The role of exposure in the analysis of road accidents: a Belgian case-study

F. Van den Bossche, G. Wets and T. Brijs

PROMOTOR ▶ Prof. dr. Geert Wets
ONDERZOEKSLIJN ▶ Kennis verkeersonveiligheid
ONDERZOEKSGROEP ▶ LUC, PHL, VUB, Vito
RAPPORTNUMMER ▶ RA-2005-54

UNIVERSITAIRE CAMPUS
GEBOUW D
B 3590 DIEPENBEEK
T ▶ 011 26 87 05
F ▶ 011 26 87 00
E ▶ info@steunpuntverkeersveiligheid.be
I ▶ www.steunpuntverkeersveiligheid.be

The role of exposure in the analysis of road accidents: a Belgian case-study

RA-2005-54

F. Van den Bossche, G. Wets and T. Brijs

Onderzoekslijn Kennis verkeersonveiligheid

DIEPENBEEK, 2005.
STEUNPUNT VERKEERSVEILIGHEID.

Documentbeschrijving

Rapportnummer:	RA-2005-54
Titel:	The role of exposure in the analysis of road accidents: a Belgian case-study
Ondertitel:	
Auteur(s):	F. Van den Bossche, G. Wets and T. Brijs
Promotor:	Prof. dr. Geert Wets
Onderzoekslijn:	Kennis verkeersonveiligheid
Partner:	Limburgs Universitair Centrum
Aantal pagina's:	23
Trefwoorden:	exposure, road safety, ARMA models, regression, time series
Projectnummer Steunpunt:	1.3
Projectinhoud:	Analyse van de impact van mobiliteit op de verkeersveiligheid

Uitgave: Steunpunt Verkeersveiligheid, January 2005.

Steunpunt Verkeersveiligheid
Universitaire Campus
Gebouw D
B 3590 Diepenbeek

T 011 26 87 05
F 011 26 87 00
E info@steunpuntverkeersveiligheid.be
I www.steunpuntverkeersveiligheid.be

Samenvatting

Blootstelling (verkeersintensiteit) is een cruciale variabele in onderzoek naar verkeersveiligheid. In de literatuur wordt blootstelling beschouwd als de voornaamste determinant van de verkeersveiligheid. Vaak is er echter geen goede maat voor blootstelling beschikbaar. In België hebben we maandelijkse verkeerstellingen op snelwegen voor een periode van 12 jaar. Dit biedt ons de mogelijkheid om de toegevoegde waarde van een blootstellingsmaat in onze modellen te testen, naast variabelen in verband met wetgeving, economie en klimaat. Een meervoudige regressie met ARMA foutentermen werd opgesteld om de impact van deze factoren op de geaggregeerde verkeersveiligheid te kwantificeren. Voor elke afhankelijke variabele werd een model met en zonder blootstelling opgesteld.

De modellen tonen aan dat blootstelling significant gerelateerd is tot het aantal ongevallen met doden en zwaargewonden en tot het overeenkomstige aantal slachtoffers, maar niet tot het aantal (ongevallen met) lichtgewonden. Bovendien heeft het toevoegen of weglaten van de maat van blootstelling amper invloed op het effect van de overige variabelen in het model. Het gemeten effect van blootstelling hangt duidelijk af van de maat die men hiervoor gebruikt en van de tijdspanne die men beschouwt. Een regressiemodel met ARMA foutentermen laat toe het effect van ontbrekende variabelen in de foutenterm op te nemen. Zelfs zonder een variabele als blootstelling kunnen degelijke modellen worden opgesteld.

Summary

Exposure is a key variable in traffic safety research. In the literature, it is noted as the first and primary determinant of traffic safety. In many cases, however, no valid exposure measure is available. In Belgium, we have access to monthly traffic counts for 12 years. This offers the opportunity to investigate the added value of exposure in our models, next to legal, economic and climatologic variables. Multiple regression with ARMA errors is used to quantify the impact of these factors on aggregated traffic safety. For each dependent variable, a model with and without exposure is constructed.

The models show that exposure is significantly related to the number of accidents with persons killed and seriously injured and to the corresponding victims, but not to the lightly injured outcomes. Moreover, the addition or deletion of exposure does not influence the effects of the remaining variables in the model. The effects of exposure clearly depend on the type of measure used, and on the time horizon considered. The framework of a regression model with ARMA errors allows for missing variables being accounted for by the error term. Even without a variable like exposure, valid models can be constructed.

Inhoudsopgave

1. INTRODUCTION

Despite the attention given to the problem of traffic safety, traffic is still one of the main causes of death in the world (1). In Belgium, traffic safety and mobility are main issues on today's political agenda. Accidents are the result of various influences at a certain location and time. The traffic safety problem is often studied in three dimensions (2). The first is the magnitude of the activity which results in accidents (the exposure), measured in terms of number of trips, number of vehicle-kilometers or trip duration. The second dimension is the probability of an accident or the risk, given a certain level of exposure. The third dimension is the accident consequence. Changes in one of these dimensions will change the entire safety situation.

In traffic safety, exposure is considered as key information. In a previous paper (3), we presented a model with explanatory variables concerning laws, weather and economic conditions. According to the reviewers, not including exposure was a serious limitation of the model. At that time, we did not include exposure for two reasons. First, we could not dispose of an exposure measure for a long time period. Second, we wanted to test how good a safety model can perform without any measure of exposure. Often the variable is not available on a monthly basis or for a reasonable period of time. For that reason, it is interesting to know whether accident data can be studied without exposure measure.

We agree with the reviewers on the added value of exposure. When exposure is not included in the model, it is impossible to quantify its impact. But it does not necessarily imply that models are of poor quality. In our study, econometric models were used to quantify the impact of a number of explanatory variables on traffic safety. These models deal with a large number of factors and incorporate the stochastic nature of accidents. We constructed models for the number of accidents with persons killed or seriously injured and the number of accidents with light injuries, as well as for the corresponding numbers of victims. Monthly data from January 1974 up to December 2000 were used. The models showed a good fit and high predictive power. At the time of writing this paper, we have the possibility to include monthly traffic counts on highways in Belgium for 12 years. Using this variable as a proxy measure of exposure, we will test whether the inclusion of this variable significantly changes the results.

The objective of this study is twofold. First, we rebuild the models of our previous paper, and we additionally include an exposure measure. It is investigated how exposure, weather conditions, economic growth and laws influence the number of accidents and victims. A multiple regression model with ARMA (Auto-Regressive Moving Average) errors is used to quantify the impact of these factors on traffic safety. Second, we will compare this model with one without an exposure measure.

This text is organized as follows. First, some background information is given on the use of exposure in traffic safety research, especially in time series models for traffic safety outcomes. Next an overview of the data and a summary of the methodology is given. In the results section, the outcomes for the models with and without exposure are presented and discussed. Also some general conclusions and topics for further research are provided.

2. BACKGROUND

2.1 Measures of exposure

The basic concepts in traffic safety research are exposure, risk and loss. The two main sources of information to quantify these dimensions are road accident records and exposure data (2). There are two basic methods for the collection of exposure data (4). The first is to obtain data while trips are in progress, as is done with mechanical traffic counters. Also human observations and automatic cameras can be used. With the second method, data is gathered after the trips are completed, using in-person interviews, telephone interviews and mail questionnaires. This is usually done in travel habit surveys. Researchers use various indicators of exposure like total distance travelled, travel duration or number of trips. For practical purposes, other definitions like the number of inhabitants, registered vehicles, traffic counts or fuel consumption are used.

Unfortunately, exposure measures are often the primary source of annoyance for many traffic safety researchers. First, as mentioned by Wolfe (4), the most easily obtained exposure measures are rarely the most desirable ones. Total population, the number of registered vehicles or the number of licensed drivers are not always good proxies for accident risk. Depending on the scope of the analysis, some exposure measures may be more or less relevant. Second, some measures of exposure are simply not available. In Belgium, before the first travel surveys were conducted in the late nineties (6), there was nothing else than the traffic counts, transformed in a yearly index of distance travelled. Moreover, the exposure data are rarely in a format that can be used in whatever type of analysis. Often historical exposure measures are not available over a long time period. For studies on specific groups of road users, the traffic counts are either not relevant or too aggregated. Third, typically in aggregated studies, it is not easy to find a measure of exposure that matches the traffic accidents studied. Traffic counts are usually available on a regular basis for highways, but not for the whole road network. This measure is only a rough proxy of the real exposure, especially since most of the accidents do not occur on highways. More aggregated measures of exposure are easier to find, but sometimes too rough to analyse the relationship between accident occurrence and exposure. Another problem is that road safety is generally not the primary objective when gathering travel data. Traffic surveys are not designed for the analysis of accident risk but for planning or road maintenance purposes (5).

Apart from the practical issues, the quality and reliability of exposure data is often low. This limits the level of detail of road accident analysis (5). First, since traffic data are often based on surveys, they are only representative under restrictive assumptions. Second, they are mostly not available for all kinds of traffic. In traffic counting systems, the distinction between light and heavy traffic is either not available or not reliable. Third, whereas traffic counts for motorways are mostly known, counts for regional and local roads are less frequently provided. Fourth, exposure data on a sufficiently low level of aggregation are almost never available. Even if we have data for a longer period of time, we are never sure that the data collection method remained unchanged for the considered period.

2.2 Measures of exposure in econometric traffic safety models

Econometric models, like regression and time series models, are very useful in enhancing the understanding of trends in traffic safety. The main advantages are the possibility to test the impact of a large number of factors on traffic safety outcomes and the probabilistic view on the accident process. The combination of regression models and time series analysis has been frequently applied to traffic safety data. The main ideas of traffic safety development and time series models are described in (5). An overview of macro models using aggregated explanatory variables is given in (7).

One class of explanatory models is known as the DRAG family. These are structural models, including a relatively large number of explanatory variables, whose effects on exposure, the frequency and the severity of accidents are estimated by econometric methods (5). In the DRAG models, a separate layer is included to model exposure (8). An overview of macro models and DRAG models can be found in (9). In most of these models, the relationship between traffic risk and exposure had a positive sign (7). However, it is obvious from literature that exposure measures vary considerably, and it is usually not possible to judge the quality of these variables. A plausible assumption is that the severity of accidents will at first increase with exposure, but at a certain point it will decrease, as higher exposure will decrease speed and reduce the severity. This results in an inverted U-shaped relationship between exposure and severity (8).

In the DRAG-2 model for Quebec (10), exposure is expressed as the total distance travelled, based on fuel sales and energy efficiency data and corrected for cold winters. Also changes in the type of vehicles were considered. The number of vehicles on the road was used as an explanatory variable to model the distance travelled, together with other influential factors. The effects of exposure on traffic safety were not as clear as stated in (7). A 10% increase in distance travelled resulted in an 8% increase in injury accidents and a 7.2% increase in victims injured. However, the effect on morbidity was a 0.8% decrease, although less statistically significant. For fatal accidents and victims killed, an inverted U-shaped relationship was found, indicating that traffic safety can increase with exposure. For mortality, the relation was also inversely U-shaped, although not statistically significant. Fridstrøm et al. (11) used data on gasoline sales as a proxy for exposure. They suggested the hypothesis that average severity of accidents decreases with traffic volume. The TRULS-1 model for Norway (12) used measures of exposure for various groups of road users. The injury accident frequency had an elasticity of 0.911 with respect to motor vehicle kilometres. In the DRAG-Stockholm-2 Model (13), exposure (vehicle-kilometres) for gasoline driven passenger cars was based on monthly gasoline sales within the Stockholm County and on fuel efficiency. For the periods where no data were available, estimates were obtained from a multiple regression on total gasoline sales, Gross National and Regional Product and population. In the first models, the number of bodily injury accidents was not proportional to exposure. The number of road accidents increased with the number of vehicle-kilometres, but was reduced in congested situations. At low levels of exposure, the number of light and severe injuries and fatalities decreased at first, but in congested situations the proportion of severe injuries and fatalities seemed to increase. Since these results were not logical, a new model was formulated, unfortunately resulting in a non-plausible U-shaped relationship for severe accidents and fatalities. For the severity models, the correct inverted U-shape was obtained. However, the overall performance of these new models was so poor that the authors considered the first models as superior, even with the wrong U-shaped structure for exposure. In the TAG-1 Model (14) for France, the number of kilometres travelled by all road vehicles was calculated on the basis of petrol and diesel sales. Total mileage has a significant positive impact on both injury and fatal accidents. Similar results are found for the number of fatalities, serious injuries and light injuries. On the other hand, the gravity rates for minor and serious severity were not significantly linked with exposure. Exposure to risk was positively correlated with the number of accidents and deaths. According to the authors, speed was the most important risk factor. The TRACS-CA Model for California (15) used the total number of vehicle miles travelled on state highways as an index of risk exposure for traffic on all roads in California, based on the assumption that traffic on state highways is highly correlated with the total vehicle miles travelled. Risk exposure was an important and statistically highly significant determinant of highway safety. Fatal and materials only crashes had relatively high elasticities with respect to exposure. Non-fatal injury crashes were less sensitive to risk exposure. Risk exposure increased crash frequencies, as well as mortality and morbidity. Also the fatality rate and the non-fatal injury rate were increased by exposure. In the SNUS-2.5 model for Germany (16), road demand was expressed by the kilometres driven with gasoline and diesel consumption, based on monthly gasoline and diesel consumption and the consumption rates. The number of accidents clearly depended on exposure, with a positive sign.

Also other time series models in traffic safety include in some way an exposure index. In (17), accident risk was modelled as the ratio of accident counts and exposure level. It was assumed that the coefficient of exposure is equal to 1. Another approach is presented in (18). Because exposure was not known, proxies like gas deliveries or the number of registered vehicles were used. But the authors recognized that each proxy was measured with error, and that exposure was likely to be exogenous. Therefore they modelled exposure as a latent (or unobserved) variable, determined by gasoline prices, disposable income and an error term.

Apparently, the effect of exposure and the way it should be constructed is not as clear as one would expect from theory. The assumption that traffic safety decreases with exposure is not a general truth. Instead, the inverse U-shaped relationship between exposure and accidents is more plausible. However, this assumption is not always tested, and if it is, the results are not necessarily in accordance with theory. For some traffic safety outcomes, exposure seems to have illogical signs or no impact. Moreover, very different exposure measures are used. Sometimes huge efforts are done to construct a valid exposure measure, while in other models only very simple indicators are used. All models use a variable named "exposure", but they are all differently constructed. As shown in (2), different exposure measures can lead to other results.

3. Methodology

In this study, dependent traffic safety variables are expressed in terms of independent explanatory variables. Multiple linear regression with ARMA errors can be used to model a relationship between a dependent variable and one or more independent variables, and to correct at the same time for remaining patterns in the error term. Regression models with ARMA errors are described in Pankratz (22).

3.1 Multiple Regression

The multiple regression model can be written as $Y_t=\beta_0+\beta_1X_{1,t}+\beta_2X_{2,t}+\ldots+\beta_kX_{k,t}+N_t$, where Y_t is the t-th observation of the dependent variable, and $X_{1,t},\ldots,X_{k,t}$ are the corresponding observations of the explanatory variables. The parameters $\beta_0,\beta_1,\beta_2,\ldots,\beta_k$ are fixed but unknown, and N_t is the unknown random error term. Using classical estimation techniques, estimates for the unknown parameters are obtained. If the estimated values for $\beta_0,\beta_1,\beta_2,\ldots,\beta_k$ are given by $b_0,b_1,b_2,\ldots,b_k$, then the dependent variable is estimated as $Y_{est,t}=b_0+b_1X_{1,t}+b_2X_{2,t}+\ldots+b_kX_{k,t}$, and the estimate $N_{est,t}$ for the error term N_t is calculated as the difference between the observed and predicted value of the dependent variable: $N_{est,t}=Y_t-Y_{est,t}$.

In the theoretical model, several assumptions are made about the explanatory variables and the error term. First, the model should be checked for multicollinearity. In this study, Variance Inflation Factors (23) and Variance Decomposition (24) are used to assure that multicollinearity is at an acceptable level. Second, the error terms should be uncorrelated over time. This assumption is likely to be violated in regression with time series data, giving rise to autocorrelation. The Autocorrelation Function and the Partial Autocorrelation Function are used to detect autocorrelated residuals (21). Autocorrelation can be taken into account by adding more complex structures to the regression equation, as will be explained further in this text. Third, the error terms should be identically (normally) distributed with mean zero and constant variance. Non-constant variance is called heteroscedasticity. In time series, constant variance in the regression error terms is often achieved by transforming the data (22). In this text, log-transformations are used.

3.2 ARMA Modeling

When the error terms in a regression model are autocorrelated, ARMA models can be used to describe the remaining patterns. The resulting model is a combination of a multiple regression and an ARMA model in the error terms. The ARMA modeling approach expresses a variable as a weighted average of its own past values. The model is in most cases a combination of an autoregressive (AR) part and a moving average (MA) part. Suppose a variable N_t is modeled as an autoregressive process, AR(p). Then, N_t can be expressed as a regression in terms of its own passed values:

$$\begin{aligned} N_t &= C+\varphi_1N_{t-1}+\varphi_2N_{t-2}+\ldots+\varphi_pN_{t-p}+a_t = C+\varphi_1BN_t+\varphi_2B^2N_t+\ldots+\varphi_pB^pN_t+a_t \\ &= \frac{C+a_t}{\left(1-\varphi_1B+\varphi_2B^2+\ldots+\varphi_pB^p\right)} \end{aligned}$$

In this expression, C is a constant term, φ_i (i = 1, …, p) are the weights for the autoregressive terms and at is a random term which is assumed to be normally distributed "white noise", containing no further information. The backshift operator B_i, applied on N_t, is defined as $B_iN_t=N_{t-i}$ (i=1,2,…). The series N_t can also be expressed in terms of the random errors of its past values, which is then a moving average MA(q) model:

$$N_t = C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q} = C + a_t - \theta_1 B a_t - \theta_2 B^2 a_t - \ldots - \theta_q B^q a_t$$
$$= C + \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) a_t$$

Now the θ_j (j=1,…,q) are the weights for the moving average terms. In a more general setting, it is possible to include autoregressive and moving average terms in one equation, leading to an ARMA(p, q) model:

$$\left(1 - \varphi_1 B + \varphi_2 B^2 + \ldots + \varphi_p B^p\right) N_t = C + \left(1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q\right) a_t$$

Here at is again assumed to be "white noise". An ARMA model cannot, however, be applied in all circumstances. It is required that the series be stationary. For practical purposes, it is sufficient to have weak stationarity, which means that the data is in equilibrium around the mean and that the variance around the mean remains constant over time (21). If a series is non-stationary in the variance, it often helps to log-transform the data, as is done in this text. To have a series that is stationary in the mean, differencing is used. Instead of working with the original series, changes in the series are modeled. When an ARMA model is built on differenced data, it is called an ARIMA(p, d, q) model, where I indicates the differencing and d is the period of differencing. In our models, a 12-period differencing is used.

3.3 Regression with ARMA errors

The ARMA modeling approach can now be applied to the multiple regression equation to model the information that remains in the error terms. Assume a regression model with one explanatory variable, denoted as $Y_t=\beta_0+\beta_1 X_{1,t}+N_t$. Suppose further that the error terms are autocorrelated, and that they can be appropriately described by an ARMA(1,1) process. This model can then be written as: $Y_t=\beta_0+\beta_1 X_{1,t}+N_t$, where $(1-\varphi_1 B)N_t=(1-\theta_1 B)a_t$, and at is assumed to be white noise. Substituting the correction for the error term into the regression equation gives:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} a_t$$

Because of the specific form in the error terms, the classical least squares methods are not appropriate to estimate the parameters of this equation. Instead, the SAS-ARIMA procedure with Maximum Likelihood estimation is used to set up the models. The Likelihood function is maximized using Marquardt's method via non-linear least squares estimation (25).

4. DATA

A database for Belgium, with variables on exposure, weather, laws and economic conditions has been created. Monthly observations from January 1990 up to December 2001 are used. A detailed description of the database is given in Table 1. Four dependent variables will be modeled: the (log-)number of accidents with persons lightly injured (LNACCLI) and killed or seriously injured (LNACCKSI), the (log-)number of lightly injured (LNPERLI) and killed or seriously injured persons (LNPERKSI).

4.1 Exposure

The exposure variable is a monthly indicator of the number of vehicles on highways in Belgium, based on traffic counts. It is created on the basis of average daily counts for weekdays, Saturdays and Sundays. These averages are based on counting results on various highway locations. To our knowledge, it is the only available measure of exposure on a monthly basis for Belgium. In the past, various travel surveys have been done for Belgium (20) and for the Flemish part of the country (6, 19). Usually the travel survey data allow much more detail in the exposure variables than is possible with traffic counting systems. The frequency and the number of involved road users of these travel habit surveys is, however, too low to be used on a monthly basis and for a reasonable period of time.

On the other hand, we know that our monthly traffic counts are by no means an ideal representation of monthly traffic. First, we use highway traffic counts as an exposure measure for the whole country. Although this approach has been used elsewhere (15), we think that this is a simplification of the real traffic situation. Second, the period of analysis is drastically limited by the availability of exposure data. Whereas the previous study (3) was done on data from 1974 onwards, we can now only analyze traffic safety starting from 1990. Third, it is clear that the traffic counts are not the same as the number of vehicle kilometers, based on fuel sales and car efficiency. However, it is the only measure we can use at this time, and consequently it is also the best one.

4.2 Laws, climate and economic conditions

Four dummy variables are included in the model to study the effect of the introduction of laws. These variables are equal to zero before the introduction and equal to one as from the moment of introduction. Climatologic variables were registered in the climatologic center in Ukkel (in the center of Belgium). Also some indicators are used to measure economic climate. The number of car registrations and the percentage of second hand car registrations can be seen as indications of economic welfare.

TABLE 1 Description of the data

NAME	MEANING	MEAN	STD.	MIN.	MAX.
Dependent variables					
NACCKSI	Number of accidents with persons killed or seriously injured$^{(*)}$	972	207	589	1451
NPERKSI	Number of persons killed or seriously injured$^{(*)}$	1188	270	688	1825
NACCLI	Number of accidents with persons lightly injured$^{(*)}$	3418	380	2360	4345
NPERLI	Number of persons lightly injured$^{(*)}$	5010	548	3407	6298
Measure of exposure					
Exposure	Measure of exposure (traffic counts on highways, $\times 10^6$)$^{(*)}$	389	176	118	745
Laws and regulations					
LAW0191	Mandatory seat belt use in rear seats	-	-	0	1
LAW0192	50 km/h in urban areas, 90 km/h on 2 by 2 lanes without separation, regulations on vehicle load and on cycling tourists	-	-	0	1
LAW1294	0.05% minimum alcohol level, higher fines for 0.08% or higher	-	-	0	1
LAW0496	Regulations on traffic at zebra crossings	-	-	0	1
Weather conditions					
PDAYPREC	Percentage (× 100) of days with precipitation	52	18	10	90
PDAYFROST	Percentage (× 100) of days with frost	12	18	0	75
PDAYSNOW	Percentage (× 100) of days with snow	18	22	0	79
PDAYTHUN	Percentage (× 100) of days with thunderstorm	24	16	0	65
PDAYSUN	Percentage (× 100) of days with sunlight	81	16	32	100
Economic conditions					
NUNEMP	Number of unemployed people (× 1000)$^{(*)}$	426	56	332	524
NCAR	Number of car registrations (× 1000)$^{(*)}$	90	15	57	133
POLDNCAR	Percentage of second hand car registrations	0.86	0.06	0.43	0.72
$^{(*)}$ *For these variables, the logarithm is used in the models. To indicate the logarithm, an "L" is added in front of the name (see table 2).*					

5. RESULTS

5.1 Explanatory model

Since we want to test whether the exposure variable significantly influences the results, we compare for each dependent variable the models with and without exposure. In Table 2, the parameter estimates are presented. Only variables significant at a 90% or higher confidence level were retained (each model was re-estimated after dropping the non-significant variables one by one). In the models with exposure, this variable is always kept, even if it turns out to be insignificant. For each variable, the parameter estimate and the approximate absolute t-value (between brackets) are reported. Also the Akaike Information Criterion (AIC) and the error standard deviation are reported. The AIC is smaller when less parameters are used or when the likelihood increases. The lower the AIC, the better the model.

5.1.1 Laws and Regulations

The laws on the obligatory use of seat belts in the rear seats (LAW0191), the laws on speed limits (LAW0192) and the laws on alcohol (LAW1294) are significant for all safety outcomes. For example, in the model with exposure, the introduction of the laws on alcohol resulted in a decrease in NACCKSI of 1-exp(-0.1481)=13.77%. The laws for pedestrians (LAW0496) are only significant for LNACCKSI and LNPERKSI. The addition to or deletion from the model of exposure does not alter the significance of these law variables. Even in the models where exposure turns out to be clearly significant, the effects of the laws do not change.

5.1.2 Weather Conditions

As we already found in our previous study, traffic safety is highly influenced by climatologic conditions. The percentage number of rainy days (PDAYPREC) has an influence on LNACCKSI and LNPERKSI. The percentage number of days with thunderstorm (PDAYTHUN) has a highly significant relationship with all traffic safety variables, except with LNPERKSI. Thunderstorm generally decreases traffic safety. For example, the model with exposure shows an increase of 0.12% in NACCKSI when the percentage of days with thunderstorm increases by 1. Further, a higher monthly percentage of days with frost (PDAYFROST) decreases all dependent variables. PDAYSNOW is not significant in our models, almost surely because snow is not common in Belgium. Also a higher percentage of days with sunshine does not significantly alters the traffic safety outcomes. These results confirm again that the effect of weather is related to the geographical properties of the area and the considered time period. In almost all cases, including the exposure measure has no effect on the significance of the weather variables.

TABLE 2 Results for the four models, with and without exposure

	LNACCKSI		LNPERKSI		LNACCLI		LNPERLI	
Exposure included?	Yes	No	Yes	No	Yes	No	Yes	No
Measure of Exposure								
LExposure	-0.0707 (-2.54)	-	-0.0741 (-2.74)	-	0.0207 (1.08)	-	0.0096 (0.51)	-
Laws and Regulations								
LAW0191	-0.0661 (-2.65)	-0.0714 (-2.42)	-0.0755 (-3.13)	-0.0780 (-2.37)	-0.0371 (-2.00)	-0.0369 (-1.99)	-0.0337 (-1.83)	-0.0336 (-1.83)
LAW0192	-0.1418 (-4.30)	-0.1679 (-4.70)	-0.1585 (-5.08)	-0.1850 (-4.64)	-0.0779 (-3.91)	-0.0730 (-3.76)	-0.0580 (-2.59)	-0.0551 (-2.55)
LAW1294	-0.1481 (-5.55)	-0.1728 (-5.94)	-0.1368 (-5.32)	-0.1666 (-5.11)	-0.0485 (-2.86)	-0.0373 (-2.80)	-0.0500 (-2.91)	-0.0442 (-3.42)
LAW0496	-0.0702 (-2.64)	-0.0843 (-2.72)	-0.0729 (-2.91)	-0.0848 (-2.48)				
Weather conditions								
PDAYPREC	-0.0014 (-3.85)	-0.0013 (-3.41)	-0.0011 (-2.61)	-0.0007 (-1.94)				
PDAYFROST	-0.0022 (-4.69)	-0.0023 (-4.71)	-0.0027 (-4.89)	-0.0021 (-3.78)	-0.0017 (-3.95)	-0.0017 (-4.05)	-0.0017 (-3.95)	-0.0017 (-4.02)
PDAYSNOW								
PDAYTHUN	0.0012 (2.49)	0.0010 (2.06)	0.0012 (2.03)		0.0018 (4.31)	0.0019 (4.49)	0.0019 (4.65)	0.0020 (4.77)
PDAYSUN								
Economic Conditions								
LNUNEMP	0.2567 (3.10)	0.3042 (2.20)	0.4610 (4.59)	0.4827 (3.97)				
LNCAR								
POLDNCAR			-0.6905 (-2.60)	-0.6097 (-2.27)			-0.2696 (-1.74)	-0.2807 (-1.83)
Goodness of Fit (AIC and Standard Error)								
AIC	-310	-305	-285	-285	-349	-350	-351	-352
Stand. Error	0.0715	0.0735	0.0791	0.0791	0.0628	0.0628	0.0623	0.0621

5.1.3 Economic Conditions

From the indicators of economic condition, unemployment (LNUNEMP) and the percentage of second hand car registrations (POLDNCAR) are significant. Unemployment increases LNACCKSI and LNPERKSI, while POLDNCAR only affects the number of victims and not the number of accidents. For example, if the percentage of second hand car registrations increases by 1, then NPERLI decreases by 0.27%. In a better economic climate, both unemployment and the percentage of second hand car registrations are expected to be lower. However, their impact on traffic safety seems to be different. The results on LNUNEMP and POLDNCAR are not completely in line with literature, although effects in both directions can be found in other studies. The positive sign for unemployment might be explained by a corresponding drop in income. The demand for safer cars will decrease, which in turn leads to more accidents and victims. On the supply side, bad economic conditions can imply lower budgets for road infrastructure, leading to lower quality roads and thus more accidents (7). Also, bad economic conditions can have a negative influence on the driver behaviour, in terms of aggression or lack of attention (14). The negative sign for POLDNCAR can be the result of risk compensation. When drivers know that the quality of their vehicle is low, they might adapt their driving style accordingly. It is clear that the effects of economic conditions on traffic safety are not straightforward, and various forces are working in opposite directions. Further analysis of these results is necessary. Again, adding or deleting the exposure variable does not alter the results of the economic variables.

5.1.4 Exposure

Exposure makes the difference between the two models that were developed for each of the dependent variables. The exposure variable is significant for the killed and seriously injured outcomes, with a negative sign. A higher exposure level results in a lower number of accidents with persons killed or seriously injured, as well as a lower number of victims. Since both the dependent and independent variable are in logarithms, we can easily give an interpretation in terms of elasticities. If exposure is increased by 1%, then NACCKSI and NPERKSI are both reduced by about 0.07%. The decrease in severity of accidents may be related to an improved infrastructure, lower speed limits, a higher enforcement level and better car technology, without, however, automatically reducing the number of accidents. In some specific cases, more traffic also reduces speed, and consequently lighten accident severity. For NACCLI and NPERLI, exposure does not seem to have a significant impact. This may be related with the level of under-registration, which is known to be quite high for accidents with light injuries. However, it is interesting to see that the sign of the effect is opposite. Higher exposure will increase the number of accidents with lightly injured persons and the corresponding number of light injuries. More traffic may increase the probability of accidents, but the severity of accidents will be lower.

These results are not counterintuitive at all. As indicated by Oppe (26), the evolution of exposure may be approximated by a logistic function, leveling off at a certain point in time. We expect traffic to be higher than in the past, but with a lower growth rate. This is in line with the conclusions in (27), where the authors state that the number of kilometres driven in Sweden remained unchanged during the nineties. The effects of exposure on traffic safety will therefore also be different from some decades ago. In many DRAG models, a quadratic relationship between exposure and accidents was assumed. However, given our limited time horizon, it is quite possible that we will not find a clear quadratic relation, especially if all observed exposure levels are in the increasing or decreasing part of the parabola.

Further, we also recognize that the quality of the exposure measure leaves room for improvement. However, it is the only variable we can use at the moment, and the literature review shows that exposure is a problem child in many countries. The quality of our variable may be improved by modeling it as a function of fuel sales and fuel efficiencies. On the other hand, the results bring us to some interesting findings. First, our measure of exposure is not always significant. It is influential in models for killed and seriously injured persons, and less for lightly injured outcomes. Second, even if exposure is significant, the goodness of fit of the models, in terms of AIC values, does not change. The combination of regression models with ARMA errors therefore is able to capture the remaining effect of omitted variables.

Even if we do not have a good exposure measure, we can draw sensible conclusions on the other variables in the model. Of course, these models will not provide any insights in the effects of exposure, and are therefore more limited in their explanatory power. But the quality of the estimates is equally high. Mostly, the value of the coefficients in the models without exposure is slightly higher than in the models with exposure. They partly take over the effect of exposure. On the other hand, it is clear from the models that the changes in the values of the coefficients are not that large that the conclusions would become invalid.

5.2 Error model

The error terms of the regression equations should be corrected for possible autocorrelation. As explained in the methodology section, both Autoregressive (AR) and Moving Average (MA) corrections are possible. The estimated error structures are summarized in the table below. Here, Nt is the original regression error term, while at is the corrected ("white noise") error term, which contains no further information. The backshift operator B is the same as defined before. All coefficients in the error structure are significant on a 95% confidence level.

TABLE 3 Error Structures for the Models

	Exposure	Error Structure
LNACCKSI	**Yes**	$(1 - 0.3121B^2)(1 + 0.2169B^4)\,N_t = (1 - 0.7903B^{12})\,a_t$
	No	$(1 - 0.3207B^2)\,N_t = (1 - 0.8070B^{12})\,a_t$
LNPERKSI	**Yes**	$N_t = (1 - 0.7297B^{12})\,a_t$
	No	$N_t = (1 - 0.2086B + 0.2601B^2)(1 - 0.7529B^{12})\,a_t$
LNACCLI	**Yes**	$N_t = (1 - 0.7743B^{12})\,a_t$
	No	$N_t = (1 - 0.7689B^{12})\,a_t$
LNPERLI	**Yes**	$N_t = (1 - 0.8249B^{12})\,a_t$
	No	$N_t = (1 - 0.8232B^{12})\,a_t$

The models for the killed and seriously injured outcomes, where exposure has a significant contribution, have a more complex error structure. In the models for the lightly injured outcomes, only moving average terms were needed. Also note the difference between the models with and without exposure. For the lightly injured outcomes, the difference is negligible, but for the killed and seriously injured outcomes the inclusion or deletion of exposure alters the error structure. Although it is difficult to give a meaningful interpretation to the error terms, we can assume that when exposure is not explicitly included in the model, it is partly reflected in the error term. For these models, more information has to be filtered from the error term. These results are also in line with the conclusions on the explanatory variables. Since the model structure allows for a correction of the error term, the effect of missing variables can be canalized to the error term, without drastically changing the effect of other variables.

6. CONCLUSIONS AND FURTHER RESEARCH

In this study, the use of an exposure variable is evaluated in models for accident frequency and severity in Belgium. Next to exposure, the impact of weather, laws and economic conditions was tested. The impetus for this study was the review of a paper for the previous TRB conference, where the lack of an exposure variable was mentioned as a serious limitation. We agree with the reviewers that the inclusion of a measure of exposure may improve the explanatory power of the model, in the sense that the effect of exposure cannot be isolated when it is not included.

At the time of writing our first paper, we did not have any exposure measure, and we tested whether good models could be obtained by using only the available explanatory variables. The models were quite convincing, both from the explanatory and the predictive point of view. In the meantime, we obtained monthly traffic counts for 12 years for highways in Belgium. With this variable, our models were re-estimated. In the new models, exposure seems to have a significant negative impact on the number of accidents with persons killed and seriously injured, and on the corresponding number of victims. The effect on the outcomes for lightly injured persons is not significant.

In spite of the prominent place of exposure in traffic safety research, results in literature are quite unclear and not always comparable. It is, to our knowledge, not uncommon that researchers have difficulty in finding a good exposure measure. In many studies, it is difficult to judge the quality of the exposure measure. Also the time horizon considered in the models can change the view on the effect of this variable. It seems logical to include a quadratic relationship between traffic safety and exposure, but it is not certain that the available data would cover the whole quadratic surface. Conclusions are therefore only valid within the range of the available data and extrapolation is not without danger. It is also possible that different exposure measures show a different relation with traffic safety. Furthermore, the traffic safety indicators can be different. In our study, we looked at accidents and victims as a function of explanatory variables concerning exposure, laws, economic conditions and climate, but also morbidity and mortality indices could have been studied. It is possible that the effect of exposure may differ with the kind of data that is used. Instead of using time series, traffic safety and exposure can be expressed with cross-sectional data. Michener (28) postulates that the most important variable in cross-sectional studies of accidents is some measure of scale, like kilometres driven, registered vehicles or licensed drivers. The effect of exposure in a cross-sectional study may differ from the effect of the nationwide exposure on accidents over time.

The relatively small impact of exposure on the other variables in the model is interesting. Even if exposure is significant, the deletion of this variable will only slightly alter the effects of the other variables in the model. Also the model quality is insensitive for the inclusion of an exposure measure. The effect of exposure is partly filtered by the other variables, but the main part of the effect will be found in the error term. This indicates that the framework of the regression model with an ARMA error structure can deal with missing exposure.

Some parts of the analysis should be further investigated. First, there is room for improving the exposure measure. Instead of using only highway traffic counts, we could include fuel sales and fuel efficiency. Using this information, a measure of exposure in terms of vehicle kilometres could be constructed. This is an important topic for policy makers. If there is an interest in knowing how exposure influences traffic safety, efforts should be put in defining and constructing a useful measure of exposure. This will enable a uniform treatment and straightforward interpretation of the results. Second, the exposure should be better tuned to the traffic safety indicators. When nationwide accidents are used, the exposure index should reflect traffic on all roads. Third, it is necessary to investigate the form of the relationship between exposure and traffic safety together with the time horizon of the data. The relationship may show an inverted U-shaped relationship only over a long period of time, which will not be visible from our sample. Fourth, from the modelling point of view, the variable selection procedure might be refined. Although the error structure is capable of filtering out remaining patterns that are not accounted for by the included variables, it is advisable to make a good selection of variables. The combination of a reliable variable selection and the powerful framework of regression with ARMA errors may enlarge the insights in traffic safety.

7. References

(1) *World Report on Road Traffic Injury Prevention*. World Health Organization, Geneva, 2004.

(2) *OECD Road Transport Research. Road Safety Principles and Models: Review of Descriptive, Predictive, Risk and Accident Consequence Models*. OCDE/GD(97)153, OCDE, Paris, 1997.

(3) Van den Bossche, F., Wets,G. and Brijs,T. A Regression Model with ARMA Errors to Investigate the Frequency and Severity of Road Traffic Accidents. In: *Proceedings of the 83rd annual Meeting of the Transportation Research Board*, USA, 2004.

(4) Wolfe, A.C. The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis and Prevention*, 14(5), pp. 337-340, 1982.

(5) *COST329: Models for Traffic and Safety Development and Interventions. Final Report of the Action*, European Commission, 1999.

(6) Hajnal, I. and Miermans, W. *Onderzoek Verplaatsingsgedrag Vlaanderen. Analyseopdracht. Eindverslag*, Provinciale Hogeschool Limburg, Departement architectuur, Diepenbeek. 172 p., 1996.

(7) Hakim, S., Shefer, D. and Hakkert, A.S. A Critical Review of Macro Models for Road Accidents. *Accident Analysis and Prevention*, 23, 5, 1991, pp. 379-400.

(8) Gaudry, M. and Lassarre, S. *Structural Road Accident Models: The International DRAG Family*. Elsevier Science, Oxford, 2000.

(9) Van den Bossche, F. and Wets, G. *Macro Models in Traffic Safety and the DRAG Family: Literature Review*. Report RA-2003-08, Flemish Policy Research Center for Traffic Safety, Diepenbeek, 2003.

(10) Fournier, F. and Simard, R. The DRAG-2 Model for Quebec. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 2, pp. 37-66, Elsevier Science, Oxford, 2000.

(11) Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R. and Thomsen, L.K. Measuring the contribution of randomness, exposure, weather and daylight to the variation in road accident counts. *Accident Analysis and Prevention*, Vol. 27, 1995, pp. 1-20.

(12) Fridstrøm, L. The TRULS-1 Model for Norway. In: Structural Road Accident Models: The International DRAG Family (Gaudry, M. and Lassarre, S. Eds.), Chap. 4, pp. 97-126, Elsevier Science, Oxford, 2000.

(13) Tegnér, G., Holmberg, I., Loncar-Lucassi, V. and Nilsson, C. The DRAG-Stockholm Model. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 5, pp. 127-156, Elsevier Science, Oxford, 2000.

(14) Jaeger, L., Lassarre, S. The TAG-1 Model for France. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.), Chap. 6, pp. 157-184, Elsevier Science, Oxford, 2000.

(15) McCarthy, P. The TRACS-CA Model for California. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.). Chap. 7, 185-204, Elsevier Science, Oxford, 2000.

(16) Blum, U. and Gaudry, M. The SNUS-2.5 Model for Germany. In: *Structural Road Accident Models: The International DRAG Family* (Gaudry, M. and Lassarre, S. Eds.). Chap. 3, pp. 67-96, Elsevier Science, Oxford, 2000.

(17) Ledolter, J. and Chan, K.S. Evaluating the impact of the 65 mph maximum speed limit on Iowa rural interstates. *American Statistician*, 50, 1996, pp. 79-85.

(18) Johansson, P. Speed limitation and motorway casualties: a time series count data regression approach. *Accident Analysis and Prevention*, Vol. 28, 1996, pp. 73-87.

(19) Zwerts, E. & Nuyts E. (2002), *Onderzoek Verplaatsingsgedrag Vlaanderen (januari 2000- januari 2001). Deel 3A: Analyse Personenvragenlijst*, Provinciale Hogeschool Limburg, Departement Architectuur, 214 p.

(20) Hubert, J.-P. and Toint Ph. La mobilité quotidienne des Belges, Presses Universitaires de Namur, Namur, 2002.

(21) Makridakis, S., Wheelwright, S. and Hyndman, R. *Forecasting: Methods and Applications*. Third edition, John Wiley and Sons, 1998.

(22) Pankratz, A. *Forecasting With Dynamic Regression Models*. John Wiley & Sons, 1991.

(23) Marquardt D.W. Generalized inverses, ridge regression, biased linear estimation and non-linear estimation. *Technometrics*, 12, 1970, pp. 591-612.

(24) Belsley D.A., Kuh E. and Welsh R.E. *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley and Sons, New York, 1980.

(25) SAS Institute Inc. *SAS OnlineDoc™*, Version 7-1, 1999.

(26) Oppe, S. Macroscopic models for traffic and traffic safety. *Accident Analysis & Prevention*, Vol. 21, 1989, pp. 225-232.

(27) Edwards, H., Nilsson, G. and Vorwerk, P. *Traffic amount in vehicle kilometres on Swedish roads, 1950–1997*, VTI rapport 439, 2001.

(28) Michener, R. and Tighe, C. A Poisson Regression Model of Highway Fatalities, *American Economic Review*, American Economic Association, vol. 82(2), 1992, pp. 452-456.