HANDLING INTRA-HOUSEHOLD CORRELATIONS IN MODELING TRAVEL: A COMPARISON OF HIERARCHICAL (RANDOM EFFECT) MODELS AND MARGINAL (GEE) MODELS

Mario Cools, Corresponding Author
University of Liège, ARGENCO, Local Environment Management & Analysis (LEMA)
Quartier Polytech 1, Allée de la Découverte 9, BE-4000 Liège, Belgium
Tel: +32 4 366 4813
Fax: +32 4 366 2909
Email: mario.cools@ulg.ac.be

Elke Moons
Statistics Netherlands (CBS)
CBS-weg 11, Postbus 4481, 6401 CZ Heerlen, The Netherlands
Tel: +31 45 570 7421
Email: e.moons@cbs.nl

Word count: 5,202 words text + 7 tables x 250 words (each) + 35 references = 6,952 words

Submission Date: November 14, 2015

ABSTRACT
In this paper, the necessity for treating intra-household correlation is investigated by analyzing two travel behavior indices, i.e. travel time and travel distance, for three important travel motives (commuting, shopping, and leisure). Data stemming from the 2010 Belgian National Household Travel Survey are used in the analysis. Two model approaches that accommodate for intra-household correlation are compared, namely the generalized linear mixed model (GLMM) and GEE model approach. Both model approaches show that high levels of intra-household correlation are present, and therefore the use of models that take into account intra-household correlation, is strongly recommend. Results indicate that this requirement is the most urgent for non-commuting trips. Moreover, the results show that GLMM and GEE yield comparable estimates in the case of normally distributed data. Furthermore, evidence was provided that the more the estimates of the intra-household correlation provided by the two approaches differ, the less the homogeneity of the parameters is assured. In this regard, if one has to choose between the GLMM and GEE methodology, especially the negative consequences of choosing an inappropriate covariance model in the case of a GLMM model favor the selection of the GEE methodology. Further research is needed to compare the two approaches in the context of non-normally distributed travel behavior data.

Keywords: multilevel modeling, GLMM, generalized estimation equations, GEE, travel distance, travel time expenditure

1 BACKGROUND

Modeling travel behavior has always been a major area of concern in transportation research. In this regard, the human activity approach (*1-3*), commonly referred to as activity-based or time-use-based approach (*4*) has become the dominating paradigm for analyzing and modeling travel behavior (*5*). In studies, where the focus lies on the identification and assessment of influencing factors, travel behavior is commonly quantified in terms of three indices, i.e. (*i*) the number of trips, (*ii*) the total distance travelled, and the travel time expenditure (*6-12*). Studies focusing on travel distances often assess the impact of the built-environment and urban development (*9,13*), whereas studies on travel time expenditure often focus on (the lack) of dynamics (*14-17*).

To collect information about travel behavior, travel surveys are still one of the most important ways of obtaining the critical information exhibiting the required level of behavioral detail. Typically, these surveys collect information about the demographic, socio-economic, and trip-making characteristics (e.g. transport mode, motive, duration, distance) of individuals and households. Diaries mainly form the basis of a travel survey, and next to these individual questionnaires, also a household survey needs to be filled out. Given the particular nature of repeated measurements (i.e. multiple trips per person), and variables that relate to different levels of measurement (i.e. individual versus household), an appropriate methodology needs to adopted to avoid statistical problems (*18-20*). Moreover, different studies have underlined the importance of household interactions in decision making processes with respect to travel behavior *(21-24*). In this regard, a recent literature review is provided by Ho and Mulley (*22*).

In travel behavior analysis, a commonly used methodology to overcome this problem is the use of multilevel models - also referred to as hierarchical, conditional or mixed-effect models - for which early applications were realized in the beginning of 2000 (*8, 20, 25-26*). A less common approach is the use of generalized estimation equations (GEE), which with a few exceptions (e.g. *27-28*) has barely been used in the field of travel behavior analysis, and to the best of our knowledge never been compared to multilevel models in this context. Hubbard et al. (*29*) argue that in general multilevel models involve unverifiable assumptions on the data generating distribution, which lead to potentially misleading estimates and biased inference, and conclude that the GEE approach provides a more useful approximation of the truth. Therefore, the main goal of this paper is to compare the suggested modeling strategies and to underline their need by implementing them to model the travel time and distance for the commuting (work/school), shopping and leisure trips.

2 BELGIAN NATIONAL HOUSEHOLD TRAVEL SURVEY DATA

2.1 Belgian National Household Travel Survey

For the comparison of multilevel models and generalized estimating equations (GEE), data stemming from the 2010 Belgian Household Travel (*30*) are used. This survey was selected, as it was the last travel survey that was conducted in Belgium of which official results were released.
The data collection effort of this survey, which was spread over the period December 2009 – December 2010, encompassed the enquiry of 8,532 households. In total, 15,821 respondents, who were more than 6 years old, were asked to record their trips for a predefined day. For each trip, the purpose, location, timing, and duration were queried, as well as the chain of transport mode that was used to arrive at the destination. In total 37,680 trips were registered. Besides this information, some general personal (e.g. age, professional status) and household information (e.g. household composition, vehicle possession) were gathered as well.

To get a general idea of travel behavior in Belgium, some basic features will be provided next. The weighted mean number of trips carried out on one day equals 3.3, encompassing about 42.5 km. It is important to emphasize that people who do not travel, often referred to as 'immobiles', are not incorporated in these figures. As noted in the introduction, people tend to travel for different reasons. The most important trip motive, accounting for 28.8% of all trips, is commuting, defined as work and school related trips. Shopping (19.8%) is the second most important motivation for traveling, followed by escorting trips, where something or someone is either being picked up or dropped off (13.2%). Leisure trips (including touring trips such as walking with the dog) account for 12.5% and visit trips (11.4%) wind up the top 5 of the most frequent trips.

2.2 Data Description

Recall that the primary objective of this paper is to compare multilevel models (in particular generalized linear mixed models (GLMM)) and generalized estimation equations (GEE). The idea is to investigate how personal and household characteristics influence travel time and distance on a regular day. Since the focus is to determine how much time is actually spent on our transport network daily, people who do not travel are excluded from the analysis.

2.2.1 Dependent Variables

To quantify travel behavior, the total distance travelled and the total time spent during one day is modelled for three different alternatives (i.e. commuting, shopping and leisure). Given the fact that distributions for the three travel time expenditures and the three total distances travelled are quite skewed to the left, the choice was made to take a (natural) log-transform of these response variables. Taking the natural logarithm of the distances and times travelled per trip motive has as implication that only travelers who realized trips for that particular trip motive are taken into account in the final analyses (the natural logarithm of 0 is not defined). Table 1 provides insight on the basic descriptive statistics of the 2 × 3 transformed outcome variables that are retained for the statistical analysis.

TABLE 1 Description of the Transformed Dependent Variables

Variable	Mean	Std Dev	N[1]
Natural logarithm of daily travel time expenditure: commuting (work/school)	3.66	0.89	3625
Natural logarithm of daily travel distance: commuting (work/school)	2.85	1.42	3150
Natural logarithm of daily travel time expenditure: shopping	3.13	0.89	2642
Natural logarithm of daily travel distance: expenditure: shopping	1.89	1.39	2641
Natural logarithm of daily travel time expenditure: leisure	3.62	0.97	1771
Natural logarithm of daily travel distance: expenditure: leisure	2.31	1.43	1771

[1]N: number of observations retained after excluding undefined transformed value

2.2.2 Explanatory Variables

The explanatory variables can be divided in two categories: predictors that behave at the level of the individual and predictors that behave at the level of the household. Table 2 shows the variables that are used in the analyses, together with a short explanation.

TABLE 2 Explanatory Variables: Abbreviation and Short Description

Variable	Description
Explanatory variables at household level	
HHSIZE	Household size (Mean: 2.8, Std Dev: 1.4, Min: 1.0, Max: 6.0)
NRBIKE	Number of bicycles owned (Mean: 2.2, Std Dev: 2.0, Min: 0.0, Max: 10.0)
NRCARS	Number of cars owned (Mean: 1.4, Std Dev: 0.8, Min: 0.0, Max.: 5.0)
HINC	Net monthly household income: 1: <1499 € (17.0%), 2: 1500-4999 € (68.1%), 3: >5000 (8.3%), 4: Did not disclose income (6.7%)
URBAN	Urbanization of city/municipality where the residence is located: 1: Urban (47.3%), 2: Sub-urban (28.6%), 3: Rural (24.2%)
REGION	Region where the residence is located: 1: Brussels (22.2%), 2: Flanders (27.5%), 3: Wallonia (50.3%)
Explanatory variables at individual level	
AGE	Age (Survey year 2010) (Mean: 43.5, Std Dev: 19.7, Min: 6.0, Max: 93.0)
NRTRIPS	Total number of trips during reporting day (Mean: 3.4, Std Dev: 2.0, Min: 1.0, Max: 12.0)
GENDER	Gender: 1: Male (49.8%), 2: Female (50.2%)
EDU	Education level: 0: No higher education: none/primary/secondary education (57.8%), 1: Higher education: university/university college (42.2%)
STATUS	Professional Status: 1: Pupil, student (18.2%), 2: White-collar worker (non-executive) (31.6%), 3: White-collar worker (executive) (4.7%), 4: Blue-collar worker (8.3%), 5: Liberal profession/independent (5.5%), 6: Retired, unemployed, incapacitated (27.7%), 7: Housewife/househusband (3.9%), 8: Other (0.3%)
DRIVLIC	Car driving license: 0: No (25.0%), 1: Yes (75.0%)
ABOPT	Season ticket public transport: 0: No (77.3%), 1: Yes (22.7%)
FREQCAR	Frequent[1] car (driver/passenger) user: 0: No (14.0%), 1: Yes (86.0%)
FREQPT	Frequent[1] public transport user: 0: No (70.4%), 1: Yes (29.6%)
FREQBIKE	Frequent[1] bike user: 0: No (79.3%), 1: Yes (20.7%)
MOBRESTR	Physical mobility restraints: 0: No (85.0%), 1: Yes (15.0%)
PARTNER	Partner (married/officially living together): 0: No (45.4%), 1: Yes (54.6%)

[1] Frequent is defined as travelling at least once per week with this transport mode.

With regard to the explanatory variables at the household level, two factors have been included to capture the effect of the built environment. First, the urbanization level of the residential location is taking into account. Secondly, the region of the residential location is incorporated. The latter could be considered as an indicator for the unobserved heterogeneity in urban structures. The high share of persons residing in Wallonia is an artifact of the sampling procedure, as discussed by Cornelis et al. (*30*).

Concerning the explanatory variables at individual level, one could depict a small difference in the average number of trips reported in this table (3.4) in comparison to the one reported in Section 2.1. This small difference is due to suppression of some erroneous and outlying (people travelling more than 8 hours during the day of reporting) data.

3 METHODOLOGY

In this section, the different modeling strategies that tackle intra-household correlations are highlighted. First the methodology concerning multilevel models is explained. Afterwards, the use of generalized estimation equations is highlighted.

3.1 Multilevel Models

In this paper, a classical example of a two-level hierarchical structure will be shown. Level 1 is the level of the smallest unit (individual) whereas the second level denotes the clusters of the units (household). The main idea of a Generalized Linear Mixed Model (GLMM) is to examine the behavior of the level 1 outcome as a function of predictors that behave both on level 1 and on level 2. Two models are tested. The first model is an unconditional means or a one-way random effects ANOVA model. This shows how much of the variation in the data can be captured by allowing solely a separate intercept for each household. The second model includes both predictors at the household (level 2) level, and at the individual level (level 1).

3.1.1 Unconditional Means Model

At first, the variation in the response variables across households is examined by means of an unconditional means (UCM) or a one-way random effects ANOVA model. The ANOVA-way of writing down this model expresses the outcome, Y_{ij} , as a linear combination of the grand mean (μ), household deviations from that mean (α_j) and a random error (ε_{ij}) associated with the *i*-th individual in household *j*:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad \text{with } \alpha_j \sim_{iid} N\left(0, \sigma_H^2\right) \text{ and } \varepsilon_{ij} \sim N\left(0, \sigma^2\right). \qquad \text{(eq. 1)}$$

This model will now be re-parameterized to multilevel notation, because this notation can be generalized more easily to the more complex models. It expresses the level 1 outcome by means of set of linked models: one at the individual level and one at the household level. At level 1, the outcome can be denoted as the sum of an intercept for the individual households (β_{0j}) and a random error (ε_{ij}) associated with the *i*-th individual in household *j*:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \varepsilon_{ij} \quad \text{with } \varepsilon_{ij} \sim N\left(0, \sigma^2\right). \qquad \text{(eq. 2)}$$

At the second level, the household's intercept is expressed as a sum of the overall mean (μ) and a series of random deviations from that mean (α_j):

$$\text{Level 2: } \beta_{0j} = \mu + \alpha_j \quad \text{with } \alpha_j \sim_{iid} N\left(0, \sigma_H^2\right). \qquad \text{(eq. 3)}$$

Substituting the level 2 model (eq. 3) in the level 1 equation (eq. 2) yields the multilevel model:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad \text{with } \alpha_j \sim_{iid} N\left(0, \sigma_H^2\right) \text{ and } \varepsilon_{ij} \sim N\left(0, \sigma^2\right). \qquad \text{(eq. 4)}$$

Note that it is also assumed that α_j and ε_{ij} are independent of one another. One can notice that there is a direct equivalence between the one-way random effects ANOVA notation and the multilevel notation. This model can be partitioned into two separate parts: a fixed part that contains the single effect μ (the overall intercept) and a random part that contains two random effects (the intercepts α_j and the within-household residuals ε_{ij}). The fixed effect μ provides information about the average outcome in the population, the parameter of the first random effect σ_H^2 offers insight about the variability in the household means, while σ^2 tells something about the variability of the outcome within the households.

The unconditional means model considered in this study postulates that the variance-covariance structure takes a special form, i.e. that of compound symmetry. This means that the variance for each individual is assumed to be $\sigma_H^2 + \sigma^2$, the covariance of the outcome for two individuals of the same household equals σ_H^2, while the covariance of the outcomes for two individuals belonging to a different household is 0.

3.1.2 Conditional Model with Effects at Household (Level 2) and Individual (Level 1) Level

The unconditional means model provides a baseline against which more complex models can be compared. The conditional model envisaged in this study accommodates the unconditional means model by incorporating explanatory variables both at the level of the household as well as the level of the individual. In its most general form, the conditional model predicting the outcome of individual *i* in household *j* can be written as:

$$Y_{ij} = \beta_{0j} + \sum_{k=1}^{K} \beta_{kj} \mathbf{X}_{\mathbf{k}ij} + \varepsilon_{ij}, \quad \text{with } \beta_{kj} = \zeta_{k0} + \sum_{p=1}^{P} \zeta_{kp} \mathbf{Z}_{\mathbf{p}j} + u_{kj}, \text{ for } k = 0,\dots,K,$$

$$\varepsilon_{ij} \sim N\left(0,\sigma^2\right) \text{ and } \begin{pmatrix} u_{0j} \\ \vdots \\ u_{Kj} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \cdots & \sigma_0\sigma_K \\ \vdots & \ddots & \vdots \\ \sigma_K\sigma_0 & \cdots & \sigma_K^2 \end{pmatrix}\right]. \qquad \text{(eq. 5)}$$

The *X*-variables refer to predictors at the first level (individual), whereas the *Z*'s are explanatory variables at the second level (household). If one only wants to account for predictors at the second level, it shows that all other β-parameters, except β_{0j} can be set equal to zero. Similarly, only accounting for explanatory variables at the first level can be carried out by setting all ζ-parameters, except ζ_{k0} equal to zero. The index *p* behaves at household level, while index *k* is associated with the individual level in this paper.

3.2 Marginal Model Approach: Generalized Estimating Equations (GEE)

In the previous subsection, the use of multilevel models was proposed as a technique for handing intra-household correlation. In this subsection, a second approach for dealing with clustered data is elucidated, namely the use of generalized estimating equations, which could be seen as an extension of the generalized linear model (GLM) approach. The method of GEE is a marginal (or population-averaged) approach to estimation with correlated data. Marginal models differ from conditional (cluster-specific) approaches such as the multilevel model described earlier in this paper: while conditional approaches model the probability distribution of the dependent variable as function of the covariates and a parameter specific to each cluster, marginal models, by contrast, model the marginal expectation of the dependent variable as a function of the covariates (*31*).

A GEE-model uses the same link function and linear predictor setup as in the independence case. In this study, this predictor setup is the one of a linear regression function, which assumes the identity link between the dependent variables (log travel time, log travel distance) and the explanatory variables. The random component is described by the same variance functions as in the independence case, but the covariance structure of the correlated measurements must also be modeled. For discrete data, this implies specification of the first-order moments, as well as of all higher-order moments. For normally distributed data, such as the data analyzed in this paper, the full-model specification reduces to modeling the first- and second-order moments only, a situation much simpler than in the non-normal case. Notwithstanding, even then can the choice of inappropriate covariance models seriously invalidate inferences for the mean structure. Consequentially, a technique like generalized estimating equations is still preferred (*32*).

Thus, a key difference between the GEE and other approaches to correlated data is the necessity of specifying a correlation matrix called "working correlation matrix" to define the covariance matrix. In a classical GLM-model an independent working correlation structure is assumed: let Y_{hj} and Y_{ij} be the outcomes of individuals *h* and *i* in household *j*, then the independence structure is defined as:

$$Corr\left(Y_{hj},Y_{ij}\right)=\begin{cases}1 & h=i\\ 0 & h\neq i\end{cases}. \qquad \text{(eq. 6)}$$

In the GEE models that are built in this paper an exchangeable working correlation matrix is estimated. This working correlation matrix assumes that the correlations between different members of the household are the same:

$$Corr\left(Y_{hj},Y_{ij}\right)=\begin{cases}1 & h=i\\ \alpha & h\neq i\end{cases}. \qquad \text{(eq. 7)}$$

A more general structure could have been the unstructured working correlation matrix, which estimates the correlations between different members of the household separately (e.g. correlation between person 1 and 2 of a particular household is allowed to be different from the correlation between person 1 and 3), defined as:

$$Corr\left(Y_{hj},Y_{ij}\right)=\begin{cases}1 & h=i\\ \alpha_{hi} & h\neq i\end{cases}. \qquad \text{(eq8).}$$

This general structure of the working correlation is not adopted as this resulted in estimation problems: the number of response pairs for estimating correlation was less than the number of regression parameters, in which a more restricted correlation model, such as the exchangeable working correlation, is more appropriate. To compare the model performance of the GEE model using this exchangeable working correlation with the model assuming an independence structure, the QIC (Quasi-likelihood under the Independence model Criterion) is analogous to the AIC (Akaike's Information Criteriion), lower values indicating a better model fit. Note however, that the model results from the independence models are not presented in this paper.

3.3 Model Comparison

Caution is needed when comparing conditional models with marginal models. Recall that parameters from conditional models (e.g. GLMM) have a cluster-specific interpretation, while the parameters of marginal models (e.g. GEE) have a population average interpretation. For non-normally distributed data (i.e. binomial and Poisson), the parameter estimates of the conditional model should be larger than the estimates of the marginal model. In this case, however, normally distributed data are modeled, and both sets of parameters can be compared without making transformations or integration over all random effects (*33*). To ease model comparability, the same explanatory variables that are obtained in the GLMM model are used for the GEE model.

In terms of model fit, it should be noted that likelihood based criteria such AIC can be calculated for the GLMM model, but cannot be computed for the GEE model, as the parameter estimation of the GEE model is based on quasi-likelihood instead of likelihood. To facilitate an assessment of model fit, for the GLMM models the AIC values are calculated for the intercept only model, which in practice corresponds to the AIC value of the UCM model. For the GEE model, the QIC values for the models with an independence correlation structure and the models with an exchangeable working correlation structure are tabulated.

4 RESULTS

In this Section, the results of the different model approaches are discussed. First, the necessity for taking into account intra-household correlations is assessed. Afterwards, the model parameters of the different model strategies are compared.

4.1 Intra-household Correlation

A first approach for examining the need for tackling intra-household correlation is to look at the intra-class correlation ρ estimated by means of the unconditional means (UCM) model. This intra-class correlation indicates what portion of the total variance occurs between households, and is estimated by:

$$\hat{\rho} = \frac{\hat{\sigma}_H^2}{\hat{\sigma}_H^2 + \hat{\sigma}^2},$$

where $\hat{\sigma}_H^2$ represents the estimated variability in the household means, and $\hat{\sigma}^2$ corresponds to the predicted variability of the outcome within the households. The estimated intra-class correlations, displayed in Table 3, clearly indicate clustering is present between different members of the same household and – consequently – that travel times and travel distances between different members of the same household tend to be very similar. The higher the value of this intra-class correlation, the more an ordinary least squares (OLS) regression analysis of the data is likely to yield misleading results. This risk is especially high for non-commuting trips, the most clustering being present in leisure trips.

A second way to investigate the urgency for handling intra-household correlation, is to look at the estimated working correlation α in the generalized estimating equations (GEE) model. The estimates of these working correlations $\hat{\alpha}$ yield the same conclusion as the intra-class correlations estimated using the UCM model: there is a clear necessity for treating intra-household correlation in a proper way. Homogenous to the results of the UCM model, this necessity is highest for non-commuting trips. The contrast between commuting and non-commuting trips can be (partially) accounted for by the fact work and school activities are obligatory activities which are performed almost always by one specific household member individually, while non-work/school activities are often performed jointly with other household members (*34*).

TABLE 3 Measures for Intra-household Correlation

Trip Purpose	Travel Parameter	UCM $\hat{\sigma}_H^2$	UCM $\hat{\sigma}^2$	UCM $\hat{\rho}$	GEE $\hat{\alpha}$
Commuting	Time	0.111	0.683	0.140	0.148
	Distance	0.543	1.458	0.271	0.275
Shopping	Time	0.482	0.304	0.613	0.788
	Distance	1.237	0.677	0.646	0.670
Leisure	Time	0.607	0.322	0.653	0.751
	Distance	1.313	0.695	0.654	0.751

Finally, the necessity for explicitly taking into account the intra-household correlation is assessed by comparing the QIC values off all the GEE models with an exchangeable working correlation structure, with the GEE models assuming an independence structure. For Table 4, one can see that for all 6 GEE models, the model with the exchangeable working correlation structure (final model) outperforms - lower QIC vales - the ones assuming an independence structure (independence model), underling the necessity for explicitly tackling intra-household correlations.

4.2 Model Results

Having stressed the need for models that account for intra-household correlations in the previous paragraphs, the remainder of this section focuses on the comparison of the model results of the different model approaches. For the three considered trip motives (commuting, shopping and leisure), both travel time expenditures and travel distances are investigated. An overview of which explanatory variables that are accounting for heterogeneity in the travel indices, is displayed in Table 4. From this table one could depict different interesting issues by interpreting the p-values of the Wald Type III-tests. These tests indicate the overall contribution of a variable to the model. In the case of a continuous or dummy variable, this is equivalent to the significance test of the parameter, whereas for categorical variables with more than two categories these tests assess the simultaneous contribution of the different parameters representing this categorical variable.

First, one could notice that the explanatory variables have a significant impact, especially in the commuting related models, whereas they play a considerably less important role in the models estimating shopping and leisure related indices. One the one hand, this is an indication that household travel surveys are especially tailored for capturing heterogeneity with respect to commuting behavior. On the other hand, this is a sign that the variability in non-mandatory trips such as shopping and leisure trips is much larger in comparison to mandatory trips such as commuting, and as a result more difficult to capture with traditional household travel surveys. The influence of social networks on leisure trips for instance (*35*), supports the latter hypothesis.

Second, one can see that all the variables that were indicated as significant in the GLMM model, were also significant in the corresponding GEE models, with the exception of the number of trips [NRTRIPS], which was not significant at the 5% level of significance in the GEE model predicting the total distance travelled for leisure trips. The high correspondence between the p-values of both modeling approaches supports the hypothesis that these approaches indeed are adequate in capturing and correctly dealing with the intra-household correlation.

Third, two variables that were considered did not play at all a significant role in predicting daily travel indices, i.e. net household income [HINC] and physical mobility restraints [MOBRESTR]. A possible explanation for the non-significance of income is the fact that income effects on daily travel are for a large extent due to increased car ownership. The latter does have an effect in most of the models. Note here that the correlation among car ownership and income, and more generally among the different explanatory variables did not significantly affect the results, as the variance inflation factors were all below the critical value of 4.

The non-significance of physical mobility constraints is an indication that, people, who have such constraints and who are performing trips, are not negatively affected by it, in terms of travel times and travel distances. However, contrary to this finding, physical mobility constraints do induce a higher share of immobility. Given that immobile persons (persons making no trips during the survey day) were excluded from the analysis, caution is needed in generalizing the non-significance of physical mobility constraints.

TABLE 4 Wald Type III P-Values of the 2 (Travel Indices) × 3 (Trip Motives) × 2 (Methodologies) Models

	Commuting				**Shopping**				**Leisure**			
	Time		**Distance**		**Time**		**Distance**		**Time**		**Distance**	
Variable	**GLMM**	**GEE**	**GLMM**	**GEE**	**GLMM**	**GEE**	**GLMM**	**GEE**	**GLMM**	**GEE**	**GLMM**	**GEE**
Explanatory variables at household level												
HHSIZE	----	----	0.021	0.011	----	----	0.004	0.005	----	----	0.001	<0.001
NRBIKE	----	----	----	----	<0.001	<0.001	----	----	----	----	0.024	0.024
NRCARS	0.012	0.020	<0.001	<0.001	----	----	<0.001	<0.001	----	----	0.001	0.002
HINC	----	----	----	----	----	----	----	----	----	----	----	----
URBAN	0.037	0.035	0.026	0.039	----	----	0.001	0.001	----	----	----	----
REGION	0.013	0.007	<0.001	<0.001	----	----	<0.001	<0.001	----	----	<0.001	<0.001
Explanatory variables at individual level												
AGE	----	----	<0.001	<0.001	----	----	----	----	----	----	0.003	0.002
NRTRIPS	<0.001	<0.001	<0.001	<0.001	----	----	----	----	<0.001	0.001	0.013	0.082
GENDER	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	----	----	----	----	0.025	0.015
EDU	<0.001	<0.001	----	----	----	----	----	----	----	----	----	----
STATUS	<0.001	<0.001	<0.001	<0.001	<0.001	0.002	----	----	<0.001	<0.001	0.016	0.004
DRIVLIC	0.003	0.004	<0.001	<0.001	----	----	----	----	----	----	----	----
ABOPT	<0.001	<0.001	<0.001	<0.001	----	----	----	----	----	----	0.008	0.033
FREQCAR	0.048	0.036	<0.001	<0.001	----	----	0.003	0.004	----	----	0.022	0.044
FREQPT	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	----	----	0.002	0.003	----	----
FREQBIKE	0.023	0.028	0.003	0.002	----	----	0.039	0.050	----	----	----	----
MOBRESTR	----	----	----	----	----	----	----	----	----	----	----	----
PARTNER	----	----	----	----	<0.001	0.001	<0.001	<0.001	----	----	----	----
Model fit criteria												
AIC intercept only model	9421.8	n.a.	11069.2	n.a.	6621.6	n.a.	8932.0	n.a.	4635.0	n.a.	5999.0	n.a.
AIC final model	8624.2	n.a.	10478.3	n.a.	6564.1	n.a.	8755.2	n.a.	4593.2	n.a.	5955.6	n.a.
QIC independence model	n.a.	3652.8	n.a.	3177.6	n.a.	2656.3	n.a.	2655.9	n.a.	1782.9	n.a.	1796.7
QIC final model	n.a.	3652.0	n.a.	3175.6	n.a.	2655.3	n.a.	2653.3	n.a.	1778.1	n.a.	1788.8

---- indicates that the variable is not included in the final model, n.a.: not applicable

4.2.1 Model Results for Commuting

The parameter estimates for the GLMM and GEE models predicting the time and distance traveled on commuting trips are shown in Table 5. Recall that for building the GEE models, the same explanatory variables were chosen as in the 'best' GLMM model (according to the Akaike Information Criterion). From this Table, one can see that, although a large amount of variables are used to predict both outcomes, the best model for each travel index (i.e. travel time and travel distance) contains unique variables: household size and age only appear in the distance models, whereas education was only included in the models predicting travel time.

TABLE 5 Parameter Estimates for GLMM and GEE Models for Commuting Trips

	Time				Distance			
	GLMM		GEE		GLMM		GEE	
Variable	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
INTERCEPT	*3.574*	*0.074*	*3.574*	*0.075*	*3.126*	*0.159*	*3.129*	*0.169*
Explanatory variables at household level								
HHSIZE	----	----	----	----	*-0.049*	*0.021*	*-0.051*	*0.020*
NRCARS	*0.051*	*0.020*	*0.051*	*0.022*	*0.226*	*0.039*	*0.233*	*0.040*
URBAN=1	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
URBAN=2	*0.089*	*0.040*	*0.086*	*0.039*	*0.180*	*0.068*	*0.166*	*0.065*
URBAN=3	0.003	0.040	-0.005	0.041	0.127	0.069	0.104	0.070
REGION=1	*-0.111*	*0.044*	*-0.116*	*0.042*	*-0.602*	*0.076*	*-0.620*	*0.077*
REGION=2	0.029	0.035	0.025	0.037	-0.073	0.060	-0.085	0.059
REGION=3	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
Explanatory variables at individual level								
AGE	----	----	----	----	*-0.010*	*0.002*	*-0.010*	*0.002*
NRTRIPS	*-0.082*	*0.008*	*-0.080*	*0.008*	*-0.099*	*0.012*	*-0.097*	*0.013*
GENDER=1	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
GENDER=2	*-0.192*	*0.027*	*-0.191*	*0.026*	*-0.432*	*0.045*	*-0.435*	*0.044*
EDU=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
EDU=1	*0.142*	*0.035*	*0.144*	*0.036*	----	----	----	----
STATUS=1	*-0.397*	*0.053*	*-0.400*	*0.053*	*-0.934*	*0.127*	*-0.935*	*0.150*
STATUS=2	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
STATUS=3	*0.184*	*0.055*	*0.187*	*0.055*	*0.243*	*0.086*	*0.256*	*0.083*
STATUS=4	*-0.123*	*0.048*	*-0.122*	*0.049*	*-0.302*	*0.070*	*-0.309*	*0.075*
STATUS=5	-0.099	0.055	-0.095	0.065	*-0.193*	*0.086*	*-0.193*	*0.093*
STATUS=6	-0.111	0.073	-0.113	0.074	*-0.501*	*0.093*	*-0.505*	*0.108*
STATUS=7	-0.353	0.213	*-0.367*	*0.161*	*-0.545*	*0.198*	*-0.574*	*0.195*
STATUS=8	*-0.622*	*0.202*	*-0.628*	*0.299*	2.882	1.234	2.880	0.113
DRIVLIC=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
DRIVLIC=1	*0.148*	*0.049*	*0.141*	*0.049*	*0.336*	*0.088*	*0.336*	*0.092*
ABOPT=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
ABOPT=1	*0.292*	*0.042*	*0.292*	*0.044*	*0.314*	*0.075*	*0.313*	*0.081*
FREQCAR=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
FREQCAR=1	*0.084*	*0.043*	*0.089*	*0.042*	*0.391*	*0.077*	*0.385*	*0.084*
FREQPT=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
FREQPT=1	*0.541*	*0.041*	*0.537*	*0.046*	*0.305*	*0.072*	*0.314*	*0.082*
FREQBIKE=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
FREQBIKE=1	*-0.080*	*0.035*	*-0.078*	*0.035*	*-0.192*	*0.063*	*-0.189*	*0.062*

----: not included in the final model, Ref.: reference category, italic: p-values < 0.05

The estimates should be interpreted in the following way. The estimates of the total number of cars owned by the household [NRCARS] indicate that according to the GLMM model the average slope representing the relationship between this number and (log) travel time is equal to 0.051, or in other words that for each extra car owned by the household, the log travel time increases with 0.051. Note that the GEE model produced exactly the same slope.

When parameter estimates of the GLMM and GEE models are compared, one could conclude that the parameter estimates of the two model approaches are pointing in the same direction: the sign of the significant GLMM and GEE is the same in most cases. Moreover, the absolute values lie very close to each other.

With respect to household level variables, one can observe that household size has a decreasing effect on distance travelled, but has no effect on time expenditure. The number of cars has an increasing effect, as explained above. Concerning urbanization, one could derive that households, residing in urban agglomerations (reference category), spend less time and travel less far then households residing in suburban municipalities. Households, residing in the Brussels capital area, spend less time and travel less far, which is in accordance to the particularly high number of job opportunities in that region.

Regarding variables at individual level, one could notice that elder people are travelling less far. The number of trips realized during the journey is negatively correlated to both travel indices, indicating the existence of a travel time frontier (*14*). Higher educated people travel longer to their work, but with respect to distance this effect was not confirmed. Concerning professional status, especially executives travel longer and further in comparison to people with another status. Finally, regarding transport options and mode uses, one could notice that increased transport options and mode frequencies coincide with longer travel distances and travel times, with exception of the frequent users of bike. The latter effect could be interpreted as a sign of self-selection in residential location choice.

4.2.2 Model Results for Shopping Trips

Parameter estimates obtained by the models predicting the travel time and travel distance of shopping trips are presented in Table 6. In contrast to the models estimating commuting time and distance, for shopping trips considerable differences exist in terms of significant variables. In the model predicting time expenditure especially individual parameters play a significant role, whereas the distance model in mainly modeled using household level attributes.

Contrary to the results of the commuting models, the parameter estimates of the GLMM and GEE shopping models show a (slightly) higher level of heterogeneity, especially in the case of the model predicting time expenditure. This is in line with Hubbard et al. (*29*), who concluded that when the correlation estimates (displayed in Table 3) differ more, larger discrepancies in estimates can be noticed. Notwithstanding, even though the differences are larger, overall parameters still can be considered equivalent.

Similar to the commuting models, bike-related variables (i.e. bike possession in the time expenditure model, and frequent cycling in the distance model) have a decreasing effect. With respect to the variables at individual level in the time expenditure model, one can notice that females travel longer. Among the different professional states, especially retired and students spend longer times travelling for shopping in comparison to their counterparts. Having a partner increases both travel distance and travel times.

Regarding household level attributes, household size has a decreasing effect, which to some extent counterbalances the effect of having a partner. Besides, car ownership has an increasing effect. Parameter estimates of urbanization and region highlight the same effect; locations with a higher number of shopping opportunities have a decreasing effect on distance travel for shopping. In contrast, these parameters did not play a significant role on shopping related travel time expenditure, showing that on a daily basis more nearby shopping destinations results in a higher number but much shorter (in terms of distance) trips.

TABLE 6 Parameter Estimates for GLMM and GEE Models for Shopping Trips

	Time				**Distance**			
	GLMM		**GEE**		**GLMM**		**GEE**	
Variable	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
INTERCEPT	*2.887*	*0.050*	*2.894*	*0.053*	*1.398*	*0.095*	*1.398*	*0.097*
Explanatory variables at household level								
HHSIZE	----	----	----	----	*-0.078*	*0.027*	*-0.080*	*0.028*
NRBIKE	*-0.043*	*0.011*	*-0.048*	*0.012*	----	----	----	----
NRCARS	----	----	----	----	*0.266*	*0.047*	*0.266*	*0.046*
URBAN=1	----	----	----	----	0.000	Ref.	0.000	Ref.
URBAN=2	----	----	----	----	*0.160*	*0.078*	*0.159*	*0.076*
URBAN=3	----	----	----	----	*0.295*	*0.081*	*0.293*	*0.079*
REGION=1	----	----	----	----	*-0.356*	*0.083*	*-0.356*	*0.082*
REGION=2	----	----	----	----	*-0.216*	*0.072*	*-0.219*	*0.070*
REGION=3	----	----	----	----	0.000	Ref.	0.000	Ref.
Explanatory variables at individual level								
GENDER=1	0.000	Ref.	0.000	Ref.	----	----	----	----
GENDER=2	*0.105*	*0.030*	*0.097*	*0.031*	----	----	----	----
STATUS=1	*0.192*	*0.065*	*0.195*	*0.072*	----	----	----	----
STATUS=2	0.000	Ref.	0.000	Ref.	----	----	----	----
STATUS=3	0.052	0.082	0.048	0.096	----	----	----	----
STATUS=4	-0.006	0.071	-0.013	0.070	----	----	----	----
STATUS=5	0.000	0.078	0.012	0.099	----	----	----	----
STATUS=6	*0.183*	*0.041*	*0.167*	*0.046*	----	----	----	----
STATUS=7	0.088	0.070	0.101	0.075	----	----	----	----
STATUS=8	-0.487	0.390	-0.634	0.518	----	----	----	----
FREQCAR=0	----	----	----	----	0.000	Ref.	0.000	Ref.
FREQCAR=1	----	----	----	----	*0.237*	*0.078*	*0.237*	*0.083*
FREQPT=0	0.000	Ref.	0.000	Ref.	----	----	----	----
FREQPT=1	*0.196*	*0.039*	*0.191*	*0.045*	----	----	----	----
FREQBIKE=0	----	----	----	----	0.000	Ref.	0.000	Ref.
FREQBIKE=1	----	----	----	----	*-0.141*	*0.068*	*-0.133*	*0.068*
PARTNER=0	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
PARTNER=1	*0.146*	*0.040*	*0.162*	*0.046*	*0.219*	*0.056*	*0.222*	*0.057*

----: not included in the final model, ref: reference category, italic: p-values < 0.05

4.2.3 Model Results for Leisure Trips

Similar to the shopping models, the models predicting travel distance and travel time of leisure trips contain different explanatory variables, as could be noticed from Table 7. In the travel time expenditure model, only the number of trips, professional status and the usage frequency of public transit play a role, whereas in the distance based model both household level and an even greater number of individual level characteristics play a role.

Analogous to the results of the shopping model, parameter of GLMM and GEE are similar, yet slightly more different when compared to the commuting models. This again confirms the statement of Hubbard et al. (*29*).

In analogy with the distance-based commuting and shopping models, a lower total distance can be observed for the Brussels capital region. Besides, the increased transport options considered at both household and individual level all have an increasing effect. With respect to professional status, one could observe that especially retired and unemployed people spend more time and travel further in comparison to the other groups.

TABLE 7 Parameter Estimates for GLMM and GEE Models for Leisure Trips

	Time				Distance			
	GLMM		GEE		GLMM		GEE	
Variable	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
INTERCEPT	*3.763*	*0.066*	*3.737*	*0.077*	*2.852*	*0.216*	*2.860*	*0.236*
Explanatory variables at household level								
HHSIZE	----	----	----	----	*-0.135*	*0.040*	*-0.148*	*0.045*
NRBIKE	----	----	----	----	*0.060*	*0.027*	*0.063*	*0.028*
NRCARS	----	----	----	----	*0.190*	*0.059*	*0.194*	*0.062*
REGION=1	----	----	----	----	*-0.517*	*0.095*	*-0.505*	*0.098*
REGION=2	----	----	----	----	-0.123	0.091	-0.132	0.093
REGION=3	----	----	----	----	0.000	Ref.	0.000	Ref.
Explanatory variables at individual level								
AGE	----	----	----	----	*-0.009*	*0.003*	*-0.010*	*0.003*
NRTRIPS	*-0.045*	*0.010*	*-0.043*	*0.012*	*-0.037*	*0.015*	-0.029	0.017
GENDER=1	----	----	----	----	0.000	Ref.	0.000	Ref.
GENDER=2	----	----	----	----	*-0.123*	*0.055*	*-0.129*	*0.053*
STATUS=1	*-0.204*	*0.056*	*-0.176*	*0.061*	*-0.474*	*0.118*	*-0.442*	*0.120*
STATUS=2	0.000	Ref.	0.000	Ref.	0.000	Ref.	0.000	Ref.
STATUS=3	-0.162	0.095	-0.182	0.104	-0.110	0.143	-0.126	0.137
STATUS=4	0.017	0.097	0.049	0.101	-0.189	0.145	-0.122	0.181
STATUS=5	-0.052	0.107	-0.045	0.117	0.031	0.159	0.048	0.169
STATUS=6	*0.124*	*0.056*	*0.145*	*0.056*	0.039	0.101	0.090	0.095
STATUS=7	0.022	0.107	0.064	0.097	-0.087	0.163	-0.011	0.159
STATUS=8	-0.360	0.349	-0.320	0.100	-0.671	0.530	*-0.649*	*0.162*
ABOPT=0	----	----	----	----	0.000	Ref.	0.000	Ref.
ABOPT=1	----	----	----	----	*0.203*	*0.076*	*0.190*	*0.089*
FREQCAR=0	----	----	----	----	0.000	Ref.	0.000	Ref.
FREQCAR=1	----	----	----	----	*0.235*	*0.102*	*0.235*	*0.117*
FREQPT=0	0.000	Ref.	0.000	Ref.	----	----	----	----
FREQPT=1	*0.146*	*0.046*	*0.146*	*0.050*	----	----	----	----

----: not included in the final model, ref: reference category, italic: p-values < 0.05

5 CONCLUSIONS AND FURTHER RESEARCH

In this paper, the necessity for treating intra-household correlation was acknowledged by calculating the intra-class correlation using unconditional means (UCM) models and using the working correlation estimated in the generalized estimating equation (GEE) models. Both model approaches showed that high levels of intra-household correlation were present, and therefore the use of models that take into account intra-household correlation, is strongly recommend. Results indicated that this requirement is the most urgent for non-commuting trips.

After establishing and confirming the need for models that handle intra-household correlation, two model approaches that accommodate for intra-household correlation were compared, namely the generalized linear mixed model (GLMM) and GEE model approach. For the three travel motives, travel time and travel distance were estimated using these model strategies. The results acknowledge the conclusion of Schukken et al. (*33*) that GLMM and GEE yield comparable estimates in the case of normally distributed data or equivalently data that after transformation is normally distributed. Furthermore, evidence was provided that the more the estimates of the intra-household correlation provided by the two approaches differ, the less the homogeneity of the parameters is assured, confirming the conclusion by Hubbard et al. (*29*). In this regard, if one has to choose between the GLMM and GEE methodology, especially the consequences of choosing an inappropriate covariance model in the case of a GLMM model, i.e. invalid inference for the mean structure (*32*), favor the selection of the GEE methodology. In practice this implies that the GEE methodology guarantees a correct interpretation of the significance of the factors that contribute to the different travel times and distances, whereas the GLMM only provides correct interpretation when the correct covariance model has been selected. In contrast, the adoption of an inappropriate covariance model leads to biased interpretations.

Further research is needed to compare the two approaches in the context of non-normally distributed travel behavior data. After all, for non-normally distributed data, the results of the GLMM and GEE approach cannot be directly compared, requiring an integration of the GLMM results over all random effects or the use of an approximated calculation as suggested by Schukken et al. (*33*). Besides, it would be interesting to assess the predictive performance of the different modeling approaches, especially in the context of missing data.

6 ACKNOWLEDGEMENTS

The research was funded through the Special Research Fund, financed by the Wallonia-Brussels Federation.

7 REFERENCES

1. Fox, M. Transport planning and the human activity approach. *Journal of Transport Geography,* Vol. 3, No. 2, 1995, pp. 105-116.
2. Hägerstrand, T. What about people in regional science? *Papers of the Regional Science Association*, Vol. 24, No. 1, 1970, pp. 7-21. (OLD5)
3. Jones, P.M., M.C. Dix, M.I. Clarke, and I.G. Heggie. U*nderstanding Travel Behavior*. Gower, Aldershot, 1983. (OLD6)
4. Pendyala, R. M., and K. G. Goulias. Time use and activity perspectives in travel behavior research. *Transportation*, Vol. 29, No. 1, 2002, pp. 1-4.
5. Rasouli, S., and H. Timmermans. Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences*, Vol. 18, No. 1, 2014, pp. 31–60.
6. Bottai, M., N. Salvati, and N. Orsini. Multilevel models for analyzing people's daily movement behavior. *Journal of Geographical Systems*, Vol. 8, No. 1, 2006, pp. 97-108.
7. Giuliano, G., and D. Narayan. Another look at travel patterns and urban form: the US and Great Britain. *Urban Studies*, Vol. 40, No. 11, 2003, pp. 2295-2312.
8. Goulias, K. G. Multilevel analysis of daily time use and time allocation to activity types accounting for complex covariance structures using correlated random effects. *Transportation*, Vol. 29, No. 1, 2002, pp. 31-48.
9. Maat, K., and H. J. P. Timmermans. A causal model relating urban form with daily travel distance through activity/travel decisions. *Transportation Planning & Technology*, Vol. 32, No. 2, 2009, pp. 115-134.

10. Mercado, R., and A. Páez. Determinants of distance traveled with a focus on the elderly: a multilevel analysis in the Hamilton CMA, Canada. *Journal of Transport Geography*, Vol. 17, No. 1, 2009, pp. 65-76.
11. Van Acker, V., F. Witlox, and B. Van Wee. The Effects of the Land Use System on Travel Behavior: A Structural Equation Modeling Approach. *Transportation Planning & Technology*, Vol. 30, No. 4, 2007, pp. 331-353.
12. Yang, D., H. Timmermans, and A. Grigolon. Exploring heterogeneity in travel time expenditure of aging populations in the Netherlands: results of a CHAID analysis. *Journal of Transport Geography*, Vol. 33, 2013, pp. 170-179.
13. Elldér, E. Residential location and daily travel distances: the influence of trip purpose. *Journal of Transport Geography*, Vol. 34, 2014, pp. 121-130.
14. Banerjee, A., X. Ye, and R. Pendyala. Understanding Travel Time Expenditures Around the World: Exploring the Notion of a Travel Time Frontier. *Transportation*, Vol. 34, No. 1, 2007, pp. 51–65.
15. Cools, M., E. Moons, and G. Wets. Assessing the Impact of Public Holidays on Travel Time Expenditure: Differentiation by Trip Motive. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2157, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 29-37.
16. Golob, T. F. The Dynamics of Household Travel Time Expenditures and Car Ownership Decisions. *Transportation Research Part A: Policy and Practice*, Vol. 24, No. 6, 1990, pp. 443-463.
17. van Wee, B., P. Rietveld, and H. Meurs. Is average daily travel time expenditure constant? In search of explanations for an increase in average travel time. *Journal of Transport Geography*, Vol. 14, No. 2, 2006, pp. 109-122.
18. Bolger, N., A. Davis, and E. Rafaeli. Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, Vol. 54, No. 1, 2003, p. 579-616.
19. Snellen, D., A. Borgers, and H. Timmermans. Urban form, road network type, and mode choice for frequently conducted activities: a multilevel analysis using quasi-experimental design data. *Environment and Planning A*, Vol. 34, No. 7, 2002, pp. 1207- 1220.
20. Bhat, C. R. A multi-level cross-classified model for discrete response variables. *Transportation Research Part B: Methodological*, Vol. 34, No. 7, 2000, pp. 567-582
21. Bhat, C. R., and R. M. Pendyala. Modeling intra-household interactions and group decision-making. *Transportation*, Vol. 32, No. 5, 2005, pp. 443-448.
22. Ho, C., and C. Mulley. Intra-household interactions in transport research: a review. *Transport Reviews*, Vol. 35, No. 1, 2015, pp. 33-55.
23. Kang, H., and D. Scott. Impact of different criteria for identifying intra-household interactions: a case study of household time allocation. *Transportation*, Vol. 38, No. 1, 2011, pp. 81-99.
24. Zhao, X., H. Guan, and L. Gong. Modeling Correlation of Family Holiday Activities Based on Structural Equation Model. *Journal of Transportation Systems Engineering and Information Technology*, Vol. 11, No. 4, 2011, pp. 85-90.
25. Bhat, C., and H. Zhao. The spatial analysis of activity stop generation. *Transportation Research Part B: Methodological*, Vol. 36, No. 6, 2002, pp. 557-575.
26. Schwanen, T., and M. Dijst. Travel-time ratios for visits to the workplace: the relationship between commuting time and work duration. *Transportation Research Part A: Policy and Practice*, Vol. 36, No. 7, 2002, pp. 573-592.

27. Abdel-Aty, M., and F. Abdalla. Understanding Drivers' Route Choice Under Long-Term Pretrip and Short-Term En-Route Traffic Information Using Generalized Estimating Equations. *Journal of Transportation Engineering*, Vol. 130, No. 6, 2004, pp. 777-786.
28. Heinen, E., K. Maat, and B. van Wee. Day-to-Day Choice to Commute or Not by Bicycle. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2230, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 9-18.
29. Hubbard A.E., J. Ahern, N.L. Fleischer, M. Van der Laan, S.A. Lippman, N. Jewell, T. Bruckner, and W.A. Satariano. To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health. *Epidemiology*, Vol. 21, No. 4, 2010, pp. 467-474.
30. Cornelis, E., M. Hubert, P. Hunyen, K. Lebrun, G. Patriarche, A. De Witte, L. Creemers, K. Declercq, D. Janssens, M. Castaigne, L. Hollaert, and F. Walle. *Belgian Daily Mobility (BELDAM): Enquête sur la mobilité quotidienne des belges*. SPF Mobilité & Transports, Brussels, 2012.
31. Zorn, C. J.W. Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *American Journal of Political Science*, Vol. 45, No. 2, 2001, pp. 470-490. (old 16)
32. Molenberghs, G., and G. Verbeke. *Models for Discrete Longitudinal Data*. Springer, New York, 2005. (old 17)
33. Schukken, Y.H., Y.T. Grohn, B. McDermott, and J.J. McDermott. Analysis of correlated discrete observations: background, examples and solutions. *Preventive Veterinary Medicine*, Vol. 59, No. 4, 2003, pp. 223-240. (old 22)
34. Fujii, S., R. Kitamura, and K. Kishizawa. Analysis of Individuals' Joint-Activity Engagement Using a Model System of Activity-Travel Behavior and Time Use. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1676, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 11-19.
35. Carrasco, J.-A., and E. J. Miller. The social dimension in action: A multilevel, personal networks model of social activity frequency between individuals. *Transportation Research Part A: Policy and Practice*, Vol. 43, No. 1, 2009, pp. 90–104.