

Socioeconomic and sociodemographic inequalities and their association
with road traffic injuries

Peer-reviewed author version

PIRDAVANI, Ali; DANIELS, Stijn; VAN VLIERDEN, Karin; BRIJS, Kris & KOCHAN, Bruno (2017) Socioeconomic and sociodemographic inequalities and their association with road traffic injuries. In: Journal of Transport & Health, 4, p. 152-161.

DOI: 10.1016/j.jth.2016.12.001

Handle: <http://hdl.handle.net/1942/23054>

Please cite this article as: Pirdavani, A., et al., Socioeconomic and sociodemographic inequalities and their association with road traffic injuries. Journal of Transport & Health (2016), <http://dx.doi.org/10.1016/j.jth.2016.12.001>

Socioeconomic and Sociodemographic Inequalities and their Association with Road Traffic Injuries

Ali Pirdavani (corresponding author)

Faculty of Engineering Technology of Hasselt University

Address: Agoralaan Gebouw H, BE-3590 Diepenbeek, Belgium

Phone: +32 (0) 11 29 2183 E-mail: ali.pirdavani@uhasselt.be

Stijn Daniels, Karin van Vlierden, Kris Brijs and Bruno Kochan

Transportation Research Institute (IMOB) of Hasselt University

Address: Agoralaan, BE-3590 Diepenbeek, Belgium

Abstract

Increasing evidence suggests that neighborhood-based measures of socioeconomic status are correlated with traffic injury. The main objective of this study is to determine the differences in associations between predictive variables and injury crashes (i.e. including injury and fatal crashes). This study makes a novel contribution by establishing the association between traffic casualties and socio-demographic, socioeconomic characteristics, traffic exposure data and road network variables, at the neighborhood-level while categorized by different genders and transport mode; 'car driver', 'car passenger' and 'active mode users' (i.e. pedestrians and cyclists). In this study an activity-based transportation model called FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) is utilized to produce exposure measures. Exposure measures are in the form of production/attraction trips for several traffic analysis zones (TAZ) in Flanders, Belgium. Analyzing crashes at a neighborhood-level provides important information that enables us to compare traffic safety of different neighborhoods. This information is used to identify safety problems in specific zones and consequently, implementing safety interventions to improve the traffic safety condition. This can be carried out by associating casualty counts with a number of factors (i.e. developing crash prediction models) which have macro-level characteristics, such as socio-demographic and network level exposure. The results indicate that socioeconomic variables are differently associated with casualties of different travel modes and genders. For instance, income level of residence of a TAZ is a significant predictor of male car driver injury crashes while it does not significantly contribute to the prediction of female car driver injury crashes.

Keywords: Crash prediction; Socioeconomic status; Traffic safety; Traffic analysis zone; Activity-based transportation models.

1. Introduction

The association between road crashes and socioeconomic status has been extensively documented during the last decades. A large amount of literature points to the differing road crash involvement of various socioeconomic groups within countries. Several indicators of socioeconomic status (e.g., income, educational level, and occupational status) have already been investigated to that end. Generally it is concluded that low socioeconomic status and deprivation increase the fatality risk (Chen et al., 2010) or the risk of being injured in traffic (Factor et al., 2008; Laflamme et al., 2009). Also in developing countries (Sehat et al., 2012), lower economic level seems to be associated with increased frequency and injury severity of traffic crashes.

Analyses of the link between casualties and deprivation either use information on the socioeconomic status of victims or they take an area-based approach. The typical procedure is to test for association between spatial variation in casualty counts and area levels of deprivation (Graham et al., 2013). Laflamme and Diderichsen (2000) mention two different area-based designs. Some studies group geographic areas by socioeconomic status and compare injury rates across socioeconomic groups. Other studies take the reverse approach and group areas by level of injury risk and compare socioeconomic characteristics across risk levels.

Numerous area-based studies on road crash involvement are conducted from a sociological or health oriented point of view (Chakravarthy et al., 2010; Hosking et al., 2013; Jones et al., 2008; Males, 2009). Morency et al. (2012) conclude that it is plausible to argue that along the causal pathway that leads from neighborhood socioeconomic position to road traffic injuries, the number of people at risk, traffic volume and road characteristics are intermediate variables or mediators. Spoerri et al. (2011) include in their research both variables describing individuals (education, profession, marital status, nationality and household type) and variables describing areas (language region and population density).

Also in the domain of transportation plan development, there is a growing tendency to connect crash counts not only to the transportation characteristics of a zone, but to several socioeconomic and demographic characteristics as well (Zhang et al., 2014). Besides infrastructural characteristics, De Guevara et al. (2004) found population density and number of employees to be significant variables for the injury and property-damage crash prediction models they developed. Cottrill and Thakuriah (2010) explored geographic correlates of pedestrian-vehicle crashes. They particularly emphasized spatial variations of these events in areas with large low-income and minority populations versus other areas. Other studies were conducted in France and focused on all modes (Fleury et al., 2010; Licaj et al., 2011). Fleury et al. (2010) studied the traffic crash risk of residents of deprived areas (i.e. sensitive urban areas) as compared to the risk of residents of control areas. They wanted to differentiate between these risks to guide decisions made on safety actions. More in detail, they integrated both the social- and the spatial dimensions of the safety level. Living in a sensitive urban area induced excessive risk, especially in people over the age of 20, and among both male and female residents (with a much higher relative risk for the male population). Explanations for the origin of such an excessive risk were related to behavior (attitudes and risk-taking among certain social groups, notably young residents of these neighborhoods), and to variations in mobility in the different areal types (e.g., travelling significantly less by car in the sensitive urban areas). A similar study conducted by Licaj et al. (2011), included only young people (under 25) and also considered

the severity of road trauma. It was found that the incidence of injuries as a pedestrian, cyclist or motorist was higher among young people living in deprived municipalities. The severity of road injuries was lower in deprived neighborhoods. Yet, after taking crash characteristics into account, the type of municipality no longer had a significant effect on severity. The authors concluded that deprived areas should be targeted by dedicated education programs as well as by further investigations on urban planning.

Based on the above mentioned literature, the association between socioeconomic status and crash risk may seem straightforward, with a higher risk for people of low socioeconomic status or living in deprived neighborhoods. However, that viewpoint changes when gender is taken into account. The evidence concerning socioeconomic differences in relation to gender is indeed inconsistent. While some studies find a similar social patterning, other studies show diverse or conflicting results across genders (Laflamme et al., 2009). Nolasco et al. (2009) found that motor vehicle crashes represent a greater fatality risk among men from lower socioeconomic level than among men from the most privileged socioeconomic level. Yet, for women, an inconsistent pattern has been found. Borrell et al. (2005) studied the different distributions of fatality by educational level in nine European settings. They found higher death rates in all settings for men with a low educational level. Contrary to that, female traffic fatality showed no inequalities in most of the settings. Only in Finland, Austria and Belgium was there an increase in traffic fatality for women with a lower educational level and of the youngest age group (30-49 years old). A reverse pattern was found in Norway where women with high educational levels had higher injury rates.

Besides socioeconomic variables, other network and exposure variables ought to be considered for model development completeness (Hauer, 2015). If not the most significant predictor of crashes, exposure is a key determinant of traffic safety. The relationship between crash occurrence and exposure is fairly straightforward. The higher the exposure, the greater the possibility for a crash to occur. Exposure can be expressed in different forms such as vehicle kilometers traveled or trip production/attraction (P/A). P/A numbers are basically the number of trips (NOTs) being produced or attracted by each zone. Typically, they are represented by means of origin-destination matrices.

Recently, some researchers constructed crash prediction models by associating crash counts with trip P/A and other network characteristics. Abdel-Aty et al. (2011a) identified and prioritized important variables which can be associated with crashes per traffic analysis zone (TAZ) by means of the classification and regression trees (CART) technique. They showed that considering NOTs can be helpful in predicting safety status for long term transportation planning. Abdel-Aty et al. (2011b) also developed several macroscopic prediction models for different crash severity levels using NOTs as the exposure variable. They concluded that different sets of predictors should be considered based on crash severity or type (e.g., total trip productions and attractions provided better model fit for the total and peak hour crashes while severe crashes were better predicted by different trip motive related variables). Naderan and Shahi (2010) investigated the feasibility of associating travel demand in urban areas with crash frequencies in each TAZ. They developed a series of zonal crash prediction models (ZCPMs) using the P/A trips as predictors. Their conclusion was that these models provide a suitable tool for the evaluation of travel demand management policies in urban transportation planning, also in terms of traffic safety. More specifically, the application of a specific policy may reduce trip P/A for a specific motive. Another asset of considering P/A is that an extensive range of

exposure for other transportation modes (such as pedestrians and cyclists) can also be accounted for. This of course allows a more comprehensive prediction of traffic safety.

Crash prediction models can be developed at different levels of aggregation such as the local level (i.e. road segment or intersection) or the regional level (e.g. TAZ). Recently, crash analyses at a macroscopic level received more attention from the research community. Several studies examined the association of a collection of neighborhood-level factors (e.g., traffic patterns, socio-demographic and socioeconomic variables, land use patterns and weather conditions) with crashes, aggregated according to specific spatial scales (Aguero-Valverde and Jovanis, 2006; Hadayeghi et al., 2010a, 2010b; Huang et al., 2010; Lovegrove and Litman, 2008; Pirdavani et al., 2013, 2012; Wier et al., 2009). Neighborhood-level crash analyses can provide valuable information that would enable cross-sectional neighborhood comparisons, more accurate identification of neighborhood-specific safety problems, and subsequent implementation of appropriate safety interventions (Huang et al., 2010). An additional reason to conduct neighborhood-level crash analyses is that travel demand management policies or safety strategies are commonly performed and evaluated at geographically aggregated levels rather than merely at the level of individual intersections or road sections. Hence, potential impact of these strategies should also be evaluated at a level higher rather than individual road segments. In fact, the relevance of local crash prediction models is limited to the prediction of safety effects of infrastructural improvements at specific sites. For the evaluation of traffic safety impacts of more widely implemented management policies, the application of prediction models at a higher aggregation level (e.g. neighborhood level) will be more practical (Pirdavani, 2012; Tarko et al., 2008). Based on the abovementioned arguments, this study will focus on the association between predictive variables and traffic safety at the neighborhood-level.

Subsequent sections of the paper are outlined as follows. Section two describes the preparation process of the required data and their descriptive statistics. Section three explains the prediction model development procedure, while section four demonstrates the results by means of coefficient estimates of the developed casualty prediction models and explains the associations between casualties and other explanatory variables. Section five discusses the major findings and presents conclusions together with study limitations.

2. Data Preparation

The study area in this research is confined to Flanders. This is the Dutch-speaking region in the northern part of Belgium. Flanders has over 6 million inhabitants, about 60% of the population of Belgium. The main objective is to identify the relationship between the traffic safety condition of TAZs and socio-economic variables (in this case: income level, car ownership, number of employees, number of school children, population and driving license possession). These relationships are extensively scrutinized in the previous literature, which justifies our choice for including them in the initial analysis. Each of these variables has either a causal relationship with traffic casualties or can serve as a proxy measure for other variables of interest.

2.1. Socioeconomic, network and crash data

Income-level and car ownership are socioeconomic variables that reveal the prosperity level of a neighborhood. In the literature, there is a well-established association indicating that deprived neighborhoods experience more crashes and, therefore, casualties (Aguero-Valverde and

Jovanis, 2006; Huang et al., 2010; Noland and Oh, 2004; Pirdavani, 2012; Quddus, 2008; Wier et al., 2009). In addition, average driving license ownership and number of employees in a zone are alternative measures expressing how busy a specific neighborhood is, in terms of daily commuting in general and private vehicle trips in particular. We expect that if the activities in a certain zone increase, the frequency of crash occurrence will increase accordingly. Number of children is also a proxy for the size of this age group. Presence and the magnitude of this age group play a major role in parental mode choice behavior resulting in different crash occurrence patterns (Evers et al., 2014; Hatamzadeh et al., 2016). Here the intention is to verify whether this variable is statistically associated with any type of casualty or not. All socioeconomic variables are collected for each household in the entire study area and are then geographically aggregated to a macroscopic level. This has been carried out at zonal level, comprising 2,200 TAZs in Flanders. The average size of TAZs is 6.09 square kilometers with standard deviation of 4.78 square kilometers and the average number of inhabitants equal to 2416 persons. Moreover, for each TAZ a set of road network variables were collected to construct the ZCPMs. The crash data used in this study consist of casualties derived from a geo-coded set of fatal and injury crashes that occurred during the period 2010 to 2012. The data was provided by Statistics Belgium and by the Flemish Ministry of Mobility and Public Works. Table 1 shows a list of selected variables, together with their definition and descriptive statistics as they have been used in developing the ZCPMs presented in this article.

2.2. Travel demand

In addition to the abovementioned variables, an activity-based model within the FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) framework (Janssens et al., 2007) is applied to the Flemish population. This is done to derive more in-depth information on Flemish peoples' travel behavior and demand. A sequence of 26 decision trees is used in the scheduling process, which is derived from observed travel behavior, by means of the chi-squared automatic interaction detector (CHAID) algorithm. Decisions are based on a number of attributes of the individuals (e.g. age, gender), of the household (e.g. available number of cars) and of the geographical zone (e.g. population density, number of shops). The model simulates whether an activity (e.g., shopping, working, leisure activity, etc.) will be carried out or not for each individual with its specific attributes. Subsequently, amongst others, the location, transport mode, start time and duration of the activity are determined, taking into account the attributes of the individual (Kochan et al., 2008). FEATHERS produces for each individual in the population a schedule consisting of a listing of all consecutive activities and trips during the day. When schedules of all individuals are predicted, it becomes possible to derive traffic demand in the form of origin-destination (OD) matrices. These OD matrices include the number of trips for each traffic mode at different segmentation levels (i.e. age, gender, day of the week, time of day, and trip motive). Within the context of this study, the FEATHERS model output is labeled as "Exposure variable" and are presented in Table 1. Detailed explanation of the input and output data of FEATHERS is provided in the following sections.

Table 1: List of explanatory variables with their definition and descriptive statistics

	Variable	Definition	Average	Min	Max	SD ^a
Dependent variables	Cas_CD_M	total casualties of male car drivers observed in a TAZ	23.52	0	274	28.73
	Cas_CD_F	total casualties of female car drivers observed in a TAZ	14.16	0	156	17.10
	Cas_CP_M	total casualties of male car passengers observed in a TAZ	2.97	0	37	3.94
	Cas_CP_F	total casualties of female car passengers observed in a TAZ	4.70	0	53	6.13
	Cas_SL_M	total casualties of male active mode users observed in a TAZ	9.29	0	156	14.39
	Cas_SL_F	total casualties of female active mode users observed in a TAZ	5.74	0	142	11.11
Exposure variables	Exp_CD_M	daily produced/attracted trips by male car drivers in a TAZ	1620.73	1	15414	1546.7
	Exp_CD_F	daily produced/attracted trips by female car drivers in a TAZ	1129.36	1	7236	1099.2
	Exp_CP_M	daily produced/attracted trips by male car passengers in a TAZ	202.42	1	1210	194.36
	Exp_CP_F	daily produced/attracted trips by female car passengers in a TAZ	481.94	1	2876	451.76
	Exp_SL_M	daily produced/attracted trips by male active mode users in a TAZ	523.13	1	4016	610.84
	Exp_SL_F	daily produced/attracted trips by female active mode users in a TAZ	580.27	1	4846	706.56
Network variables	Speed	average speed limit in a TAZ (km/hr)	69.4	30	120	10.91
	Capacity	hourly average capacity of links in a TAZ	1790.1	1200	7348.1	554.6
	Area	total area of a TAZ in square kilometers	6.09	0.09	45.22	4.78
	Link Length	total length of the road network in a TAZ (km)	15.86	0.39	87.95	10.79
	Intersection	total number of intersections in a TAZ	5.8	0	40	5.9
	Motorway	presence of motorway in a TAZ describes as below: 'No' represented by 0 'Yes' represented by 1	0	0	1	- ^b
	Urban	Is the TAZ in an urban area? 'No' represented by 0 'Yes' represented by 1	0	0	1	-
	Suburban	Is the TAZ in a suburban area? 'No' represented by 0 'Yes' represented by 1	0	0	1	-
Socio-economic variables	Driving License	average driving license ownership in a TAZ describes as below: 'No' represented by 0 'Yes' represented by 1	1	0	1	-
	Income Level	average income of residents in a TAZ describes as below: 'Monthly salary less than 2249 Euro' represented by 0 'Monthly salary more than 2250 Euro' represented by 1	1	0	1	-
	School Children	total number of school children in a TAZ	364.09	0	9245	772.4
	Population	total number of inhabitants in a TAZ	2614.52	0	15803	2582.1
	Employees	total number of employees in a TAZ	888.73	0	16286	1574.9
	Car Ownership	average car ownership per household in a TAZ	1.13	0	14	0.47
	Total Cars	Total private cars owned by households of a TAZ	1119.26	0	6772	1057.6

a: Standard deviation

b: Data not applicable.

2.3. FEATHERS input data

The FEATHERS system is equipped with a dedicated data module. The data module provides access to the data that needs to be accessible throughout all other modules' activities inside the FEATHERS. Four major types of data are provided by the data module, namely: observed travel behavior data, level-of-service data, land-use data, and microscopic population data (Bellemans et al., 2010).

Observed travel behavior

Observed travel behavior input data is used in order to estimate FEATHERS' sub-models so that person's travel behavior in the study area can be captured and used in future forecasts. The observed travel behavior data used in FEATHERS was collected by the Flemish government and is called "Onderzoek VerplaatsingsGedrag" (OVG) (Moons, 2009).

The OVG data has been collected from August 2007 till June 2008 and comprises four random samples of the Flemish population. Each sample was a two-step stratified sample of inhabitants of six years or more from the National Register. In order to prevent the survey office from having to go to all different towns to do one or only a few interviews, respondents were clustered by town. So if one potential respondent was sampled in a town, at least fifteen persons in that town were sampled. Strictly speaking the postal code was the primary sampling unit.

The number of clusters from a postal code was determined randomly taking into account the number of inhabitants (the higher the number of inhabitants, the higher the probability that a cluster was selected for that postal code). First the number of clusters from a postal code was determined proportionally to the number of inhabitants. This sampling was done independently for the four quarters. The second step was to select 15 persons for each cluster. For instance, if two clusters were to be taken from a specific postal code, this meant that for that postal code $2 \times 15 = 30$ persons were selected at random. The raw sample consists of 12.000 persons' information. That number was distributed in sets of 3000 persons per quarter. Hence, 200 clusters of 15 persons were selected per quarter.

The government decided to abandon the contact procedure by telephone and instead, utilized a face-to-face contact practice where a representative of the government passed by the person's house. The survey procedure was also set up in such a way that in case of drop-outs of respondents, new respondents were selected and contacted. The face-to-face survey was done with the help of a computer. Each respondent was asked to fill in information on household and person characteristics questionnaires, and also general questionnaire on travel behavior. After this first contact, the respondent was handed over a booklet that had to be used to register all trips during a randomly chosen day. After this was done, all trip information was entered into the computer during a second face-to-face contact. In total 8932 and 8934 persons filled out the personal questionnaire and the household questionnaire, respectively (i.e. the response rate was about 75%).

The raw data format of the OVG data is not directly readable by FEATHERS. Hence, a pre-processing step is needed to create so-called diaries, a term commonly used in the domain of activity-based modelling. These diaries are the input format for FEATHERS. In case FEATHERS receives such diaries, it is able to estimate sub-models based on these data in order to capture travel behavior that can be used in a later stage to predict travel behavior for the full Flemish population.

LOS data

Level-of-service data is a qualitative measure used to describe the quality of traffic service for different kinds of available transport modes. These level-of-services have an impact on choices people make when choosing a transport option or alternative for their commute trips, leisure trips, and etc. The availability of LOS data allows us implement LOS related scenarios where one can alter measurements of LOS. These alterations will then have impacts on personal

choices. The LOS data used in FEATHERS was derived by Significance, a Dutch research institute located at The Hague with strong expertise in quantitative research on mobility and transport.

Land-use data

Land-use data contains information on geographical zones in the study area such as the attractiveness of a zone for conducting certain activities. In FEATHERS, land-use data helps us determine where people can perform their activities. Individuals can often choose between various locations to conduct their activities. At the same time, however, the physical system restricts human behavior. The fact that a particular location only has a single function, implies that only one activity type can be conducted at that location. Land-use data is, therefore, required to represent the opportunities available to travelers. During the activity-travel scheduling process inside FEATHERS, the availability of locations is determined based on sector employment size at the lowest possible geographical level.

Microscopic population data

Microscopic population data is an important input for FEATHERS as it is the starting point for predicting person activity/trip schedules. It is typically used in activity-based modelling as opposed to traditional models where person information is commonly aggregated on the zonal level. In FEATHERS, individuals are characterized by a set of attributes such as their age, gender, driving license possession, and etc. These person attributes are then used to start modelling the activities and the trips that people perform throughout the day. In Flanders, we employed the CENSUS data provided by the government, which was collected for all individuals residing in Flanders.

2.4. FEATHERS output data

FEATHERS is an implementation of an activity-based demand model that simulates individuals' daily activity-travel schedules, which are consistent in time and space (Bellemans et al., 2010). FEATHERS succeeds in this by gradually building a schedule for a day starting from scratch by sequentially modelling various activity-related decisions such as: decision to go to work, at what time, where, with which transport mode, whether or not to have a lunch break, and so on. When making these decisions, characteristics of the individual together with the information of its household, and the environment (including the access to different transport modes) are taken into account.

FEATHERS makes all of these decisions while taking into account constraints in terms of time (no overlapping activities), space (you can only perform activities in one place at a time), time-space (travel takes time and is not instantaneous, so typically only a subset of locations are within reach given a finite time window of opportunity), and social constraints (household composition, availability of transport modes to household members). The result is a detailed listing of what individuals do (in terms of activities and trips) throughout the day, and this for each member of the population in the study area. As a result, travel demand can be derived for the entire population and be segmented according to any desired dimension.

3. Model development

Crash data consist of non-negative integers (i.e. number of casualties). Since ordinary least-squares regression serves continuous dependent variables (e.g. crash rate), it is not an

appropriate technique to model crash counts or casualties (Lord and Mannering, 2010). For decades, researchers applied Poisson regression models for crash prediction analysis. Because of the natural characteristic of crash data that variance does not necessarily equals to mean, application of Poisson regression models becomes risky as it might bias the results by making parameter estimates inconsistent (Lord and Mannering, 2010). To overcome this problem, the negative binomial (NB) model which allows the variance to differ from the mean, was applied as an extension of the Poisson model. The NB model is the most commonly used model in crash data modeling when over-dispersion is observed (Lord and Mannering, 2010).

Crash analysis can be conducted at different levels of aggregation. Microscopic crash analysis usually concerns an infrastructural element such as a specific road segment or an intersection. Different from that, macroscopic crash analysis concerns a relatively large area and associates the probability of crash occurrence with a set of variables that is generally macroscopically characterized. Several efforts have been made to develop macro-level crash prediction models (Abdel-Aty et al., 2011a; Huang et al., 2010; Lovegrove and Litman, 2008; Pirdavani et al., 2014a, 2014b, 2013, 2012; Quddus, 2008) based on the association of crashes with predictive variables that have macro-level characteristics, like socio-economic-, exposure-, and network variables.

Application of crash prediction models at TAZ level was initially introduced by Levine and colleagues who based it on a linear regression (Levine et al., 1995). Notwithstanding, as already mentioned, the application of NB models in crash prediction analysis became increasingly popular (Aguero-Valverde and Jovanis, 2006; De Guevara et al., 2004; Naderan and Shahi, 2010; Pirdavani et al., 2013, 2012). This was due to the fact that crash data usually have a greater variance compared to mean and NB models being better able to handle this over-dispersion. In this study, due to the observed over-dispersed crash data, the NB models were developed within the generalized linear modeling (GLM) framework.

A literature review showed that the model form below has been widely exercised by different researchers (Abdel-Aty et al., 2011b; Pirdavani, 2012). In this formulation the exposure is expressed separately from other potential variables. This is done because in case of 0 exposure, 0 crashes are expected.

$$E(C) = \beta_0 \times (Exposure)^{\beta_i} \times e^{\sum \beta_i x_i} \quad (1)$$

Where;

$E(C)$: expected crash frequency,

β_0 and β_i : model parameters,

$Exposure$: exposure variable (e.g. produced/attracted trips), and

x_i : other explanatory variables.

Logarithmic transformation of equation 1 when considering only one exposure variable yields:

$$\ln[E(C)] = \ln(\beta_0) + \beta_1 \ln(Exposure) + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n \quad (2)$$

Several models are constructed to associate the relationships between casualties of different road users and the explanatory variables. In this study ZCPMs are developed for ‘car driver’,

‘car passenger’ and ‘active mode user’ (i.e. pedestrians and cyclists) as transportation modes and for different genders (i.e. male vs. female) separately.

4. Results

As previously mentioned, distinct models were constructed by regressing the casualty counts of different road users in each TAZ together with other socio-economic network variables on the natural logarithmic transformation of produced/attracted trips, with the latter operating as the exposure variables. The results show that among the considered predictive variables, exposure measures together with ‘Income level’, ‘Car ownership’, ‘Capacity’, ‘Motorway’, ‘Urban’ and ‘Suburban’ were statistically significant in predicting certain types of casualties.

For all models, exposure variables were the most significant predictors and positively associated with the number of casualties in each TAZ. As the number of trips increases, casualties also tend to increase. Similar associations between the number of produced/attracted trips and traffic safety have been reported by other researchers (Abdel-Aty et al., 2011a, 2011b; Naderan and Shahi, 2010; Pirdavani et al., 2012). In order to incorporate the impact of moving motorized traffic, some network variables are considered to serve as proxy measures. For all car occupant casualties (i.e. car drivers and car passengers) the variable ‘Motorway’ turned out to be a significant predictor. The positive association between this variable and casualties means that for TAZs with motorways, a higher number of car driver and passenger casualties are expected. For active mode users’ casualty models, ‘Capacity’ was found to be a significant predictor. The positive sign indicates that TAZs with greater road network capacity are expected to have more traffic in addition to the number of produced/attracted trips. Consequently, potential conflicts between motorized vehicles and active mode users are expected to be higher, leading to more vulnerable road user casualties.

The degree of urbanization was categorized into three different levels and thus, represented by two dummy variables, i.e., ‘Urban’ and ‘Suburban’. When ‘Urban’ and ‘Suburban’ metrics in a TAZ were both 0, then this TAZ was located in a rural area. For all models, the coefficient estimate of the variable ‘Urban’ was larger than the one for ‘Suburban’. This means that the models predict more casualties for more urbanized TAZs. This is in line with the findings of Huang et al. (2010) who also found that counties with a higher level of urbanization were associated with higher crash occurrence probability.

Given that our main objective was to assess the link between socio-economic variables and traffic safety, these associations are discussed in more detail. As can be seen in Table 2, all constructed models showed a negative association between casualties and ‘Income Level’. These results are in line with many other studies where it has been shown that lower income level or poverty has a positive relationship with crashes occurring in a TAZ (Aguero-Valverde and Jovanis, 2006; Pirdavani, 2012). The negative sign for the ‘Income Level’ variable indicates that TAZs with higher income level are expected to have fewer casualties compared to less prosperous TAZs. Despite this negative association for all models, ‘Income level’ was not a significant predictor at 90% confidence level for each type of casualty. Average household income level at the TAZ level was found to be a significant predictor for all types of male casualties. Although we found a negative association between income level and female casualties, this association was less considerable in predicting female car driver and active mode road user casualties. This indicates that active mode or car driver casualties among females are

not significantly predicted by their household income level. Car ownership, as another socioeconomic variable, was also found to be a significant predictor of male active mode road users only. This suggests that increased car ownership level leads to less male active mode road user casualties. In general, these results show that female casualties are overall less well explained by household socioeconomic conditions in comparison to male casualties.

Table 2: Parameter estimates for casualty prediction models

Model results for male car driver casualties					Model results for female car driver casualties			
Variable	Estimate	Std. Error	z value	Pr(> z)	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.4276	0.1785	7.996	0.000	0.9219	0.1778	5.184	0.000
Ln(Exp_CD_M)	0.2449	0.0242	10.134	0.000	-	-	-	-
Ln(Exp_CD_F)	-	-	-	-	0.2504	0.0251	9.963	0.000
Income level	-0.219	0.0824	-2.658	0.008	-0.1184	0.0874	-1.354	0.176
Urban	0.7383	0.1167	6.325	0.000	0.5796	0.1236	4.689	0.000
Suburban	0.2552	0.0576	4.433	0.000	0.2126	0.0613	3.470	0.000
Motorway	0.233	0.0529	4.405	0.000	0.2181	0.056	3.894	0.000
Model results for male car passenger casualties					Model results for female car passenger casualties			
Variable	Estimate	Std. Error	z value	Pr(> z)	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3918	0.1621	2.418	0.016	0.3541	0.1822	1.943	0.052
Ln(Exp_CP_M)	0.1777	0.029	6.122	0.000	-	-	-	-
Ln(Exp_CP_F)	-	-	-	-	0.2152	0.0289	7.449	0.000
Income level	-0.3593	0.0967	-3.715	0.000	-0.2192	0.0955	-2.294	0.022
Urban	0.5261	0.135	3.898	0.000	0.6638	0.1329	4.996	0.000
Suburban	0.1842	0.0689	2.674	0.007	0.1584	0.0673	2.355	0.019
Motorway	0.1947	0.0631	3.084	0.002	0.1411	0.0617	2.288	0.022
Model results for male active mode user casualties					Model results for female active mode user casualties			
Variable	Estimate	Std. Error	z value	Pr(> z)	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.4436	0.2018	2.198	0.028	-0.5847	0.2424	-2.412	0.016
Ln(Exp_SL_M)	0.2589	0.025	10.345	0.000	-	-	-	-
Ln(Exp_SL_F)	-	-	-	-	0.3018	0.0297	10.175	0.000
Income level	-0.1584	0.0953	-1.661	0.097	-0.0803	0.1157	-0.694	0.488
Car ownership	-0.1209	0.0587	-2.06	0.039	-0.879	0.0697	-1.26	0.208
Capacity	0.000228	0.000047	4.883	0.000	0.00027	0.000056	4.718	0.000
Urban	0.7241	0.1375	5.261	0.000	1.02	0.1651	6.179	0.000
Suburban	0.2295	0.0678	3.386	0.000	0.3661	0.0821	4.459	0.000

5. Discussion and conclusion

In this study we associated the traffic safety condition of neighborhoods (i.e. at the traffic analysis zone level) with socioeconomic, network and traffic exposure variables. On the one hand epidemiologists and sociologists primarily focus on socioeconomic and sociodemographic characteristics of the defined population when analyzing the patterns, potential causes and health related impacts of transportation. On the other hand, traffic exposure is the main variable that is predominantly employed by traffic engineers when studying traffic safety. In either of the approaches little attention is given to the key explanatory variables of the other approach. To overcome this shortcoming we used the combination of these factors in the most detailed way possible where the traffic exposure and casualty data are disaggregated by gender and transport mode. To establish this association, several macroscopic crash prediction models were developed to predict the number of casualties of different road users (including car drivers, car passengers and active mode road users) for male and female separately. This helps us distinguish how certain variables may explain male and female casualties of different road users differently. Model development was carried out at the zonal level comprising of 2200 TAZs in Flanders, Belgium. Given that the crash data were over-dispersed, negative binomial models were developed within the generalized linear modeling framework.

Developing prediction models at a macro level has its merits. The main reason for taking this approach is that several explanatory variables are of macroscopic nature. Trip production/attraction, average household income or car ownership are all characteristics of zones. Therefore, if prediction models are fit on these variables, this should also be done at that level. An additional advantage of zonal crash prediction models as developed in this study, is that they are more suitable for the assessment of safety interventions and travel management policies that are defined, implemented and evaluated at a macroscopic level.

The results revealed that the most important predictor of casualties for all travel mode types is exposure. Urbanization degree is also a significant predictor. An important notice is that, in this study, we try to explain the number of casualties instead of the risk (i.e., the average number of crashes per unit of exposure). Although the risk of certain crash types seems to be higher in less urbanized areas, the number of casualties tends to be higher in urbanized zones as they are more densely inhabited compared to rural areas. Concerning the socioeconomic variables, average household income level and car ownership were found to be significant predictors of certain types of casualties. Concerning the inclusion of the variable “Car ownership” only in active mode users’ models and its exclusion in the car occupant models, we should indicate that this variable was not found to be a statistically significant predictor of those casualty types. One potential reason could be that car ownership level of a certain zone, as a more socioeconomic type of variable, has a minor contribution to the motorized traffic density of that zone. In other words, due to the fact that most of the car trips are not intra-zonal trips, average car ownership level of a zone does not significantly define the magnitude of motorized vehicle exposure of that zone whereby its contribution to the motorized vehicle crash occurrence is insubstantial. On the contrary, level of car ownership plays an important role in the active mode users’ crash occurrence. Car ownership is a decisive variable when considering travel mode choice. Arguably, low level of car ownership leaves household members with the choice to either cycle or walk, particularly for intra-zonal or short distance trips to the adjacent neighborhoods. The negative association between car ownership and active mode casualties – reported in Table 2 – confirms this hypothesis.

Although the magnitude of the association between socioeconomic variables and casualties varies in function of gender, the directions are the same and meet the expectations. In general, higher income levels or car ownership rates are associated with fewer casualties. However, these effects are less pronounced for women as compared to men. The underlying cause of this difference could be that mode choice is gender biased. There is indeed clear evidence in the literature indicating that travel mode choice behavior is significantly influenced by gender. Polk (2004) and Vance et al. (2005) reported that women use private cars considerably less than men and are more inclined to reduce car usage as compared with men. Moreover, in households with fewer cars than drivers, men tend to use the available car(s) more than women, implying a gender inequality in access to available household cars (e.g., Vance and Iovanna, 2007). This enhances the understanding of why certain variables are not significantly predicting female casualties. The aforementioned argument upholds the general perception of gender biased car usage resulting in more male drivers as opposed to more female passengers. In a broader perspective, this indicates that household socioeconomic status has little impact on women mode choice behavior in general. This is confirmed by the model results presented in Table 2, where the average household income level is only a significant predictor of male driver casualties compared with female drivers, while there is not much of substantial difference observed for the associations with car passenger casualties. Car ownership as another variable

is found to have a negative and yet significant association only with male active mode users. Based on the discussion above, the increased car availability in a household is expected to increase car usage mostly for men, and thus result in fewer male active mode users. When in most cases men have the priority to take the available car of the household, then “Car ownership” does not affect women’s choice of travel mode anymore. This statement best justifies the triviality of the “Car ownership” variable in explaining female active mode user casualties.

The most important policy relevant finding concerns the gender inequality we observed in the developed casualty prediction models. This contributes to the general notion of gender unfairness in transportation accessibility. This finding implies the importance of mainstreaming gender difference considerations and incorporating gender equality into transport policy. This can be manifested by means of tailor-made policies targeting the population of interest exclusively. In addition, government can utilize this knowledge to support priority setting in investments, such as road reconstruction, (re)design of infrastructure, and awareness raising campaigns to name a few. Moreover, the macroscopic approach that has been followed in this study helps policy makers appropriately evaluate transport policies. Since the majority of transport policies concerns relatively large geographical entities (e.g. a neighborhood or a collection of them) rather than local road infrastructure, we should aim for a conforming model structure. The macroscopic characteristic of the developed models is in compliance with this prerequisite.

References

- Abdel-Aty, M., Siddiqui, C., Huang, H., 2011a. Zonal Level Safety Evaluation Incorporating Trip Generation Effects. Presented at the Transportation Research Board (TRB) 90th Annual Meeting, Washington D.C. USA.
- Abdel-Aty, M., Siddiqui, C., Huang, H., 2011b. Integrating Trip and Roadway Characteristics in Managing Safety at Traffic Analysis Zones. Presented at the Transportation Research Board (TRB) 90th Annual Meeting, Washington D.C. USA.
- Agüero-Valverde, J., Jovanis, P.P., 2006. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention* 38, 618–625. doi:10.1016/j.aap.2005.12.006
- Bellemans, T., Kochan, B., Janssens, D., Wets, G., Arentze, T., Timmermans, H., 2010. Implementation Framework and Development Trajectory of FEATHERS Activity-Based Simulation Platform. *Transportation Research Record: Journal of the Transportation Research Board* 2175, 111–119. doi:10.3141/2175-13
- Borrell, C., Plasència, A., Huisman, M., Costa, G., Kunst, A., Andersen, O., Bopp, M., Borgan, J.-K., Deboosere, P., Glickman, M., Gadeyne, S., Minder, C., Regidor, E., Spadea, T., Valkonen, T., Mackenbach, J.P., 2005. Education level inequalities and transportation injury mortality in the middle aged and elderly in European settings. *Inj. Prev.* 11, 138–142. doi:10.1136/ip.2004.006346
- Chakravarthy, B., Anderson, C.L., Ludlow, J., Lotfipour, S., Vaca, F.E., 2010. The relationship of pedestrian injuries to socioeconomic characteristics in a large Southern California County. *Traffic Inj Prev* 11, 508–513. doi:10.1080/15389588.2010.497546
- Chen, H.Y., Senserrick, T., Martiniuk, A.L.C., Ivers, R.Q., Boufous, S., Chang, H.Y., Norton, R., 2010. Fatal crash trends for Australian young drivers 1997–2007: geographic and socioeconomic differentials. *J Safety Res* 41, 123–128. doi:10.1016/j.jsr.2009.12.006
- Cottrill, C.D., Thakuriah, P. (Von), 2010. Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis & Prevention* 42, 1718–1728. doi:10.1016/j.aap.2010.04.012
- De Guevara, F.L.D., Washington, S., Oh, J., 2004. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board* 1897, 191–199. doi:10.3141/1897-25
- Evers, C., Boles, S., Johnson-Shelton, D., Schlossberg, M., Richey, D., 2014. Parent safety perceptions of child walking routes. *Journal of Transport & Health* 1, 108–115. doi:10.1016/j.jth.2014.03.003
- Factor, R., Mahalel, D., Yair, G., 2008. Inter-group differences in road-traffic crash involvement. *Accident Analysis & Prevention* 40, 2000–2007. doi:10.1016/j.aap.2008.08.022
- Fleury, D., Peytavin, J.F., Alam, T., Brenac, T., 2010. Excess accident risk among residents of deprived areas. *Accident Analysis & Prevention* 42, 1653–1660. doi:10.1016/j.aap.2010.04.004
- Graham, D.J., McCoy, E.J., Stephens, D.A., 2013. Quantifying the effect of area deprivation on child pedestrian casualties by using longitudinal mixed models to adjust for confounding, interference and spatial dependence. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176, 931–950. doi:10.1111/j.1467-985X.2012.01071.x
- Hadayegehi, A., Shalaby, A., Persaud, B., 2010a. Development of Planning-Level Transportation Safety Models using Full Bayesian Semiparametric Additive Techniques. *Journal of Transportation Safety & Security* 2, 45–68. doi:10.1080/19439961003687328
- Hadayegehi, A., Shalaby, A.S., Persaud, B.N., 2010b. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis & Prevention* 42, 676–688. doi:10.1016/j.aap.2009.10.016
- Hatamzadeh, Y., Habibian, M., Khodaii, A., 2016. Walking behavior across genders in school trips, a case study of Rasht, Iran. *Journal of Transport & Health*. doi:10.1016/j.jth.2016.08.011
- Hauer, E., 2015. *The Art of Regression Modeling in Road Safety*. Springer International Publishing, Cham.

- Hosking, J., Ameratunga, S., Exeter, D., Stewart, J., Bell, A., 2013. Ethnic, socioeconomic and geographical inequalities in road traffic injury rates in the Auckland region. *Aust N Z J Public Health* 37, 162–167. doi:10.1111/1753-6405.12034
- Huang, H., Abdel-Aty, M., Darwiche, A., 2010. County-Level Crash Risk Analysis in Florida. *Transportation Research Record: Journal of the Transportation Research Board* 2148, 27–37. doi:10.3141/2148-04
- Janssens, D., Wets, G., Timmermans, H.J.P., Arentze, T.A., 2007. Modelling Short-Term Dynamics in Activity-Travel Patterns: Conceptual Framework of the Feathers Model. Presented at the 11th World Conference on Transport Research, Berkeley CA, USA.
- Jones, A.P., Haynes, R., Kennedy, V., Harvey, I.M., Jewell, T., Lea, D., 2008. Geographical variations in mortality and morbidity from road traffic accidents in England and Wales. *Health Place* 14, 519–535. doi:10.1016/j.healthplace.2007.10.001
- Kochan, B., Bellemans, T., Janssens, D., Wets, G., 2008. Assessing the Impact of Fuel Cost on Traffic Demand in Flanders Using Activity-Based Models. Presented at the Travel Demand Management TDM, Vienna, Austria.
- Laflamme, L., Burrows, S., Hasselberg, M., 2009. Socioeconomic differences in injury risks. A review of findings and a discussion of potential countermeasures. World Health Organization, Regional Office for Europe, Copenhagen, Denmark.
- Laflamme, L., Diderichsen, F., 2000. Social differences in traffic injury risks in childhood and youth—a literature review and a research agenda. *Inj Prev* 6, 293–298. doi:10.1136/ip.6.4.293
- Levine, N., Kim, K.E., Nitz, L.H., 1995. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis & Prevention* 27, 675–685. doi:10.1016/0001-4575(95)00018-U
- Licaj, I., Haddak, M., Hours, M., Chiron, M., 2011. Deprived neighborhoods and risk of road trauma (incidence and severity) among under 25 year-olds in the Rhône Département (France). *Journal of Safety Research* 42, 171–176. doi:10.1016/j.jsr.2011.05.004
- Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* 44, 291–305. doi:10.1016/j.tra.2010.02.001
- Lovegrove, G.R., Litman, T., 2008. Using Macro-Level Collision Prediction Models to Evaluate the Road Safety Effects of Mobility Management Strategies: New Empirical Tools to Promote Sustainable Development. Presented at the Transportation Research Board (TRB) 87th Annual Meeting, Washington D.C. USA.
- Males, M.A., 2009. Poverty as a determinant of young drivers' fatal crash risks. *J Safety Res* 40, 443–448. doi:10.1016/j.jsr.2009.10.001
- Moons, E., 2009. Onderzoek Verplaatsingsgedrag Vlaanderen 3. Vlaamse Overheid -Departement Mobiliteit en Openbare Werken - Afdeling Beleid, Mobiliteit en Verkeersveiligheid.
- Morency, P., Gauvin, L., Plante, C., Fournier, M., Morency, C., 2012. Neighborhood social inequalities in road traffic injuries: the influence of traffic volume and road design. *Am J Public Health* 102, 1112–1119. doi:10.2105/AJPH.2011.300528
- Naderan, A., Shahi, J., 2010. Aggregate crash prediction models: Introducing crash generation concept. *Accident Analysis & Prevention* 42, 339–346. doi:10.1016/j.aap.2009.08.020
- Noland, R.B., Oh, L., 2004. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. *Accident Analysis & Prevention* 36, 525–532. doi:10.1016/S0001-4575(03)00058-7
- Nolasco, A., Melchor, I., Pina, J.A., Pereyra-Zamora, P., Moncho, J., Tamayo, N., García-Senhermes, C., Zurriaga, O., Martínez-Beneito, M.A., 2009. Preventable avoidable mortality: Evolution of socioeconomic inequalities in urban areas in Spain, 1996–2003. *Health & Place* 15, 732–741. doi:10.1016/j.healthplace.2008.12.003
- Pirdavani, A., 2012. Proactive Approaches in Quantitative Micro- and Macroscopic Crash Analysis: Focus on Model-Based Safety Evaluation of Traffic Policy Measures (Doctoral dissertation). Hasselt University, Belgium.
- Pirdavani, A., Bellemans, T., Brijs, T., Kochan, B., Wets, G., 2014a. Assessing the road safety impacts of a teleworking policy by means of geographically weighted regression method. *Journal of Transport Geography* 39, 96–110. doi:10.1016/j.jtrangeo.2014.06.021

- Pirdavani, A., Bellemans, T., Brijs, T., Wets, G., 2014b. Application of Geographically Weighted Regression Technique in Spatial Analysis of Fatal and Injury Crashes. *Journal of Transportation Engineering* 140, 04014032. doi:10.1061/(ASCE)TE.1943-5436.0000680
- Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2013. Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling. *Accident Analysis & Prevention* 50, 186–195. doi:10.1016/j.aap.2012.04.008
- Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., Wets, G., 2012. Application of Different Exposure Measures in Development of Planning-Level Zonal Crash Prediction Models. *Transportation Research Record: Journal of the Transportation Research Board* 2280, 145–153. doi:10.3141/2280-16
- Polk, M., 2004. The influence of gender on daily car use and on willingness to reduce car use in Sweden. *Journal of Transport Geography* 12, 185–195. doi:10.1016/j.jtrangeo.2004.04.002
- Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention* 40, 1486–1497. doi:10.1016/j.aap.2008.03.009
- Sehat, M., Naieni, K.H., Asadi-Lari, M., Foroushani, A.R., Malek-Afzali, H., 2012. Socioeconomic Status and Incidence of Traffic Accidents in Metropolitan Tehran: A Population-based Study. *Int J Prev Med* 3, 181–190.
- Spoerri, A., Egger, M., von Elm, E., 2011. Mortality from road traffic accidents in Switzerland: Longitudinal and spatial analyses. *Accident Analysis & Prevention* 43, 40–48. doi:10.1016/j.aap.2010.06.009
- Tarko, A., Inerowicz, M., Ramos, J., Li, W., 2008. Tool with Road-Level Crash Prediction for Transportation Safety Planning. *Transportation Research Record: Journal of the Transportation Research Board* 2083, 16–25. doi:10.3141/2083-03
- Vance, C., Buchheim, S., Brockfeld, E., 2005. Gender as a Determinant of Car Use: Evidence from Germany, in: *Transportation Research Board Conference Proceedings. Presented at the Conference on Research on Women's Issues in Transportation*, Washington, D.C., pp. 59–67.
- Vance, C., Iovanna, R., 2007. Gender and the Automobile: Analysis of Nonwork Service Trips. *Transportation Research Record: Journal of the Transportation Research Board* 2013, 54–61. doi:10.3141/2013-08
- Wier, M., Weintraub, J., Humphreys, E.H., Seto, E., Bhatia, R., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention* 41, 137–145. doi:10.1016/j.aap.2008.10.001
- Zhang, C., Yan, X., Ma, L., An, M., Zhang, C., Yan, X., Ma, L., An, M., 2014. Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones, Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones. *Mathematical Problems in Engineering*, Mathematical Problems in Engineering 2014, 2014, e987978. doi:10.1155/2014/987978, 10.1155/2014/987978