for prediction.

ACKNOWLEDGEMENTS

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), co-financed by the European Social Fund (ESF) and the Greek State.

REFERENCES

1. Das S., Manski C.F., Manuszak M. 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., Papadimitriou E., Golias J., Yannis G. Measuring accident risk exposure for pedestrians in different micro-environments. *Accident Analysis & Prevention,* Vol.39, No.6, 2007, pp. 1226-1238.
4. Chu X., Guttenplan, M., Baltes M. Why People Cross Where They Do - The Role of the Street Environment. In *Transportation Research Record No 1878*, Transportation Research Board of the National Academies, Washington, D.C., 2003, 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., Soole, D., & Ghafourian, A. Illegal pedestrian crossing at signalized intersections: incidence and relative risk. *Accident Analysis & Prevention*, Vol. 41, No. 3, 2009, 485-490.
7. Sun D., Ukkusuri S.V.S.K., Benekohal R.F., Waller S.T. Modeling of Motorist-Pedestrian Interaction at Uncontrolled Mid-block Crosswalks. In *the Proceedings of the 82nd TRB Annual Meeting*, CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2003.
8. Oxley, J., Fildes, B., Ihsen, E., Charlton, J., and Days, R. Crossing roads safely: An experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, Vol.37, 2005, pp. 962–971.
9. Sarkar, S. Evaluation of safety for pedestrians at macro- and microlevels in urban areas. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1502*, Transportation Research Board of the National Academies, Washington, D.C., 1995, pp. 105-118.
10. Baltes M., Chu X. Pedestrian level of service for mid-block street crossings. In *the Proceedings of the 81st TRB Annual Meeting*, CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2002.
11. Muraleetharan, T., Takeo A., Toru H., Kagaya S., Kawamura S. Method to Determine Overall Level-of-Service of Pedestrians on Sidewalks and Crosswalks based on Total Utility Value. In *the Proceedings of the 83rd TRB Annual Meeting*, CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2004. Vol. 23, pp. 727-735
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., Norman P. 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., Carstensen G. Preferences and behavior of pedestrians and cyclists by age and gender. *Transportation Research Part F,* Vol. 11, 2008, pp. 83-95.
16. Papadimitriou, E., Lassarre, S., and Yannis, G. Introducing Human Factors in Pedestrian Crossing Behavior Models. *In the Proceedings of the 94th TRB Annual Meeting*, CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2015.

17. Temme, D., Paulssen, M., & Dannewald, T. Incorporating latent variables into discrete choice models—a simultaneous estimation approach using SEM software. *BuR-Business Research*, Vol. 1, No. 2, 2008 220-237.
18. Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., & Polydoropoulou, A. Integration of choice and latent variable models. *Perpetual motion: Travel behavior research opportunities and application challenges*, 2002, pp. 431-470.
19. Kitrinou, E., Polydoropoulou, A., & Bolduc, D. Development of integrated choice and latent variable (ICLV) models for the residential relocation decision in island areas. *Discrete choice modelling: State of the art, state of the practice*, 2010, pp. 593-617.
20. Politis, I., Papaioannou, P., & Basbas, S. Integrated choice and latent variable models for evaluating flexible transport mode choice. *Research in Transportation Business & Management*, Vol. 3, 2012, pp. 24-38.
21. Varotto, S. F., Glerum, A., Stathopoulos, A., Bierlaire, M., & Longo, G. (2015, December). Modelling travel time perception in transport mode choices. *In the Proceedings of the 94th TRB Annual Meeting*, CD-ROM, Transportation Research Board of the National Academies, Washington, D.C., 2015.
22. Granié M.A., Pannetier M., Guého L. 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., Akin, D. 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, Proceedings of the *3rd Swiss Transportation Research Conference*, Ascona, Switzerland, 2013.