MULTILEVEL COMPARATIVE ANALYSIS OF ROAD SAFETY IN EUROPEAN CAPITAL CITIES

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

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

Marianthi Mermigka  
Civil – Transportation Engineer  
National Technical University of Athens  
Department of Transportation Planning and Engineering  
5 Heroon Polytechniou st., GR-15773 Athens  
Tel: +302107721380, Fax: +302107721454  
E-mail: marianthimg@gmail.com
ABSTRACT

The objective of this research is the comparative road safety analysis in selected European capital cities, aiming to a better understanding of road accident characteristics and causes in European megacities. Despite the continuous urbanization and the shift of population to large urban areas, this research question has received little attention in the existing literature. A database was developed for this analysis containing data regarding the number and the characteristics of road fatalities, the population and other demographic, socioeconomic and transport indicators of nine selected European capital cities for the period 2007 - 2011. Multilevel Poisson statistical models were developed, allowing for a more accurate representation of the hierarchical structure of road safety data, and they led to the identification of several factors affecting the road safety level in the selected European capital cities, revealing some additional aspects of road safety performance in these cities. Factors found with a statistically significant effect concerned city characteristics (road network length, population density, public transport use) and accident characteristics (road user and vehicle type). The comparison between the European capital cities showed that the larger the city’s road network is, the higher the level of road safety is in this city.

Key words: road safety, European capitals, multilevel model.
INTRODUCTION

Because of the shift of population from rural to large urban areas, future trends in road transportation in megacities are of increasing interest. The ever-increasing urbanization of nations around the world results in implications for road safety, due to the more complex traffic problems prevailing in cities. In 2010, 10,837 people were killed in traffic accidents on urban roads in the EU-19. This is 38% of all traffic accident fatalities in 2010. In the last decade, urban road fatalities have reduced by more than a third (39%), a little bit less than the total number of fatalities that has reduced by 42% (1).

One earlier European project examining three major cities (Copenhagen, Amsterdam and Barcelona) concluded that lack of safety was an inhibiting factor to cycling in Barcelona, but not in the other two cities (2). More recent data for a selection of capital cities, averaged over 2004 to 2006, indicated that vulnerable road user deaths (pedestrian, cyclist and powered two-wheeler) accounted for between 35 and 85% of all road deaths in capital cities (3). The availability of public transport and dedicated infrastructure together with climatic conditions leading to differences in modal split can partly explain the recorded distribution of road user deaths. Another more recent study (4), English, German and Dutch cities were compared in terms of the relationship of pedestrian and cyclists fatalities and city size; a positive relationship was found, but the Dutch data in particular showed that the effect differs for different casualty severities (i.e. for more severe injuries, rates initially rise with town size but then decrease for the larger cities).

Another recent study examined crash data for two U.S. megacities (New York and Los Angeles), discussing the notable differences between these cities and the nation in general (5). A subsequent study extended the analysis investigating crash patterns in the megacities of London and Paris in comparison with crash patterns for the entire United Kingdom and France, respectively (6). Both studies showed that crashes and fatal crashes in these cities tend to differ in several aspects from typical crashes and fatal crashes in the respective nations. The main differences are on when and where the accident occurred, the weather and lighting conditions, who was involved in the accident and what were the driver actions.

The literature review showed that little research investigating thoroughly road safety characteristics has been conducted in cities level (4, 7), and most related studies focus on vulnerable road users.

Within this context, the objective of the present research is the comparative analysis of road safety in selected European capital cities, aiming to a better understanding of road accident characteristics and causes in European megacities. The selected European capital cities are three representative cities from each basic European geographic region (southern, northern and eastern Europe), namely: Athens, Lisbon and Madrid from southern Europe, Brussels, London and Paris from northern Europe and Bucharest, Budapest and Prague from eastern Europe.

Due to the hierarchical nature of geographically structured road safety data, multilevel models were developed in order to handle appropriately the resulting dependences among the observations and provide a more comprehensive analysis of road safety in the examined European capital cities.
METHODS AND DATA

Data collection

The data collected for the objectives of this research concern the number and the characteristics of road fatalities, the population and other indicators of the selected European capital cities for a five-year period from 2007 to 2011.

The road accidents data derived mainly from CARE, the European Database on Road Accidents, which has developed a Common Accident Data Set, named CADaS, consisting of a minimum set of standardised data elements for all EU countries. The city road fatalities were extracted on the basis of the NUTS (Nomenclature of Territorial Units for Statistics) classification of EUSROSTAT used within CADaS. However, in the case of Greece, CADaS contains data only for the entire Attica district. Therefore, road accident data for the city of Athens were collected from the database of the Department of Transportation Planning and Engineering of the NTUA.

Regarding the capital cities data, population data of the European cities were obtained from EUROSTAT and for the city of Athens from the Hellenic Statistical Authority (EL.STAT.). Moreover, variables referring to other characteristics or indicators of the capital cities were derived from the database “Mobility in Cities”, which has been created by the International Association of Public Transport (UITP).

The dependent variable of the analysis is the number of persons killed (at 30 days from the accident) for each capital city. The independent variables that are examined concern the accident (accident date, weather and lighting conditions), the traffic unit (traffic unit type) and the person involved in the road accident (age, gender, road user type). The variables examined concerning the capital cities are the population, the urban population density, the length of road per thousand inhabitants and the rate of annual private motorised passenger kilometres per annual public transport passenger kilometre.

A first exploration of the data is presented in Figure 1. The top left panel represents road fatalities per million inhabitants by capital city, showing that Southern and Eastern European cities have higher fatality rates than Northern ones, a pattern which is well known from national analyses (8). Following the same city order that resulted from the ranking on the basis of the fatality rate, the cities characteristics are graphically presented in the other three panels. It can be seen that the cities rankings change significantly for each characteristic or indicator.

For instance, in the top right panel, that presents the length of road per thousand inhabitants by capital city, the city of Lisbon holds the lowest place and the city of Madrid the highest. The city of Lisbon also has the lowest population density (bottom left panel), whereas Brussels has the highest. The city of Madrid has relatively low rate of private to public passenger transport (bottom right panel). Eastern European cities have the lowest rates of private to public passenger transport, while Athens has the highest rate. It appears therefore that the road safety level of European cities can not be explained by a single characteristic of the cities; cities have variable characteristics and hence, it is important to take this variation into consideration in the analysis.
Analysis Method

Most of the data of interest for road safety research happen to be hierarchically organized, i.e., to belong to structures with several hierarchically ordered levels. For a part, these hierarchical structures result from the spatial spread of the data: Observations belong to larger geographical areas or units (these can be as various as road sites, segments, or intersections, counties, regions...). One of the main problems associated with such hierarchical data organisation is the dependence that generates among the observations (9).

Observations that are sampled from the same geographical units have in common a series of unobserved characteristics that are proper to these larger geographical areas. When using classical modelling techniques, there is no other way to include these higher-level characteristics as predictors in the model than either aggregating or disaggregating them at the level at which the observation units are defined. However, this may often result in considerable information loss (10).

The estimations obtained from most standard analysis techniques rest on the assumption that the observations are sampled from a single homogeneous population, and that the residuals resulting from the model are independent. However, the hierarchical organisation of data fundamentally challenges these assumptions. Hence, applying traditional statistical techniques (linear or generalized linear models) to hierarchically organised data typically results in underestimated standard errors and exaggeratedly narrow confidence intervals. The risk is consequently that incorrect conclusions be derived about the significance of the parameters whose effects are investigated (11).

Statistical models have been developed that allow accounting for hierarchical data structures, and taking into account the dependency they introduce among the data. Because the hierarchical structure is specified in the model, predictors that characterize the different levels considered can also be correctly defined (no need for aggregation or disaggregation). These models are labelled multilevel models.

The present research is based on a geographical data hierarchy. More precisely, the first level of the hierarchy is the accident level, which contains data about each road accident.
case with all its defining characteristics, like traffic unit and accident related characteristics (date, weather conditions) and person in the accident involved related characteristics (age, gender, road user type). All these data are nested into the capital cities that the accidents occurred, and which constitute the second level of the hierarchy, that is to say the city level. The cities characteristics are for example, the population, the length of road and the annual kilometrage. This type of hierarchy is graphically presented in Figure 2.

**FIGURE 2 Geographical hierarchies of road safety data of European cities**

Regarding the multilevel models equations, it can be generally considered to fall within the broader family of spatial analyses, which are characterized by the fact that they aim at accounting for the spatial dependence between data. Spatial dependence in this case refers to a general co-variation of properties within a geographical space. Typically, a Poisson distribution is assumed for accident counts, with an exposure estimate (e.g. the population Ni) incorporated as an offset term.

In case of Poisson multilevel modelling, the lower level unit is a count of events and there is a higher level classification of the counts across which the probability response is considered to vary. The multilevel model fitted to the data is based on iterative generalized least squares estimation (12). A Poisson distributed response vector (Oij) of observed cases is assumed, in which (i) refers to the accident level and (j) refers to the city level, and therefore it is necessary to include an offset of expected numbers of cases Eij in the model so that:

\[ O_{ij} \sim \text{Poisson}(\pi_{ij}E_{ij}) \]
\[ \log(\pi_{ij}) = \beta_{0j} + \beta_{1j}x_{ij} \]
\[ \beta_{0j} = \beta_{0} + u_{0j} \]
\[ \beta_{1j} = \beta_{1} + u_{1j} \]

where Eij represents the expected numbers of cases for each level-1 unit, and \( \beta_{0j} \) \( \beta_{1j} \) are parameters to be estimated; these are considered to vary between cities, and therefore can be further decomposed into a fixed effect (\( \beta_{0}, \beta_{1} \)) and a normally distributed random departure (\( u_{0j}, u_{1j} \)) with mean equal to zero and variances equal to (\( \sigma_{0j}, \sigma_{1j} \)) respectively. These departures from the overall mean are known as the level 2 residuals, and in the present research these are the city effects (13).

It should be underlined that no random structure can be specified at the lowest level of a Poisson multilevel model, as in the Poisson model the relationship between mean and variance is known, so that there is no need to separately estimate the latter. The assumption of
Poisson variation of the cases of counts can be estimated with a dispersion parameter. If the counts examined come from significantly heterogeneous populations, the expected values may vary significantly. In order to handle the overdispersion, one option is to consider an additional parameter \( \alpha \), resulting to an Extra-Poisson or quasi-Poisson distribution, so that:

\[
\text{var}(\theta_{ij} \pi_{ij}) = \alpha \sigma_i^2 \pi_{ij} E_{ij}
\]

It should be noted that, ignoring extra-Poisson variation might not significantly affect parameter estimates; however the related statistical significances may be slightly affected \((14)\).

The review of multilevel analyses in road safety research shows that statistical and conceptual consequences may occur when ignoring a hierarchical structure in the data. In the majority of cases, multilevel model formulations \((10, 15)\):

i. allow improving the fit of the model to the data
ii. allow identifying and explaining random variation at specific levels of the hierarchy considered
iii. can yield different (more correct) conclusions than single-level model formulations with respect to the significance of the parameter estimates.

MODEL DEVELOPMENT

The multilevel modelling analysis was carried out by using the MLwiN 2.27 dedicated statistical software. The variables tested in the statistical model of the analysis are presented below:

- **Dependent variable:** Fatally Injured (at 30 days)
- **Independent variables:**
  - Month
  - Day of Week
  - Person Age Group
  - Person Gender
  - Road User Type
  - Traffic Unit Type
  - Weather
  - Urban population density (persons/ha)
  - Annual private motorised passenger kilometres / Annual public transport passenger kilometres
- **Levels of Analysis:**
  - 1st Level: Case ID
  - 2nd Level: City
- **Offset of Poisson regression:** LogLength, i.e. the natural logarithm of the variable Length of Road per 1.000 inhabitants

The development of the statistical model of the analysis was carried out in two phases. In the first phase, only the variables associated with road accidents data were included in the statistical model. In the second phase, the variables referring to characteristics and indicators of the capital cities were added to the statistical model, in order to test whether they improve the explanatory power of the model.

First, all the variables included in the statistical model were separated in categorical and quantitative variables and their correlation was tested in advance, in order to ensure that multicollinearity would not affect the modelling results. After the multilevel statistical models were formed, statistical tests were performed to examine and evaluate their accuracy.
The statistical tests included t-test for every coefficient to confirm every variable’s statistical significance and likelihood ratio test in order to test the model’s fit.

The final multilevel Poisson model developed is presented in Table 1.

**TABLE 1 Parameter estimates of the multilevel Poisson model**

| Poisson multilevel model - Road accidents and capital cities data |
|------------------|------------------|
| **Dependent variable** | Fatally Injured (at 30 days) |
| **Offset** | Natural logarithm of Length of Road per 1.000 inhabitants |
| **Independent variables** | Coefficient $\beta_i$ | S.E. | $t$-test = $\beta_i$ / S.E. |
| **Fixed Part** | | | |
| cons | -6,322 | 0,591 | -10,697 |
| Month_1 | 0,937 | 0,073 | 12,836 |
| Day of Week_1 | 0,540 | 0,058 | 9,310 |
| Person Age Group_1 | -0,854 | 0,132 | 6,470 |
| Person Age Group_3 | 0,224 | 0,068 | 3,294 |
| Person Age Group_4 | 0,228 | 0,075 | 3,040 |
| Person Gender_2 | -0,416 | 0,062 | -6,710 |
| Road User Type_1 | -0,399 | 0,081 | -4,926 |
| Road User Type_2 | -1,009 | 0,113 | -8,929 |
| Traffic Unit Type_1 | -0,254 | 0,074 | -3,432 |
| Traffic Unit Type_3 | -1,151 | 0,148 | -7,777 |
| Traffic Unit Type_4 | -0,322 | 0,094 | -3,426 |
| Weather_1 | 1,095 | 0,086 | 12,733 |
| Urban population density(persons/ha) | -0,043 | 0,015 | -2,867 |
| Annual private motorised passenger kilometres / Annual public transport passenger kilometres | 0,242 | 0,136 | 1,779 |
| **Random Part** | | | |
| 2nd Level: City | | | |
| $\sigma^2$u$_0$ (Variation coefficient) | 0,123 | 0,063 | 1,952 |
| 1st Level: Case ID | | | |
| Extra-Poisson coefficient | 3,795 | 0,133 | 28,534 |

First, it is noted that the random city variation is statistically significant, suggesting that there is indeed random variation due to unobserved common characteristics of the cities. This effect was significant from the first steps of the model’s development (e.g. empty model) and remained significant even after introducing several characteristics of the cities as explanatory variables. This suggests that there are both fixed and random city effects on the number of fatalities per population.
As regards the fixed parameter estimates, the positive value of the variable Month_1, which corresponds to winter months, indicates that the number of road fatalities is higher during winter months than summer months (July, August). This may be related to the fact that city mobility is reduced during summer due to the holiday season.

The positive value of the variable Day of Week_1, which corresponds to the days from Monday to Friday, indicates that the number of road fatalities is higher in working days than in the weekend. This may be related to the increased mobility during weekdays due to work obligations in all these cities, compared to the lower travel demand during weekends.

The positive values of the variables Person Age Group_3 and Person Age Group_4, corresponding to categories 30-59 and >59 respectively, indicate that the more these variables increase, the more the number of road fatalities increases. This may be due to the fact that more drivers belong to the 30-59 age group and also that people of these ages travel more frequently. The >59 age group, on the other hand, is probably related to the elderly, which are more frequent pedestrians in cities, and also more physically vulnerable as car occupants.

The negative value of the variable Person Gender_2, corresponding to the female category, indicates that the number of road fatalities is higher in men than women. This may be due to the more aggressive driving behaviour of male drivers compared to female drivers, as it is confirmed by international literature. Moreover, the number of male drivers is generally larger than the number of female drivers.

The negative values of the variables Road User Type_1, corresponding to the driver category, and Road User Type_2, corresponding to the passenger category, indicate that the number of road fatalities in these categories is lower than in the pedestrian category. Indeed, pedestrians are among the most vulnerable road users in cities, and each pedestrian involvement in road accidents has increased mortality likelihood due to their difference in mass and speed and their inadequate protection compared to motorists (16).

The negative values of the variables Traffic Unit Type_1, corresponding to the passenger car category, Traffic Unit Type_3, corresponding to the category “bus, goods vehicle”, and Traffic Unit Type_4, corresponding to the “other category” (road tractor, pedal cycle, pedestrian), indicate that the number of road fatalities in the category “motorcycle, moped” is higher than in these categories. This may be due to the combination of increased mobility of two-wheelers in cities, and the high risk exposure that such vehicles have, also due to their less protection compared to motorists.

The positive value of the variable Weather_1, corresponding to the category of dry/clear weather, indicates that the number of road fatalities is higher in good weather conditions compared to adverse weather conditions (rain, snow). This partly reflects the fact that rainy days are less than good weather days in most of the examined cities, but also because drivers tend to be more careful and reduce vehicle speed in conditions of rain or snow (17).

Two city characteristics were found to explain fatality rates, namely the urban population density, and the rate of private to public passenger travel. More specifically, urban density appears to be negatively correlated with fatality rates, and this may be attributed to the fact that cities with higher urban density may be also more congested, resulting in fewer fatal accidents due to lower vehicle speeds.

On the other hand, a higher private-to-public passenger kilometres rate is associated with higher fatality rates, which is rather intuitive; indeed, private passenger transport is more dangerous than public transport, and private cars, two-wheelers etc. have significantly higher fatality rates than public transport vehicles. It is noted that the two variables (i.e. private and public transport passenger kilometres) were also tested separately in the models but were not found to be statistically significant as standalone variables.
Figure 3 shows the interaction between the variable Road user type and the model predictions for the dependent variable (i.e. road fatalities per length of road per 1.000 inhabitants) for each capital city. The fatality rate is higher for pedestrians in all of the capital cities, confirming that pedestrians are among the most vulnerable road users. However, the city of Athens is the only one that differs from this case and has a higher rate in the driver than the pedestrian category. This may be due to the particularly increased private vehicle traffic in the city of Athens, in relation to pedestrian traffic.

**FIGURE 3 Modelling results – Fatality rates per Road User Type and City**

Comparing the capital cities, it appears that the city of Lisbon holds the lowest road safety position among all cities in all road user types. The cities of Athens, London, Brussels and Paris follow in the driver category and London, Bucharest, Brussels and Paris in the pedestrian category. It is reminded that these cities have the lower values of length of road per 1.000 inhabitants, whereas capital cities such as Madrid, Prague and Budapest, which have lower predictions of fatality rates, have higher values of the length of road per 1.000 inhabitants.

**DISCUSSION**

The objective of the present research was the comparative analysis of road safety in selected European capital cities. A database was developed for this analysis containing data regarding the number and the characteristics of road fatalities, the population and other city characteristics and indicators of nine selected European capital cities for the period 2007 - 2011. Multilevel Poisson statistical models were developed, taking into account of the hierarchical structure of road safety data, and they led to the identification of several factors affecting road safety level in the selected European capital cities, including both accident, road user and city characteristics.

The results indicated that the capital cities with the highest road fatalities per road length (Lisbon, London, Athens, Brussels) have the lowest values of the indicator length of road per 1.000 inhabitants, suggesting that the larger the city’s road network is, the higher the level of road safety is in this city.
Especially as regards the city characteristics, it was found that when urban population density (persons/ha) increases, the number of road fatalities decreases. This may be due to higher road congestion in densely populated cities because of the large numbers of moving vehicles. The congestion leads to the reduction of driving speed, whereas in less populated areas there is higher driving speed, which is one of the most common road accident contributory factors.

The indicator “annual private motorised passenger kilometres / annual public transport passenger kilometres” has a positive correlation with road fatalities. This also seems reasonable, since the increase of this indicator equals the increase of annual private motorised passenger kilometres in comparison to annual public transport passenger kilometres, and it is known from the literature that private vehicles have lower road safety level than public transport.

As the results of this research suggest, creating a safe road environment for all road users, but especially for the most vulnerable ones i.e. pedestrians and motorcyclists in cities can be quite challenging. There appear to be several factors, related to the road users characteristics, the cities characteristics, but also external ones (e.g. weather) that interact resulting in the observed level of road safety in cities.

In particular, given the effect that urban population density can have on road accidents, it seems to be highly significant to consider road safety in urban mobility plans. Moreover, given the effect of a low proportion of public passenger transport on the road safety level of different cities, it is confirmed that the promotion of public transport and the shift from private transport to safer modes may be of considerable contribution to road safety.

However, there may be other characteristics of the cities that were not examined in this research and may affect road safety, such as the amount of pedestrian and powered-two-wheeler travel, the mean speed of vehicles, the proportion of different road types, the presence of traveler information systems etc. Unfortunately, this type of data is often not available, especially in a comparable international level.

The potential for analysis at city level is also limited by the lack of definition of the city itself. The city is delimited by its administrative borders, which may include different types of settlements (urban, semi-urban etc.), especially in large agglomerations (e.g. Paris). The cities examined are most often of different size in terms of area covered and number of inhabitants, different modal-split and may include zones which have speed limits above 50 km/h. All these may affect the quality of the analyses (18).

Urban safety policies may also play an important role in the road safety level, and there are clear differences in the policies adopted by different cities. London for example has been able to introduce transport policies, such as congestion charging, which do not exist elsewhere, while Paris and other French cities have been developing and applying specific road safety policies aimed at protecting vulnerable road users and encouraging the use of ecological friendly modes of transport (4). The safety performance in a city will be thus influenced both by transport and safety policies. In turn, policies will affect and be affected by, the spatial layout of the town, the road network provided, and the demand for travel by different modes.

Overall, the analysis of road safety at city level becomes a priority, when considering the increasing and continuous urbanization globally, and the increased share of road fatalities occurring at cities, especially for vulnerable groups. Observing and following road safety data and policies of cities with good road safety level might assist in the better understanding of the factors that contribute to a good safety performance.
REFERENCES


