**Wachnicka Joanna**
*Gdańsk University of Technology, Poland*

# Modelling selected road safety measures at the regional level in Europe

**Keywords**

regions, road safety, factors of influence, modelling, fatalities

**Abstract**

Regions are Europe's basic levels of management. The literature was reviewed to identify regional safety analyses and some of the factors that are important for road safety in the regions. Next, data were collected at the regional NUTS 2 level in Europe for the years 1999-2008. An analysis of the data helped identify factors which have the strongest bearing on fatalities and other safety measures. This paper presents the initial results of a broader research programme on road safety at the regional level.

## 1. Introduction

International road safety analyses have been conducted for a number of years and discussed in numerous publications. The complexity of the problem can be seen in the literature. The analyses are carried out by specialists representing a diversified range of theoretically unrelated disciplines such as highway engineering, economics, mathematics, medicine and transport. In addition, they tend to draw conflicting conclusions. Because they are based on averaged national data, global road safety models may prove too general to be useful for road safety management. As a consequence, analyses must be conducted at a lower level. EU data systems identify three categories of regions called NUTS1, NUTS2 and NUTS3. NUTS1 are major socio-economic regions (3-7 m inhabitants), NUTS 2 are basic regions for the application of regional policies (0.8-3 m inhabitants) and NUTS3 are small regions for specific diagnoses (150-800 thou inhabitants). This classification is not restrictive and regions from a specific NUTS level may be below or above the theoretical population ranges. The research, partly presented in this article, analysed NUTS2 data.

## 2. Previous methodology

Previous analyses of road safety at levels below the national level have usually compared selected rates of road safety (such as the road fatality rate in relation to demography). Fatality is calculated as follows:

$$FATALR_i = \frac{FATAL_i}{POP_i} \quad (1)$$

where:

$FATALR_i$ – traffic fatality rate in an i-th region

$FATAL_i$ – number of fatalities of an i-th region in i-th year

$POP_i$ – number of population of an i-th region in i-th year

This is the first method for classifying regions for their safety.

The second method involves analysis of time series which are used to study the trend of how road safety changes [4].

The third method can be called the factor analysis. Lassare and Thomas [7] have pointed out that population density as a variable gives a good explanation of mortality from crashes on Europe's regional roads. The authors believe that the variable takes account of a number of factors which are frequently not available in regional databases. These include vehicle kilometres travelled, road network structure or level of urbanisation. The authors assumed that a region with a higher population density has a bigger and safer structure of its road network and a greater demand for public transport. Other researchers point to the differences between urban and rural areas. The advantages of urban areas include better availability of rescue services [3],[7], lower vehicle speeds, less tolerance of risky behaviour such as drinking and driving and shorter distances covered by the inhabitants [1]. According to the researchers low population density areas will typically have a different age structure and unsanitary life style which increases risk levels [1], [2]. Zhang [8] suggests that road safety in the provinces is influenced by the gross domestic product.

As mentioned before road safety modelling at the international level has been covered extensively. In 2011 in [5] a hierarchic assessment of road traffic strategic risk was used. The method assumes that risk factors and measures change depending on the country's level of social development, the main determinant of how the road transport system operates - according to the author. Having proved the thesis, the author used the national product per capita as the main parameter of the scale of the forecast. Partial risk models were used in relation to a number of variables grouped into spatial, demographic, economic, social, motorisation, infrastructural and transport parameters of a given country. The models can be used to analyse the effects of the changing factors on the consequences of accidents at the strategic level. This is a new approach to road safety modelling and the author is testing the method at the regional level. This paper includes an analysis of road safety factors in Europe's NUTS2 regions and an attempt to build a multiple factor descriptive model of a selected safety measure.

## 3. Characteristics of the data

To ensure consistency of data collection methodology, the data come from Eurostat. They cover the period 1999-2008. *Table 1* lists the factors which may affect road safety, the abbreviations and units.

*Table 1*. Independent variables collected for the analysis.

| **Independent variables** | | | |
|---|---|---|---|
| **Group of variables** | **Name** | **Abbreviation** | **Unit** |
| Demography | Population | POP | million |
| | Density of population | DPOP | people/km$^2$ |
| | Percentage of urban population | URB | % |
| Geography | Area | AREA | thou. km$^2$ |
| Motorisation | Total number of vehicles | VEH | million |
| | Total number of passenger cars | CAR | million |
| | Total vehicles | VEHR | veh./pop. |
| | Passenger vehicle rate | CARR | car/pop. |
| | Total vehicle density | VEHD | veh./km$^2$ |
| | Passenger car density | CARD | car/km$^2$ |
| | Percentage of passenger cars | CARP | % |
| Infrastructure | Total length of roads | ROAD | thou. km |
| | Length of motorways | MOTOR | thou. km |
| | Total road concentration | ROADC | km/100 pop. |
| | Motorway concentration | MOTORC | km/100 pop. |
| | Total road density | ROADD | km/km$^2$ |
| | Motorway density | MOTORD | km/km$^2$ |
| | Percentage of motorways | MOTORP | % |
| Economy | Unemployment rate | UNEMP | % |
| | National product per capita | NPPC | thou. ID |
| Society | Percentage of secondary school and university students | STUD | % |
| | Percentage of children aged 4 and more in pre-schools | PRES | % |

*Table 2* shows some basic statistics for the independent variables such as the number, minimum, maximum and mean value and standard deviation. Because some variables vary significantly, a logarithm was introduced to normalise them. *Table 2* gives the statistics of the variables after transformation.

*Table 2.* Basic statistics for the independent variables.

| Basic statistics for the independent variables | | | | | |
|---|---|---|---|---|---|
| *Abbreviation* | *Size of sample* | *Minimum value* | *Maximum value* | *Average value* | *Standard deviation* |
| POP | 1846 | 0.065 | 11.636 | 2.015 | 1.5393 |
| DPOP | 1822 | 3.297 | 9362.871 | 428.270 | 971.0890 |
| URB | 176 | 0.404 | 0.796 | 0.596 | 0.1009 |
| AREA | 2084 | 0.013 | 154.312 | 14.346 | 16.0680 |
| VEH | 1938 | 0.027 | 6.456 | 1.134 | 0.9677 |
| CAR | 1906 | 0.028 | 5.709 | 0.987 | 0.8436 |
| VEHR | 1712 | 0.057 | 2.171 | 0.548 | 0.1846 |
| CARR | 1688 | 0.140 | 1.918 | 0.472 | 0.1562 |
| VEHD | 1906 | 1.723 | 3753.846 | 192.740 | 372.2837 |
| CARD | 1882 | 1.578 | 2876.923 | 167.180 | 318.8772 |
| CARP | 1898 | 68.571 | 98.336 | 85.902 | 4.4275 |
| ROAD | 1828 | 0.026 | 91.526 | 14.833 | 15.5198 |
| MOTOR | 2015 | 0.000 | 2.316 | 0.265 | 0.2805 |
| ROADC | 1622 | 0.006 | 4.571 | 0.825 | 0.7300 |
| MOTORC | 1806 | 0.000 | 0.104 | 0.014 | 0.0135 |
| ROADD | 1828 | 0.094 | 13.839 | 1.478 | 1.6746 |
| MOTORD | 1983 | 0.000 | 0.186 | 0.028 | 0.0277 |
| MOTORP | 1799 | 0.000 | 75.532 | 3.571 | 6.2548 |
| UNEMP | 1817 | 1.600 | 28.000 | 8.627 | 5.0912 |
| NPPC | 2132 | 7.666 | 101.568 | 26.419 | 9.548 |
| STUD | 961 | 2.300 | 36.900 | 21.262 | 4.5337 |
| PRES | 891 | 11.500 | 100.000 | 86.646 | 22.3889 |
| lnPGDP | 2132 | 2.037 | 4.621 | 3.20 | 0.3814 |
| lnVEHD | 1906 | 0.544 | 8.231 | 4.544 | 1.0964 |
| lnCARD | 1882 | 0.456 | 7.964 | 4.388 | 1.1190 |
| lnDPOP | 1822 | 1.193 | 9.145 | 5.231 | 1.1184 |
| lnROADD | 1828 | -2.363 | 2.628 | 0.015 | 0.8316 |

The main focus of this paper is on modelling the FATALR mortality rate as the most common measure in the literature.

## 4. Analysis of the relations

First, the simplest multi-dimensional analyses were conducted to study the linear correlations between the features under examination. Unfortunately, independent variables do not have a linear effect on mortality because the highest degree of correlation $R^2$ did not exceed 40%.

In an effort to find the best estimators, the next step was to carry out an analysis using the Data Mining method. As you can see in Chart 1 the programme chose the following as the best descriptive variables: population density logarithms, passenger car density, logarithm of total vehicle density, gross national product per capita and the concentration of motorways and total road density. Social variables such as the unemployment rate only had a slight effect on the safety measure in question.

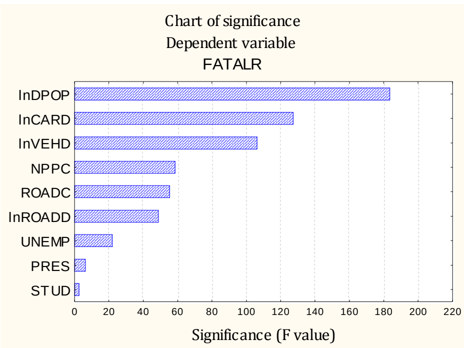


*Figure 1.* Example of how Data Mining was used to identify the best predictors of road fatality rate in relation to demography (FATALR).

While Data Mining analysis is an interesting method for selecting variables for the model, it does not show the shape of the function which could be used for the descriptive model.

As a result, more analyses were conducted. They were single factor analyses of spread diagrams to select a function that will best describe the relations in question. As you can see in *Figures 2* and *3* some data deviate which makes it difficult to select one function which will adequately describe the effect of the variable on the mortality rate.

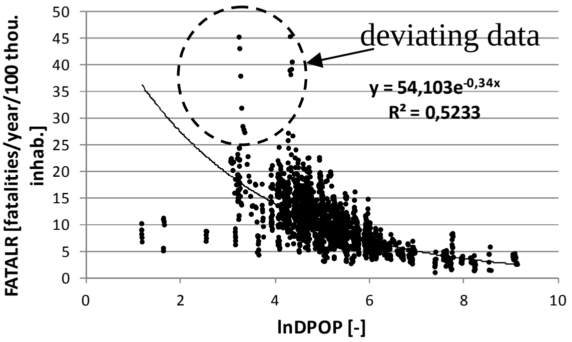


*Figure 2.* Relationship between the road fatality rate in relation to demography and the logarithm of the lnDPOP population density.

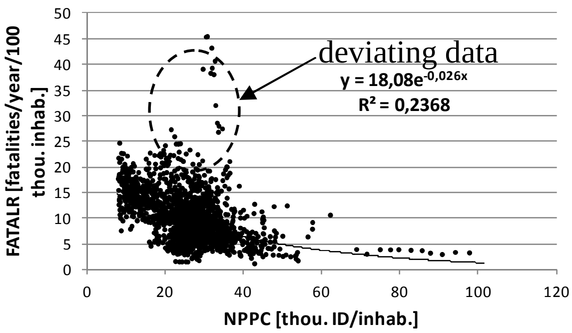


*Figure 3.* Relationship between the road fatality rate in relation to demography and the NPPC gross national product per capita.

A number of relationship charts were analysed and it was established that the data come from two NUTS2 regions in Spain: the Basque Country and Navarra. Because they deviate significantly from the others, the data were removed and were not used in the models. With the deviating data removed, the relationship chart in *Figure 2* was repeated. The figure suggested that the relationship could be described with the exponential function. However, a review of other analyses of national data showed that the relationship has a slightly different characteristics. As an example Koopitz [6] suggested the use of a spline. According to the authors the function should resemble the shape shown in *Figure 4*. As we can see in *Figure 5* the function inflects around 8 thou. ID/population and what is a growing risk becomes a downward trend described with the exponential function.

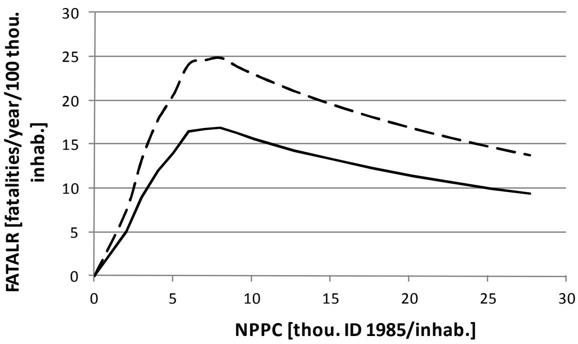


*Figure 4.* Relationship between the road fatality rate in relation to demography and the NPPC gross national product per capita as proposed in the paper [6].

As a result, the power-exponential function was tested followed by the exponential function. Figure 5 shows the results.

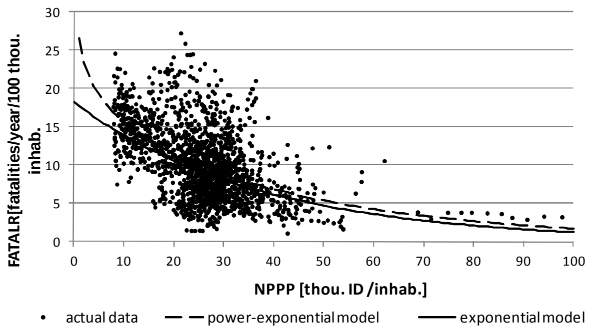


*Figure 5.* Relationship between the road fatality rate in relation to demography and the NPPC gross national product per capita. An attempt to match the exponential and power-exponential function.

Unfortunately, given the scope of data used for the power-exponential model, the inflection point could not be found. The minimum value of NPPC is almost 8thou. ID which is close to what Koopitz identified as the inflection point. Although the deviating data were excluded from the analyses, the Q factors of the models do not exceed 0.3. Because the NPPC has a wide range of values compared to the scope studied by Koopitz, the data were normalised by using a logarithmic value of NPPC. The relationship was then described with the power-exponential function. While the logarithm helped to achieve the inflection point in the power-exponential function in question, the Q factors are still less than 0.3.

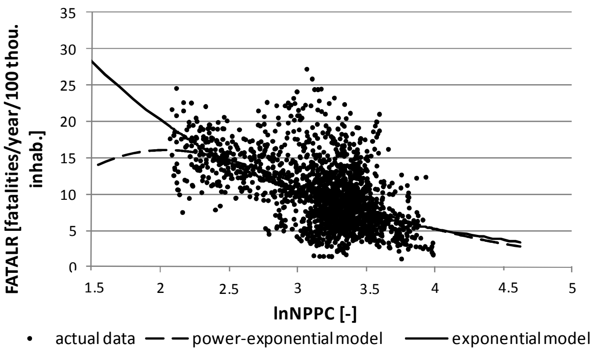


*Figure 6*. Relationship between the road fatality rate in relation to demography and the natural logarithm from the lnNPPC gross national product per capita. An attempt to match the exponential and power-exponential function.

## 5. Relationship modelling

Based on the preliminary analyses of variables and studies of the literature, the power-exponential model was selected for the modelling.

The software STATISTICA was used to analyse more than ten models of the relationship between traffic mortality and the factors in question. Model (3) was chosen as one of the best descriptive models. *Table 3* presents the parameters of the algorythm depending on the number of independent variables.The models' Q factors range between 50 and 60% which is a good result for the set comprising yearly averages. As you can see in *Table 3* the significant explanatory factors include the logarithmic gross national product per capita lnNPPC and logarithmic population density.

The additional variables which differentiate the regions and ensure a better match of the model inlude logarithmic total vehicle density lnVEHD, percentage of passenger cars in total fleet CARP and total road concentration ROADC. Unfortunately, the level of social welfare, a factor covered in other publications, could not be included in the model. When the percentage of children going to pre-schools was included in the model, the $R^2$ was less than in models without the above factor.

$$y = \alpha_1 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot e^{(\gamma_1 \cdot x_1 + \ldots + \gamma_n \cdot x_n)} \qquad (2)$$

$$FATALR_i = \alpha_1 \cdot \ln NPPC^{\beta_1} \cdot e^{(\beta_2 \cdot \ln DPOP + \beta_3 \cdot \ln VEHD + \beta_3 \cdot CARP + \beta_5 \cdot ROADC + \beta_6 \cdot PRES + \beta_7 \cdot \ln NPPC)} \qquad (3)$$

*Table 3*. Model parameters after calculation.

| Model FATALR$_i$ | $\alpha_1$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ | $\beta_6$ | $\beta_7$ | p | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | (lnNPPC) | (lnNPPC) | (lnDPOP) | (lnVEHD) | (CARP) | (ROADC) | (PRES) | (lnNPPC) | | |
| 1 | 2117.008 | -1.842 | -0.758 | 0.586 | -0.023 | - | - | - | 0.00 | 0.56 |
| 2 | 2166.961 | -1.845 | -0.806 | 0.610 | -0.021 | -0.036 | - | - | 0.00 | 0.57 |
| 3 | 1692.812 | -1.294 | -0.789 | 0.676 | -0.027 | - | -0.003 | - | 0.00 | 0.39 |
| 4 | 1026.21 | -2.778 | -0.453 | 0.281 | -0.022 | - | - | 0.468 | 0.06 | 0.56 |

*Figure 7* shows an illustration of the $FATALR_4$ model with a variation of the logarithmic annual national product per capita lnNPPC. The other variables are taken as minimum, mean and maximum values from the set under analysis. As you can see from the chart when population income increases especially from minimum values to around 8 thou. ID/year, traffic mortality in theory decreases the fastest. However, it is possible that the model produces false results (as suggested by the literature referred to in the introduction). Because the data available for European regions do not include NPPC below 7 thou. ID/year, it is more likely that the descending function FATALR in relation to lnNPPC is correct for data from the range 8-100 thou. ID/year (2-4.6 on the logarithmic scale in the chart respectively).

*Figure 8* shows an illustration of the $FATALR_4$ model with a variation of the logarithmic annual national product per capita lnNPPC, and population density DPOP. The other variables are taken as minimum, mean and maximum values from the set under analysis. As you can see from the chart when population density increases FATALR decreases, but it the impact is bigger in case of small density.

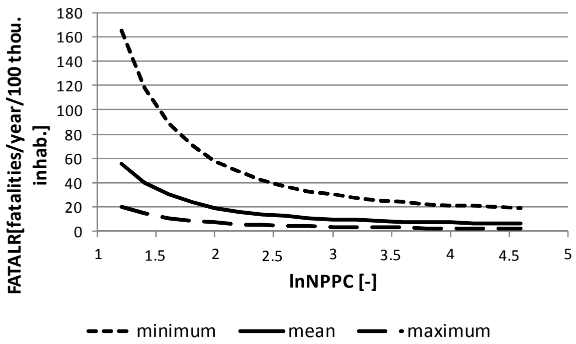


*Figure 7*. Visualisation of the $FATALR_4$ model in relationship to the natural logarithm from gross national product per capita and the other variables at minimum, mean and maximum values respectively.

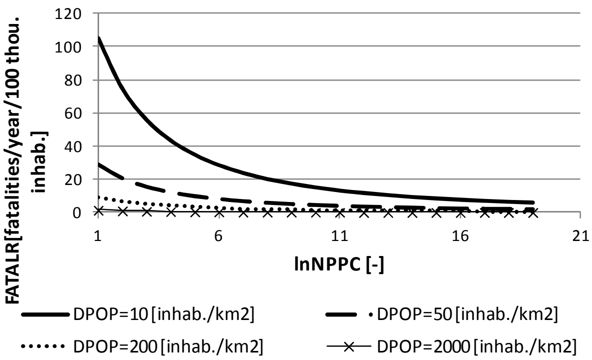


*Figure 8*. Visualisation of the $FATALR_4$ model in relationship to the natural logarithm from gross national product per capita, population density DPOP and the other variables at minimum, mean and maximum values respectively.

As you can see in *Figure 9* smaller percentage of passenger car results in higher FATALR, especially in case of lower NPPC.

## 6. Conclusion

The objective of the research was to identify the most significant factors affecting regional road safety performance and develop respective models. Studies of the literature showed that this area of road safety is poorly researched. As part of her own research the author developed regional databases, conducted preliminary analyses and started work on the models. This paper presents an example of a specific descriptive model. The results will be described in detail and more extensively in the author's doctoral thesis, currently under development. The analyses and work on developing reliable models show that the problem is complicated.

As stated in the literature, population density is in fact a significant variable and affects the FATALR variable in the models. However, more factors must

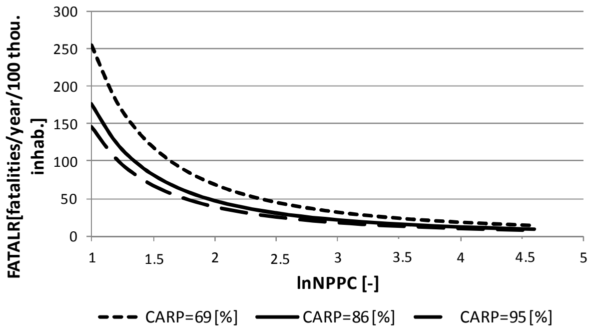


*Figure 9*. Visualisation of the $FATALR_4$ model in relationship to the natural logarithm from gross national product per capita, percentage of passenger cars and the other variables at minimum, mean and maximum values respectively.

be included to produce a better description. Available data are frequently collected for economists or demographers and do not give sufficient insight into the diverse levels of risk across regions.

To ensure proper road infrastructure management and an effective delivery of road safety programmes, a variety of data must be collected such as seatbelt rates, exceeding speed limits, drinking and driving or drugs and driving. Sadly, the data are not commonly available and sometimes never collected at the regional level. As an example, Poland ran a study on seatbelt wearing between 2000 and 2004, but it has never been repeated. The US collects information about the average time it takes to transport a car accident casualty to a hospital from the time of notification to show regional differences. No one in Europe does that.

As a consequence, further work will be aimed at developing descriptive and prognostic models using data already collected. In a long-term perspective there are plans to add more detailed data to the database. This is to be achieved by sending letters to the competent authorities and collaborating with researchers from other countries who work on similar problems.

## References

[1] Baker, S., Whitfield, R.A. & O'Neill, B. (1987). Geographic variations in mortality from motor vehicle crashes. *N Engl J Med*, 316, 1384-7.

[2] Clark, D.E. (2003). *Effect of population density on mortality after motor vehicle collisions.* Accid Anal Prev, 35, 965-971.

[3] Clark, D.E., Cushing. B. (1999). Predicting regional variations in mortality from motor vehicle crashes. *Acad Emerg Med*, 6, 125-130.

[4] Ewans, A.W. (2003). Estimating transport fatality risk from past accident data. *Accid Anal Prev*, 35, 459-472.
[5] Jamroz, K. (2011). *Metoda zarządzania ryzykiem w inżynierii drogowej*. Wydawnictwo Politechniki Gdańskiej, ISBN 978-83-7348-342-2.
[6] Koopitz, E. & Cropper, M. (2005). *Traffic fatalities and economic growth. Accid Anal Prev,* 37, 169-178.
[7] Lassarre, S. & Thomas, I. (2005). *Exploring road mortality ratios in Europe: national versus regional realities. J R Stat Soc* [Ser A] 168, 127-144.
[8] Zhang, W. et al. (2010). Road safety in China: Analysis of current challenges. *Journal of Safety Research* 41, 25 –30.