Developing Zonal Crash Prediction Models with a Focus on Application of Different Exposure Measures

Ali Pirdavani, Tom Brijs,
Tom Bellemans, Bruno Kochan and Geert Wets[1]

Transportation Research Institute (IMOB)
Hasselt University
Wetenschapspark – Building 5
B-3590 Diepenbeek, Belgium

Tel: +32(0)11 26 91{39, 55, 27, 47, 58}
Fax: +32(0)11 26 91 99

Email: {ali.pirdavani, tom.brijs, tom.bellemans,

bruno.kochan, geert.wets}@uhasselt.be

Total number of words: 6524 (2 Tables and 2 Figures included)
Date of 1st submission: 1/08/2011
Date of 2nd submission: 30/10/2011

[1] Corresponding author

ABSTRACT

Assessing the safety impacts of Travel Demand Management (TDM) policies is essential to be carried out by means of a proactive approach. Since TDM policies are typically implemented at an aggregate level, Crash Prediction Models (CPMs) should also be developed at a similar level of aggregation. These models should match better with the resolution at which TDM evaluations are performed. Therefore Zonal Crash Prediction Models (ZCPMs) are considered to construct the association between observed crashes and a set of predictor variables in each zone. This is carried out by the Generalized Linear Modeling (GLM) procedure with the assumption of Negative Binomial (NB) error distribution. Different exposure, network and socio-demographic variables of 2200 Traffic Analysis Zones (TAZs) are considered as predictors of crashes in the study area, Flanders, Belgium. To this end, an activity-based transportation model framework is applied to produce exposure measurements. Crash data used in this study consist of recorded injury crashes between 2004 and 2007. The network and socio-demographic variables are also collected from other sources. In this study, different ZCPMs are developed to predict the Number of Injury Crashes (NOICs); including fatal, severely and slightly injury crashes. These models are classified into three different groups, i.e. 1) flow-based models, 2) trip-based models and 3) a combination of the two. The results show a considerable improvement of the model performance when both trip-based and flow-based exposure variables are used simultaneously in the model's formulation. The main purpose of this study is to provide a predictive tool at the planning-level which can be applied on different TDM policies to evaluate their traffic safety impacts.

INTRODUCTION

For many years, researchers have attempted to investigate the negative impacts of growing travel demand on traffic safety by predicting the Number of Crashes (NOCs) based on the patterns they have learnt from crashes that occurred in the past. Traditionally, this reactive approach consists of different phases such as; identification, diagnosis and improvement of unsafe locations, so called hot-spots. From the ethical point of view, this reactive approach is not acceptable because it requires several years of crashes to occur in order to identify and treat safety problems. Thus, providing a more proactive approach, which is capable of evaluating road safety at the planning-level, is essential. This proactive approach is increasingly being paid attention to by researchers and practitioners in the last few years. Dealing with traffic safety at the planning-level requires the ability to integrate TDM policies into a crash predicting context. TDM policies are usually performed and evaluated at a more aggregate level than just on the level of individual intersections or road section. However, local TDM implementations like adding capacity to a segment of road may also be conducted. Typically the before/after analysis of such an infrastructure adjustment is carried out locally despite the fact that such an adjustment may have broader consequences. Thus, the impact of adopting a TDM strategy on transportation or traffic safety should be evaluated at a higher level rather than merely the local consequences. Therefore, application of CPMs at a zonal level like TAZ leads to ZCPM.

The main goal of this study is therefore to develop ZCPMs that can be used to evaluate the traffic safety effects of conducted TDM policies. Exposure is an important determinant of traffic safety. Therefore, it is needed to assess the exposure under different TDMs to be able to evaluate their traffic safety impacts. To this end, the FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) activity-based transportation model is applied on the Flemish population. The FEATHERS framework (1) was developed in order to facilitate the development of activity-based models for transportation demand in Flanders, Belgium. Currently, the framework is fully operational at the level of Flanders. The real-life representation of Flanders is embedded in an agent-based simulation model which consists of over six million agents, each agent representing one member of the Flemish population. A sequence of 26 decision trees is used in the scheduling process and decisions are based on a number of attributes of the individual (e.g. age, gender), of the household (e.g. number of cars) and of the geographical zone (e.g. population density, number of shops). For each agent with its specific attributes, the model simulates whether an activity (e.g. shopping, working, leisure activity …) is going to be carried out or not. Subsequently, amongst others, the location, transport mode and duration of the activity are determined, taking into account the attributes of the individual (*2*). As such, the FEATHERS activity-based model can provide the exposure measure, Number of Trips (NOTs), by means of (time-of-day dependent) Origin-Destination (OD) matrices. Assigning these OD matrices of car trips to the Flemish road network provides other exposure variables like Vehicle Kilometers Traveled (VKT) and Vehicle Hours Traveled (VHT). These three different exposure variables together with network and socio-demographic variables are then used to construct the ZCPMs. Since the exposure which comes out of the activity-based model is sensitive to TDM policies, these ZCPMs are also TDM sensitive.

The structure of this paper is as follows. Initially, the literature about crash prediction modeling at the zone level will be reviewed. In the next sections the data preparation and model development will be demonstrated. Finally, the results of ZCPMs will be shown followed by the final conclusions.

LITERATURE REVIEW

CPMs can be categorized in two different levels: the local level (e.g. road and intersection) and the regional level (e.g. TAZ). Usually CPMs at the local level aim to predict the safety benefits/detriment of infrastructure improvements. These models are not typically designed to evaluate traffic safety impacts of TDM policies; thus, application of CPMs at a higher aggregation level will be more practical (*3*). Recently, the application of ZCPMs became more popular amongst researchers because of their ability to estimate the effect of different TDM policies on traffic safety. This has been initially introduced by Levine et al (*4*). In their study a set of both socio-economic and network variables were chosen to predict the NOCs in different TAZs. They estimated a linear relationship between different explanatory variables and the NOCs. Several researchers have examined the association of a collection of network infrastructure, socio-demographic and socio-economic variables and weather conditions with the NOCs at the level of TAZs (*5-12*). De Guevara et al. (*10*) developed planning-level ZCPMs for the city of Tucson, Arizona. They considered many socio-demographic and network variables in their model construction. They concluded that predictors such as population density, the number of persons younger than 17 years old as a percentage of the total population, the number of employees, the intersection density, the percentage of miles of principal arterials, the percentage of miles of minor arterials and the percentage of miles of urban collectors are significant predictors for the NOCs. In a study carried out by Wier et al (*11*) it was shown that traffic volume, population size, the proportion of arterial streets without public transit, the proportion of the population living in poverty, and the number of people aged 65+ as a percentage of the total population, were significantly good predictors. Moreover, Noland and Quddus (*9*) concluded that TAZs with high employment density had more traffic crashes, while in urbanized more densely populated TAZs fewer crashes have been observed.

Hadayeghi et al (*13-18*) have been working on ZCPMs for several years. In one of their first studies, it was shown that the number of accidents in a TAZ increases when the VKT, major and minor road length, total employed labor force, household population, and intersection density increase and it decreases with a higher posted speed and a higher level of congestion in the TAZ (*14*). Hadayeghi et al (*15*) investigated the temporal transferability of the ZCPMs by applying models constructed on 1996 data to predict the NOCs for each TAZ in 2001 for the City of Toronto. In another research, twenty-three regression models were developed to examine the relationships between several types of transportation planning variables and collision frequency. The results showed the potential of planning-level safety models to serve as decision support tools for planners to consider safety in the planning phase (*16*). Hadayeghi et al (*17*) conducted the same research but this time they applied Geographically Weighted Poisson Regression (GWPR) instead of taking Generalized Linear Modeling (GLM) approach. The major difference between these two types of models is that GWPR models allow the model coefficient estimates to vary spatially for each TAZ. This very important additional attribute of these models provides some extra information as it takes the spatial location of a crash into consideration. Lovegrove and Sayed (*19*) concluded that quantifying the relationship between the zonal characteristics such as exposure, network, socio-demographic and TDM variables (e.g. Total commuters from each zone or commuter density) and crashes at a zonal level provides a predictive tool to predict the NOCs in a TAZ. They have used GLM techniques to develop ZCPMs for both urban and rural areas across the Greater Vancouver Regional District (GVRD). The results of their study show that increasing signal density, intersection density per unit area and lane kilometers, arterial-local intersections in rural areas and total arterial road lane kilometers will lead to an increase in the NOCs. On the contrary, an increase in the number of three-leg intersections and local road lane kilometers will decrease the NOCs in a TAZ. Lovegrove and Sayed (*20*) further developed a set of ZCPMs for a "black-spot" study in GVRD. These sets of ZCPMs consist of an exposure variable (i.e. VKT) and other network, socio-demographic and TDM variables. The results of this study also confirmed that ZCPMs have the ability to round out traditional reactive road safety improvement programs. An et al (*21*) found VHT, the number of intersections and the number of households with low income level to be correlated with the NOCs in TAZs.

To conclude, many variables such as traffic volume, VHT, VKT, population, employment, level of income, urbanization degree, traffic intensity, number of intersections and intersection density, speed and road length are confirmed in different studies to be significant predictors of crashes at the zonal level.

Recently, some researchers constructed ZCPMs by associating the NOCs in a TAZ with trip production/attraction and other network characteristics. Abdel-Aty et al (*22*) identified and prioritized important variables which can be associated with crashes per TAZ by means of the Classification and Regression Trees (CART) technique. It was shown that this methodology will be helpful in incorporating proactive safety measures for long range transportation planning. Abdel-Aty et al (*23*) also developed different ZCPMs for different crash severity levels using the NOTs as the exposure variable. They concluded that different sets of predictors should be considered based on the type or severity of crashes (e.g. total trip productions and attractions provide 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 (*24*) investigated the feasibility of associating travel demand in urban areas with crash frequencies in each TAZ. They developed a series of ZCPMs using the NOTs produced/attracted as predictors. They concluded that these models provide the basic tool for evaluating TDM policies in urban transportation planning in terms of traffic safety as the application of a specific TDM policy may reduce trip productions of a specific motive.

The drawback of considering only trips as an exposure variable is that the impact of trip time, trip length, route choice, intrazonal traffic and transit traffic on a TAZ will be neglected. The number of produced/attracted trips might be an acceptable indicator of how busy or active a TAZ is or how much people are exposed to unsafe situations, but it always leaves out the effects of through traffic which is just passing through a TAZ neither having their origin or destination in that TAZ. It is a well known relationship in literature that road crashes are tightly linked to traffic exposure (*13-24*); therefore, having a more informative measure of exposure, is expected to result in a better crash prediction. This study demonstrates the impact of applying different exposure variables on the performance of ZCPMs.

DATA PREPARATION

The required information to construct ZCPMs consists of exposure, network and socio-demographic data accompanied with the crash data. These data should be collected for the whole study area and also be aggregated to the zonal level. The study area in this research is the Dutch speaking region in northern Belgium, Flanders. Flanders has over 6 million inhabitants, about 60% of the population of Belgium. As already mentioned, an activity-based model within the FEATHERS framework is applied on the Flemish population to derive the in depth information of Flemish peoples' travel behavior and travel demand for a null-scenario (current situation). The basic outputs of FEATHERS are activity-travel schedules/diaries. These can then be aggregated to OD matrices. These OD matrices include the NOTs for each traffic mode at different disaggregation levels (i.e. age, gender, day of the week, time of a day and motive). This traffic demand is then assigned to the network to obtain detailed exposure measures at the network level (i.e. VHT and VKT). These network level exposure measures are then aggregated to TAZ level. This has been carried out at the zonal level comprising of 2200 TAZs. The average size of TAZs is 6.09 square kilometers with standard deviation of 4.78 square kilometers. In addition, for each TAZ a set of variables including socio-demographic and network variables were derived to construct the ZCPMs.

The crash data used in this study consist of a geo-coded set of injury crashes that have occurred during the period 2004 to 2007 and were provided by the Flemish Ministry of Mobility and Public Works. Table 1 shows a list of selected variables, together with their definition and descriptive statistics, which have been used in developing the ZCPMs presented in this paper.

MODEL DEVELOPMENT

Crash data consist of non-negative integers, so using ordinary least-squares regression which serves continuous dependent variables (e.g. time) is not an option (*25*). For decades researchers applied Poisson Regression model for crash prediction analysis. Because of the natural characteristics 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 (*25*). 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 (*25*).

Application of CPMs at TAZ level has been initially introduced by Levine et al based on a Linear Regression (*4*). As mentioned before, application of NB model in crash prediction analysis became popular amongst many researchers (*6-10*, *13-16*, *19-21*, *23-27*) . This is due to the fact that usually crash data have a greater variance compared to mean, therefore NB model can handle this over-dispersion better. In this study, NB models were developed within the Generalized Linear Modeling GLM framework.

Reviewing the literature for different model forms showed that the following model has been widely used by different researchers (*13*, *21*, *23*, *27*):

TABLE 1 List of Explanatory Variables for the ZCPMs with Their Definition and Descriptive Statistics

	Variable	Definition	Average	Min	Max	SD[a]
	Crash	total NOICs observed in a TAZ	36.03	0	326	41.58
Exposure variables	Number of Trips	average daily number of trips originating/destined from/to a TAZ	2765.8	0	18111.4	2869.8
	Total Flow	Average Annual Daily Traffic (AADT) in a TAZ (vehicle)	96414.5	70.9	4423325	181695
	VHT	total daily vehicle hours traveled in a TAZ	608.26	1.50	9998.6	930.29
	VKT	total daily vehicle kilometers traveled in a TAZ	52533.8	84.06	985192	90715.2
	Motorway Flow	AADT of motorways in a TAZ (vehicle)	37724.96	0	3881777	146757.5
	Motorway VHT	total daily vehicle hours traveled on motorways in a TAZ	260.52	0	9762.5	832.97
	Motorway VKT	total daily vehicle kilometers traveled on motorways in a TAZ	27471.82	0	946152.8	84669.53
	Other Roads Flow	AADT of other roads in a TAZ (vehicle)	58690.29	0	734152.5	73632.5
	Other Roads VHT	total daily vehicle hours traveled on other roads in a TAZ	348.51	0	3777.69	358.76
	Other Roads VKT	total daily vehicle kilometers traveled on other roads in a TAZ	26662.85	0	303237.6	28133.04
	V/C	average volume to capacity in a TAZ	0.0478	0	0.5697	0.0422
Network variables	Speed	average speed limit in a TAZ (km/hr)	69.4	31	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
	No. of Links	number of links in a TAZ	39.27	1	230	30.46
	Link Length	total length of the links in a TAZ (km)	15.86	0.39	87.95	10.79
	Link Density	link length per square kilometers in a TAZ	3.37	0.03	20.44	2.41
	Intersection	total number of intersections in a TAZ	5.8	0	40	5.9
	Intersection Density	number of intersection per square kilometers	1.76	0	50.63	3.39
	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-demographic 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	-
	Work Status	average work status of the residents in a TAZ describes as below: "Don't work" represented by 0 "Work" represented by 1	1	0	1	-
	Population	total number of inhabitants in a TAZ	2614.52	0	15803	2582.6
	Population Density	population per square kilometers	774.14	0	14567.4	1398.4
	Adults Population	total number of adult inhabitants in a TAZ	1796.06	0	12014	1823.5
	Adults Population Density	adults population per square kilometers	542.85	0	10444.8	1013.4

a: Standard deviation b: Data not applicable.

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

Where;

$E(C)$: expected crash frequency,
β_0 and β_i : model parameters,
$Exposure$: exposure variable (e.g. VHT, VKT or NOTs), and
x_i : other explanatory variables.

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

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

Several models were constructed to associate the relationship between crash frequency and the explanatory variables while the main focus is on the application of different exposure measures. The models can be categorized into three different groups based on the type of exposure measure that was utilized, i.e. 1) flow-based models, 2) trip-based models and 3) models based on a combination of the two. Flow-based models were constructed by regressing the NOICs in each TAZ on VHT or VKT, as the exposure variables, and a selection of network and socio-demographic variables listed in Table 1. Trip-based models use the same network and socio-demographic variables but use NOTs as the exposure variable. In the third type of models, both flow and trip based variables are included simultaneously as measures of exposure.

Initially in the flow-based models, VKT and VHT were computed based on all types of roads together. The Preliminary results showed that all models overestimated the NOICs in the TAZs in which a significant length of motorway was present. Based on this observation it was concluded that it is necessary to make the distinction between different road types (i.e. motorways and non-motorway roads) and consider each of their exposure values separately.

Coefficients were estimated by using a forward selection procedure by taking the intercept and one of the exposure variables for the starting point and then additional candidate variables were selected from the available data described in Table 1. When the exposure variable is included in the model, the next step is to include the second variable with the smallest p-value. This procedure continues until the remaining candidate variables have a p-value of higher than 0.05. At this point the final model is obtained. For each model, the multicollinearity phenomenon was also checked for by means of computing the variance inflation factor (VIF) for all variables which are present in the model. The results of this test did not show any multicollinearity problem for any of the developed models.

The data used in this study consists of the information from 2200 TAZs. This provides a sufficient number of cases with respect to sample size. Therefore for model development, 70% of the TAZs were chosen randomly as training set and the rest of 30% were used as a test set. Applying the developed models to the test sets shows a significant correlation between the observed and the predicted NOICs for all developed models (see Figure 1). The Pearson Correlation Coefficients (PCC) of all developed models shown in Table 2 indicate that all models are capable of predicting the NOICs quite well; however, more statistical tests are needed to be able to select the best fitted model (from the alternatives). This analysis will be carried out in the next section. The regression results of the developed models are presented in Table 2.

MODELING RESULTS

As already mentioned before, models were constructed by regressing the NOICs in each TAZ on the natural logarithmic transformation of NOTs, VHT or VKT, as the exposure variables, in addition to other network and socio-demographic variables. Different statistical tests were used to assess the goodness-of-fit of each developed model. The results of the analysis show that one or two exposure variables together with the variables (See Table 1 for a more detailed description of the variables) V/C, Capacity, Speed, Intersection, Income level, Population, Urban and Suburban were statistically significant at 95% confidence level.

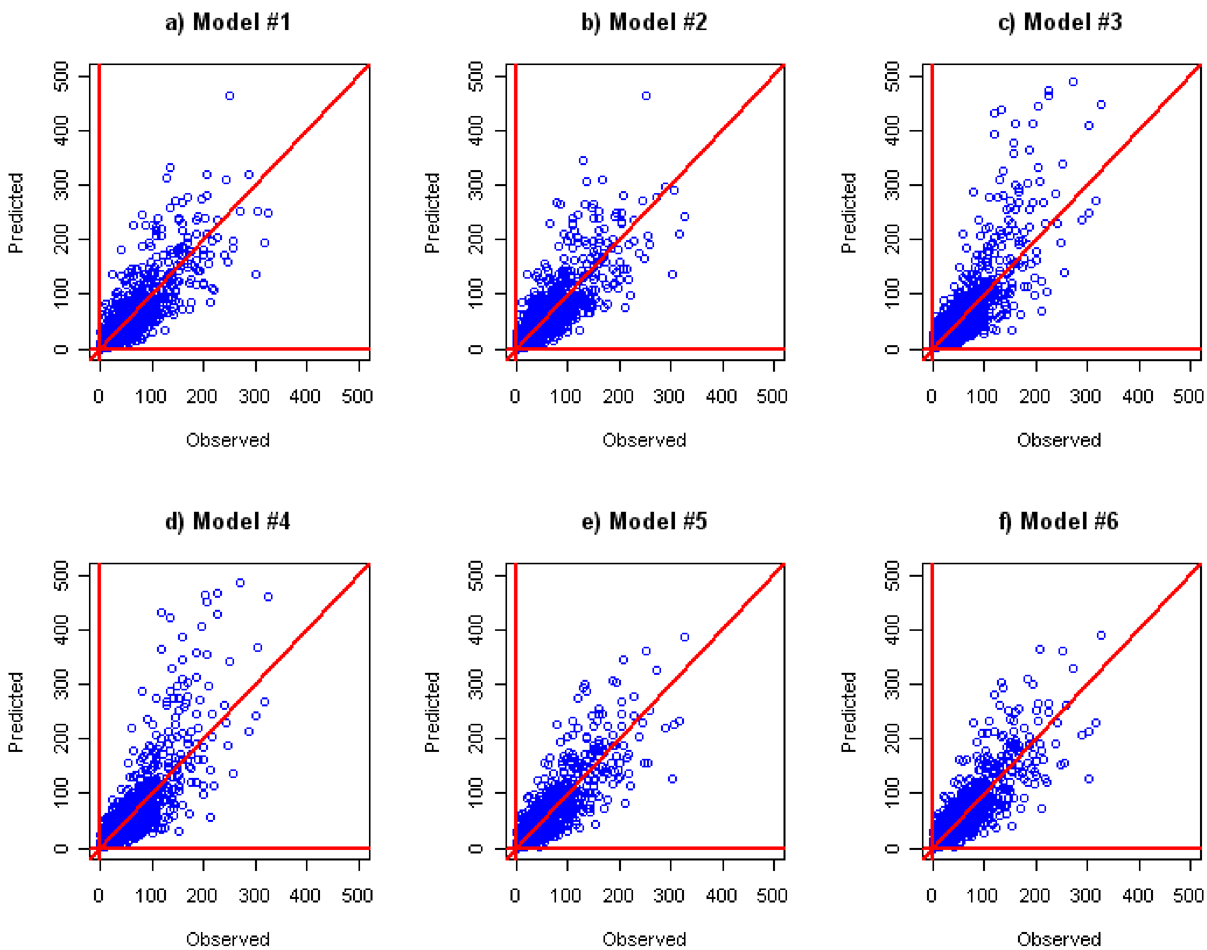

FIGURE 1 Correlations between the observed and the predicted NOICs.

For all models, exposure variables were positively associated with the NOICs in each TAZ. As the NOTs, VHT or VKT increases, total NOICs also tends to increase. Many studies found similar association between VKT (*13-20*, *27*, *28*), VHT (*21*), NOTs (*22-24*) and NOICs per TAZ. For the trip based model V/C is also positively associated with the NOICs, whereas capacity shows a greater contribution in predicting the NOICs for other models (Table 2).

TABLE 2 Regression Results of the Developed ZCPMs

	Model #1	Model #2	Model #3	Model #4	Model #5	Model #6
Coefficients	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
(Intercept)	-3.713e+00	-1.885e+00	-3.744e-01	-2.226e+00	-2.886e+00	-4.141e+00
log(Number of Trips)	7.338e-01	6.659e-01	-[a]	-	4.676e-01	4.520e-01
log(Motorways VKT)	-	-	-	9.472e-03	-	7.744e-03
log(Other Roads VKT)	-	-	-	4.267e-01	-	3.132e-01
log(Motorways VHT)	-	-	1.025e-02	-	7.717e-03	-
log(Other Roads VHT)	-	-	4.143e-01	-	3.040e-01	-
log(V/C)	-	1.883e-01	-	-	-	-
Speed	9.806e-03	-	-	-	-	-
Capacity	3.157e-04	3.190e-04	4.317e-04	3.913e-04	4.220e-04	3.894e-04
Intersection	4.184e-02	4.433e-02	3.220e-02	3.271e-02	2.844e-02	2.888e-02
Income level	-1.160e-01	-1.308e-01	-5.401e-02	-5.879e-02	-1.056e-01	-1.071e-01
Urban	3.111e-01	-	3.182e-01	4.848e-01	2.287e-01	3.520e-01
Suburban	3.487e-02	-	1.624e-01	2.047e-01	5.712e-02	9.095e-02
Population	-	-	1.372e-04	1.299e-04	2.340e-05	2.293e-05
Deviance/DF[b]	1.1386	1.1365	1.1331	1.1302	1.1365	1.1355
AIC[c]	17353	17255	17196	17166	16921	16918
MSPE[d]	581.04	583.59	1141.17	1089.96	482.41	489.74
PCC[e]	0.8458	0.8448	0.8247	0.8276	0.8709	0.8697
R^2	0.7155	0.7138	0.6802	0.6848	0.7584	0.7564

a: Data not applicable
b: Degree of Freedom (DF)
c: Akaike Information Criterion (AIC)
d: Mean Squared Prediction Error (MSPE)
e: Pearson Correlation Coefficient (PCC) between observed and predicted crash values

Positive correlation of average speed limit and number of intersections with NOICs per TAZ can be observed for all models. This positive relationship has also been reported in other studies (*10*, *14*, *19*, *21*, *23*). This can be explained as injury crashes are more likely to occur at higher speeds. In general, intersections have a higher risk of experiencing conflicts compared to road links because of their natural design, therefore there are more crashes expected to occur in TAZs that have a higher number of intersections. Population is also found to have a positive association with the NOICs. It can be explained as in the TAZs with a higher number of inhabitants, there will be more people exposed to unsafe situations compared to TAZs with fewer inhabitants. The same association has been recognized by other researchers (*4*, *5*, *11*). As it can be observed in Table 2, all of the constructed models showed negative association with "Income Level" unlike other explanatory variables, which have positive signs. These results are in line with other studies' findings. It has been shown in many studies that poverty has a positive relationship with the crashes that occurred in a TAZ (*5-9*, *11*, *12*, *21*). The negative sign for the "Income Level" variable indicates that TAZs with higher income level are expected to have fewer crashes compared to less prosperous TAZs. In all of the models, coefficient estimate of the variable “Urban” is more than the coefficient estimate of “Suburban”. This means that the model is correctly predict more crashes for more urbanized TAZs. If a TAZ is located in a rural area, both variables relevant to urbanization degree will be zero and subsequently there will be less crashes predicted for rural areas. This is in line with the findings of Huang et al (*5*) that counties with a higher level of urbanization are associated with higher crash risk. In general, it can be concluded that all of the parameters’ signs are along the lines of theoretical expectations and findings of other previously published studies.

To select the best fitted model, different criteria were taken into consideration. The Akaike Information Criterion (AIC) is a measure of the relative goodness of fit. AIC is defined as:

$$AIC = 2k - 2ln(L)$$

Where;
k : the number of parameters in the model and
L : the maximized value of the likelihood function for the estimated model.

Models may be ranked according to their AIC values in a sense that the preferred model is the one with the minimum AIC value.

Another measure that has been used in comparative analysis between different models is Mean Squared Prediction Error (MSPE) (*13*). The MSPE is the sum of the squared differences between predicted and observed crashes divided by the sample size. The MSPE is defined as:

$$MSPE = {\sum_{i=1}^{n} (\mathrm{P_i} - \mathrm{O_i})^2}\Big/_{\mathrm{n}}$$

Where;
$\mathrm{P_i}$: predicted NOICs for $\mathrm{i_{th}}$ TAZ,
$\mathrm{O_i}$: observed NOICs for $\mathrm{i_{th}}$ TAZ and
n : total number of TAZs

Comparing different models performance shows a significant improvement in models which are developed based on both types of exposure variables; NOTs and assigned traffic on the network (i.e. VHT or VKT). The maximum AIC value stands for the trip-based model (Model #1) (See Table 2). It indicates that this model achieves the poorest fit on the data compared to other models. The model which includes V/C as an exposure related model also performs badly (Model #2). Flow-based models (Models #3 and #4) provide a better fit compared to trip-based model according to their AIC values; however their MSPE values are the greatest among other models. Goodness-of-fit measures for combined exposure models (Models #5 and #6) signify the better performance of these models compared to the ones that include only one of the exposure variables. As it can be seen in Table 2, Models #5 and #6 provide almost equally the minimum MSPE and AIC values, therefore it can be concluded that these models are the best fitted models. In Figure 2, the observed and the predicted NOICs are displayed for each TAZ. Darker TAZs have higher observed/predicted NOICs compared to brighter TAZs. By comparing Figures 2(a), 2(b) and 2(c), a relatively similar pattern can be noticed. This is an indication that the final chosen models (Models #5 and #6) are capable of predicting the NOICs quite well.

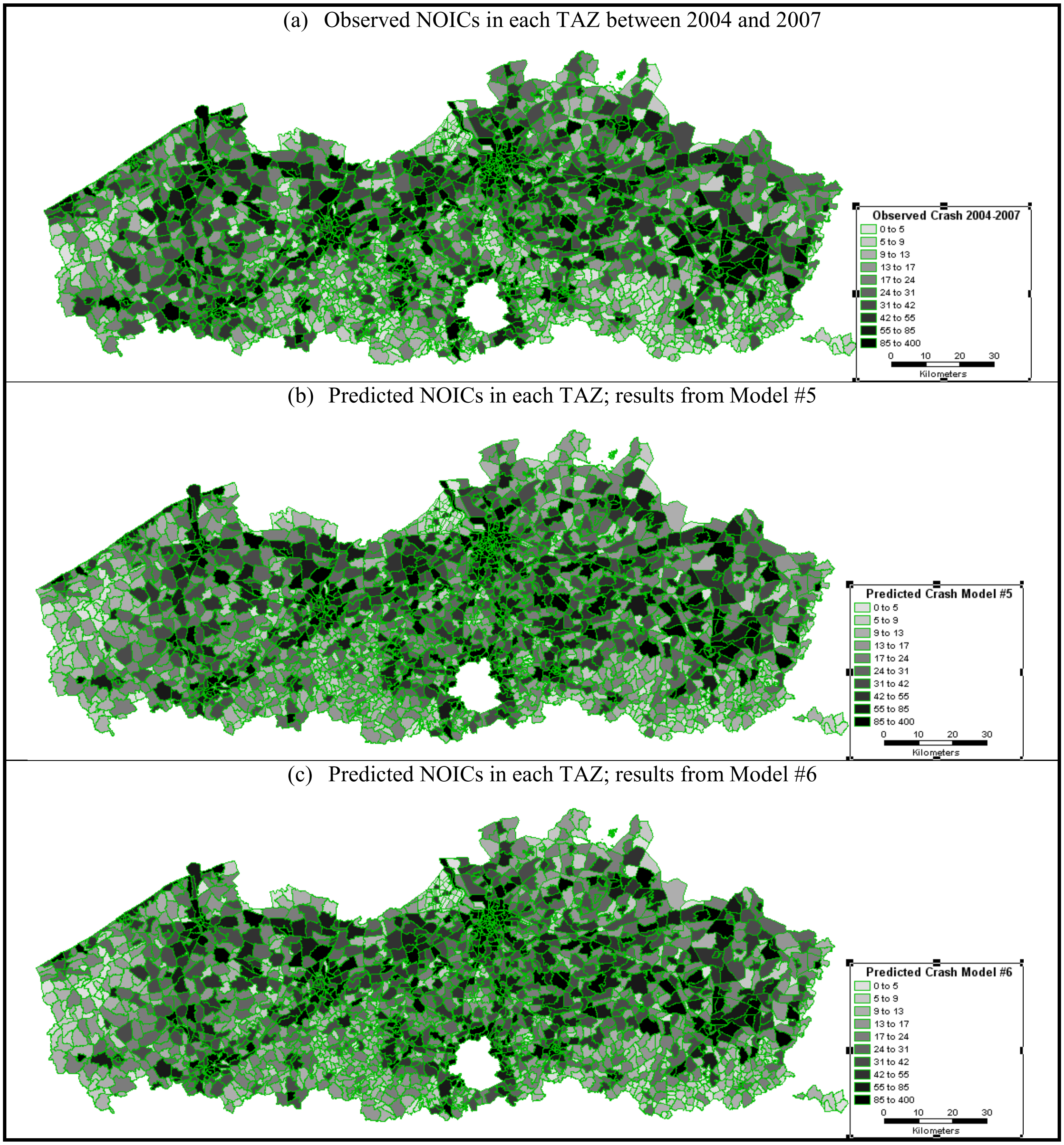

FIGURE 2 Graphical representation of observed and expected NOICs for each TAZ in Flanders.

CONCLUSIONS

In this study a proactive approach was presented, which can be applied in transportation safety planning. Different ZCPMs were developed in order to associate Number of Injury Crashes (NOICs) with different exposure, network and socio-demographic variables at the zonal level comprising of 2200 TAZs in Flanders, Belgium. To this end, Negative Binomial (NB) models were developed within the Generalized Linear Modeling (GLM) framework. This approach has been widely followed by other researchers when the crash data is over-dispersed. The results of the analysis showed the different performance of the models when different exposure variables are considered to be included in model development. The contributing variables and their measure of effectiveness also provide some information that can be used by researchers and practitioners to evaluate the impact of each variable in predicting crashes. This can serve as a useful traffic safety input into transportation projects at the planning-level.

In order to assess TDM's traffic safety implications, it is essential to have TDM sensitive exposure measures. Activity-based transportation models provide an adequate range of in-depth information about individuals' traveling behavior. The advantage of using activity-based transportation models is that these models can be adjusted to simulate different TDM scenario and therefore a wide range of traffic safety evaluation studies can be carried out based on their output. In this study, traffic demand was prepared by the activity-based transportation model's outputs, FEATHERS.

Based on the results presented in this paper the following conclusions can be drawn:

- Different exposure, network and socio-demographic variables have been considered for model development. Coefficients were estimated by using a forward selection procedure. Variable selection has been carried out until all remaining variables have a p-value of higher than 0.05. For each developed model multicollinearity was checked and the results didn't show any multicollinearity issues. Positive or negative association of all selected variables with the NOICs has been checked by comparing them with the results of other studies reported in the literature. The results found are in line with what can be found in literature.
- Sole use of NOTs originating/destining from/to a TAZ for crash prediction will result missing some important information about the characteristics of travel demand; i.e. NOTs, as an exposure variable, is not sensitive to trip time, trip length and route choice. Moreover, transit traffic which is just passing through a TAZ can have a significant share of the exposure observed in a TAZ. This part of the exposure is left out by only using the NOTs. Thus, other exposure variables which are sensitive to the impacts of trip assignment should be taken into account. This has been carried out by assigning the traffic demand to the network using an equilibrium assignment and by computing exposure variables that are sensitive to the assignment like VHT and VKT.
- Different models were developed based on the different measures of exposure that were generated. These models were categorized into three groups according to the exposure measures used as independent variables, i.e. 1) flow-based models, 2) trip-based models and 3) a combination of the two. The results of the model comparison showed that the models that contain the combination two exposure variables outperform the models which only have one of the exposure variables (NOTs or VHT/VKT) in their formulation. Therefore considering the application of both flow-based and trip-based exposure variables in ZCPM construction is recommended.

Spatial and temporal transferability of the model and application of other model types would also be some directions for future studies.

REFERENCES

(1) Janssens, D., G. Wets, H. J. P. Timmermans, and T. A. Arentze. 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, 2007.

(2) Kochan, B., T. Bellemans, D. Janssens, and G. Wets. Assessing the Impact of Fuel Cost on Traffic Demand in Flanders Using Activity-Based Models. Presented at the Travel Demand Management TDM, Vienna, Austria, 2008.

(3) Tarko, A. P., M. Inerowicz, J. Ramos, and W. Li. Tool with Road-Level Crash Prediction for Transportation Safety Planning. In *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2083, no. 1, 2008, pp. 16-25.

(4) Levine, N., K. E. Kim, and L. H. Nitz. Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. *Accident Analysis & Prevention*, vol. 27, no. 5, 1995, pp. 675-685.

(5) Huang, H., M. Abdel-Aty, and A. Darwiche. County-Level Crash Risk Analysis in Florida. In *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2148, no. 1, 2010, pp. 27-37.

(6) Amoros, E., J. L. Martin, and B. Laumon. Comparison of road crashes incidence and severity between some French counties. *Accident Analysis & Prevention*, vol. 35, no. 4, 2003, pp. 537-547.

(7) Aguero-Valverde, J. and P. P. Jovanis. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, vol. 38, no. 3, 2006, pp. 618-625.

(8) Noland, R. B. and L. Oh. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. *Accident Analysis & Prevention*, vol. 36, no. 4, 2004, pp. 525-532.

(9) Noland, R. B. and M. A. Quddus. A spatially disaggregate analysis of road casualties in England. *Accident Analysis & Prevention*, vol. 36, no. 6, 2004, pp. 973-984.

(10) de Guevara, F. L., S. P. Washington, and J. Oh. Forecasting crashes at the planning level. A simultaneous negative binomial crash model applied in Tucson, Arizona. In *Transportation Research Record: Journal of the Transportation Research Board*, 2004, pp. 191-199.

(11) Wier, M., J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, vol. 41, no. 1, 2009, pp. 137-145.

(12) Quddus, M. A. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention*, vol. 40, no. 4, 2008, pp. 1486-1497.

(13) Hadayeghi, A. Use of Advanced Techniques to Estimate Zonal Level Safety Planning Models and Examine Their Temporal Transferability. University of Toronto, 2009.

(14) Hadayeghi, A., A. Shalaby, and B. Persaud. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. In *Transportation Research Record*: *Journal of the Transportation Research Board*, vol. 1840, no. 1, 2003, pp. 87-95.

(15) Hadayeghi, A., A. S. Shalaby, B. N. Persaud, and C. Cheung. Temporal transferability and updating of zonal level accident prediction models. *Accident Analysis & Prevention*, vol. 38, no. 3, 2006, pp. 579-589.

(16) Hadayeghi, A., A. Shalaby, and B. Persaud. Safety Prediction Models: Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. In *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2019, no. 1, 2007, pp. 225-236.

(17) Hadayeghi, A., A. S. Shalaby, and B. N. Persaud. Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, vol. 42, no. 2, 2010, pp. 676-688.

(18) Hadayeghi, A., A. Shalaby, and B. Persaud. Development of Planning-Level Transportation Safety Models using Full Bayesian Semiparametric Additive Techniques. *Journal of Transportation Safety & Security*, vol. 2, no. 1, 2010, pp. 45-68.

(19) Lovegrove, G. R. and T. Sayed. Macro-Level Collision Prediction Models for Evaluating Neighbourhood Traffic Safety. *Canadian Journal of Civil Engineering*, vol. 33, 2006, pp. 609-621.
(20) Lovegrove, G. R. and T. Sayed. Macrolevel Collision Prediction Models to Enhance Traditional Reactive Road Safety Improvement Programs. In *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2019, no. 1, 2007, pp. 65-73.
(21) An, M., C. Casper, and W. Wu. Using Travel Demand Model and Zonal Safety Planning Model for Safety Benefit Estimation in Project Evaluation. Presented at 90th Annual Meeting of the Transportation Research Board, Washington D.C. USA, 2011.
(22) Abdel-Aty, M., C. Siddiqui, and H. Huang. Zonal Level Safety Evaluation Incorporating Trip Generation Effects. Presented at 90th Annual Meeting of the Transportation Research Board, Washington D.C. USA, 2011.
(23) Abdel-Aty, M., C. Siddiqui, and H. Huang. Integrating Trip and Roadway Characteristics in Managing Safety at Traffic Analysis Zones. Presented at 90th Annual Meeting of the Transportation Research Board, Washington D.C. USA, 2011.
(24) Naderan, A. and J. Shahi. Aggregate crash prediction models: Introducing crash generation concept. *Accident Analysis & Prevention*, vol. 42, no. 1, 2010, pp. 339-346.
(25) Lord, D. and F. Mannering. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, 2010, pp. 291-305.
(26) Lovegrove, G. R. and T. Litman. 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 87th Annual Meeting of the Transportation Research Board, Washington D.C. USA, 2008.
(27) Lovegrove, G. R. Community-Based, Macro-Level Collision Prediction Models. University of British Columbia, 2005.
(28) Karlaftis, M. G. and A. P. Tarko. Heterogeneity considerations in accident modeling. *Accident Analysis & Prevention*, vol. 30, no. 4, 1998, pp. 425-433.