D7.8
The CARE data in perspective

Please refer to this report as following:
Martensen, H. & Dupont, E. (Eds.) (2008). The CARE data in perspective. Deliverable D7.8 of the EU FP6 project SafetyNet.

Contract No: TREN-04-FP6TR-S12.395465/506723
Acronym: SafetyNet
Title: Building the European Road Safety Observatory

Integrated Project, Thematic Priority 6.2 "Sustainable Surface Transport"

Project Co-ordinator:
Professor Pete Thomas
Vehicle Safety Research Centre
Ergonomics and Safety Research Institute
Loughborough University
Holywell Building
Holywell Way
Loughborough
LE11 3UZ

Organisation name of lead contractor for this deliverable: IBSR

Due Date of Deliverable: 31/10/2008

Submission Date: 31/10/2008

Report Author(s): Brandstaetter, C. (KfV), Chaziris, A. (NTUA), Dupont, E. (IBSR), Eksler, V. (INRETS), Lassare, S. (INRETS), Martensen, H. (IBSR), Papadimitriou, E. (NTUA), Yannis, G. (NTUA)

Project Start Date: 1st May 2004 **Duration:** 4.5 years

Project co-funded by the European Commission within the Sixth Framework Programme (2002 -2006)
Dissemination Level: PU

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

Table of Contents

Executive Summary

This deliverable presents the CARE data in the light of different augmenting data. In Chapter 1, the Introduction, different types of augmenting data sources are described and an overview is given over the results.

In Chapter 2, a spatial analysis of Greek CARE data is presented. For this purpose the accident and fatality numbers for Greece were disaggregated to the county level and it is shown how these data can be used to create a road-safety map for the whole country. It is explained how differences between counties with respect to the number of accidents or fatalities per inhabitant are for some part determined by their location: Neighbouring counties tend to be more similar than counties located far away from each other. The method of spatial analysis is introduced and it is demonstrated how the systematic “neighbourhood structure” in the accident/fatality data can be disentangled from those differences that occur purely at random. By comparing different spatial models, it is tested whether regions that are connected by ferry only could be seen as neighbouring regions just like those connected by roads. The answer for the case of islands is that ferry connections do not facilitate neighbourhood effects in the way that road connections do.

In Chapter 3, the information about injury accidents in CARE is compared to that from hospital data. This chapter capitalizes on work done in task 1.5 of SafetyNet, in which the CARE data were compared to and matched with hospital data, and the proportion of accidents that were not reported in CARE subsequently estimated for 7 countries. On this basis, correction factors for the injuries reported in the CARE database were determined. In Chapter 3, these correction factors themselves were then introduced into an analysis to investigate whether there are general trends which sort of injury accidents tend to become subject to under- or misreporting. This way, it was not only shown that injuries in CARE are not always correctly reported, but also that the way this happens differs widely between countries, injury types and road-user types.

In Chapter 4, an example is shown how the CARE data can be used as a reference for another database. More specifically, the Fatal Accident Investigation Database was compared to the CARE Database on a number of key variables. It was investigated whether the distribution of accidents in both samples was the same for the categories of several important variables. It was investigated whether they involved the same person and vehicle types in the same proportions, whether they took place at the same places (type road, junction, area type) and under the same circumstances (weather, lightning conditions). For the vehicle and person type variables it can be concluded that the distributions are very similar in the CARE and FAI data. For the other variables the general patterns (which categories occur often which seldom) in the CARE data are reproduced by the FAI data, but the exact proportions can differ.

In Chapter 5, the relation between accident and fatality data from the CARE database and attitude data from the SARTRE project is investigated. Both databases were disaggregated to the level of country, gender, and age. On the basis of group membership, the CARE and the Sartre data were joined together. The attitude and behaviour data from Sartre were analysed in a Principal Component Analysis and three main components were identified: Aggressiveness and Speeding, (2) Drink driving and none-use of seatbelt, and (3) Perceived likelihood of control.

It was shown that a positive attitude towards speeding and aggressiveness occurred more frequently in groups that also have a higher number of accidents and fatalities. This is true for men as well as for women.

Chapter 1 - Introduction

CARE is a Community database created with the aim of centrally recording road accidents that resulted in death or injury throughout Europe. It comprises detailed data on individual accidents as collected by the member states. When national data sets are integrated into CARE they are transformed according to a series of transformation rules, allowing the data from all member states to be compatible despite original differences between countries in terms of structure and definition. The purpose of the CARE system is to provide a powerful tool which makes it possible to identify and quantify road safety problems throughout the European roads and evaluate the efficiency of road safety measures.

As demonstrated in the Annual Statistical Reports (e.g., 2008) which are derived from the CARE database and published on the website of the European Road Safety Observatory (www.erso.eu), the accident and fatality data in themselves can give detailed insights into the distribution of different types of accidents across 18 EU Member States.

For a more complete picture of road-safety, these data have to be augmented with other information. Generally speaking, the most important type of augmenting information is risk exposure, for example, the number of km driven or the size of the population (see D7.10, Stipdonk & Van Norden, 2008, that is dedicated to the joint modelling of accident and exposure numbers). In the present document, the CARE data is put in relation with information from other sources: spatial information, injury information gathered from hospitals, and attitude and (selfreported) behaviour information from the Sartre project (http://sartre.inrets.fr/).

1.1 Spatial modelling of road safety in Greece

In Chapter 2, on the example of Greece, it is presented how the spatial structure of a country can be integrated in an analysis of its data. For this purpose the accident and fatality numbers for Greece were disaggregated to the county level (Nuts-3). The differences between counties with respect to the number of accidents or fatalities per inhabitant are for some part determined by their location: Neighbouring counties tend to be more similar than counties located far away from each other. The method of spatial analysis is introduced and it is demonstrated how the systematic “neighbourhood structure” in the accident/fatality data can be disentangled from those differences that occur purely at random. These data can be used to create a road-safety map for the whole country.

Spatial modelling can also be used to compare different descriptions of the spatial structure of a country. To set up a spatial model, one has to define which counties are neighbours and which are not. In road-safety this is not always straight forward and in it self an interesting question: Is it the arithmetic distance that makes some pairs of counties more similar than others? Or is it the fact that they are connected by a road? Does it matter what kind of a road it is?

In our Greek example, different spatial models are compared to test whether regions that are connected by ferry only could be seen as neighbouring regions just like those connected by roads. It was concluded that ferry connections do not facilitate neighbourhood effects in the way that road connections do.

The Greek example illustrates the principle of spatial modelling which can also be applied to other countries or scaled up to larger regions (e.g. countries within Europe). It can be used to identify borders in road-safety. Such borders can be political or natural borders … or anything that might make a difference to the occurrence of accidents or fatalities on either of its sides. Spatial models can be used to test which regions affect each others' accident and fatality occurrences and which do not. Questions like this can be important when determining how far a road-safety measure has to reach to be effective or which areas can be candidate for isolated measures.

1.2 Injury reporting

Chapter 3 forms an example for using external information to identify problems with the information in CARE as so far. The injury classifications (seriously, slightly) for non-fatal accidents in CARE is compared to those from hospital data (MAIS 1-6). This chapter capitalizes on work done in task 1.5 of SafetyNet, in which the CARE data were compared to and matched with hospital data. The proportion of accidents that were not reported in CARE was subsequently estimated and formed the basis for correction factors for the injuries reported in the CARE database.

In Chapter 3, these correction factors themselves were then introduced into an analysis explaining them as a function of the severity indicated in the police report, the severity reported in the hospital record, the country, and the type of road user. On the basis of the resulting log-rate model the actual number of casualties for a particular category can be estimated, given the number of casualties recorded by the police. This can be done separately for severely and slightly injured victims, for different road user types and for different countries.

The analysis demonstrated that countries differ widely in their degree of underreporting. Moreover they do so in different ways for the different road-user types and for the different degrees of severity. For example, the results suggest that injury underreporting is higher in France, in Greece and in Hungary compared to the underreporting in the other examined countries. However, the more disaggregate results reveal additional effects, which may assist in the understanding and improvement of the increased underreporting in these countries. For example, in France there is higher underreporting of serious injuries, whereas in Greece there is higher underreporting of slight injuries, compared to the other countries, and therefore national authorities should focus on these particular groups respectively.

1.3. CARE as a Reference: Comparison with the Fatal Accident Investigation Database

In Chapter 4, the CARE data is used as a reference for another database. The Fatal Accident Investigation Database (FAI database, Reed & Morris, 2006) was compared to CARE on a number of key variables. It was investigated whether the distribution of accidents in both samples was the same across the categories of several important variables.

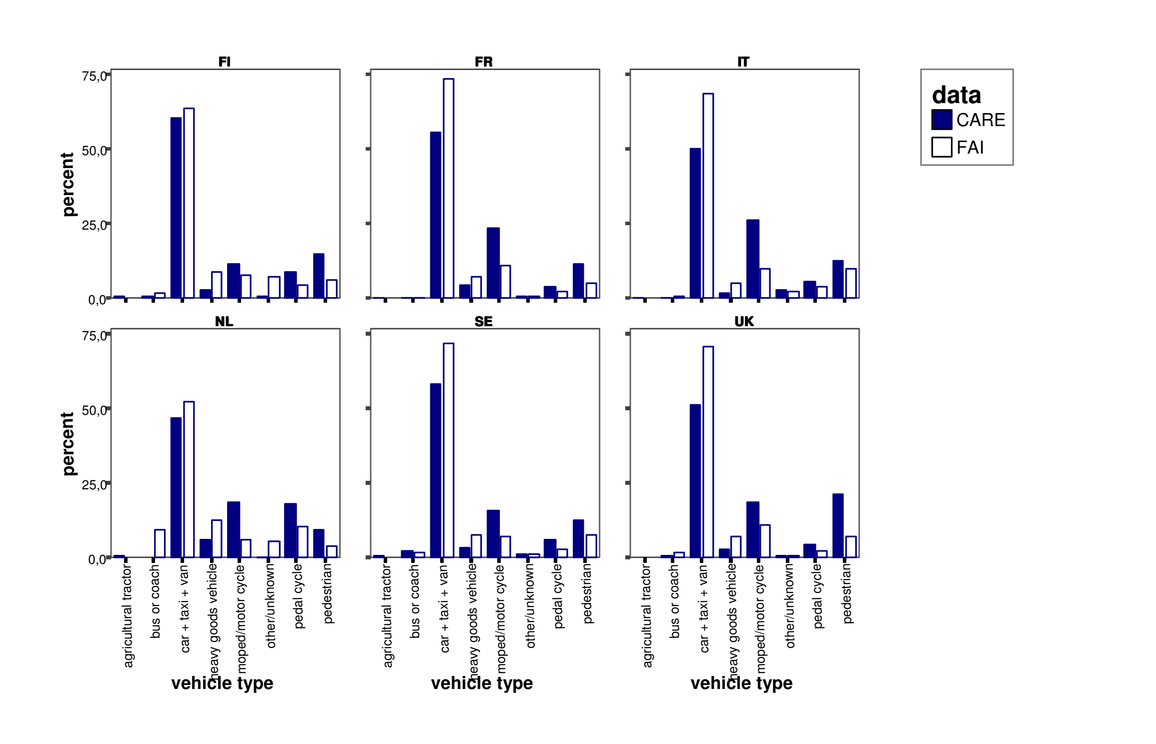


Figure 1.1 Distribution of vehicle types in CARE and FAI data

For the distribution of age and gender, both samples were very similar. With respect to person class (driver, passenger, pedestrian) the samples were also very similar, however, with a slight overrepresentation of drivers and an underrepresentation of pedestrians in the FAI data as compared to CARE. A similar tendency can be seen with respect to vehicle type (see Figure 1.1): The proportions for each vehicle type in each country are very similar in the FAI data and in CARE, but almost all countries show an overrepresentation of cars and an underrepresentation of vulnerable road users (especially motor cycles).

For variables that describe the location (road-type, junction vs. road-segment, rural vs. urban are) or the circumstances of the accidents (weather, lightening conditions) the general patterns (which categories occur often which seldom) in the CARE data are reproduced by the FAI data, but the exact proportions can differ. The only variable that shows markedly different distributions in both samples is one describing the use of protective systems. In CARE, this variable is almost always unknown, while the FAI data contain information about seat-belt and helmet use.

Generally, it can be concluded that the FAI data are a good sample of the fatal accidents collected in CARE.

1.4 Attitudes in Road Safety: Joint analysis of CARE and SARTRE data

One of the important factors in road safety is the driver's behaviour and underlying the attitudes guiding their behaviour. It is therefore interesting to see whether there is a relation between an accident database like CARE and a database containing data about road-safety attitudes in Europe, like the one resulting from the SARTRE project. However, CARE and the SARTE data are inherently different: SARTRE collects attitudes from people who did not have an accident while CARE is a collection of accidents involving people of whom we do not know the attitudes. To overcome this difference, both databases were disaggregated to the level of country, gender, and age. As an example, the number of accidents, the number of fatalities and the answers of questions on the SARTRE questionnaire were determined for 18 year old male Austrians.

The attitude and behaviour data from Sartre were analysed in a Principal Component Analysis and three main components were identified: Aggressiveness and Speeding, (2) Other unsafe behaviour (seat belt, drink driving), and (3) Perceived likelihood of control. This means that in groups where many people admit to show aggressive behaviour in traffic also many people admit to speeding, whereas other unsafe behaviours like drink driving and not using a seatbelt is not necessarily shown in these groups.

As a result of the Principal Component analysis, each age gender and country group (e.g., the 18 year old male Austrians) gets three scores, one for each of the three components. Subsequently these scores were related to the number of accidents and fatalities for each of these groups. It was shown that a positive attitude towards speeding and aggressiveness occurred more frequently in groups that also have a higher number of accidents and fatalities. This is true for men as well as for women. It is important to note that this statistical relation does not prove that the groups' members were indeed aggressive and speeded, nor that their attitudes actually caused the accidents. However, it shows that a positive attitude towards speeding and aggressiveness is most typical for the problematic groups and might therefore be the most promising attitude to be addressed in campaigns.

1.5 Outlook

To conclude, this deliverable shows a wide range of approaches to analysing the CARE data. On the one hand the CARE data still need to be augmented themselves, as will be demonstrated in Chapter 3, where it is not only shown that injuries in CARE are not correctly reported, but also that the way this happens differs widely between countries, injury types and vehicle types. On the other hand they can be used as reference data for other databases, as will be demonstrated in Chapter 4, where the CARE data are used to asses the representativity of the accidents collected in Fatal Accident Investigation Database. Finally it is demonstrated how the CARE data can be used at different levels of disaggregation. In Chapter 1, the data disaggregated to the county level are considered jointly with spatial information. In Chapter 5, the country data are disaggregated by age and gender and considered jointly with attitude data from the Sartre Project.

Chapter 2 – Modelling spatial effects in road safety analysis: an example from Greece

George Yannis (NTUA), Eleonora Papadimitriou (NTUA), Vojtech Eksler (INRETS), Sylvain Lassare (INRETS)

2.1 Background and objective

Spatial analysis includes a broad range of techniques, which may be applied in different fields, in order to study entities or observations in relation to their topological, geometric or geographical properties. The basic idea around which these techniques have been developed is that entities or observations that are closer to one another are more likely to be related or similar than entities that are distant from one another. In other words, a part of the properties of an entity are due to the properties of the nearby entities.

Spatial dependence is therefore a covariation of properties within a geographical space: characteristics at proximal locations appear to be correlated, either positively or negatively. As a consequence, spatial correlation can be understood as the spatial analogue of temporal correlation in time series analysis, where sequential observations are correlated in time. In both cases, this dependence, if not accounted for, leads to a violation of the independence among observations assumption, used by most standard statistical analysis methods. On that purpose, different techniques have been developed in order to capture the spatial autocorrelation among observations.

In road safety research, spatial dependences may be involved in different research areas. For example, when comparing countries in terms of road safety features, the geographical relationship between countries needs to be accounted for, in order to test whether neighbour areas are similar in terms of road safety. Moreover, the spatial dependences within each country may affect each country's overall road safety outcomes. For instance, it is likely that the poor road safety performance of a country is due to the disappointing performance of particular areas, with specific demographic or transport characteristics. In order to identify and quantify these effects, it is necessary to work on a more disaggregate spatial level than the country level and examine regions or counties within a country.

Road safety research traditionally focused on the identification of explanatory effects on the variation of road accident risk, through the incorporation of different (and often numerous) variables in the models.
In the recent years, there was an increasing interest in identifying and modelling the spatial dependence among road safety outcomes (Amoros et al. 2003;

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

Noland and Quddus, 2004, Aguero-Valverde and Jovanis 2006, Yannis et al. 2007). These studies revealed important spatial effects that, once accounted for, allowed for a more accurate estimation and interpretation of the additional explanatory effects (Eksler and Lassarre, 2008).

In most of these studies, a relatively small spatial scale of regions, counties, or even municipalities is considered, and the spatial dependence is defined by means of the neighbourhood structure of the spatial units i.e. the physical borders between regions, counties or municipalities. From a transportation / road safety viewpoint, neighbouring counties may share common infrastructure (e.g. road or railway connections) and present similar demographic and economic characteristics, resulting in similar road and traffic environments and road users' behaviour, which in turn result in similar road safety outcomes. Systematic county-specific effects (e.g. local road safety regulations, enforcement etc.) can then also be investigated in light of the given neighbourhood structure.

The added value of using such an approach, either within a country or across Europe was demonstrated in recent research (Eksler, 2008), where a number of interesting examples highlight the importance of investigating the spatial structure of road safety outcomes:

- At European level, a series of neighbouring (spatially dependent) regions was identified, stretching from north-west Spain to north-east Poland, which presents increased mortality risk. This set of regions appears to be homogenous and not affected by the presence of national borders. In particular, this set of regions appeared to correspond to areas of international transit corridors for heavy goods vehicles.
- A clear east-west spatial pattern can be identified in Germany, accounting for part of the random variation in mortality risk. In particular, the eastern regions of Germany present increased mortality risk compared to the western ones, and this pattern can only be identified once the spatial dependence among regions is examined separately.
- Accordingly, a distinct south-north pattern is observed in Belgium, where the spatial analysis revealed increased mortality risk in the southern region of Wallonia.

Within this context, the objective of this section is to provide an additional example of spatial effects in the analysis of road accidents and fatalities. Greece is selected as a country for which spatial effects have not been previously explored, and also as a country presenting some spatial particularity making such an analysis interesting. More specifically, a large part of Greece lies on the Aegean and Ionian Seas, including several clusters of small or quite big islands. Therefore, many neighbouring counties may share a physical border.

Therefore, four types of models are presented in this section:

1. Models without spatial effects, i.e. standard Poisson models for accidents counts

2. Models with spatial effects in accidents counts, in which the spatial structure is defined, as in most studies, by means of the physical borders between counties. This is translated to counties connected through a road connection.
3. Models with spatial effects in accidents counts, in which the spatial structure is defined by means of either physical or sea borders between counties. More specifically, counties connected through either road or ferry connections are considered as neighbouring counties.
4. Models with spatial effects in accidents counts and explanatory variables.

The third case is considered as an interesting test on whether spatial effects in road safety are indeed effects rising not only from similar infrastructure and socio-economic features, but also from similar users' behaviour characteristics (e.g. it is likely that the fact that two counties are connected, even by ferry, implies an exchange of local road safety behaviours and attitudes). The fourth case serves as a typical example of the consequences of ignoring possible spatial effects on the estimated effects of explanatory variables.

Through these examples, the advantages of this approach in road safety research are discussed and the usefulness of applying such methods at European level is demonstrated, both as regards the CARE data and national data.

2.2 Methods

2.2.1 Spatial modelling

The spatial effects in the studies mentioned above (Germany, Belgium, whole of Europe - Eksler, 2008, Eksler and Lassarre, 2008) would have never been identified by simply comparing the mortality rates between different counties, because the spatial dependence is an unobserved (latent) component, "hidden" into the total variation in the mortality rates.

The basic idea behind spatial modelling techniques is the decomposition of the variation of road accident risk into two distinct components: a "structured" component, which is assumed to represent the spatial structure of the road safety outcomes, and an "unstructured" component, which is assumed to be random. Therefore, the road safety outcome Y_i of county (i) is considered as a result of systematic variables effects $\beta_i x_i$, and random variation ε_i, which is further decomposed into "structured" variation u_i and unstructured variation v_i. Typically, a Poisson distribution is assumed for road safety outcomes, and the logarithm of the respective counts is used as dependent variable, with an exposure estimate (e.g. the population N_i) incorporated as an offset term, in order to model road accident rates instead of counts:

$$Y_i / \varepsilon \sim \text{Poisson}(\lambda_i N_i)$$

where λ_i is the Poisson parameter, so that:

$$\log(\lambda_i^{\varepsilon}) = \log(N_i) + \beta_i x_i + \varepsilon_i$$

With $\varepsilon_i = u_i + v_i$

Such a model, incorporating a random effect within λ_i (i.e. combining fixed and random effects), is sometimes called a mixture model. The Bayesian modelling (see next section 2.2.2) may be implemented for rates (Y_i/N_i) and counts (Y_i), assuming that the corresponding counts follow a Poisson distribution while conditioned on random effects.

The random effect ε_i captures all the uncertainty regarding the differential mortality rate in each geographical unit, such as that arising due to reporting error, missing covariates, over-dispersion or even genuine underlying differences in mortality rate. Then the exponential of random effect $\exp(\varepsilon_i)$ represents the local area mortality ratio, adjusted for population (also called Bayes relative risk).

There are alternative ways to estimate the structured component of the random variation in statistical modelling, however all of these can be considered to fall within the family of hierarchical (multilevel) models. The most common models used in spatial analysis are the MMMC model (Multiple Membership Multiple Classification model), the CAR model (Conditional Auto-Regressive model) and the CAR convolution model. A detailed presentation of the formulation and statistical properties is beyond the scope of this section, and only a summary of basic assumption is presented here. The reader is referred to Appendix 2.1 for the formulation of these models. For a complete presentation, please see Eksler and Lassarre (2008).

Spatial models are based on a "neighbourhood matrix", in which the spatial structure is defined through the identification of neighbouring regions. Obviously, the neighbourhood matrix needs to be symmetrical (i.e. if region i is indicated as a neighbour of region j, then region j should also be indicated as a neighbour of region i). The way this neighbourhood matrix is used differs between different methods.

The MMMC model is a multilevel model, which combines a cross-nested structure and a multiple membership structure (for details see Dupont and Martensen, 2007). The cross-nested structure is used to describe the fact that for each county there is part of the variation coming from the county itself (i.e. overdispersion in the accident counts) and a part of the variation coming from the neighbourhood structure (i.e. spatial effects). The multiple membership structure is used to describe the fact that each county has more than one neighbour. The MMMC model assumes that counties are separate entities, adopting thus a rather strict consideration. Two types of random effects are estimated in an MMMC model: an "exchangeable county effect" to account for overdispersion, and a "neighbourhood" effect to account for spatial dependence. The remaining random variation is the unexplained part of the model.

The CAR model uses a slightly different approach, in which counties are no longer separate entities (Browne, 2004) and a conditional auto-regressive distribution is initially assumed for the structured variation. Moreover, while in the MMMC model there are (r) random effects for each county, with (r) equal to the number of neighbours of the county, the CAR model assumes one (global) random neighbourhood effect for each observation. Consequently, in the CAR model, the set of neighbours of each county is initially examined as a whole.

Moreover, no exchangeable county random effect is initially assumed in the CAR model and consequently only the structured part of the random variation can be estimated. Incorporating such an effect results in the CAR convolution model, in which both the structured and the unstructured components are estimated, through a mixture of an exchangeable normal and a conditional autoregressive distribution.

In general, the CAR method results are considered to be more reliable, for both theoretical and practical reasons. First, the independence among counties assumption is more relaxed in the CAR models, which are anyway a more standard method for spatial analyses than the MMMC method. Secondly, CAR models can fitted through standard software packages for spatial analysis (e.g. WinBUGS), whereas the MMMC models can be fitted through the related option of the MLwiN software for multilevel modelling. For these reasons, CAR family modelling results are considered to be more reliable and therefore MMMC results are not presented in this section.

Regardless of the method used, the estimated unstructured variance and spatial variance are not directly comparable, given that different distributional assumptions are used for each component. However, the relative contribution of spatial variance versus the unstructured variance can be used instead. The estimation of the relative importance of each component of the total random effect variations can be done by comparing the marginal variability of the structured part of random effects and the variance of the unstructured part of random effects. This is done through a ratio called spatial fraction (k), which stands for the fraction of total variation in log relative risks due to spatial effects.

$$\kappa = \sigma^2(u_i)/[\sigma^2(u_i) + \sigma^2(v_i)]$$

A value of the spatial fraction close to 1 indicates the domination of spatial variation over the unstructured variation, while a value close to 0 indicates the domination of unstructured variation.

2.2.2 Bayesian modelling

The above models are estimated by means of Bayesian modelling (for details see Dupont and Martensen, 2007), which allow for an estimate of the "posterior distribution" of a parameter to be obtained, after a large number of successive draws from an initially assumed "prior" distribution. This could not be done analytically for the examined spatial models. However, unlike many numerical algorithms, this is a simulation approach, which is not guaranteed to converge around the correct solution. Hence, in addition to models fit and diagnostics, it is necessary to carry out convergence diagnostics as well. In this section, models are run for 20,000 iterations, from which the first 15,000 are considered as "burn-in" and are therefore discarded. The presented values are the averages of the chains 15,000-20,000.

The Bayesian Deviance Information Criterion (DIC) is used to assess the models fit. A reduced DIC indicates an improved model. Moreover, the parameter estimates presented in the following sub-sections are interval estimates, for which the values on various quantiles are available, however only the mean is reported here. For details on Bayesian modelling assumptions and tests see Appendix 2.1.

2.3 Data

A NUTS-3 spatial classification is opted for the Greek road safety data (accidents, fatalities, and population) of year 2002. According to the NUTS-3 classification, 51 counties are considered, and their neighbourhood structure is defined in two ways: first, according to the road connections between counties (see Appendix 2.2) and second, according to the road and/or ferry connections between counties (see Appendix 2.3).

The variables and values used in the analysis are summarized in Table 2.1:

County	The 1-51 counties of Greece
Accidents	The number of accidents of each county on year 2002
Fatalities	The number of accident fatalities of each county on year 2002
lnpop (offset)	The natural logarithm of the population of each county on year 2002
Cons	The constant term
Lnalcpop	The natural logarithm of the number of alcohol controls per population

Table 2.1: Variables and values used in the analysis

2.4 Results

2.4.1 Models without spatial effects

We start from the basic "empty" models, in which no spatial dependence is considered and only a constant term is included. Both log-accidents and log-fatalities are examined. In these models, presented in Table 2.2, the constant term corresponds to the mean accident and fatality rate of all counties of Greece.

Project co-financed by the European Commission, Directorate-General Transport and Energy

		Fatalities		Accidents	
	Parameter	Estimate	Sd	Estimate	Sd
	Constant	-8.812	0.025	-6.518	0.008
DIC		658.41		2939.88	

Table 2.2: Single level Poisson model (15,000-20,000 iterations)

2.4.2 Models with spatial effects based on road connections

The first type of spatial dependence to be examined is the one resulting from neighbourhood effects of counties connected by roads (see neighbourhood matrix in Appendix 2.2). In this case, the minimum number of neighbours for one county is zero (corresponding to island counties) and the maximum number of neighbours for one county is 7.

		Fatalities		Accidents	
CAR model	Parameter	Estimate	Sd	estimate	Sd
	Constant	-8.615	0.048	-6.839	0.057
Spatial structure	$\sigma^2(u)$	0.775	0.028	0.540	0.021
DIC		349		525	
CAR convolution model					
	Constant	-8.629	0.067	-6.827	0.120
Unstructured	$\sigma^2(v)$	0.088	0.028	0.108	0.197
Spatial structure	$\sigma^2(u)$	0.001	0.000	0.280	0.520
Spatial fraction	κ	0.116	0.170	0.623	0.091
DIC		341		489	

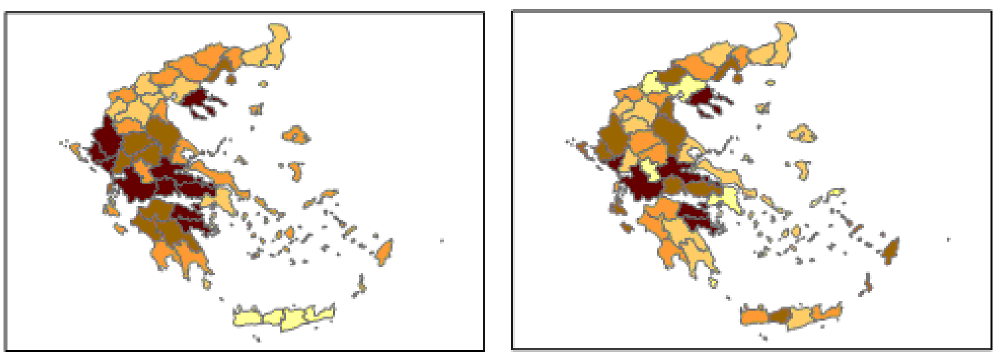
Table 2.3: Spatial models based on road connections (15,000-20,000 iterations)

The models are presented in Table 2.3. An initial remark concerns the impressive reduction of the DIC in both spatial models, especially in the accidents model, indicating that the "structured vs. unstructured" components, and the overdispersion, account for a very important part of the variation of both road safety outcomes.

Both the CAR and CAR convolution methods indicate a significant part of the variation in accident and fatality rates attributable to the spatial structure of the examined counties. It is noted that spatial effects become non significant in the CAR convolution model for accidents, most likely due to the fact that overdispersion is accounted for, which may be the principal part of random variation. However, the spatial fraction is significant, indicating that the spatial effect is important when compared to the remaining random variation. This is not the case for fatalities, in which the related counts are quite lower and therefore less subject to overdispersion.

The CAR method results may serve as a useful demonstration of the role of the spatial structure of road safety outcomes in Greece. Moreover, the CAR and CAR convolution models are considered equivalent in this case, given that the DIC reduction is similar. However, it is also indicated that the CAR convolution model is far more appropriate when overdispersion is present. These results allowed for the estimation of the spatial fraction, which was found equal to 0.62 for accident rates and 0.12 for fatality rates, and therefore only a small part of the variation of fatality rates is attributed to the spatial structure, whereas a non negligible part of the variation in accident rates is due to the spatial structure among Greek counties.

As an example, the structured and unstructured components of the variation in fatality rates in Greece are demonstrated in Figure 2.1. It can be seen that the structured component (left part of Figure 2.1) shows a relatively smooth evolution of fatality risk among neighbouring counties, in which an interesting pattern can be identified. More specifically, increased fatality risk is observed in central Greece, and mainly on counties including or surrounding large urban areas. On the other hand, the unstructured component (right part of Figure 2.1) shows no clear pattern, indicating that this part of the variation is random indeed.



<u>Figure 2.1:</u> Structured (left panel) and unstructured (right panel) heterogeneity of Bayes relative risk for fatalities per population (exp(U), exp(V))

Finally, the Bayes relative risk for fatalities and accidents on the basis of the CAR models is presented in Figure 2.2. Two remarks can be made on this Figure:

- First, the fatality rates (left part of Figure 2.2) are very similar to those represented in the unstructured component of fatality rates (right part of Figure 2.1), confirming thus that the unstructured component is predominant in the variation of fatality rates, as indicated by the respective low spatial fraction.
- On the other hand, the accident rates (right part of Figure 2.2) present a clear north / south spatial pattern. It is reminded that a high spatial fraction was estimated for accident rates, indicating a predominance of the structured variation over the unstructured variation. Consequently, in the overall risk rates, the spatial effect is quite evident.

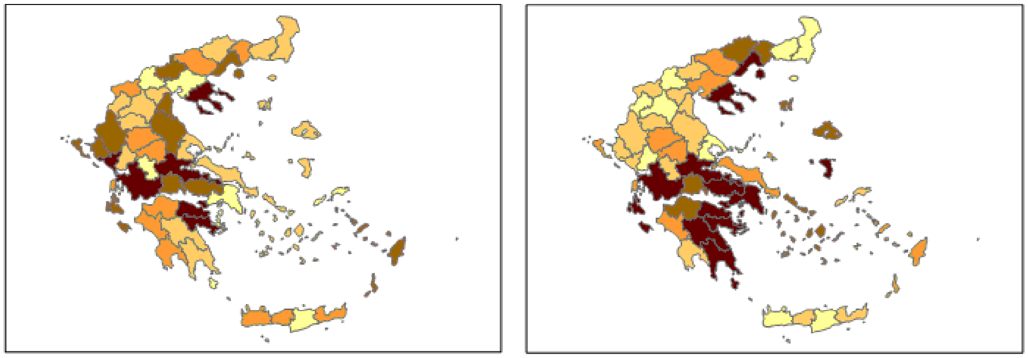


Figure 2.2: Bayes relative risk for fatalities (left panel) and accidents (right panel) per population

2.4.3 Models with spatial effects based on road and ferry connections

Given the above encouraging results as per the role of the neighbourhood structure of the Greek counties in their road safety outcomes, the next step is to examine whether this spatial structure goes beyond the presence of road connections to the existence of other type of connections e.g. ferry connections. The general idea is that neighbourhood is not only a matter of proximity but also a matter of connectivity.

Therefore, the neighbourhood matrix of Greece was extended to include road and ferry connections between counties, as show in Appendix 2.3. In this case, the minimum number of neighbours for one county is one (corresponding to island counties having a single ferry connection to a mainland county) and the maximum number of neighbours for one county is 14 (corresponding to the Athens area which has ferry connections with a high number of islands).

Due to the fact that the structured component and the spatial fraction for fatalities counts were very low in the road connections, in the present example only accidents are examined. The results are presented in Table 2.4.

		Accidents	
CAR model	Parameter	Estimate	Sd
	Constant	-6.817	0.018
Spatial structure	$\sigma^2(u)$	0.699	0.190
DIC		453	
CAR convolution model			
	Constant	-6.821	0.066
Unstructured	$\sigma^2(v)$	0.061	0.096
Spatial structure	$\sigma^2(u)$	0.142	0.190
Fraction	κ	0.540	0.184
DIC		467	

Table 2.4: Spatial models based on road and ferry connections (15,000-20,000 iterations)

The spatial effect based on road and ferry connections in the CAR model is significant, suggesting that the spatial structure in underlying risk factors represented by road and ferry connections accounts for important part of the variation of accident rates in Greece.

In the CAR convolution model, however, the random parameters (structured and unstructured) become non significant, possibly due to the fact that oversidpersion in the accident counts is accounted for, and is possibly the main source of variation. However, a significant spatial fraction was estimated equal to 0.54, indicating a non negligible contribution of this spatial structure to the variation of accident rates in Greece.

It is noted though that the fraction of this spatial structure (road and ferry connections) is slightly reduced in relation to the initial one presented in Table 2.3 (road connections). Moreover, the structured and unstructured components of these models did not present significant differences from those of Figure 2.1. On the other hand, the DIC of both models is also reduced in relation to those of Table 2.3. It appears that the extended neighbourhood structure explains a larger part of the variation, but not in favour of the structured component. These results do not allow concluding that considering a more complex spatial structure does really add to the explanation of the random variation of accidents counts.

It is likely, however, that this more complex spatial structure was not properly defined. More specifically, road and ferry connections were assigned equal weights in the neighbourhood structure, which is the default estimation process. A more realistic consideration would involve assigning higher weights to road connections in relation to ferry connections. However, such a consideration would require particular effort and caution and is therefore beyond the scope of the present demonstration.

2.4.4 Models with spatial effects and explanatory variables

For demonstration purposes, it will be shown in this section how the magnitude and significance of explanatory variables can be modified once spatial effects are considered in a counties road accidents model. This is based on the comparison of a single level Poisson model with explanatory variables, to the related CAR convolution model. The neighbourhood structure based on road connections between counties is used. The number of alcohol controls per population is used as an explanatory variable; according to previous research (Yannis et al. 2007) police enforcement is a main explanatory factor of road safety development in Greece from year 2001 onwards.

		Accidents	
Single level Poisson model	Parameter	Estimate	sd
	Constant	-7.014	0.044
Alcohol controls / population	Beta	-0.193	0.017
DIC		2767	
CAR convolution model			
	Constant	-7.175	0.177
Alcohol controls / population	Beta	-0.155	0.070
Spatial structure	$\sigma^2(u)$	0.213	0.489
Unstructured	$\sigma^2(v)$	0.091	0.188
Fraction	κ	0.646	0.092
DIC		480	

Table 2.5: Models with explanatory variables (15,000-20,000 iterations)

As shown in Table 2.5, the number of alcohol controls per population has a negative effect on accidents i.e. an increase of alcohol controls per population leads to a decrease in accident rates, which is intuitive. It can be seen that, once the spatial structure random parameter is introduced in the model, the effect of alcohol controls per population is quite reduced. This suggests that, if the effect of the spatial structure was not accounted for, the effect of this explanatory variable would be overestimated. In other words, the effect of the explanatory variable in the single level Poisson model also included part of the structured effect.

Although in the CAR convolution model the random components are non significant, possibly due to the magnitude of overdispersion in this data, a spatial fraction of around 0.65 is obtained, indicating a predominance of the structured component over the unstructured component. When comparing this spatial model with the CAR convolution model of Table 2.3, it can be seen that the introduction of the explanatory variable leads to a small increase of the spatial fraction, as well as a decrease in the DIC. Given that the contribution of the extended neighbourhood matrix of road and ferry connections could not be validated, overall, the CAR convolution model with a neighbourhood structure based on roads and the specific explanatory variable may be the best model for explaining the variation of accident rates in Greece on year 2002.

2.5 Conclusion

Road safety data are often available at more disaggregate level than country level. First of all, the CARE road accidents and fatalities data are available and may be obtained according to the European NUTS regional classification. Moreover, in most European countries, several other road safety related data may be available at NUTS level, as they are often collected by regional or local authorities (e.g. demographic and economic data, traffic data, police enforcement data etc.). Moreover, in recent research, there was particular interest in analysing road safety outcomes at a more disaggregate level than the global country level. It is argued that the overall road safety performance of a country may be better understood and improved by such a disaggregation, which may not only reveal the highest risk areas but also provide a comprehensive picture of the evolution of risk in space.

Spatial analysis allows for such questions to be investigated, by explicitly taking into account the spatial structure of road safety outcomes within the context of the various exogenous effects. The main idea behind spatial analysis approaches is that no reliable conclusions can be drawn by simply comparing the risk rates of e.g. NUTS-2 regions, unless the spatial structure among these regions is considered. To do this, the overall variation in road safety outcomes is decomposed into structured and unstructured variation. The structured component reveals the spatial dependence between the observations, whereas the unstructured component constitutes a random parameter.

Several applications of spatial analysis techniques in road safety have been published in the recent years, presenting some interesting findings, as well as examples of the consequences of ignoring the spatial correlation between observations (see references in section 2.1). In the present section, an additional example was provided, concerning road accidents and fatalities in Greek counties on year 2002, through which the main principles, techniques and tools for spatial analysis have been presented.

The application of spatial analysis techniques on the road accident and fatality risk in Greece provided some interesting results. First of all, spatial effects were found to be significant, even though Greece has a relatively loose spatial structure compared to other countries, due to the existence of many islands. This suggests that the spatial dependences need to be examined also in non typical spatial structures. Two types of spatial structure were examined, one based on road connections between counties and one based on road or ferry connections between counties. The results showed limited contribution of the extended neighbourhood structure in the variation of road safety outcomes, also suggesting that the basic spatial dependences in Greece are important and do not need to be strengthened by an extended consideration.

In the present analysis, explanatory effects were not a priority, as particular focus was given on the identification of spatial effects among the random effects in road safety outcomes. However, a basic demonstration of spatial modelling results with explanatory variables revealed that, if the spatial structure was not accounted for, the explanatory effects would be quite overestimated.

For these reasons, it is important to explicitly model the spatial dependence among observations, when working with regional road safety data. Once the spatial effects are taken into account, the (often more interesting) explanatory effects can be estimated more accurately.

Finally, it is noted that extensions of the models discussed in this section may concern the consideration of unequal weights in the definition of the neighbourhood structure; in particular, the road connections considered could be weighted according to the amount of traffic accommodated by these connections, if detailed origin / destination data for road transport demand were available at national level. Another useful extension of the model concerns the incorporation of time effects, resulting in a model capturing both space and time dependences of the road accident data.

Chapter 3 – Modelling injury underreporting in seven European countries

George Yannis, Eleonora Papadimitriou, Antonis Chaziris, NTUA

3.1 Introduction

The under-reporting in national accident databases and the differences between countries with respect to the definitions used to classify injury severity are important limitations in European comparisons concerning non-fatal road accidents and related injuries. Currently, the only comparable road safety units available in the CARE database are the numbers of fatal accidents and of persons killed, which form only a small minority of the total road safety figures. In this case, there are common classifications (definitions); the low level of under-reporting allows for meaningful comparisons to be carried out in most European countries and appropriate correction factors exist.

The same is not true, however, of non-fatal accidents and of persons with non-fatal injuries. In particular, the definition of injury severity differs among countries, so that a casualty which would be recorded in one country might not be recorded in another. Moreover, a casualty which might be recorded as 'serious' in one country might be recorded as 'slight' in another.

In order to overcome the inconsistencies in the reporting of non-fatal casualties and to extend European comparisons including the full range of injury severities, Task 1.5 of SafetyNet has estimated the under-reporting level for non-fatal casualties in several EU countries. For this purpose, a uniform methodology was applied, based on linking and matching Police and Hospital injury records. Moreover, Task 1.5 has established a new common definition for injury severity, based on an internationally accepted measurement unit, namely the MAIS. This measurement unit summarises injury severity and probability of survival that takes account of multiple injuries. The comparison between police and hospital injuries resulted in a conversion factor quantifying the extent of underreporting. This conversion factor allowed for the estimation of the number of serious casualties per country according to the MAIS.

The linking and matching methodology was tested by means of national studies in 8 European countries (7 of which were finally considered in the present analysis). The coverage of the studies, the size of the datasets, the period included and the details of the medical data varied (in some cases widely) among countries, not allowing to consider all the results as being nationally representative. Relevant national studies would be required in all European countries in order to generalize the results. However, the work of Task 1.5

allowed for the estimation of appropriate conversion factors (coefficients) for the true number of injuries, disaggregated by injury severity and road user type, those being an important and useful contribution towards more reliable comparisons of road accident injuries.

On the basis of these results, the present analysis aims to assess the extent and the variation of the underreporting problem in different countries. It is envisaged through this analysis additional insight will be provided concerning the causes of under-reporting as well as the better understanding of its constituent components. In particular, the analysis has a twofold objective: first, to compare the conversion factors of injury underreporting among different countries; and second, to analyze the variation of the estimated conversion factors of injury underreporting for different types of road users and different types of injuries, both within countries and among countries. On that purpose, a log-rate analysis technique is used, allowing examining both the magnitude and the statistical significance of single and combined effects of different factors on the conversion factors.

The analysis has the following structure: first, the methodology and results of Task 1.5 of SafetyNet are presented, providing the complete background of the analysis. Then, the data and modelling techniques are presented. Finally, following the modelling results, statistically significant effects on underreporting are compared in detail. Particular emphasis is given on possible limitations in the interpretation of the modelling results, due to the characteristics and particularities of the data.

3.2 Estimation of correction factors for injury underreporting (overview of SafetyNet Task 1.5)

3.2.1 Overview of the methodology of SafetyNet Task 1.5

In order to identify a common international standard to provide a benchmark with which to compare accident data from each country, a method was applied, which has been widely used in several countries to study the level of under-reporting. This method consists of comparing those road accident victims who have been recorded by the police in the national accident database, with those who have been recorded in medical records maintained by hospitals. Therefore, the basic approach adopted within Task 1.5 of SafetyNet consists of (Broughton et al, 2007):

- linking road accident and medical databases, to investigate the level of under-reporting and, equally importantly, to copy details from the medical record of each linked casualty to the corresponding record in the accident database
- comparing the distributions of linked casualties by a standard injury score (namely the MAIS), usually available in the hospital databases
- defining a new injury classification based on the most appropriate medical variable(s) and calculating national coefficients to estimate 'true' casualty totals from the numbers recorded by the police

At this point it is important to define the international MAIS score for injury severity. This is based on the Abbreviated Injury Scale (AIS), a specialized trauma classification of injuries based mainly on anatomical descriptors of the tissue damage caused by the injury. The AIS classification system was designed to distinguish between types of trauma of clinical importance as well as types of trauma of interest to vehicle designers and research engineers. It has been shown to provide a good basis for valid measurement of probability of death. The AIS has two components: (1) the injury descriptor which is a unique numerical identifier for each injury description; and (2) the severity score which ranges from 1 (relatively minor) to 6 (currently untreatable), and is assigned to each injury descriptor. The severity scores are consensus assessments assigned by a group of experts and implicitly based on four criteria: threat to life, permanent impairment, treatment period, and energy dissipation. The MAIS is the maximum AIS, out of all injury diagnoses[1].

It is noted that linkage of road accident and medical databases on the basis of MAIS is a quite demanding task, involving subjective decisions, to some extent, when specifying the differences that may be tolerated when deciding whether a pair of records actually refer to the same casualty. Moreover, road accident casualties with only slight injuries may not require significant medical treatment and hence would not be recorded in any medical database; the common methodology does not cover such cases. Several techniques, including probabilistic linking have been developed for record matching from different databases (SWOV 2001). In these methods, a generalised distance function is defined which quantifies the similarity between pairs of records in the two databases. This quantified similarity can be used to assess the probability of the correctness of a match.

More specifically, medical records were cross-checked with the police accident records. The checking took account of the catchment[2] area of each hospital, comparing the hospital records with police accident records only for that area, in order to identify all cases where the same person is present in both sets of records. Given that personal names were seldom available in both data sets, this process needed to be based on factors common to both data sets such as the casualty's age (or date of birth), gender and mode of travel, together with accident circumstances such as date, time and location.

The combined police and medical data set was used to compile a 3-dimensional matrix of casualty counts (Table 3.1), based on the severity of their injuries as

[1] The MAIS can be estimated directly by trained staff, but alternatively it can be derived from more detailed codes. The International Classification of Diseases (ICD) is a system designed to promote international comparability in the collection, processing, classification, and presentation of mortality statistics. The ICD is developed collaboratively between the World Health Organization (WHO) and 10 international centres, to ensure that medical terms reported on death certificates are internationally comparable, and it has been revised approximately every 10 years since 1900. The recent ICD-10 more closely reflects current medical knowledge than the previous ICD-9

[2] The area around the hospital from which accident victims are normally brought to the hospital for initial treatment

summarised by the MAIS score. Road user type was identified from police data, as it was often poorly recorded in medical records.

road user type	X	MAIS[3]	X	police severity score
car occupant **pedestrian** **pedal cyclist** **motorcyclist** **other**		**1-6** **not coded *(not matched in medical records)***		**fatal** **serious** **slight** **not coded *(not matched in police records)***

Table 3.1: Matrix of casualties counts for linking police and hospital data (Broughton et al. 2007)

On the basis of the common methodology described above, national studies were carried out in 8 European countries. The features of each study are summarized in Table 3.2.

Country	Study area	Period	Sample size	Coding of MAIS
Austria	National	2001	69233	From ICD10
Czech Republic	Kromeriz, central Moravia	2003-2005	1649	From ICD10
France	Département of the Rhône	1996-2003	90457	Coded directly
Greece	Island of Corfu	1996-2003	11915	From ICD9
Hungary	Part of Budapest	Aug 2004 - Jan 2006	3459	From ICD10
Netherlands	National	1997-2003	129616	From ICD9
Spain	Castilla y Leon	2nd semester of 2005	8113	From ICD9
UK	Scotland	1997-2005	201006	From ICD10

Table 3.2: Characteristics of the national studies (Broughton et al. 2007)

3.2.2 An example: the UK national study of SafetyNet Task 1.5

In this section, an analytical example is presented, based on the UK national study (Broughton et al, 2007), so that the process for estimating the conversion factors for the number of casualties can be demonstrated. Table 3.3 summarizes the results of the linkage of police and hospital data on the basis of MAIS.

[3] It is noted that a similar matrix was developed using the casualty's length of stay at the hospital, instead of the MAIS score. Related conversion factors were developed for this case as well; however, given that the proposed common injury classification is based on the MAIS score, this case is not examined in the present analysis.

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

	MAIS	Police records serious	Police records slight	Hospital, not police	% not reported by police
Hospital	1	3823	3642	6294	46%
	2	8336	2473	8050	43%
	3	3139	412	2227	39%
	4	226	33	220	46%
	5	75	1	44	37%
	6	197	19	134	38%
	9 (unknown)	1638	2184	3703	49%
Total	1-9	17434	8764	20672	
Police, not hospital		12831	138334		

Table 3.3 Cross-classification of police and hospital records per severity score (UK national study, Broughton et al. 2007)

The top part of Table 3.3 shows the classification of the number of road accident casualties recorded by hospitals, for different MAIS scores. In particular, each hospital casualty of a given MAIS score is classified according to whether it was recorded by the police as 'serious' or 'slight', or whether it was not recorded by the police. Consequently, the highlighted part of Table 3.3 corresponds to the matched hospital and police casualties. It can be seen that, for all MAIS scores, a percentage of casualties recorded by hospitals ranging from 37% to 49% was not recorded by the police. It is worth mentioning that 19 out of the 216 matched casualties (nearly 10%) with MAIS=6 were recorded by the police as slight injuries.

The bottom part of Table 3.3 shows the classification of accident casualties recorded by the police that were not available in the hospitals. Interestingly, a number of both serious and slight police recorded casualties were not available in the hospital data. It is noted however that, although these casualties were recorded by the police within the catchment area of the examined hospitals, it is possible that some of them have been treated in other hospitals, beyond the study area. Moreover, these non-reported casualties might be attributed to hospital under-reporting as well as to problems in the data linking. Especially as regards slight injuries, in specific countries it is also possible that these casualties did not receive any hospital treatment.

Therefore, for an efficient estimation of the appropriate conversion factors, a number of issues needed to be addressed. The first issue concerns the treatment of the 'unknown' MAIS 9 scores in the hospital data. These casualties appear on the whole to have relatively minor injuries, for example the incidence of MAIS 9 is lower among serious casualties than among slight and the percentage not reported by police is greater than for MAIS 1. To exclude these cases would introduce one type of bias, to treat them all as MAIS 1 would introduce another type as they may well include cases with an actual MAIS>2. On the whole, it appeared preferable to treat the MAIS 9 cases as MAIS 1.

A second issue concerned the fact that the MAIS of police recorded casualties, which were not available in the hospital data, needed to be estimated. As mentioned above, these casualties have not required in-patient treatment, so it seems unlikely that MAIS will have exceeded 3. On the other hand, some MAIS 2 casualties could well be treated as outpatients, or in local doctors' surgeries. It is reasonable to assume that all of these casualties had MAIS 1 or 2, but that they cannot be distributed reliably between 1 and 2. These casualties were therefore distributed pro rata at each MAIS level to simulate the police severity coding. The classification resulting from these treatments is presented in Table 3.4.

	reported by police		**Hospital, not police**		**police, not hospital (estimated distribution)**		**estimated total**	
MAIS	serious	slight	serious	slight	Serious	slight	serious	slight
1 or 2	13797	8299	11019	7028	12831	138334	37647	153661[1]
3	3139	412	1969	258	0	0	5108	670
4	226	33	192	28	0	0	418	61
5	75	1	43	1	0	0	118	2
6	197	19	122	12	0	0	319	31
all	17434	8764	13345	7327			43610	1544251

[1] *likely to be underestimated*

Table 3.4 Adapted cross-classification of police and hospital records per severity score (UK national study, Broughton et al. 2007)

Consequently, it can be calculated that for each serious casualty in the police records (including the records not matched with hospitals) there is actually 43610/(17434+12831)=1.44 serious casualties overall.

This methodology also allows for the calculation of underreporting with reference to specific actual MAIS scores. For instance, it can also be calculated that for each serious casualty in the police records, there is actually (5108+418+118+319) / (17434+12831)=0.20 casualties with MAIS higher than 2. Furthermore, a small proportion of casualties recorded as 'slight' by the police, actually had an MAIS score higher than 2 in the hospital records; in particular, there are (670+61+2+31)/(8764+138334)=0.005 casualties with MAIS higher than 2 per slight casualty recorded by the police. Consequently, if actual serious casualties were to be defined as those with MAIS higher than 2, then the actual total number of serious casualties in the examined area could be estimated as:

$N_{serious}$ = 0.20 × number of serious casualties reported by the police + 0.005 × number of slight casualties reported by the police

The first part of the above equation refers to the actual police under-reporting (casualties not reported by the police), while the second part refers to casualties that were mis-reported by the police (serious casualties that were reported as slight). The opposite would apply if the real number of slight casualties were to be calculated:

N_{slight} = a × number of serious casualties reported by the police + b × number of slight casualties reported by the police

where: (a) would represent mis-reporting of slight casualties as serious
and (b) would represent the real under-reporting of slight casualties.

By applying the same process for all MAIS categories, the appropriate conversion factors in each case can be estimated.

3.2.3 Summary of results from all national studies of SafetyNet Task 1.5

On the basis of the uniform methodology described above, national studies were carried out in 8 European countries, and conversion factors for the number of serious and slight casualties recorded by the police were developed, disaggregated by road user type and MAIS score. Table 3.5 summarizes these results for two MAIS categories: lower than three and higher or equal to three (the detailed conversion factors per MAIS score are presented in Appendix 3.1). It is noted that usable results were eventually obtained for 7 countries only, whereas the maximum level of disaggregation was achieved in 5 countries.

All studies used accident data from national accident databases that had been compiled from police accident reports. Most studies used files of medical data compiled by national or regional authorities from hospital records. The various linkage approaches were applied rigorously and by using the same variables to identify potential matches, so there is no reason to suppose that the results would have differed significantly if a common, optimised technique had been applied (Broughton et al, 2007).

As indicated by Table 3.2, the extent of the 8 studies varies widely in time and space. Similarly, the size of the linked datasets varies widely, from 1600 records from the Czech study to 201000 records from Scotland. It is inevitable that the strength of the results achieved by the various studies differs, if only on statistical grounds. Additionally, different injury codings were used in some countries for calculating the MAIS scores, and a number of transformations were carried out to enhance compatibility. Overall, however, the results achieved represent an important step forward in comparing the numbers of road accidents and casualties across a range of countries.

Czech Republic												
	car occupant		motorcyclist		pedal cyclist		pedestrian		Other		**all**	
MAIS	serious	slight	serious	slight	serious	slight	serious	slight	Serious	slight	serious	slight
1- 2	0.97	1.11	1.03	1.17	1.11	3.50	1.05	1.77	0.88	1.00	1.07	1.56
>2	0.11	0.01	0.09	0.01	0.50	0.04	0.35	0.04	0.13	0.00	0.21	0.02
All	1.08	1.12	1.12	1.18	1.61	3.54	1.40	1.80	1.00	1.00	1.28	1.58
France												
	car occupant		motorcyclist		pedal cyclist		pedestrian		Other		**all**	
MAIS	serious	slight	serious	slight	serious	slight	serious	slight	Serious	slight	serious	slight
1- 2	1.32	2.38	1.35	3.13	4.69	10.39	1.01	1.90	1.52	2.67	1.43	2.69
>2	0.51	0.029	0.83	0.123	1.97	0.275	0.57	0.097	1.06	0.058	0.68	0.061
All	1.84	2.41	2.18	3.25	6.67	10.66	1.58	2.00	2.58	2.73	2.11	2.75
Greece												
	car occupant		motorcyclist		pedal cyclist		pedestrian		other/unknown		**all**	
MAIS	serious	slight	serious	slight	serious	slight	serious	slight	Serious	slight	serious	slight
1- 2	4.08	6.09	6.89	10.72	7.50	23.75	2.49	3.91	11.45	15.41	5.92	9.10
>2	0.57	0.17	0.68	0.19	0.17	1.17	0.45	0.14	0.69	0.09	0.60	0.17
All	4.65	6.25	7.57	10.91	7.67	24.92	2.93	4.04	12.14	15.50	6.52	9.28
Hungary												
	vehicle occupant						pedestrian				**all**	
MAIS	serious	slight					serious	slight			serious	slight
1- 2	0.83	1.28					0.86	1.16			0.84	1.27
>2	0.52	0.04					0.35	0.03			0.48	0.04
All	1.35	1.33					1.21	1.19			1.32	1.31
Netherlands												
	car occupant		motorcyclist		pedal cyclist		pedestrian		Other		**All**	
MAIS	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight
1- 2	1.07	1.02	1.21	1.04	1.90	1.10	1.23	1.04	1.24	1.01	1.29	1.04
>2	0.22	0.006	0.37	0.016	0.73	0.041	0.36	0.023	0.21	0.007	0.39	0.016
All	1.29	1.02	1.59	1.05	2.63	1.14	1.59	1.06	1.45	1.02	1.68	1.05
Spain												
											All	
MAIS											Serious	Slight
1- 2											1.22	1.06
>2											0.26	0.018
All											1.48	1.07
UK												
	car occupant		motorcyclist		pedal cyclist		pedestrian		other		**All**	
MAIS	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight
1- 2	1.15	1.03	1.34	1.13	2.54	1.24	1.05	1.03	1.62	1.06	1.24	1.04
>2	0.15	0.004	0.27	0.012	0.29	0.009	0.23	0.008	0.26	0.006	0.20	0.005
All	1.30	1.03	1.61	1.14	2.83	1.25	1.28	1.04	1.88	1.07	1.44	1.05

Table 3.5 Conversion factors per country, road user type, MAIS score and police severity score (Broughton et al. 2007)

Most importantly, the work of Task 1.5 allowed for a new common definition of serious casualties to be proposed: it was concluded that the optimal definition of serious casualty for use with the CARE database should be a non-fatal casualty with MAIS between 3 and 5 (inclusive). According to this definition, the actual number of serious casualties in the examined countries was estimated for the period 2003-2005 (Table 3.6). It is interesting to note that for all countries except Greece, the actual number of serious casualties according to the new proposed definition is lower than the number of serious casualties in CARE (those taken as casualties recorded as 'serious' by the police). The fact may be attributed to a higher degree of police misreporting slight casualties as serious, more than in the other examined countries.

	Serious casualties			Slight casualties			Serious*	Serious*
	CARE	factor 1	N1	CARE	factor 2	N2	(N1+N2)	Serious
France	19,898	0.68	13612	100,587	0.061	6157	19768	0.99
Greece	2,338	0.46	1081	18,650	0.121	2259	3339	1.43
Hungary	8,320	0.83	6897	19,185	0.069	1325	8223	0.99
Netherlands	10,881	0.39	4254	29,608	0.016	474	4728	0.43
Spain	23,323	0.26	6084	117,286	0.018	2059	8143	0.35
UK	32,445	0.34	11130	254,253	0.009	2298	13428	0.41

Serious* refers to the estimated actual serious casualties

Table 3.6 Conversion of the CARE casualties according to the common proposed definition of injury severity (Broughton et al. 2007)

3.3 Investigation of variations on injury under-reporting

3.3.1 Methodology and data

In this section, systematic differences in road accident injury underreporting between different countries, road user categories and police records are investigated. In particular, differences in underreporting are examined both within countries and between countries. Moreover, additional parameters affecting the level of underreporting are analyzed both individually and jointly. In this analysis, the dependent variable is the rate of actual number of casualties of a given level of MAIS score (N*) per the related number of casualties recorded by the police (N).

The available main factors influencing the magnitude of the actual number of casualties are:

- Characteristics of the countries
- Characteristics of the road user, e.g. it is likely that bicyclists' injuries are more underreported compared to other road users
- The way that the police records injury severity, e.g. it is likely that injuries that would be defined by the police as 'slight' are more underreported
- The MAIS scores, e.g. it is likely that injuries corresponding to higher MAIS are less underreported

- Combinations of the above parameters, e.g. it is likely that the injuries of certain road user types are more underreported in certain countries.

The variables and values examined are presented in Table 3.7.

Dependent variable	**Type**	**Values**
Number of actual casualties	Numerical	The actual number of casualties × 100*
Explanatory variables	**Type**	**Values**
Number of casualties recorded by the Police	Numerical	The number of casualties recorded by the Police × 100*
Country	Categorical	1: Czech Republic 2: France 3: Netherlands 4: UK 5: Greece 6: Hungary 7: Spain
Road User Type	Categorical	1: car occupant 2: motorcyclist 3: Pedal cyclist 4: Pedestrian 5: Other
Police score	Categorical	1: serious injury 0: slight injury
MAIS score	Categorical	1: MAIS>2 0: MAIS 1 or 2

* Multiplied by 100 because in some cases the ratio N*/N was lower than 1

Table 3.7 Variables and values used in the analysis

The analysis aims to associate underreporting conversion factors, defined as the rate of the number of actual casualties (N*) to the number of casualties recorded by the Police (N), with countries, road user types, police severity scores and MAIS scores. In order to determine the significance of all possible interactions among these factors, the analysis of a four-dimensional table through a log-rate modelling approach was attempted. This method allows the investigation of both single and combined variable effects (interactions). The models assumptions and formulation are presented in Appendix 3.2.

The objective of this modelling approach is to investigate whether the variation of injury underreporting can be explained by country, road user type, police severity scores or MAIS scores, and most importantly, by combinations of these parameters. In the framework of the present research, the third-order interaction (i.e. interaction of four variables) is the most interesting since, if this effect is significant, then there is a significant interaction of all the examined parameters with respect to injury underreporting. If (and only if) not, then the various lower-order effects can be further analysed and interpreted (Goodman, 1973).

For example, if the interaction of country and road user type is found to be significant, this would mean that injury underreporting is not only significantly different between countries, but that the way countries differ varies between user groups.

From the best-fitting log-rate model, the parameter estimates and their statistical significance are determined. The ultimate test is whether the table generated by the model closely fits the observed table. A likelihood ratio goodness-of-fit statistic is used to accept or reject the model (Everitt, 1977).

3.3.2 Modelling results

The first stage of the analysis concerned the identification of the best generating class for the model, i.e. the highest order significant interaction between variables. The third-order interaction was significant, suggesting that there is a significant joint association of all the examined variables (country, road user type, MAIS score and police severity) with respect to underreporting (see Appendix 3.3). This means that there are differences in underreporting at the most disaggregate level of the parameters examined, e.g. underreporting of pedal cyclists' injuries recorded by the Police as slight and having an MAIS score 1 or 2 in the Netherlands is significantly different from the underreporting of "other" road users injuries recorded as serious by the Police and having an MAIS score higher than 2 in Greece. Obviously, comparisons at such a disaggregate level are not always meaningful, however the important information obtained by this results is that even the most detailed combinations of casualty type may present significant differences between them.

From a modelling viewpoint, this suggests that the generating class of the log-rate analysis should be [country*road user type*police score*MAIS score], followed by all lower-order combinations (i.e. country*road user type*MAIS, road user type*MAIS*police severity, country*MAIS, road user type*police severity, etc) of these variables. According to the above, a saturated design was generated for the log-rate model i.e. a design including all possible single and combined effects (interactions). The parameter estimates of the model are presented in Appendix 3.3. It is noted that a dummy coding is used, in which all parameter estimates for each variable are estimated in relation to a value of reference, taken as the last category of each variable. Consequently, all combinations including the last category of one of the variables were set to zero.

It is noted that road user type information was not available for Hungary and Spain. For these two countries, road user type values were replaced by the respective mean values for the different road user types across countries. Consequently, all the effects concerning road user groups for these two countries were estimated as equal to zero.

The modelling results show that there are statistically significant combinations of variables in almost all effects, therefore there are significant differences in injury underreporting in almost all combinations of casualty characteristics (see Appendix 3.3). In particular, all single variable effects except "road user type" are significant; this suggests that, overall, underreporting presents significant differences between countries, police severity scores and MAIS scores, but not between road user types.

Moreover, most second-order effects are significant; interestingly, the interactions of "road user type" and severity scores are not significant, although the interactions of "road user type" and "country" are significant. This indicates that underreporting is not systematically different for different severity scores of different road user types overall, although there are systematic differences between different road user types across different countries (i.e. underreporting of pedal cyclists' injuries in France is different than in Greece, but underreporting of slight pedal cyclists' injuries recorded as serious by the police is not different overall from the underreporting of injuries recorded as slight by the police).

When examining higher-order interactions, i.e. those corresponding to combinations of three or four variables, effects presenting significant differences are still identified.

The increased number of parameter estimates and the extensive use of default dummy coding does not allow for a straightforward interpretation of parameter estimates (in direct comparisons). However, the parameter estimates of the log-rate model are in fact log-odds ratios. Therefore, the odds ratios can be easily calculated (see last column of Appendix 3.3). In the remaining of this section, the main modelling results are graphically presented and discussed. It is noted that relative underreporting rates are presented, as all effects are normalized according to appropriate reference groups in each case.

In particular, as regards first-order effects, the following results were found (see Figure 3.1):

- Underreporting, defined as the number of actual casualties of a given severity corresponding to each casualty recorded by the police, is around ten times lower (odds ratio 1:0.077) for casualties recorded by the police as 'serious' compared to those recorded by the police as 'slight', which is intuitive.
- Underreporting of casualties with MAIS higher than 2 is around 4.5 times lower compared to those with MAIS 1 or 2, which is also intuitive.

It is interesting to note that the examination of police data alone does not allow capturing the whole range of underreporting. It is reminded that, for the relatively small number of casualties recorded by the police but not available in hospitals, an MAIS 1 or 2 had been assigned, in favour of smaller underreporting of slight casualties.

Moreover:

- France and Greece present the highest underreporting rates compared to the other countries examined, namely four and three times higher compared to Spain (see Figure 3.2).
- Czech Republic and the Netherlands present the lower underreporting rates in the examined countries.

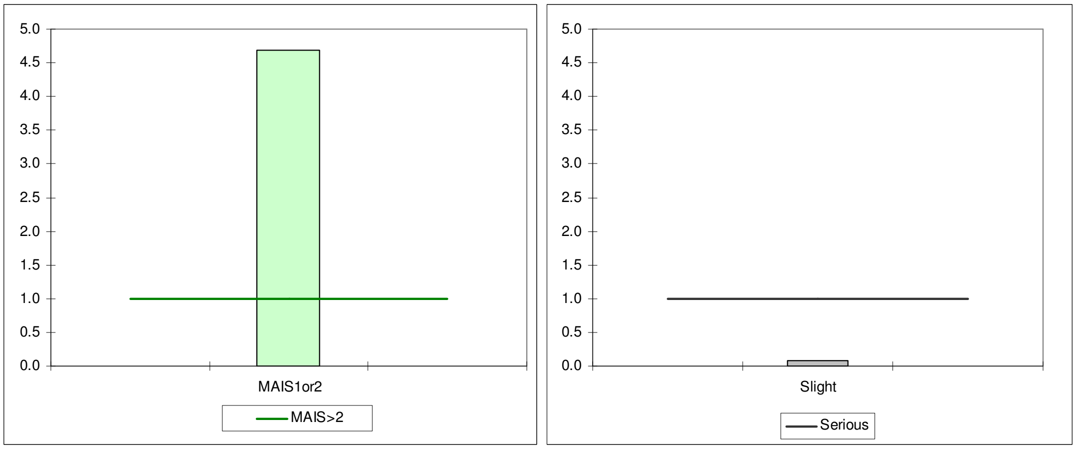


Figure 3.1 Relative underreporting rates per severity score

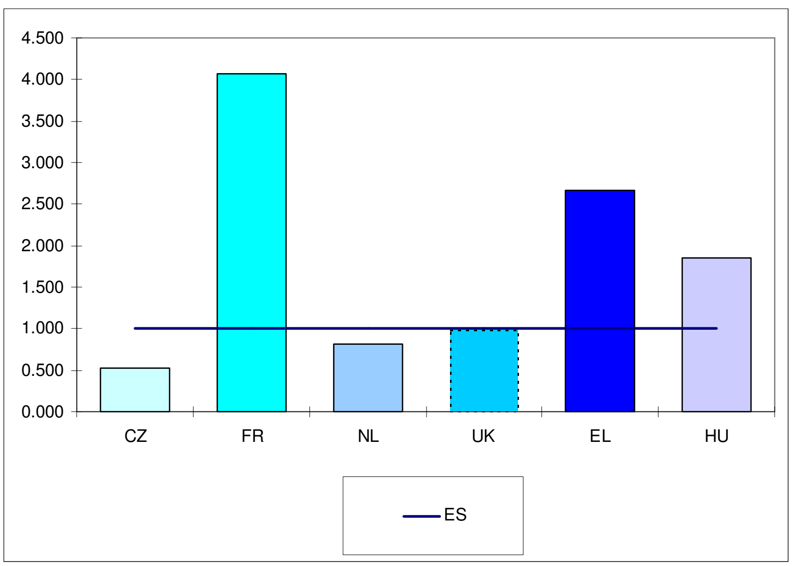


Figure 3.2 Relative underreporting rates per country

The examination of second- and third-order effects reveals additional results. It is noted that in the following figures, a bar without border indicates a non significant effect.

Figure 3.3 shows the relative underreporting rates between countries in relation to the two different severity scores, the Police severity score and the MAIS score. Comparing to the overall country results, it can be seen that the higher underreporting in Greece is mainly attributed to casualties of MAIS 1 or 2 and to casualties the police records as slight. On the contrary, the higher underreporting in France is mainly attributed to casualties of MAIS > 2 and to casualties the police records as serious. A mixed image can be identified for the Netherlands and the UK, where slightly higher underreporting of casualties with

MAIS 1 or 2 is observed. However, the significantly lower underreporting of casualties recorded as slight by the Police results in overall lower underreporting rates compared to other countries (in this case, MAIS 1 or 2 and police slight casualties do not follow the same trends in underreporting). Some related effects for the Czech Republic and Hungary were not statistically significant and are not discussed.

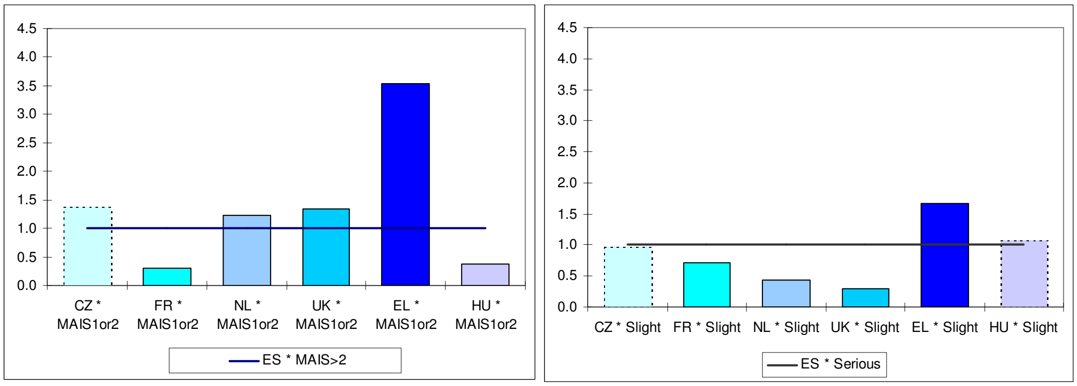


Figure 3.3 Relative underreporting rates per country and severity score (MAIS / Police)

Figure 3.4 presents the results of the joint examination of road user type and country. The reference groups are 'Spain' and 'other' road users. It is noted that Hungary and Spain did not have detailed information per road user type so the respective effects were zero. Underreporting of pedal cyclists' casualties is much higher compared to car occupants' in several countries (i.e. 3.5 times higher in the Czech Republic and the Netherlands, more than double in France, 1.5 times in the UK). Underreporting of pedestrians'casualties is higher in the Czech Republic and the Netherlands, whereas underreporting of motorcyclists' casualties is higher in the Netherlands and the UK.

Overall, it can be considered that the higher overall underreporting rates of France are due to pedalcyclists' underreporting, whereas the respective rates of Greece are probably due to the "other" road users underreporting (reference group). Moreover, the lower overall underreporting rates in the Netherlands are probably also mainly due to the "other" road users group, given that the remaining types of road users present significantly higher underreporting rates compared to the other countries. Underreporting is consistently low in the UK for all road user groups, resulting in the overall low underreporting shown in Figure 2.

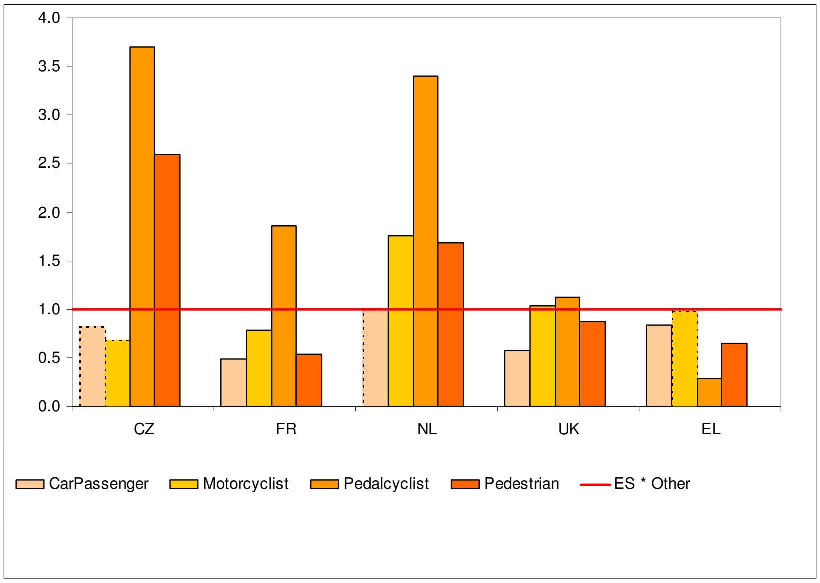


Figure 3.4: Relative underreporting rates per road user type and country

In Figure 3.5 the results are presented in further detail, concerning the joint effect of country, road user type and MAIS score, which are standardized in relation to Spain, "other road users" and MAIS>2. Almost all effects are statistically significant, apart from the ones on car occupants and motorcyclists in the Czech Republic. Interestingly, in France underreporting of casualties of MAIS 1 or 2 is more important than of MAIS > 2 for all road user types; the opposite is the case for the Netherlands, whereas in the other countries the underreporting per road user type varies for different MAIS scores.

In particular, pedal cyclists' casualties of MAIS 1 or 2 are 2.5 times more underreported in Greece in relation to pedal cyclists' casualties of MAIS higher than 2. Respective figures for France and UK are almost 2. Moreover, motorcyclists' casualties of MAIS 1 or 2 are by 10% less underreported compared to MAIS > 2 in the UK, and by 10% more underreported in France. Finally, pedestrian casualties of MAIS 1 or 2 are by around 50% less underreported in the Czech Republic and in the Netherlands, and by 25% more underreported in France, in relation to casualties of MAIS >2.

Concerning the 4th order interaction between countries, road user groups, police severity scores and MAIS scores, although most of the effects are statistically significant, it is difficult to achieve overall comparisons. However, it is important to mention that this interaction adds explanatory effect and should not be removed.

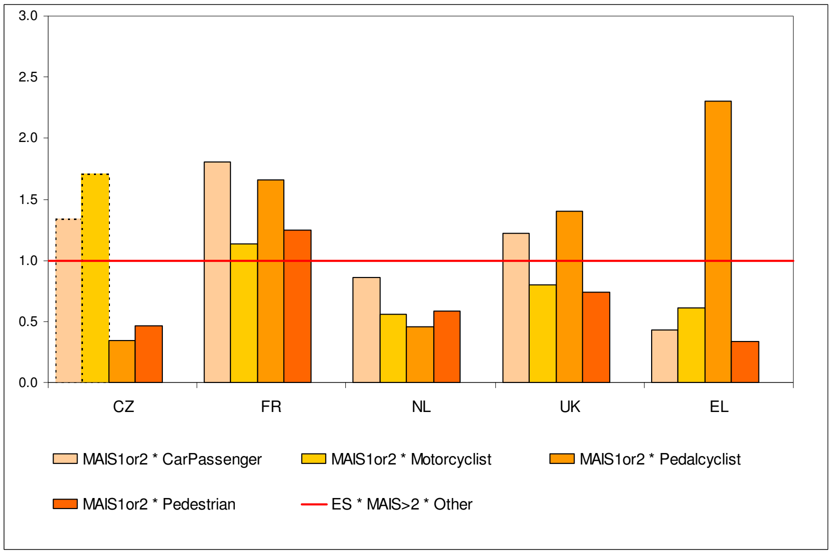


Figure 3.5: Relative underreporting rates per country, road user type and MAIS

Summarizing, the log-rate modeling allowed for the identification of significant single and combined effects on casualty underreporting, defined as the true number of casualties of a given MAIS score per each casualty recorded by the police. Important variation across countries, road user groups and severity scores was identified. Moreover, it was found that the examination of different interactions of these three main effects lead to significant and different effects on the degree of underreporting. It is important to note, however, that more detailed effects can not be always directly interpreted as disaggregations of overall effects.

3.4 Conclusion

In this section, the work carried out within Task 1.5 of SafetyNet, concerning the determination of the degree of underreporting in several European countries, was exploited in a statistical analysis of the findings. The contribution of Task 1.5 is considered to be most important, for several reasons; first, the degree of underreporting was systematically estimated in a number of European countries, through the joint examination of police and hospital data, and appropriate conversion factors were proposed. Moreover, a common definition of a 'serious' road accident casualty was proposed, according to a standard international definition of the severity of injury.

From the present analysis, two types of inappropriate reporting have been identified: injury underreporting, concerning injuries that were not recorded by the Police and could only be found in hospital records, and mis-reporting, concerning injuries that were recorded by the Police but with a wrong severity score (e.g. "slight" instead of "serious"). It is noted that a separate analysis of the second case of inappropriate injury severity reporting can be found in Deliverable 7.9 of SafetyNet, where the factors affecting the degree of injury severity mis-reporting by the Police have been identified on the basis of in-depth data.

The complexity of the methodologies and definitions related to injury severities and underreporting conversion factors does not always allow for a straightforward interpretation of results. However, the development of a log-rate model provided useful and analytical results concerning the variation of under-reporting in several European countries by several variables. More specifically, it was found that country, road user type and injury type characteristics have important effects on underreporting. Moreover, different combinations of these characteristics results in significantly different degrees of underreporting.

For example, the results suggest that injury underreporting is higher in France, in Greece and in Hungary compared to the underreporting in the other examined countries. However, the more disaggregate results reveal additional effects, which may assist in the understanding and improvement of the increased underreporting in these countries. For example, in France there is higher underreporting of serious injuries, whereas in Greece there is higher underreporting of slight injuries, compared to the other countries, and therefore national authorities should focus on these particular groups respectively.

Moreover, although the Netherlands do not present higher underreporting rates compared to the other countries overall, a particular problem concerning motorcyclists and pedal cyclists was identified in this country; these road users' injuries are much more underreported than other road users' in the Netherlands.

On the basis of this model, if the number of casualties recorded by the police is known, for a given police severity, MAIS score and road user type in a given country, the actual number of casualties of this category can be estimated. Moreover, the characteristics affecting the most the degree of underreporting of this particular injury category can be identified. Certainly, some of these effects may be due to specific features of the related national studies, whose catchment area and level of detail varied significantly in some cases. It is not possible, therefore, to consider these results as (fully) representative of national patterns. This should be kept in mind especially when examining overall trends; the detailed analysis of combined effects may make these issues more identifiable.

The modelling results revealed a number of significant effects (8 simple, 26 first order, 34 second order and 11 third order interactions) on underreporting and also allowed for the quantification of these effects through the calculation of relative ratios. This methodology was very efficient in the particular context of analysis and is considered to be promising for similar analyses on more detailed and extensive results (more countries, more harmonized national studies, more disaggregate conversion factors) once these are available.

Chapter 4 – CARE as a reference for other databases

Christian Brandstaetter (KfV)

4.1 Introduction

In Work Package (WP) 5 of the SafetyNet Integrated Project detailed data of fatal accidents were collected to develop *Fatal Accident Investigation* (FAI) database. The sampling methods are described in SafetyNet Deliverable 5.2: In-depth Accident Causation Data Study Methodology Development Report. This chapter compares these data with key data from the CARE Database. The results are discussed with respect to representativeness of the FAI database.

4.2 Data

All data available from the in-depth accident causation database is used for the following analysis. Table 4.1 gives an overview about the number of persons involved in the collection of the Fatal Accident Investigation database (FAI data.)

Jahr	FR	DE	FI	IT	NL	SE	UK	total
2006	271	233	7	640	54	129	85	1419
2007	62	155	151	763	74	179	576	1960
2008		47					18	65
total	333	435	158	1403	128	308	679	3444

Table 1: Data available in the In-depth accident causation database

Table 4.2 gives an overview about the availability of data in the CARE Database for the countries in the in-depth Database.

Code	Country	2003	2004	2005	2006
FR	France	**X**	**X**	**X**	**X**
DE	Germany				
IT	Italy	**X**	**X**		
NL	The Netherlands	**X**			
FI	Finland	**X**	**X**	**X**	**X**
SE	Sweden	**X**	**X**	**X**	**X**
UK	Great Britain	**X**	**X**	**X**	**X**
	North Ireland	**X**	**X**	**X**	

Table 4.2: Data availability in the CARE Database: X denotes data available.

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

Table 4.1 and 4.2 show, that there is no possibility for a direct relation based on year and country. Therefore the following rules for aggregating the data for the comparison tables in the next section are defined:

- For the FAI columns all data from the accident causation database as seen in table 1 is aggregated irrespective of the year.
- For the CARE Columns data from the last year available is used. So the fatal accidents for France, Finland, Sweden, from 2006 are aggregated. For UK, a synthetic 2006 data set has been build by adding the North Ireland data from 2005 to the Great Britain data from 2006. For Italy, the Data from 2004 and for The Netherlands the data from 2003 are used.
- Germany is excluded from the following analysis because there is no CARE data available.
- All Variables used from the FAI dataset are recoded and re-categorized where necessary to make it compatible with the CARE definitions.

Following the rules in the Annual Statistical Reports (ASR) of SafetyNet work package 1 for selecting the variables (e.g., 2008), the basic characteristics of fatal road accidents in the EU member states have been selected as those which might be useful for road accident analysis and where data are available for all or most of the EU countries (these are results of CAREPLUS project on making accident data comparable). More characteristics are available since the CAREPLUS 2 project. More precisely, the basic characteristics of fatal road accidents refer to (exact definitions can be found in the glossary of the ASR):

- Person class (driver, passenger, pedestrian)
- Person killed (age groups and/or gender)
- Area type (inside or outside urban area)
- Motorway (yes or no)
- Junction: (here only yes or nor, because the counts in junction type are often very small.
- Weather conditions (dry, fog or mist, rain, snow/sleet/hail, strong wind)
- Modes of transport – vehicle group (agricultural tractor, bus or coach [>8 seats], car or taxi, heavy goods vehicles, lorry under 3,5 tons, pedal cycle, moped, motorcycle, other)
- Instead of the time related variables in the ASR (Month, Day of the week and Hour) the lightning condition is used which is correlated with the hour. This decision was made because of the different time frames for the 2 datasets.
- In addition to the above list, the variable security equipment was selected.

With these rules, the results can easily be compared and linked to further information in the Annual Statistical Reports or the different Basic Fact Sheets from WP 1 of the SafetyNet project (http://www.erso.eu/data/content/accident_statistics.htm#_Accident_statistics)..

4.3 Results

For the comparison of the two datasets in this section, the most relevant variables following the rules above have been selected, which at least have reasonable distributions over the categories in some countries. The results are presented as figures in the main text. The underlying tables can be found in the Appendix 4.1 which includes the raw data in the upper part and the column percentages in the lower part. Whereas the raw counts in the upper part should be used as information, if this variable can be used in further statistical analysis (most methods based on normal or chi-square assumptions need expected counts greater than 5 as rule of thumb), the lower part facilitates the comparison of the in-depth dataset (FAI) and the CARE database.

Figure 4.1 shows the results for gender.

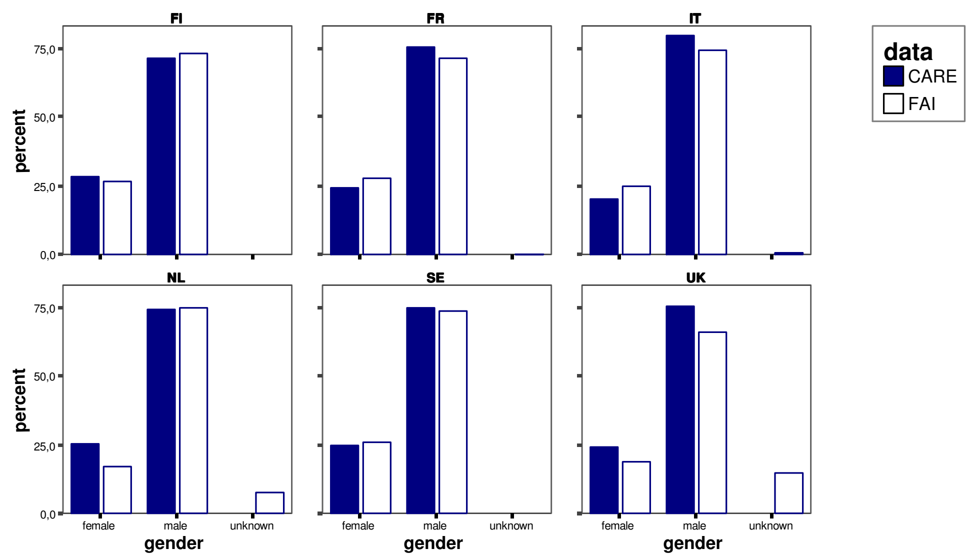


Figure 4.1: Gender

The only remarkable deviation can be found in UK, but this results from the 15% unknown in the FAI data. The percentages based only on the valid counts in male and female are 22.0% for females and 78% for males, which is close to the CARE relation.

The result for age of the victims categorized in 10 year intervals is similar, as shown in figure 4.2.

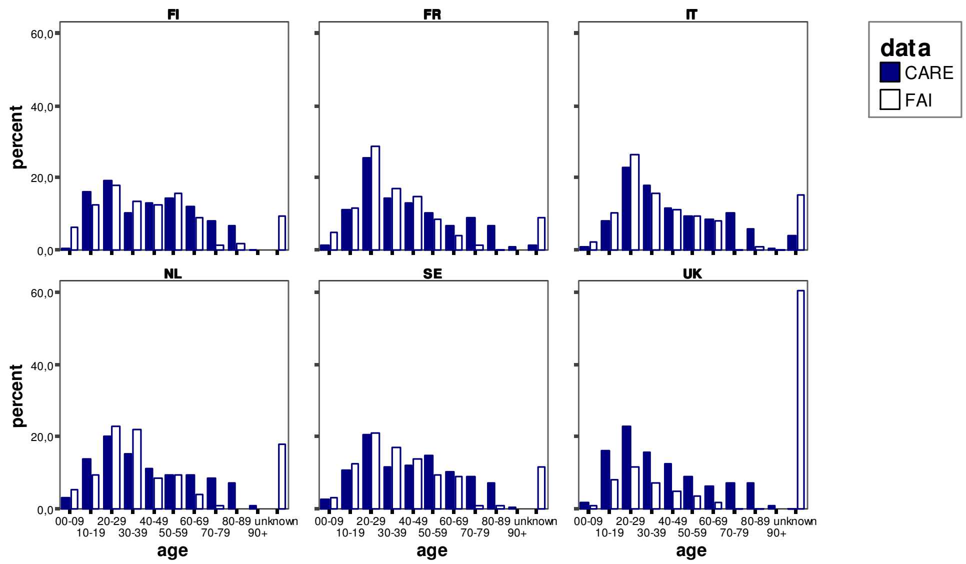


Figure 4.2: Age groups

The notable differences in the UK column are again a consequence of the large unknown part in this sample. The pattern based on the percentages of "known" ages only is very similar to the CARE distribution.

Concerning the person class, the proportion of drivers in the FAI data is a good sample from CARE. Pedestrians are under-represented whereas passengers show the opposite pattern. This bias is maybe caused by different rules of recording or different accuracy in getting information (see Figure 4.3).

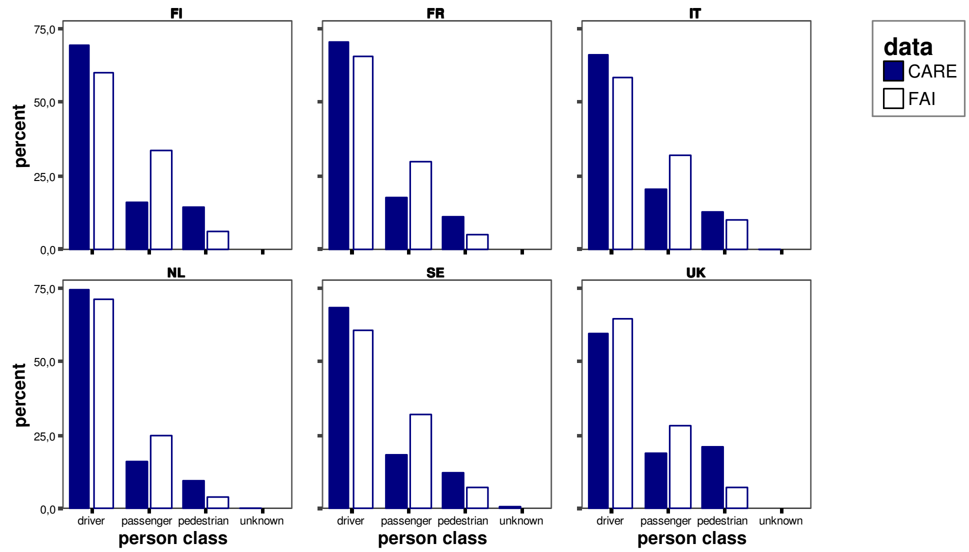


Figure 4.3: Person Class

The distribution for the categories in the variable 'vehicle group' is mostly similar in the two data sources. The percentage of involved cars is slightly higher in the FAI data and the percentage of motor cycles and mopeds slightly lower (see figure 4.4).

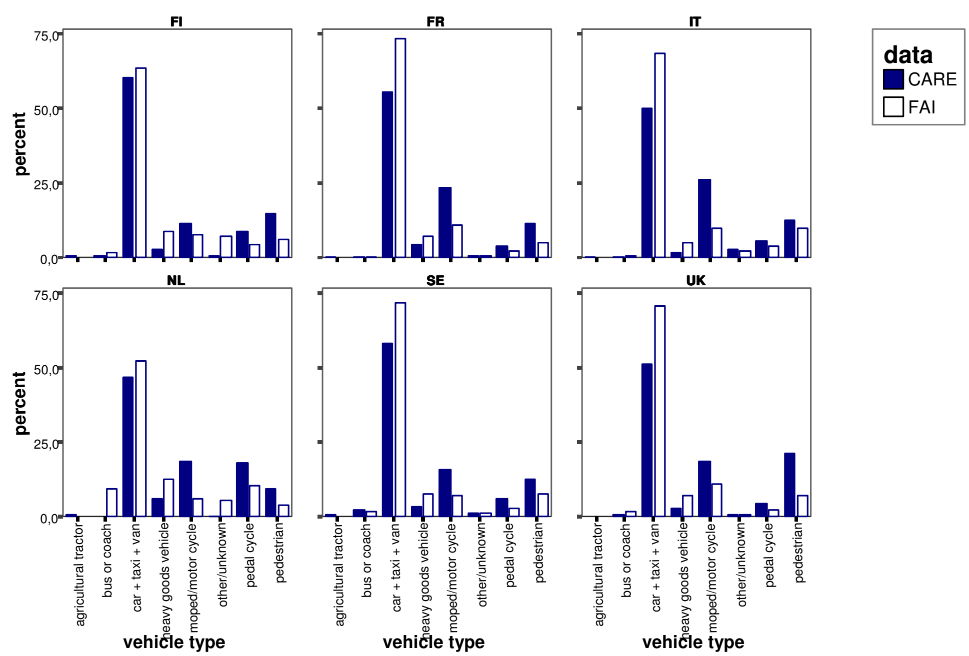


Figure 4.4: Vehicle group

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

The availability of information about security equipment differs between the countries in the CARE database as shown in the following Figure 4.5.

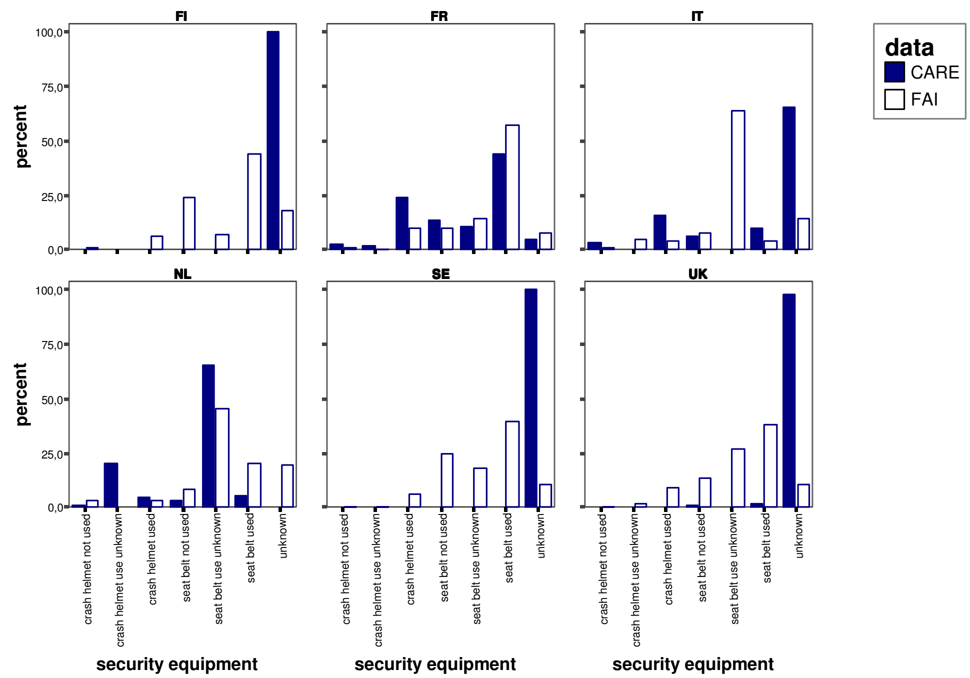


Figure 4.5: Security equipment

The following Figures 4.6 to 4.8 show variables describing the location of the accidents, figures 9 and 10 show the lightning condition and the weather at the time, the accident happened. The results are similar: the main pattern of the CARE distribution is reproduced in the FAI sample, but there are differences in the amount of the proportions and these differences also differ between the countries.

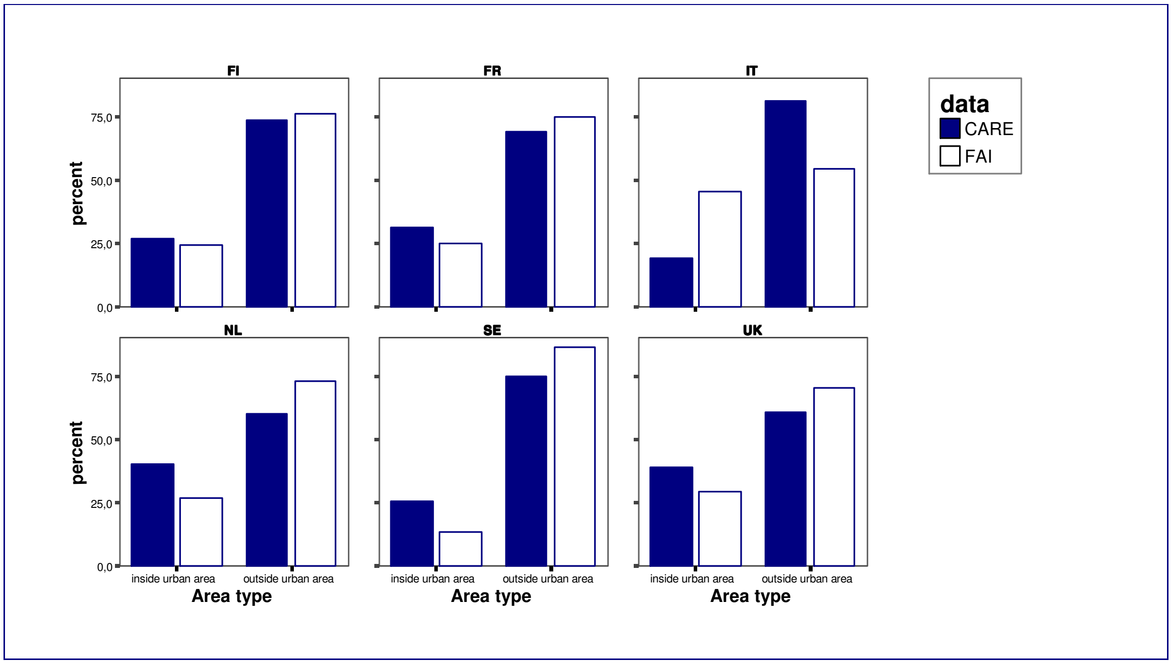


Figure 4.6: Area type

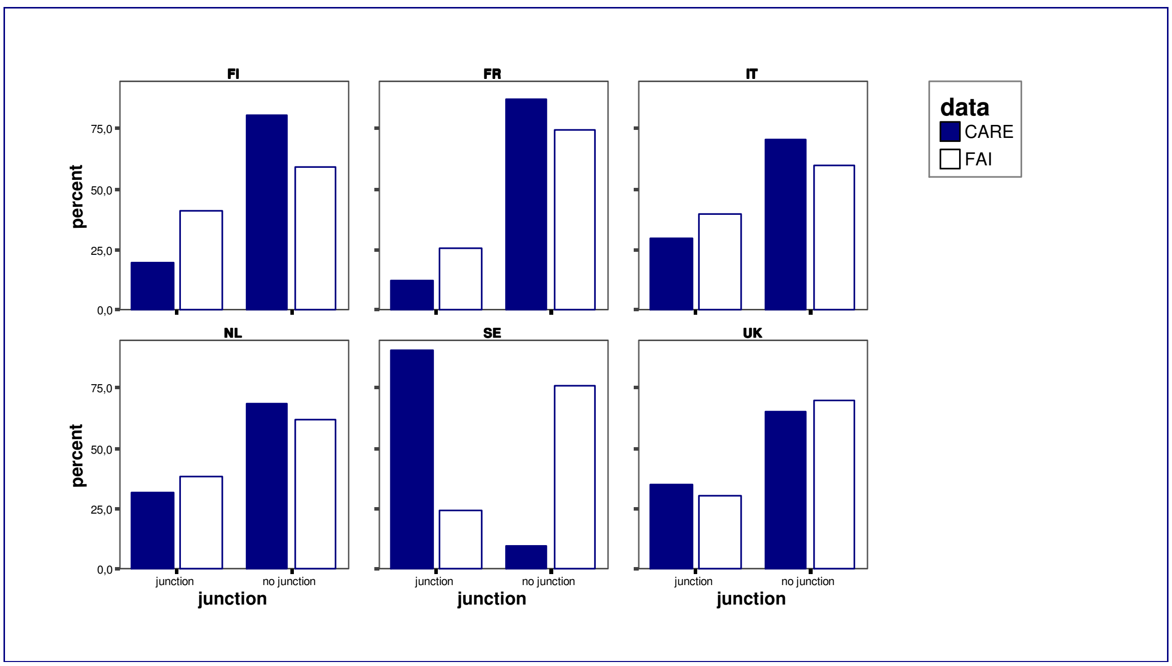


Figure 4.7: Junction

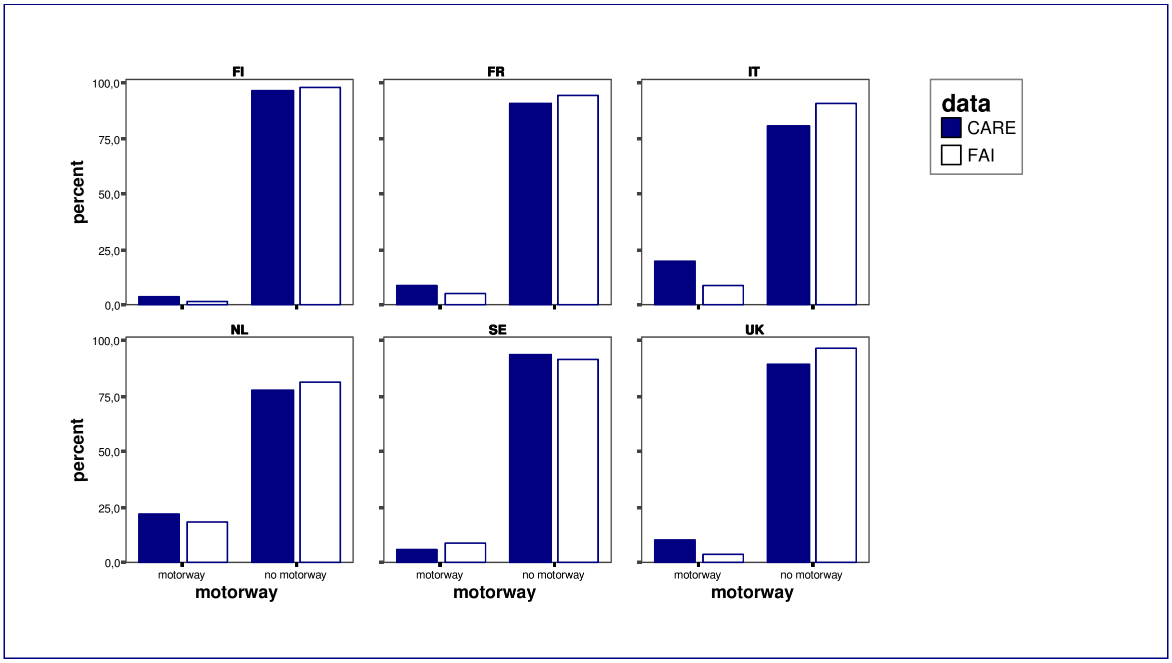


Figure 4.8: Motorway

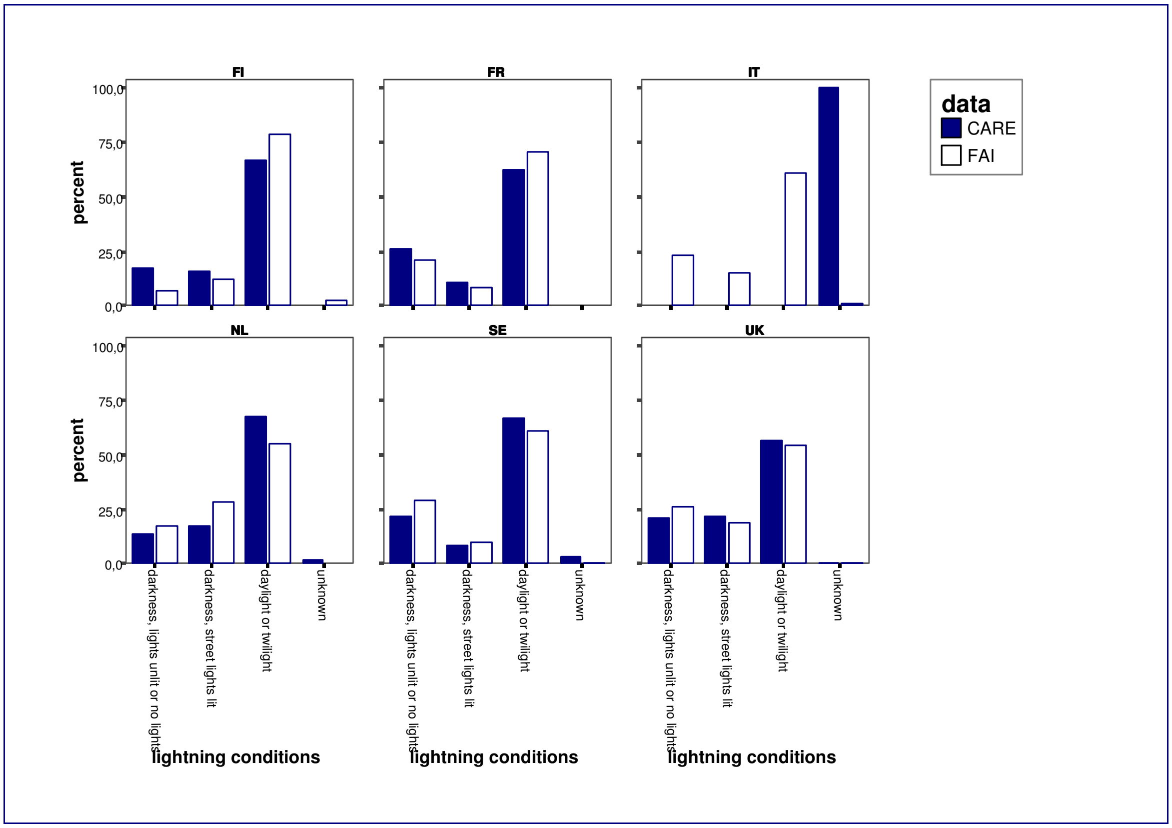


Figure 4.9: Lightning conditions

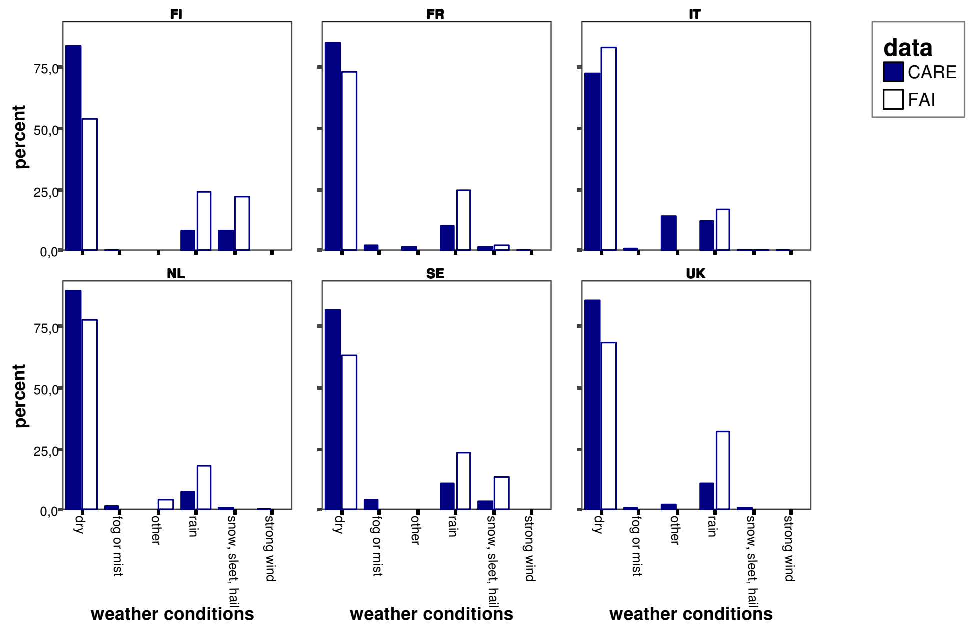


Figure 4.10: Weather conditions

4.4 Conclusion

As a conclusion, it can be said that the FAI data form a very good sample of the fatal accidents in the five countries analysed. The comparison between CARE and FAI with respect to the most important variables available in CARE shows that the distributions in both databases match relatively well. Consequently, many of the results from analysing the FAI data can be generalized to the country in question.

Representativity, however, is never a general characteristic of a given data set. Rather it has to be investigated for each particular research question by selecting a set of reference cases (e.g., from CARE) that matches as well as possible the population to which the results have to be generalized. The comparison between the dataset that will be analysed (e.g., FAI) and the reference dataset should include those variables that are most important in the analysis to be conducted. This way the test of representativity has to be tailored to the analysis to be conducted, and – if necessary – weights should be calculated on this basis.

Chapter 5 - Accidents and Road-safety Attitudes in Europe

Heike Martensen (IBSR), Emmanuelle Dupont (IBSR), &
Christian Brandstaetter (Kfv)

5.1 Objective

In Figure 5.1 we can that the number of fatalities per million inhabitants varies across Europe. However, we also see two trends that are relatively stable in all countries: (1) There are more fatalities among young people than among old and (2) there are more fatalities among men than among women. In this Chapter it is investigated whether this robust trend is related to different attitudes in these different age or gender groups. To explore a possible relation, the CARE data are related to the SARTRE data.

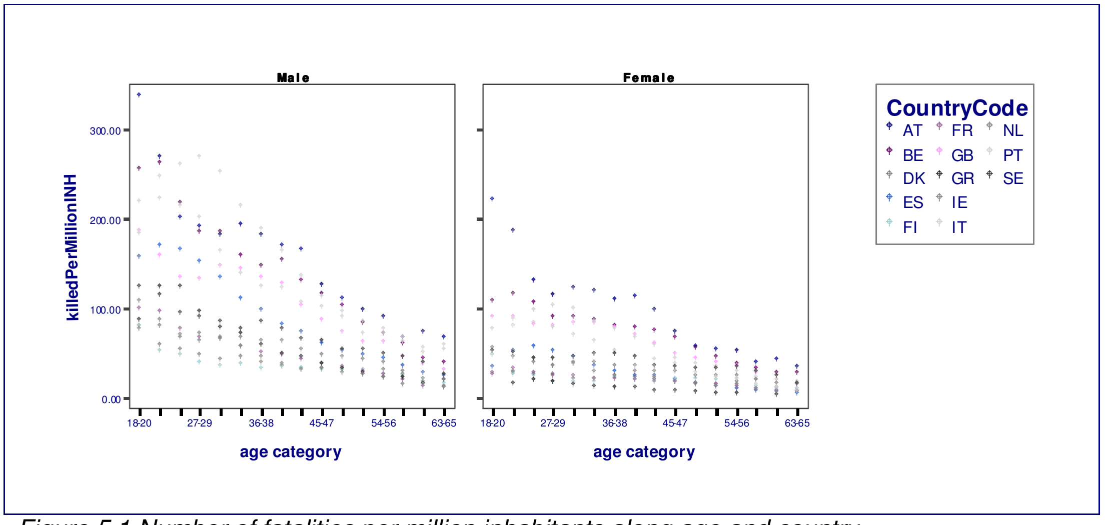


Figure 5.1 Number of fatalities per million inhabitants along age and country

The different waves of the SARTRE study have allowed collecting information about the attitudes of a large panel of European drivers. The SARTRE data themselves have already been carefully explored and scrutinized in order to identify the structure underlying the observations (SARTRE consortium 2004a & b).

The present analyses have a different goal, namely: To determine whether these attitudes data enable the prediction of the occurrence and/or severity of accidents, such as those compiled in the CARE database. The observations contained in one data file, however, do not concern the same individuals as in the other: The majority of the SARTRE respondents had never had any accident, while those people involved in the accidents stored in the CARE database had not been questioned with respect to their attitudes. This precludes

Project co-financed by the European Commission, Directorate-General Transport and Energy

answering questions such as “How do drivers’ attitudes relate to their accident scores?” It is, however, possible to aggregate both types of data at a meaningful group level in order to overcome this matching problem. For example, Biecheler and Cauzard (1996) formed homogenous groups of French drivers and linked their Sartre results to mobility and accident data. In the present study, the groups were defined on the basis of the respondents’ age, gender, and country. Age was the construct of main interest. The association between age and accident records is indeed known to be a strong one. Clearly, experience is involved to a large extent in the high accident records of the youngest drivers. But attitudes are also frequently evoked to explain the fact that young drivers are overly represented in national accident records. The present piece of work intends to focus on this specific aspect of the problem, by examining the *relationships* existing *between the attitudes characterising different age groups,* on the one hand, and these *age groups’ accidents and fatality records, on the other*. For this purpose, the SARTRE data were aggregated at the group level, reduced to a small number of dimensions, and related to the accident number, fatalities number, and the accident severity (the ratio between fatalities and accidents).

5.2 Methodology

The bases for this study were two aggregated data files. The aggregated CARE data contained accident and fatality numbers as well as the accident severity for groups defined by age, gender and country. The aggregated SARTRE data contained attitude data that was aggregated for the same groups. The data therefore varied with respect to age, gender, and country. Because age was the variable of main interest, the variation due to this variable was disentangled from that of country and gender.

Indeed, the groups’ accident records are most plausibly affected by a variety of other factors – many of them being determined by their country, such as the infrastructure, the laws and their enforcement, the climate, and the extent of reporting (or more importantly of *under*reporting) to name just a few salient ones. The influence of these country-related factors can be thought to be more or less the same for each group *within* a particular country (i.e. everybody is subject to the same laws, enforcement practices, infrastructure weather, accident registration, etc.). They can, however, vary strongly across countries.

Countries also bring about differences in exposure that need to be controlled for when interpreting accident numbers. Due to different size and population structure, the accident numbers of age groups from different countries differ from the outset. The usual procedure would consist of controlling for this difference by including an exposure measure, correcting for exposure differences while leaving all other differences between the countries in the data. As mentioned above however, this analysis focuses on variation associated with age, not with countries. The approach chosen was therefore to *take all systematic differences between the countries out of the analysis*, by including country as a random factor. In other word, the models’ intercepts will be allowed to randomly vary across countries.

Of course, differences between genders also introduce variation in the data that should be disentangled from the one of age. Again, this is true for both the attitude scores and accident records: A large difference was observed between accident numbers for men and women. This is assumed to be due – to a certain extent at least - to gender differences in exposure. This exposure difference could not be controlled for in this analysis. Therefore, the results for both genders cannot directly be compared, and should be interpreted differently. To deal with this problem all analyses were run separately for men and women.

The goal of the present analysis is to identify attitudes that are related to a high accident and/or fatality risk. Generally speaking, there are three variables that could be used as dependent variables in such analyses: the *number of accidents*, the *number of fatalities*, and the *accident severity* (the number of fatalities divided by the number of accidents). The number of fatalities is related to the number of accidents (more accidents cause more fatalities) as well as – although less strongly so – to their severity (the higher the number of fatalities, the higher the severity). In theory, the number of accidents should be independent from their severity. Therefore these two variables were selected as dependent variables in the present study.

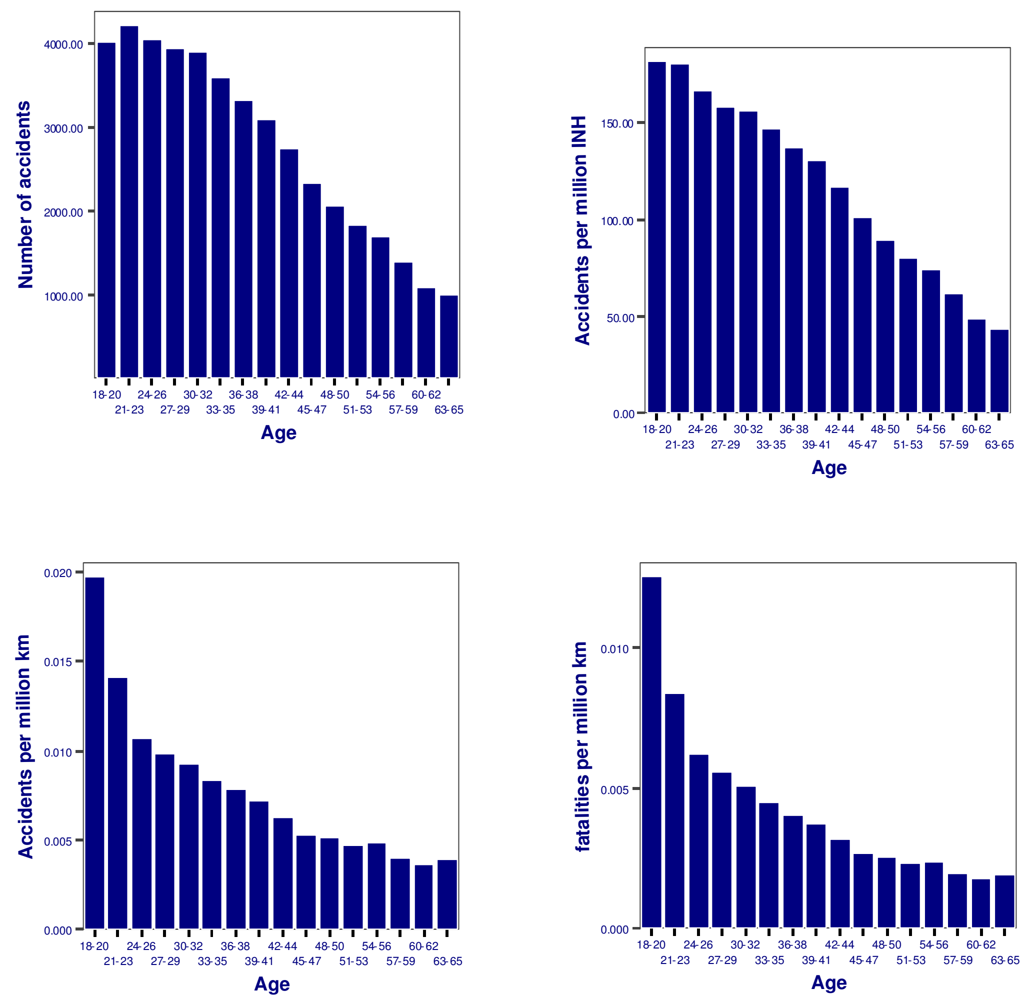
Contrary to accident and fatality numbers, accident severity is not directly dependent on exposure and might therefore seem to be comparable between countries. However, the extent of reporting strongly affects the measured accident severity (in a country in which only very serious accidents are reported at all, the ratio fatalities by accidents will be much higher than in countries where even the slightest injury accident is reported). Moreover, as mentioned above, situational factors (law, enforcement, infrastructure, climate, etc.) can also influence the accident severity in a way that is relatively homogeneous for all age groups *within* a particular country, but not across different countries. For these reasons, countries were included as random factors in the analyses of accident severity as well.

It should be noted that the approach of bundling all country differences in random factors is not unproblematic. On the one hand, different age groups within each country might differ on a number of factors that do not have anything to do with attitudes (e.g. exposure, laws that apply specifically to particular age groups…). These differences are not accounted for in the present analysis. On the other hand, the attitudes under investigation here may also vary across countries and therefore affect the overall accident number or severity score in each country, which are now removed from the analysis. However, should the assumption according to which attitudes affect accident frequency or severity be supported, then this relation should also manifest itself through differences between age groups *within* each country.

In the following, it will first be described how the data from CARE were aggregated, and subsequently how the data from SARTRE were reduced and aggregated, and how the two data sets were matched to each other.

5.2.1 Aggregating the CARE data

CARE data from 2003 were selected. This is firstly because the SARTRE data used in this analysis were also collected that year. Secondly, for the majority of the countries involved in the Sartre project, 2003 is the most recent year for which accident data are available. All data selected in CARE concerned only the drivers, given that the SARTRE data were collected exclusively from drivers. The data initially downloaded were the number of killed (30 days) and the number of accidents. They were aggregated over country, age, and gender.



<u>Figure 5.2</u> Distributions of accidents and fatalities numbers (corrected for exposure in lower graphs) along age categories

The age-groups were defined as clusters of three years, starting at 18. Although there are a few 17 year old drivers, these were excluded from the analysis because the size of these groups within each country was not large enough. Participants older than 65 were not considered in this study because the number of drivers in this group and their exposure to traffic is declining. The age categories were: 18-20; 21-23; 24-26; 27-29; 30-32; 33-35; 36-38; 39-41;42-44; 45-47; 48-50; 51-53; 54-56; 57-59; 60-62; 63-65.

The number of accidents, accidents per million KM driven (as estimated from the number of inhabitants and the SARTRE indication about km driven per age group), fatalities (killed30) and fatalities per KM driven are plotted for each age group in Figure 5.2. This Figure indicates that the development across age groups looks very similar for the number of accidents and the number of fatalities. The picture changes quite strongly once the mean number of km driven by each group is taken into account: The very young drivers stick out as having much more accidents and fatalities than what is suggested by the number of km they indicated to have driven.

The accident severity was calculated on the basis of the number of fatalities and of accidents, namely, as the log of the ratio between fatalities and accidents (= log(Fatalities) – log(Accidents)). Figure 5.3 below shows the accident severity for each age group. Although there is a clear trend for younger people to have more fatal accidents than for older ones, the relation with age is not as strong as for the sheer numbers of Accidents or Fatalities.

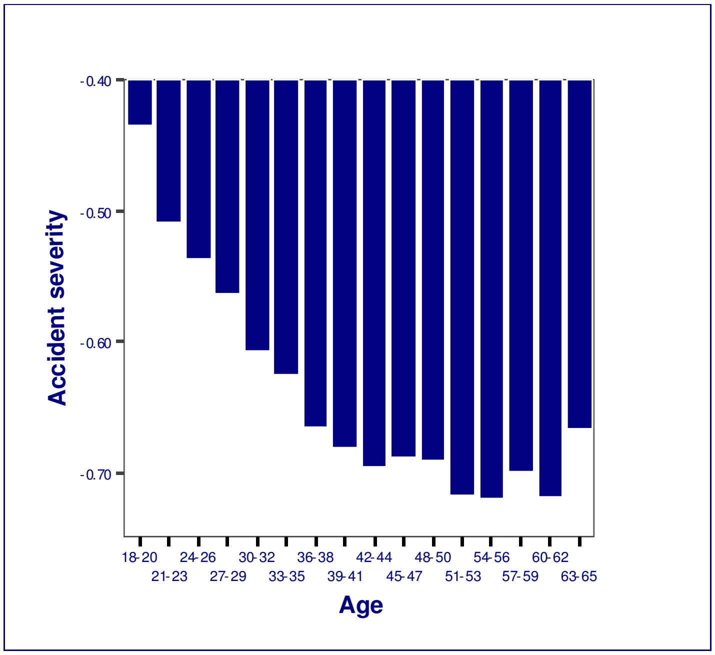


Figure 5.3 Accident severity along age categories

5.2.2 Aggregation and reduction of the SARTRE data

Only the most recent Sartre data, which have been collected in 2003 (Sartre 3) were used for the analysis. Consequently, only the countries for which CARE data for 2003 are available have been selected (namely, Austria, Belgium, Denmark, Finland, France, Greece, Ireland, Italy, The Netherlands, Portugal, Spain, Sweden, and Great Britain). The data were weighed by a factor defined by the SARTRE team to guarantee representativity with respect to the distribution of age and gender in the sample (SARTRE 3 Consortium 2004b) [4].

To produce predictors based on each groups' responses to the SARTRE attitude questionnaire, two steps were conducted, first aggregating the data per group and, in a second step, creating component scores. In the first step the variables from the original SARTRE data file were aggregated across the members of different groups. The groups were defined in the same way as those described above for the CARE data. Four groups were excluded from the analyses, because they had fewer than 4 respondents (Greece: women 18-21, women 30-32, women 62-65; Ireland, women 18-21). On the basis of the questionnaire's structure, subsets of items were selected on basis of which new, composite variables were created. See Appendix 5.1 for details of how the individual responses to each question were aggregated into group scores.

In the second step, principal component analyses (PCA) were conducted on the aggregated group scores. An initial PCA was conducted that included all composite variables listed and defined in Appendix 5.1, as well as data for all countries. This analysis did not yield very satisfying results, though (45.7% of the variance explained by 3 components).

Composite variables in PCA	*Communalities*
CauseExternalInternal	.574
MoreDangerous	*.263 x*
FasterThanOthers	.750
Speeding	.754
LikelyCheckSpeed	.578
FinedPunishedSpeed	.383
OftenDangerousBehaviour	*.177 x*
NoSeatBelt	.584
NeverTransportChild	.423
FinedPunishedSeatBelt	.411
DrinkAlcohol	*.122 x*
DrinkDrive	.580
CheckedAlcohol	.487
FinedPunishedAlcohol	.379
LikelyCheckedAlcohol	.789
ImpatientDriver	.545
NotWorried	.418
Aggressive	.435
OnlyPrimaryEducation	*.187 x*
Urban	*.335 x*
PhoneCalls	.485
DriveProfession	*.318 x*

Table 5.1: Communalities for the composite variables included in PCA on all data (x: variables that were removed because of low communalities)

Two additional steps were then taken in order to increase the amount of explained variance: First, all composite variables having low communalities

[4] The datafile used was S3x23j.sav. Actually there is a file with more recent weights S3x23j_miss.sav. The results for this file were checked and are very similar to the present ones.

(<.35, see Table 5.1) in the original 3-component-solution were removed. Second, 13 PCAs were conducted on the reduced set of variables, with one country left out at a time. With all the countries included, the solution obtained explained 56.3% of the variance with 3 components. Every time that the exclusion of a country yielded 1% or more increase of the explained variance, the same country was definitively discarded from the remaining of the analysis. Three countries were excluded on the basis of this criterion: Belgium, France, and Portugal. The data from 10 countries were consequently left in the analyses (Austria, Denmark, Finland, Greece, Ireland, Netherland, Spain, Sweden, and Great Britain). The final solution was a 3-component solution that explained 61.6% of the variance. The whole procedure and results of the PCA are described in Appendix 5.2.

The resulting components

Component 1: Aggression and Speeding.

Variable	*Loading*	*Definition*
Faster than others	.86	Percentage of drivers reporting driving a little or much faster than others
Speeding	.82	Average of self-reported frequency of speed infringements on various types of road
Impatient Drivers	.82	Percentage of drivers reporting being very to fairly annoyed by other drivers and fairly to very much enjoying speeding
Aggressive	.71	Percentage of drivers reporting having displayed aggressive behaviours towards others in the last 12 months
Phonecalls	.68	Average self-reported number of times of phoning-while-driving in the last 12 months
Not Worried	.64	Percentage of drivers reporting not being much worried or not worried at all when their family is out driving, or not agreeing with the idea that « a car is just a means of transport »
Fined/Punished for speed	.56	Percentage of drivers having been fined/punished in the last 3 years

Table 5 .2 Variable definitions for Component 1

The composite variables with the highest loadings on this components were aggressive behaviour, speeding, fined for speeding. In Figure 5.4, it can be seen that men score generally higher than women on this component.

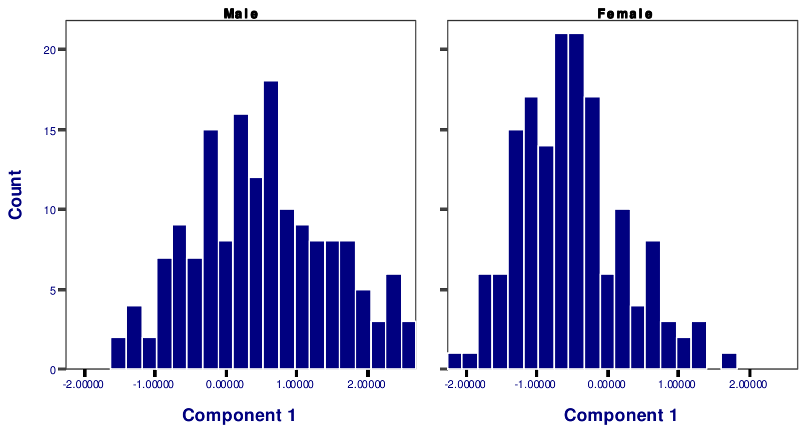


Figure 5.4: Distribution of the Agression & Speeding component in the two gender groups

Component 2: Unsafe Behaviours

Groups that scored high on this component contained many people who report a low impetus to contribute to their own or others safety by their behaviour. They admit on drink driving and not wearing a seat-belt. They see accidents as caused by external factors rather than factors that can be controlled by the driver and they typically do not have (or take care of) children.

Component name	*Loading*	*Definition*
Drink/drive	.84	Average self-reported frequency of alcohol infringements during the last week: The higher the score, the more frequent the infringement
Cause Internal/External	.75	Ratio of mean for external causes to mean for internal causes: The higher the score, the more the emphasis on external causes
Seatbelt	-.75	Average of self-reported frequency of SB use on various types of roads (the higher the score, the more frequent the SB is used)
Never Transport Child	.64	Percentage of people reporting never transporting children in a group
Fined/Punished Seatbelt	.52	Percentage of drivers having been fined/punished in the last 3 years

Table 5.3 Variable definitions for Component 2

As can be seen in Figure 5.5, the difference between the score distributions for men and women is not as large for this component as it is for Component 1. Still, there are overall more male than female groups with high values.

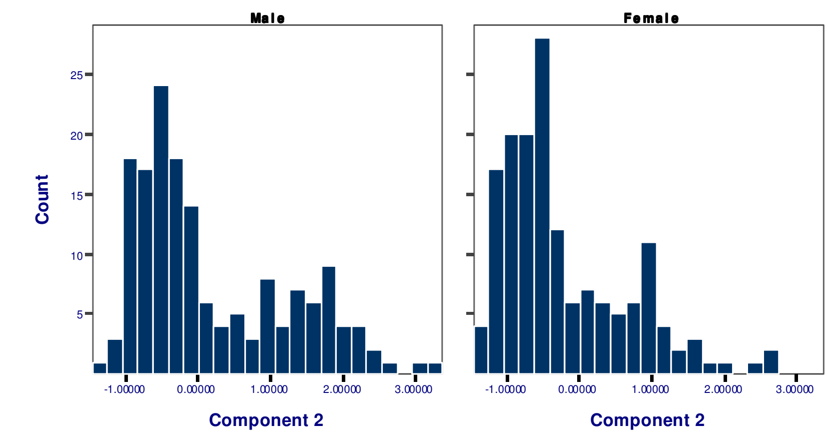


Figure 5.5 Distribution of the Other Dangerous Behaviour component in the two gender groups

Component 3: Perceived Likelihood of Controls

Groups that scored high on this component contained a high percentage of people who perceive the likelihood to be controlled by the police (either for alcohol or for speed) as moderate or high.

Component name	*Loading*	*Definition*
Likely checked alcohol	.88	*Percentage of drivers* reporting that it is likely (sometimes, often, always) that they'd be checked for alcohol on a typical journey
Likely checked speed	.76	*Percentage of drivers* reporting that it is likely (sometimes, often, always) that they'd be checked for speed on a typical journey
Checked alcohol	.67	*Percentage of drivers* who answered having been checked for alcohol in the last 3 years more than once

Table 5.4 Variables of Component 3

Figure 5.6 indicates that most male groups have a moderately high score on this component, while the scores among the female groups are spread more strongly. There are a few more female groups (as compared to the male ones) with relatively low scores but especially more female groups with a relatively high score.

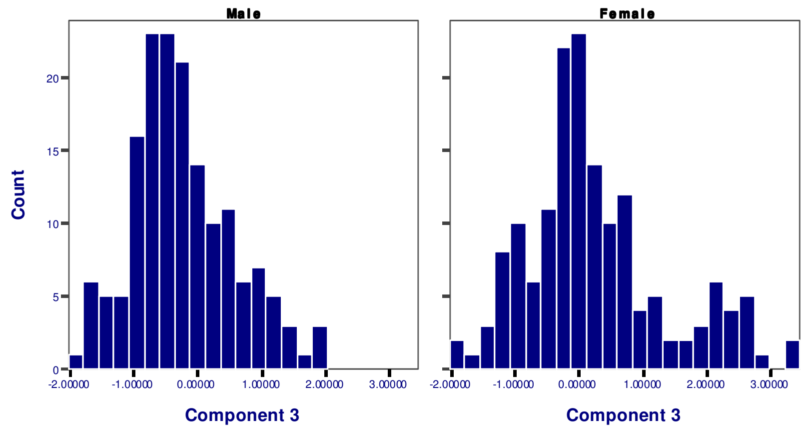


Figure 5.6 Distribution of the Perceived Likelihood of Controls component scores in the two gender groups

5.2.3 Regression between the components scores and Accident/Fatality outcomes

To investigate the relation between the components resulting from SARTRE and the accident outcome variables, linear mixed effects models were calculated in which the Component-scores were used to predict accident and victim data.

Three different dependent variables were considered:
1.) Number of accidents
2.) Number of killed drivers
3.) Accident severity

The number of accidents and the number of driver fatalities are counts. Consequently, for these two variables generalized linear models were calculated using the quasi-Poisson distribution.

The accident severity is defined as the ratio of two counts. The difference of the logarithm of (log (fatalities) – log (accidents)) was thus used as the dependent variable in a linear model. To compensate for the reduced variation resulting from using the difference between two highly correlated values, each value was weighted by the number of accidents it was based upon (Bijleveld, 2005). A detailed description of the underlying reasoning is provided in Appendix 5.3.

All models were estimated using the “R” software (package lme4), by maximising the restricted maximum likelihood (REML). For the log number of fatalities and log number of accidents a Laplacian approximation (which is more accurate than the PQL algorithm used in MLwin; see e.g. Diaz, 2007) was used. In each model, country was introduced as a random factor. As described in the introduction, the underlying reasoning is that the numbers (accidents and/or killed) are strongly determined by country-related factors that have nothing to do with attitudes, and are common to all age groups within a country. The most important of these factors is of course the size of the driving population, but many other factors play a role as well. By defining country as a random factor, the intercept is allowed to vary across countries. In this way, country-specific intercepts are estimated that reflects the variation in the dependent variable that is common to all groups within the same country.

Subsequently, it was tested whether assuming random country variation in the slope for each of the three components was also necessary. In other words, it was tested whether the relation between say, the Aggression and Speeding component varied across country, without making any attempt at explaining this variation. For each dependent variable, the full model (containing all three PCA components as fixed factors, and random factors for the intercept as well as for all three component slopes) was compared to restricted models with one of the random slopes removed . The deviances associated with the full and with the restricted models were compared on the basis of Chi-square tests. As far as the tests of random country variation are concerned, the results obtained suggested the following main conclusions:

1.) Accident severity
a. Women: no random slopes (joint $\chi 2$ (9) = 7.165)
b. Men: random slope for Component 1 ($\chi 2$ (4) = 24.24) but not for Component 2 (X2(4)=0) or Component 3 ($\chi 2$ (4)=4.41).
2.) Accidents numbers: Random slopes for all three components (all $\chi 2$s > 500).
3.) Fatality numbers: Random slopes for all three components (all $\chi 2$s >300).

In order to ensure that a common model structure was used with data from both women and men and for each of the three dependent variables, a random slope for each component was included in the model for each gender.

To test the significance of the fixed effects parameters, the Bayesian MCMC procedure in R was used (mcmcsamp in lme4). Each model was run for 200000 cycles. This procedure resulted in a posterior distribution for each of the parameters. These distributions were used to determine the probabilities for each parameter to be zero. According to Baayen, Davidson, & Bates (2008), the probabilities acquired in this way are to be preferred above the anticonservative p-values derived from the estimates of parameters and standard errors of the original t-tests.

5.3 Results

An overview of the results of the models that were run separately for men and women is provided in Table 5.5. The numbers presented are the coefficients (B) for the three component scores (Aggression and speeding, Other unsafe behaviour, Perceived control likelihood) when predicting the three dependent variables (Accident severity, Number of accidents, Number of fatalities). The level of significance is indicated by asterisks. More detailed results can be found in Appendix 5.4.

	Accident Severity		*Number Accidents*		*Number Fatalities*	
	Men	*Women*	*Men*	*Women*	*Men*	*Women*
C1 - Aggression and Speeding	.094	.043	**.443***	**.516***	**.538***	**.551(*)**
C2 – (Other) unsafe behaviour	.023	-.004	-.032	-.098	-.016	-.071
C 3 – Perceived control likelihood	.041	-.039	-.057	-.123	-.014	-.172

*Table 5.5 Separate gender models – Results of the regression of the three Components scores on AccidentFatality, Accidents numbers and Fatalities numbers –(*) p<.10; * p<.05; ** p<.01; *** p<.001*

The coefficients in Table 5.5 indicate that the Aggression and Speeding component is strongly related to the number of accidents, and the number of fatalities. The relation is positive, indicating that groups with higher scores on Component 1 have more accidents and more fatalities. The relation between Component 1 and accident severity is not significant.

The scores of Component 2 "(Other) unsafe Behaviours" and 3 "Perceived likelihood of control" were not significantly related to any of the dependent variables.

The results show more or less identical patterns for men and women. Except for two cases all coefficients are estimated in the same direction and in the same order of size. The two cases where signs changed between the coefficient for men and that for women (accident severity with component scores for other dangerous behaviour and perceived likelihood of control) the coefficients were very close to zero. The similar results in this analysis do not mean that men and women have equally high accident and fatality scores (which they don't: Men have a higher accident and fatality rate in each of the countries). They simply means that just like for male groups, those female groups with high scores on the aggressiveness- and speed-component are relatively speaking also the ones with higher occurrences of fatalities and accidents as compared to the other female groups.

5.4 Conclusion

The attitude data from Sartre were linked to the CARE accident data by investigating whether population groups that have many accidents and/or driver fatalities also show different results in the Sartre questionnaires as compared to groups with low accident or fatality numbers. To this end the responses to the Sartre questionnaires were condensed to three dimensions: Component 1 represents readiness to speed and to display aggressive behaviour. Component

2 represents the unwillingness to wear a seatbelt and readiness to drink and drive; and Component 3 represents a low perceived chance to be controlled by the police. The scores for these components were calculated for groups defined by age, gender, and country (e.g. 18-21 year old males from Austria). For the same groups the number of accidents and driver fatalities were retrieved from CARE.

Component 1, the aggressiveness- and speed-component gives relatively clear-cut results. It is well interpretable in terms of the questions that are statistically related to it, it explains the highest proportion of variance in the original Sartre scores, and it shows a strong relation with the accident and fatality numbers. Components 2 (other unsafe behaviour) and 3 (perceived likelihood of control) are less coherent in terms of the content of the questions that are statistically related to them. Moreover they explain less variance in the Sartre data. Finally there was no clear relation between Components 2 and 3 on the one hand and the accident and fatality numbers on the other hand. Not finding a relation between the accident and fatality data and the latter two components could either be a sign that there is indeed no relation or it could mean that the concepts that play a role in Components 2 and 3 (drink driving, seat belt wearing and perceived control likelihood) are not yet optimally measured.

The analyses indicated that those groups that have high accident and driver fatality number also more readily admit to speeding and to behaving aggressively in traffic. These findings need to be interpreted with extreme caution: What the analysis show is a co-occurrence of accidents/driver-fatalities and particular attitudes in different age groups. This does not prove that there is a causal relation. To illustrate this point: One would probably also find that groups with many accidents tend to contain people with very quick reflexes, because younger people have more accidents and they also have quicker reflexes. One would, however, not conclude that their quick reflexes cause accidents. In the same way, the present results do not say anything about whether the attitudes that young people display when filling in a questionnaire about traffic safety were actually the reason for the high number of accidents and fatalities. All they say is that the tendency to admit to speeding and aggressive behaviour and a high number of accidents and fatalities occur in the same groups.

Nevertheless, the results do show that a positive attitude towards speeding and impatient, aggressive behaviour is most strongly present among young people, who are also the ones with most accident. Consequently, these might be the most promising attitudes to be addressed in campaigns in order to increase road-safety.

Chapter 6 - Conclusion

In the present deliverable the use of and the need to put an accident database like the CARE database into perspective by adding other information was demonstrated. On the one hand, the contents of the CARE database were compared with external information, as the comparison with hospital data in Chapter 3, to identify needs for improvement in the CARE content (in this case the reporting of injury cases). On the other hand, information from the CARE database can be used as a reference for quality checks on other databases. For example the Fatal Investigation Database (FAI) described in Chapter 4, was compared to the CARE database and it was shown that the distribution of key variables was very similar in both databases, suggesting that the FAI data are a good sample of fatal data for their respective countries.

The major difference between CARE and many other international databases is the high level of disaggregation. CARE comprises detailed data on individual accidents as collected by the Member States. This gives the user the flexibility to disaggregate the CARE data according to a great number of variables. This way a wide variety of questions can be addressed and the CARE data can be analysed jointly with very different types of other information.

In Chapter 2, the Greek fatality and accident data were selected and disaggregated by county, to draw a road-safety map of Greece. Due to the fine-grained data structure (Nuts3) it was also possible to relate the accident data to the number of alcohol and speed controls in each county and show that those counties with most controls are those with the fewest accidents.

In Chapter 5, the data were disaggregated by age and gender. These disaggregated data – showing that young drivers have an overproportional share of the accidents and fatalities – could then be linked to attitude data from the SARTRE project that also had been aggregated for age and gender groups (e.g., all 18 year old male drivers) in each country. The results indicated that among the unsafe behaviours that were enquired about in SARTRE speed and aggression are the most characteristic for the groups of young people which have a high number of accidents and fatalities.

Generally speaking, the most important type of information that the CARE data have to be augmented with is a measure of exposure, as for example the km driven. Given that this information is not always available in the present deliverable, we have presented alternative approaches. In Chapter 2, the population numbers were included into the analyses as a proxy for the exposure to the risk of being involved in an accident. This approach is reasonable if it can be assumed that the transportation patterns (i.e., the distribution across transportation modes, the length of the routes) are similar across the regions in question. In Chapter 5, age differences were the focus of the study. The countries in this study differ widely in terms of exposure, but these differences were excluded from the analysis by including country as a random factor (which amounts to removing differences from the data that are the same for all groups in a particular country, and only retaining the differences between the age groups within the country).

To conclude, the studies presented in this deliverable give an impression of the wide range of different types of analyses that can be conducted with the CARE data. The data structure of CARE gives road-safety researchers great flexibility to disaggregate data tailored to the research question. This way, the data can also linked to augmenting data from all kinds of sources at all kinds of levels of disaggregation.

References

Aguero-Valverde, J., Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention*, 38(3), 618–625.

Annual Statistical Report (2008). Deliverable D1.20 of the EU FP6 project SafetyNet.

Amoros, E., Martin, J. L., Laumon, B. (2003). Comparison of road crashes incident and severity between some French counties. *Accident Analysis and Prevention*, 35, 537–547

Baayen, R.H., Davidson, D.J. and Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. To appear in *Journal of Memory and Language*, special issue on *Emerging Data Analysis Techniques*.

Biecheler, M.B. & Cauzard, J.-P. (1996) Road risk production: a social analysis through mobility, attitude reported behaviour and accident data. *IATSS Research 20 (2)*, 53-64.

Bijleveld, F., 2005, The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. *Accident Analysis and Prevention, 37*, 591-600.

Brace, C., 2005, *Fatal Data Methodology Development Report*. Deliverable D5.1 of the EU FP6 project SafetyNet.

Broughton J.; Amoros E.; Bos N.; Evgenikos P.; Holló P.; Pérez C.; Tecl J. (2007). *Estimating the real number of road accident casualties*. Deliverable 1.15 - Final Report on Task 1.5 of SafetyNet. TRL, Crowthorne.

Browne W.J. (2004). *MCMC estimation in MLwiN - Version 2.0*. Centre for Multilevel Modelling, Institute of Education, University of London.

Browne W.J., Goldstein, H. and Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling* 1: 103-124.

Stipdonk & Van Norden (2008), Time series applications on road safety developments in Europe. Deliverable D7.10 of the EU FP6 project SafetyNet.

Diaz, R. E., 2007, Comparison of PQL and Laplace 6 estimates of hierarchical linear models when comparing groups of small incident rates in cluster randomised trials. *Computational Statistics & Data Analysis*

Dupont, E. and Martensen, H. (Eds.) (2007) *Multilevel modelling and time series analysis in traffic research - Methodology*. Deliverable D7.4 of the EU FP6 project SafetyNet.

Eksler, V. (2008). Exploring Spatial Structure Behind Road Mortality in Europe. Article in press, *Journal of Applied Spatial Analysis and Policy*.

Eksler, V., Lassarre, S., (2008). Evolution of road risk disparities at small-scale level: Example of Belgium. Article in press, *Journal of Safety Research*, doi:10.1016/j.jsr.2008.05.008.

Everitt, B.S., (1977). *The analysis of contingency Tables*. Chapman and Hall.

Goodman, L.A., (1973). Guided and unguided methods for the selection of models for a set of T multidimensional contingency Tables. *Journal of the American Statistical Association* vol. 68, pp: 165-175.

Hays, W.L., (1981). *Statistics*, 3rd edition, Holt-Saunders International Editions.

Kim, K., Li, L., Richardson J., Nitz, L., (1998). Drivers at fault: Influences of Age, Sex and Vehicle Type. *Journal of Safety Research* Vol. 29 No 3, pp: 171-179.

Noland, R. B., Quddus, M. A. (2004).A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention*, 36(6), 973–984.

Reed, S. (2008). *A glossary of the data variables collected in both tasks 5.1 and 5.2.* Deliverable D5.5 of the EU FP6 project SafetyNet.

Raftery, A.E., and Lewis, S.M. (1992). How many iterations in the Gibbs Sampler? In J.M. Bernando, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Eds.), Bayesian Statistics 4, pp.763-777. Oxford: Oxford University Press.

SARTRE 3 consortium (2004a). European Drivers and Road Risk -- Report on principal results. *SARTRE 3 reports, Part 1*.

SARTRE 3 consortium (2004b). European Drivers and Road Risk -- Report on in-depth analyses. *SARTRE 3 reports, Part 2*.

SWOV (2001). A new linking procedure for the determination of the total number of hospitalised road traffic casualties by comparing police and hospital reports. The Netherlands.
Volume 51, Issue 6, Pages 2871-2888

Yannis G., Papadimitriou E., Antoniou, C. (2007). Multilevel modelling for the regional effect of enforcement on road accidents. *Accident Analysis and Prevention* 39 (4), 818-825.

Appendices

Appendix 2.1 Spatial models formulation and Bayesian estimation

Models without spatial effects (single level Poisson models):

Accident and fatality counts are modelled by assuming the data to be Poisson distributed:

$Y_i \sim$ Poisson $(\lambda_i N_i)$ …………………Level 1 model

where λ_i is the Poisson parameter, often referred to as the mortality rate, and N_i is the number of inhabitants in county (i).

$\log(\lambda_i)=\log(N_i)$

$\log(\lambda_i)=\log(N_i)+ \alpha + \beta.xi$

Models with spatial effects (multiple membership multiple classification MMMC multilevel models):

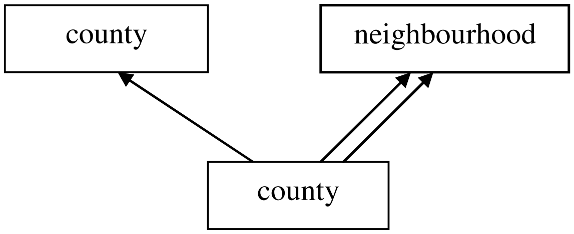
Two types of random effects are estimated: an exchangeable county random departure $u_{county\ (i)}$ (to account for overdispersion in the counts) and a neighbourhood random departure $u_{neigh1(i)}$, which is in fact a multiple set of random effects (to account for spatial effects) (Browne, 2004, Browne et al. 2001).

$\log(\lambda_i)=\log(N_i)+ \alpha_i + \beta_i\ xi$

$\alpha_i = \alpha + u_{county\ (i)}{}^{(2)} + \Sigma_{j\in neigh1(i)}\ w_{i,j}{}^{(3)}\ u_{jneigh1}(j)^{(3)}$

$u_{county\ (i)} \sim N(0, \Omega_u{}^{(2)})$ …………….unstructured component - second level
$u_{neigh1(i)} \sim N(0, \Omega_u{}^{(3)})$ ……………..structured component - third level

With respect to multilevel structure, this formulation corresponds to a MMMC (multiple membership multiple classification) model, which combines a cross-nested and multiple membership structure as follows:



In this model, although 3 levels are set, only 2 levels are practically considered, as "county" and "neighbourhood" are crossed at level 2.

Models with spatial effects (CAR and CAR convolution models):

Two models are considered: The first one containing only spatially structured residuals (CAR Conditional Autoregressive model) and the second one often called convolution model containing both structured and unstructured residuals.

In the CAR models, the individual random effects are not independent. The CAR model can be written as follows (Browne, 2004)

$Y_i \sim \text{Poisson}\ (\lambda_i N_i)$

$\log(\lambda_i)=\log(N_i)+ \alpha_i + \beta_i\ xi$

$\alpha_i = u_{county\ (i)}^{(3)}$

$u_{county\ (i)}^{(3)} \sim N\ (\bar{u}^{(3)}_{county(i)}, \sigma^2_{u(3)} / r_{county(i)})$

where $\bar{u}^{(3)}_{county(i)} = \Sigma_{\ j\in neigh(county(i))}\ \ w^{(3)}_{county(i),j}\ u^{(3)}_j / r_{county(i)})$

CAR convolution models are attempting to fit a set of spatial/neighbour effects and a set of exchangeable unstructured normal random effects. Therefore, the latter is to be included in the model formulation.

$Y_i \sim \text{Poisson}\ (\lambda_i N_i)$

$\log(\lambda_i)=\log(N_i)+ \alpha_i + \beta_i\ xi$

$\alpha_i = u_{county\ (i)}^{(3)} + u_{county\ (i)}^{(2)}$

$u_{county\ (i)}^{(3)} \sim N\ (\bar{u}^{(3)}_{county(i)}, \sigma^2_{u(3)} / r_{county(i)})$

where $\bar{u}^{(3)}_{county(i)} = \Sigma_{\ j\in neigh(county(i))}\ \ w^{(3)}_{county(i),j}\ u^{(3)}_j / r_{county(i)})$

$\bar{u}^{(2)}_{county(i)} \sim N\ (0, \sigma^2_{u(2)})$

While in the MMMC model there are (r) random effects for each county, with (r) equal to the number of neighbours of the county, the CAR model assumes one (global) random neighbourhood effect for each observation. Consequently, in the CAR model, the set of neighbours of each county is initially examined as a whole.

From a practical viewpoint, in the MMMC model the initial weights given to the neighbours of each county sum up to 1, whereas in the CAR model these are initially set equal to 1 for all neighbours of a county, because they will be divided by the number of neighbours of each county during the CAR estimation procedure.

Bayesian modelling

In the spatial models presented above, the fixed parameters are assumed to follow a uniform prior distribution, whereas the precisions (inverse value of variances) are assumed to follow a gamma distribution.

Two diagnostics can be used to ensure that the number of iterations is sufficient to obtain accurate posterior estimates: the Raftery-Lewis (Nhat), indicating the necessary number of iterations for accurate quantile estimates of the parameter posterior distribution (Raftery and Lewis, 1992), and the Brooks-Draper (Nhat), indicating the necessary number of iterations for accurate mean estimates of the parameter posterior distribution. In addition, convergence of the parameters can be estimated by means of visual examination of MCMC chains and autocorrelations plots.

Appendix 2.2 NUTS-3 neighbourhood matrix for Greece (road connections)

	Country name	direct neighbourhood matrix (road connection)						
		neigh1	neigh2	neigh3	neigh4	neigh5	neigh6	neigh7
1	Attiki	6	36					
2	Evia	6						
3	Evritania	4	5	46	41			
4	Fokida	5	6	46	3			
5	Fthiotida	3	4	6	41	46	42	43
6	Viotia	1	2	4	5			
7	Chalkidiki	13						
8	Imathia	9	10	11	13	51		
9	Kilkis	10	12	13	8			
10	Pella	8	9	13	48	51		
11	Pieria	42	51	8				
12	Serres	9	13	18	20			
13	Thessaloniki	7	8	9	10	12		
14	Chania	17						
15	Heraklio	16	17					
16	Lasithi	15						
17	Rethymno	14	15					
18	Drama	12	20	22				
19	Evros	21						
20	Kavala	12	18	22				
21	Rodopi	19	22					
22	Xanthi	18	20	21				
23	Arta	24	25	41	44	46		
24	Ioannina	23	25	26	44	49	50	
25	Preveza	23	24	26				
26	Thesprotia	24	25					
27	Kerkyra							
28	Kefalinia							
29	Lefkada	46						
30	Zakynthos							
31	Chios							
32	Lesvos							
33	Samos							
34	Arkadia	35	36	37	38	45	47	
35	Argolida	34	36					
36	Korinthia	34	35	1	45			
37	Lakonia	34	38					
38	Messinia	34	37	47				
39	Cyclades							
40	Dodekanissa							
41	Karditsa	3	5	23	42	44		
42	Larissa	41	43	44	5	11	49	51
43	Magnisia	5	42					
44	Trikala	23	24	41	42	49		
45	Achaia	34	36	47				
46	Aitoloakarnania	3	4	5	23	29		
47	Ileia	34	38	45				
48	Florina	10	50	51				
49	Grevena	24	42	44	50	51		
50	Kastoria	24	48	49	51			
51	Kozani	8	10	11	42	48	49	50

Appendix 2.3 NUTS-3 neighbourhood matrix for Greece (road + ferry connections)

county	neigh1	neigh2	neigh3	neigh4	neigh5	neigh6	neigh7	neigh8	neigh9	neigh10	neigh11	neigh12	neigh13	neigh14
1	6	13	14	15	16	17	31	32	33	35	36	37	39	40
2	5	6												
3	4	5	41	46										
4	3	5	6	46										
5	2	3	4	6	41	42	43	46						
6	1	2	4	5										
7	13													
8	9	10	11	13	51									
9	8	10	12	13										
10	8	9	13	48	51									
11	8	42	51											
12	9	13	18	20										
13	1	7	8	9	10	12	15	31	32	33	39	40	43	
14	1	17												
15	1	13	16	17										
16	1	15												
17	1	14	15											
18	12	20	22											
19	21	31	32	33	40									
20	12	18	22											
21	19	22												
22	18	20	21											
23	24	25	41	44	46									
24	23	25	26	44	49	50								
25	23	24	26											
26	24	25												
27	45	46												
28	45													
29	46													
30	46													
31	1	13	19	32	39									
32	1	13	19	31	39									
33	1	13	19	39										
34	35	36	37	38	45	47								
35	1	34	36											
36	1	34	35	45										
37	1	34	38											
38	34	37	47											
39	1	13	31	32	33	40	47							
40	1	13	19	39										
41	3	5	23	42	44									
42	5	11	41	43	44	49	51							
43	5	13	42											
44	23	24	41	42	49									
45	27	28	34	36	47									
46	3	4	5	23	27	29	30							
47	34	38	39	45										
48	10	50	51											
49	24	42	44	50	51									
50	24	48	49	51										
51	8	10	11	42	48	49	50							

Project co-financed by the European Commission, Directorate-General Transport and Energy

Appendix 3.1 Detailed conversion factors per country, road user type, MAIS score and police severity score

Czech

Conversion Factors	car occupant		motorcyclist		pedal cyclist		pedestrian		other		all	
	serious	slight	serious	slight	serious	Slight	serious	slight	serious	slight	serious	slight
1 or 2	0.97	1.11	1.03	1.17	1.11	3.50	1.05	1.77	0.88	1.00	1.07	1.56
3	0.07	0.01	0.05	0.01	0.30	0.04	0.31	0.04	0.13	0.00	0.15	0.02
4	0.01	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.03	0.00
5	0.02	0.00	0.05	0.00	0.03	0.00	0.04	0.00	0.00	0.00	0.03	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
All	1.08	1.12	1.12	1.18	1.61	3.54	1.40	1.80	1.00	1.00	1.28	1.58
>2	0.11	0.01	0.09	0.01	0.50	0.04	0.35	0.04	0.13	0.00	0.21	0.02

France

Conversion Factors	car occupant		motorcyclist		pedal cyclist		pedestrian		other		all	
	serious	slight	serious	slight	serious	slight	serious	slight	serious	slight	serious	slight
1 or 2	1.32	2.38	1.35	3.13	4.69	10.39	1.01	1.90	1.52	2.67	1.43	2.69
3	0.35	0.03	0.69	0.11	1.64	0.27	0.43	0.08	0.69	0.05	0.52	0.05
4	0.12	0.00	0.10	0.01	0.26	0.00	0.10	0.01	0.30	0.01	0.12	0.01
5	0.05	0.00	0.05	0.00	0.07	0.00	0.03	0.00	0.07	0.00	0.05	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
All	1.84	2.41	2.18	3.25	6.67	10.66	1.58	2.00	2.58	2.73	2.11	2.75
>2	0.51	0.03	0.83	0.12	1.97	0.27	0.57	0.10	1.06	0.06	0.68	0.06

Greece

Conversion Factors	car occupant		motorcyclist		Bicyclist		Pedestrian		Other-Unknown		All	
	serious	slight	serious	slight	serious	slight	serious	slight	serious	slight	serious	slight
1 or 2	4.08	6.09	6.89	10.72	7.50	23.75	2.49	3.91	11.45	15.41	5.92	9.10
3	0.57	0.15	0.60	0.17	0.00	1.00	0.31	0.13	0.53	0.07	0.52	0.15
4	0.00	0.01	0.07	0.01	0.17	0.17	0.13	0.00	0.17	0.02	0.08	0.01
5	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
All	4.65	6.25	7.57	10.91	7.67	24.92	2.93	4.04	12.14	15.50	6.52	9.28
>2	0.57	0.17	0.68	0.19	0.17	1.17	0.45	0.14	0.69	0.09	0.60	0.17

Hungary

MAIS	vehicle occupant		Pedestrian		all road users	
Conversion factors	serious	slight	Serious	Slight	serious	slight
1 or 2	0.83	1.28	0.86	1.16	0.84	1.27
3	0.43	0.04	0.22	0.02	0.38	0.037
4	0.06	0.00	0.08	0.00	0.06	0.00
5	0.02	0.00	0.05	0.00	0.03	0.00
6	0.00	0.00	0.01	0.00	0.00	0.00
All	1.35	1.33	1.21	1.19	1.32	1.31
>2	0.52	0.04	0.35	0.03	0.48	0.04

Netherlands

Conversion Factors	car/van		TWMV		pedal cycle		pedestrian		Other		All	
	serious	slight	serious	slight	serious	slight	serious	slight	serious	slight	serious	slight
1+2	1.07	1.02	1.21	1.04	1.90	1.10	1.23	1.04	1.24	1.01	1.29	1.04
3	0.18	0.01	0.32	0.01	0.66	0.04	0.31	0.02	0.18	0.01	0.33	0.01
4	0.02	0.00	0.03	0.00	0.05	0.00	0.03	0.00	0.02	0.00	0.03	0.00
5	0.01	0.00	0.02	0.00	0.02	0.00	0.02	0.00	0.01	0.00	0.02	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Grand Total	1.29	1.02	1.59	1.05	2.63	1.14	1.59	1.06	1.45	1.02	1.67	1.05
>2	0.22	0.006	0.37	0.016	0.73	0.041	0.36	0.023	0.21	0.007	0.37	0.016

UK

Conversion Factors	car occupant		motorcyclist		pedal cyclist		pedestrian		other		All	
	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight	Serious	Slight
1 or 2	1.15	1.03	1.34	1.13	2.54	1.24	1.05	1.03	1.62	1.06	1.244	1.045
3	0.13	0.00	0.25	0.01	0.26	0.01	0.18	0.01	0.23	0.01	0.169	0.005
4	0.01	0.00	0.01	0.00	0.02	0.00	0.03	0.00	0.01	0.00	0.014	0.000
5	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.004	0.000
6	0.01	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.01	0.00	0.011	0.000
all	1.30	1.03	1.61	1.14	2.83	1.25	1.28	1.04	1.88	1.07	1.441	1.050
>2	0.15	0.00	0.27	0.01	0.29	0.01	0.23	0.01	0.26	0.01	0.197	0.005

Spain

Conversion Factors	Serious	Slight
1 or 2	1.22	1.06
3	0.16	0.01
4	0.08	0.00
5	0.03	0.00
6	0.00	0.00
all	1.48	1.07
>2	0.26	0.02

Project co-financed by the European Commission, Directorate-General Transport and Energy

Appendix 3.2 Log-rate modelling formulation

A four-dimensional Table of i rows, j columns and k, l layers can be decomposed in row effects, column effects, layers effects and their interactions:

Simple	i, j. k, l
First order interaction	$i{\times}j$, $i{\times}k$, $i{\times}l$, $j{\times}k$, $j{\times}l$, $k{\times}l$
Second order interaction	$i{\times}j{\times}k$, $i{\times}j{\times}l$, $i{\times}k{\times}l$, $j{\times}k{\times}l$
Third order interaction	$i{\times}j{\times}k{\times}l$

The log-rate analysis uses an additive model that incorporates main effects and interactions between variables (1: country, 2: road user type, 3: police score, 4: MAIS score) in the following form:

$$\text{Log } F_{ijkl}= \text{Log } O_{ijkl} + u + u_{1(i)}+ u_{2(j)}+ u_{3(k)} + u_{4(l)} + u_{12(ij)}+ u_{13(ik)}+ u_{14(il)}+ u_{23(jk)}+ u_{24(jl)}+ u_{34(kl)}+ u_{123(ijk)} + u_{124(ijl)} + u_{134(ikl)} + u_{234(jkl)} + u_{1234(ijkl)}$$

Where F_{ijkl} are the expected cell frequencies, O_{ijkl} is an offset term and u are parameters to be estimated. In this case, the underreporting coefficients are modelled as the rate of the number of actual casualties to the number of casualties recorded by the police, therefore cell frequencies F_{ijkl} correspond to the actual number of casualties and the offsets O_{ijkl} correspond to the respective number of casualties recorded by the police.

The above formula for a four-dimensional Table corresponds to a saturated log-rate model, containing all possible four- and lower order effects. Moreover, it should be underlined that the models considered are hierarchical, meaning that whenever a higher order effect is included in the model, the lower order effects composed from variables in the higher effect are also included (Everitt, 1977, Kim et al., 1998).

The hypotheses of the analysis are those of mutual independence, which specifies that there are no associations of any kind between the four variables, or in other words that there are no first-order interactions between any pair of variables and no conjoint three- or four-variable interaction:

$$H_0: u = u_1= u_2= u_3 = u_4 = u_{12} = u_{13} = u_{14} = u_{23} = u_{24} = u_{34} = u_{123} = u_{124} = u_{134} = u_{234} = u_{1234} = 0$$

Main effect parameters are measured as deviations of row, column or layer means of log-rates from the overall mean. Each of the u parameters represents a deviation from the grand mean due to that effect (Hays, 1981). For example, $u_{2(j)}$ are road user type effects on the conversion factors, with a separate parameter estimate for each road user group. The term $u_{23(jk)}$ represents the interaction between road user type and police score, with a separate parameter estimate for each pair of sub-categories. Estimates of the parameters in the fitted model are obtained as functions of the logarithms of cell frequencies and the form of such estimates is very similar to those used for the parameters in analysis of variance models. It should be noted though that no dependent variable in the usual sense is designated in a log-rate model, as all variables are "factors" classifying the observations (conversion factors) according to a group in which they belong; there is no "response" variable classifying according to a description of what happens during or after an experiment (Everitt, 1977).

Appendix 3.3 Parameter estimates of the saturated log-rate model

Parameter	Estimate (B)	Sig.		ExpB
Constant	-1.347	0.000	*	
[Country = 1]	-0.655	0.016	*	0.519
[Country = 2]	1.405	0.000	*	4.075
[Country = 3]	-0.196	0.000	*	0.822
[Country = 4]	-0.005	0.694		0.995
[Country = 5]	0.979	0.000	*	2.661
[Country = 6]	0.613	0.000	*	1.846
[Country = 7]	0.000	.		1.000
[MAIS = 0]	1.546	0.000	*	4.692
[MAIS = 1]	0.000	.		1.000
[RoadUser = 1]	0.000	1.000		1.000
[RoadUser = 2]	0.000	1.000		1.000
[RoadUser = 3]	0.000	1.000		1.000
[RoadUser = 4]	0.000	1.000		0.222
[RoadUser = 5]	0.000	.		1.000
[PoliceSev = 0]	-2.562	0.000	*	0.077
[PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0]	0.325	0.238		1.384
[Country = 1] * [MAIS = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0]	-1.188	0.000	*	0.305
[Country = 2] * [MAIS = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0]	0.211	0.000	*	1.234
[Country = 3] * [MAIS = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0]	0.288	0.000	*	1.334
[Country = 4] * [MAIS = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 0]	1.261	0.000	*	3.527
[Country = 5] * [MAIS = 1]	0.000	.		1.000
[Country = 6] * [MAIS = 0]	-0.986	0.000	*	0.373
[Country = 6] * [MAIS = 1]	0.000	.		1.000
[Country = 7] * [MAIS = 0]	0.000	.		1.000
[Country = 7] * [MAIS = 1]	0.000	.		1.000
[Country = 1] * [RoadUser = 1]	-0.192	0.508		0.825
[Country = 1] * [RoadUser = 2]	-0.378	0.291		0.685
[Country = 1] * [RoadUser = 3]	1.310	0.000	*	3.706
[Country = 1] * [RoadUser = 4]	0.954	0.001	*	2.596
[Country = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 2] * [RoadUser = 1]	-0.727	0.000	*	0.483
[Country = 2] * [RoadUser = 2]	-0.246	0.000	*	0.782
[Country = 2] * [RoadUser = 3]	0.622	0.000	*	1.862
[Country = 2] * [RoadUser = 4]	-0.625	0.000	*	0.535
[Country = 2] * [RoadUser = 5]	0.000	.		1.000
[Country = 3] * [RoadUser = 1]	0.013	0.503		1.013
[Country = 3] * [RoadUser = 2]	0.562	0.000	*	1.754

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

Parameter	Estimate (B)	Sig.		ExpB
[Country = 3] * [RoadUser = 3]	1.224	0.000	*	3.401
[Country = 3] * [RoadUser = 4]	0.521	0.000	*	1.684
[Country = 3] * [RoadUser = 5]	0.000	.		1.000
[Country = 4] * [RoadUser = 1]	-0.548	0.000	*	0.578
[Country = 4] * [RoadUser = 2]	0.037	0.034	*	1.037
[Country = 4] * [RoadUser = 3]	0.111	0.000	*	1.118
[Country = 4] * [RoadUser = 4]	-0.132	0.000	*	0.876
[Country = 4] * [RoadUser = 5]	0.000	.		1.000
[Country = 5] * [RoadUser = 1]	-0.186	0.000	*	0.830
[Country = 5] * [RoadUser = 2]	-0.013	0.701		0.987
[Country = 5] * [RoadUser = 3]	-1.256	0.001	*	0.285
[Country = 5] * [RoadUser = 4]	-0.434	0.000	*	0.648
[Country = 5] * [RoadUser = 5]	0.000	.		1.000
[Country = 6] * [RoadUser = 1]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 2]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 3]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 4]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 5]	0.000	.		1.000
[Country = 7] * [RoadUser = 1]	0.000	.		1.000
[Country = 7] * [RoadUser = 2]	0.000	.		1.000
[Country = 7] * [RoadUser = 3]	0.000	.		1.000
[Country = 7] * [RoadUser = 4]	0.000	.		1.000
[Country = 7] * [RoadUser = 5]	0.000	.		1.000
[Country = 1] * [PoliceSev = 0]	-0.032	0.929		0.969
[Country = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [PoliceSev = 0]	-0.343	0.000	*	0.710
[Country = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [PoliceSev = 0]	-0.836	0.000	*	0.434
[Country = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [PoliceSev = 0]	-1.203	0.000	*	0.300
[Country = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [PoliceSev = 0]	0.511	0.000	*	1.667
[Country = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [PoliceSev = 0]	0.079	0.407		1.082
[Country = 6] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 1]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 2]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 3]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 4]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[MAIS = 0] * [PoliceSev = 0]	2.422	0.000	*	11.264

Project co-financed by the European Commission, Directorate-General Transport and Energy

Parameter	Estimate (B)	Sig.		ExpB
[MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[RoadUser = 1] * [PoliceSev = 0]	0.000	1.000		1.000
[RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[RoadUser = 2] * [PoliceSev = 0]	0.000	1.000		1.000
[RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[RoadUser = 3] * [PoliceSev = 0]	0.000	1.000		1.000
[RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[RoadUser = 4] * [PoliceSev = 0]	0.000	1.000		1.000
[RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 1]	0.295	0.314		1.344
[Country = 1] * [MAIS = 0] * [RoadUser = 2]	0.537	0.137		1.711
[Country = 1] * [MAIS = 0] * [RoadUser = 3]	-1.070	0.000	*	0.343
[Country = 1] * [MAIS = 0] * [RoadUser = 4]	-0.773	0.006	*	0.462
[Country = 1] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 1]	0.591	0.000	*	1.805
[Country = 2] * [MAIS = 0] * [RoadUser = 2]	0.127	0.000	*	1.136
[Country = 2] * [MAIS = 0] * [RoadUser = 3]	0.509	0.000	*	1.663
[Country = 2] * [MAIS = 0] * [RoadUser = 4]	0.217	0.000	*	1.243
[Country = 2] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 1]	-0.154	0.000	*	0.857
[Country = 3] * [MAIS = 0] * [RoadUser = 2]	-0.584	0.000	*	0.558
[Country = 3] * [MAIS = 0] * [RoadUser = 3]	-0.795	0.000	*	0.452
[Country = 3] * [MAIS = 0] * [RoadUser = 4]	-0.530	0.000	*	0.588
[Country = 3] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 1]	0.203	0.000	*	1.225
[Country = 4] * [MAIS = 0] * [RoadUser = 2]	-0.224	0.000	*	0.799
[Country = 4] * [MAIS = 0] * [RoadUser = 3]	0.337	0.000	*	1.401
[Country = 4] * [MAIS = 0] * [RoadUser = 4]	-0.300	0.000	*	0.741
[Country = 4] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000

Parameter	Estimate (B)	Sig.		ExpB
[Country = 4] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 1]	-0.847	0.000	*	0.429
[Country = 5] * [MAIS = 0] * [RoadUser = 2]	-0.495	0.000	*	0.610
[Country = 5] * [MAIS = 0] * [RoadUser = 3]	0.833	0.034	*	2.301
[Country = 5] * [MAIS = 0] * [RoadUser = 4]	-1.093	0.000	*	0.335
[Country = 5] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 1]	0.000	1.000		1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 2]	0.000	1.000		1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 3]	0.000	1.000		1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 4]	0.000	1.000		1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 1]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 2]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 3]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 4]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 5]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 1]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 2]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 3]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 4]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 5]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [PoliceSev = 0]	0.304	0.392		1.356
[Country = 1] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [PoliceSev = 0]	1.050	0.000	*	2.857
[Country = 2] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [PoliceSev = 0]	0.778	0.001	*	2.176
[Country = 3] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [PoliceSev = 0]	0.921	0.000	*	2.511

Parameter	Estimate (B)	Sig.		ExpB
[Country = 4] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 0] * [PoliceSev = 0]	-0.073	0.578		0.929
[Country = 5] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 5] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [MAIS = 0] * [PoliceSev = 0]	0.475	0.000	*	1.608
[Country = 6] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 6] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [MAIS = 0] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [MAIS = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.282	0.606		1.326
[Country = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.775	0.402		2.170
[Country = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.046	0.912		1.048
[Country = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.389	0.000	*	1.475
[Country = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [RoadUser = 1] * [PoliceSev = 0]	0.039	0.749		1.039
[Country = 2] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [RoadUser = 2] * [PoliceSev = 0]	0.996	0.000	*	2.707
[Country = 2] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [RoadUser = 3] * [PoliceSev = 0]	0.934	0.000	*	2.545
[Country = 2] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [RoadUser = 4] * [PoliceSev = 0]	1.143	0.000	*	3.136
[Country = 2] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [RoadUser = 1] * [PoliceSev = 0]	-0.116	0.643		0.891
[Country = 3] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [RoadUser = 2] * [PoliceSev = 0]	0.271	0.273		1.311
[Country = 3] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [RoadUser = 3] * [PoliceSev = 0]	0.522	0.034	*	1.685
[Country = 3] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [RoadUser = 4] * [PoliceSev = 0]	0.643	0.010	*	1.902
[Country = 3] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [RoadUser = 1] * [PoliceSev = 0]	0.121	0.530		1.129
[Country = 4] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [RoadUser = 2] * [PoliceSev = 0]	0.618	0.003	*	1.855

Project co-financed by the European Commission, Directorate-General Transport and Energy

Parameter	Estimate (B)	Sig.		ExpB
[Country = 4] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [RoadUser = 3] * [PoliceSev = 0]	0.298	0.177		1.347
[Country = 4] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [RoadUser = 4] * [PoliceSev = 0]	0.441	0.022	*	1.555
[Country = 4] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [RoadUser = 1] * [PoliceSev = 0]	0.810	0.000	*	2.247
[Country = 5] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [RoadUser = 2] * [PoliceSev = 0]	0.791	0.000	*	2.206
[Country = 5] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [RoadUser = 3] * [PoliceSev = 0]	3.830	0.000	*	46.046
[Country = 5] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [RoadUser = 4] * [PoliceSev = 0]	0.875	0.000	*	2.398
[Country = 5] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 5] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [RoadUser = 1] * [PoliceSev = 0]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [RoadUser = 2] * [PoliceSev = 0]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [RoadUser = 3] * [PoliceSev = 0]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [RoadUser = 4] * [PoliceSev = 0]	0.000	1.000		1.000
[Country = 6] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 6] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 6] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 7] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 7] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	0.000	1.000		1.000
[MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.		1.000

Parameter	Estimate (B)	Sig.		ExpB
[MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	-0.279	0.608		0.756
[Country = 1] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	-0.779	0.399		0.459
[Country = 1] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	0.967	0.021	*	2.630
[Country = 1] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 1] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	-0.019	0.878		0.982
[Country = 2] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	-0.721	0.000	*	0.486
[Country = 2] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	-0.706	0.000	*	0.493
[Country = 2] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	-1.077	0.000	*	0.341
[Country = 2] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 2] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000

Parameter	Estimate (B)	Sig.		ExpB
[Country = 2] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	0.259	0.300		1.296
[Country = 3] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	-0.226	0.360		0.798
[Country = 3] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	-0.875	0.000	*	0.417
[Country = 3] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	-0.610	0.014	*	0.543
[Country = 3] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 3] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	0.192	0.319		1.212
[Country = 4] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	-0.367	0.078		0.693
[Country = 4] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	-0.593	0.007	*	0.553
[Country = 4] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	-0.038	0.842		0.962
[Country = 4] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.		1.000
[Country = 4] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	-0.706	0.000	*	0.494
[Country = 5] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	-0.646	0.000	*	0.524
[Country = 5] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	-2.974	0.000	*	0.051
[Country = 5] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.		1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	-0.720	0.000	*	0.487

Parameter	Estimate (B)	Sig.	ExpB
[Country = 5] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.	1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.	1.000
[Country = 5] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.	1.000
[Country = 5] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	0.000	1.000	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	0.000	1.000	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	0.000	1.000	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	0.000	1.000	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.	1.000
[Country = 6] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.	1.000
[Country = 6] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 0] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 1] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 2] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 0]	0.000	.	1.000

Parameter	Estimate (B)	Sig.	ExpB
[Country = 7] * [MAIS = 1] * [RoadUser = 3] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 4] * [PoliceSev = 1]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 0]	0.000	.	1.000
[Country = 7] * [MAIS = 1] * [RoadUser = 5] * [PoliceSev = 1]	0.000	.	1.000

Appendix 4.1 Raw counts and percentages for CARE-FAID comparison

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
female	42	95	93	1155	127	802	351	1133	22	262	81	112	716	3559
male	116	241	239	3554	451	2505	1046	4492	96	758	227	333	2175	11883
unknown	0	0	1	0	101	0	6	0	10	0	0	0	118	0
Total	158	336	333	4709	679	3307	1403	5625	128	1020	308	445	3009	15442

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
female	26.6%	28.3%	27.9%	24.5%	18.7%	24.3%	25.0%	20.1%	17.2%	25.7%	26.3%	25.2%	23.8%	23.0%
male	73.4%	71.7%	71.8%	75.5%	66.4%	75.7%	74.6%	79.9%	75.0%	74.3%	73.7%	74.8%	72.3%	77.0%
unknown	0.0%	0.0%	0.3%	0.0%	14.9%	0.0%	0.4%	0.0%	7.8%	0.0%	0.0%	0.0%	3.9%	0.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

<u>Table 3</u>: Gender

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
00-09	10	2	17	62	8	68	30	52	7	33	10	12	82	229
10-19	20	54	39	533	55	534	146	468	12	144	39	48	311	1781
20-29	28	64	95	1199	79	756	368	1285	29	209	65	91	664	3604
30-39	21	34	56	680	50	521	220	1005	28	157	52	52	427	2449
40-49	20	43	49	605	34	410	160	645	11	117	43	54	317	1874
50-59	25	48	29	485	25	294	133	521	12	99	29	65	253	1512
60-69	14	40	13	309	14	210	113	474	5	95	28	47	187	1175
70-79	2	27	4	417	2	240	4	577	1	90	3	40	16	1391
80-89	3	23	1	311	2	232	11	326	0	73	3	33	20	998
90+	0	1	0	43	0	33	2	30	0	11	0	3	2	121
unknown	15	0	30	65	410	9	216	242	23	0	36	0	730	316
Total	158	336	333	4709	679	3307	1403	5625	128	1028	308	445	3009	15450

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
00-09	6.3%	0.6%	5.1%	1.3%	1.2%	2.1%	2.1%	0.9%	5.5%	3.2%	3.2%	2.7%	2.7%	1.5%
10-19	12.7%	16.1%	11.7%	11.3%	8.1%	16.1%	10.4%	8.3%	9.4%	14.0%	12.7%	10.8%	10.3%	11.5%
20-29	17.7%	19.0%	28.5%	25.5%	11.6%	22.9%	26.2%	22.8%	22.7%	20.3%	21.1%	20.4%	22.1%	23.3%
30-39	13.3%	10.1%	16.8%	14.4%	7.4%	15.8%	15.7%	17.9%	21.9%	15.3%	16.9%	11.7%	14.2%	15.9%
40-49	12.7%	12.8%	14.7%	12.8%	5.0%	12.4%	11.4%	11.5%	8.6%	11.4%	14.0%	12.1%	10.5%	12.1%
50-59	15.8%	14.3%	8.7%	10.3%	3.7%	8.9%	9.5%	9.3%	9.4%	9.6%	9.4%	14.6%	8.4%	9.8%
60-69	8.9%	11.9%	3.9%	6.6%	2.1%	6.4%	8.1%	8.4%	3.9%	9.2%	9.1%	10.6%	6.2%	7.6%
70-79	1.3%	8.0%	1.2%	8.9%	0.3%	7.3%	0.3%	10.3%	0.8%	8.8%	1.0%	9.0%	0.5%	9.0%
80-89	1.9%	6.8%	0.3%	6.6%	0.3%	7.0%	0.8%	5.8%	0.0%	7.1%	1.0%	7.4%	0.7%	6.5%
90+	0.0%	0.3%	0.0%	0.9%	0.0%	1.0%	0.1%	0.5%	0.0%	1.1%	0.0%	0.7%	0.1%	0.8%
unknown	9.5%	0.0%	9.0%	1.4%	60.4%	0.3%	15.4%	4.3%	18.0%	0.0%	11.7%	0.0%	24.3%	2.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

<u>Table 4</u>: Age groups

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
driver	95	233	218	3326	437	1976	818	3739	91	767	187	305	1846	10346
passenger	53	54	99	848	193	628	447	1164	32	163	98	81	922	2938
pedestrian	10	49	16	535	49	703	138	710	5	97	23	55	241	2149
unknown	0	0	0	0	0	0	0	12	0	1	0	4	0	17
Total	158	336	333	4709	679	3307	1403	5625	128	1028	308	445	3009	15450

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
driver	60.1%	69.3%	65.5%	70.6%	64.4%	59.8%	58.3%	66.5%	71.1%	74.6%	60.7%	68.5%	61.3%	67.0%
passenger	33.5%	16.1%	29.7%	18.0%	28.4%	19.0%	31.9%	20.7%	25.0%	15.9%	31.8%	18.2%	30.6%	19.0%
pedestrian	6.3%	14.6%	4.8%	11.4%	7.2%	21.3%	9.8%	12.6%	3.9%	9.4%	7.5%	12.4%	8.0%	13.9%
unknown	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.2%	0.0%	0.1%	0.0%	0.9%	0.0%	0.1%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 5: Person class

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
pedestrian	10	49	16	530	49	701	138	683	5	97	23	55	241	2115
pedal cycle	7	29	8	180	17	150	56	290	13	188	8	26	109	863
moped/motor cycle	12	39	37	1098	73	612	135	1413	8	189	22	70	287	3421
car + taxi + van	101	203	245	2578	480	1679	963	2690	67	483	221	261	2077	7894
heavy goods vehicle	14	10	23	205	47	97	72	104	16	63	24	16	196	495
bus or coach	3	2	1	8	10	28	8	19	12	0	6	10	40	67
agricultural tractor	0	2	0	12	0	0	0	20	0	5	0	3	0	42
other/unknown	11	2	3	32	3	30	31	164	7	3	4	5	59	236
Total	158	336	333	4643	679	3297	1403	5383	128	1028	308	446	3009	15133

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
pedestrian	6.3%	14.6%	4.8%	11.4%	7.2%	21.3%	9.8%	12.7%	3.9%	9.4%	7.5%	12.3%	8.0%	14.0%
pedal cycle	4.4%	8.6%	2.4%	3.9%	2.5%	4.5%	4.0%	5.4%	10.2%	18.3%	2.6%	5.8%	3.6%	5.7%
moped/motor cycle	7.6%	11.6%	11.1%	23.6%	10.8%	18.6%	9.6%	26.2%	6.3%	18.4%	7.1%	15.7%	9.5%	22.6%
car + taxi + van	63.9%	60.4%	73.6%	55.5%	70.7%	50.9%	68.6%	50.0%	52.3%	47.0%	71.8%	58.5%	69.0%	52.2%
heavy goods vehicle	8.9%	3.0%	6.9%	4.4%	6.9%	2.9%	5.1%	1.9%	12.5%	6.1%	7.8%	3.6%	6.5%	3.3%
bus or coach	1.9%	0.6%	0.3%	0.2%	1.5%	0.8%	0.6%	0.4%	9.4%	0.0%	1.9%	2.2%	1.3%	0.4%
agricultural tractor	0.0%	0.6%	0.0%	0.3%	0.0%	0.0%	0.0%	0.4%	0.0%	0.5%	0.0%	0.7%	0.0%	0.3%
other/unknown	7.0%	0.6%	0.9%	0.7%	0.4%	0.9%	2.2%	3.0%	5.5%	0.3%	1.3%	1.1%	2.0%	1.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 6: Vehicle group

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
crash helmet not used	2	0	3	98	1	0	14	147	4	8	1	0	25	253
crash helmet use unknown	0	0	2	60	10	0	62		0	146	1	0	75	206
crash helmet used	10	0	32	979	62	0	59	779	4	35	20	0	187	1793
seat belt not used	38	0	32	547	94	23	104	301	11	26	75	0	354	897
seat belt use unknown	11	0	47	437	181	0	900		58	472	56	0	1253	909
seat belt used	69	0	191	1801	259	37	60	469	26	37	122	0	727	2344
unknown	28	287	26	209	72	2497	204	3154	25		33	390	388	6537
Total	158	287	333	4131	679	2557	1403	4850	128	724	308	390	3009	12939

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
crash helmet not used	1.3%	0.0%	0.9%	2.4%	0.1%	0.0%	1.0%	3.0%	3.1%	1.1%	0.3%	0.0%	0.8%	2.0%
crash helmet use unknown	0.0%	0.0%	0.6%	1.5%	1.5%	0.0%	4.4%	0.0%	0.0%	20.2%	0.3%	0.0%	2.5%	1.6%
crash helmet used	6.3%	0.0%	9.6%	23.7%	9.1%	0.0%	4.2%	16.1%	3.1%	4.8%	6.5%	0.0%	6.2%	13.9%
seat belt not used	24.1%	0.0%	9.6%	13.2%	13.8%	0.9%	7.4%	6.2%	8.6%	3.6%	24.4%	0.0%	11.8%	6.9%
seat belt use unknown	7.0%	0.0%	14.1%	10.6%	26.7%	0.0%	64.1%	0.0%	45.3%	65.2%	18.2%	0.0%	41.6%	7.0%
seat belt used	43.7%	0.0%	57.4%	43.6%	38.1%	1.4%	4.3%	9.7%	20.3%	5.1%	39.6%	0.0%	24.2%	18.1%
unknown	17.7%	100.0%	7.8%	5.1%	10.6%	97.7%	14.5%	65.0%	19.5%	0.0%	10.7%	100.0%	12.9%	50.5%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 7: Security equipment

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
inside urban area	31	101	83	1664	177	1302	570	900	31	20	41	110	933	4097
outside urban area	97	278	245	3654	423	2034	684	3841	84	30	260	322	1793	10159
Total	128	379	328	5318	600	3336	1254	4741	115	50	301	432	2726	14256

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
inside urban area	24.2%	26.6%	25.3%	31.3%	29.5%	39.0%	45.5%	19.0%	27.0%	40.0%	13.6%	25.5%	34.2%	28.7%
outside urban area	75.8%	73.4%	74.7%	68.7%	70.5%	61.0%	54.5%	81.0%	73.0%	60.0%	86.4%	74.5%	65.8%	71.3%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 8: Area Type

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
junction	65	73	85	664	204	1162	559	1641	49	321	75	98	1037	3959
no junction	93	300	246	4654	475	2184	842	3884	79	704	233	10	1968	11736
Total	158	373	331	5318	679	3346	1401	5525	128	1025	308	108	3005	15695

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
junction	41.1%	19.6%	25.7%	12.5%	30.0%	34.7%	39.9%	29.7%	38.3%	31.3%	24.4%	90.7%	34.5%	25.2%
no junction	58.9%	80.4%	74.3%	87.5%	70.0%	65.3%	60.1%	70.3%	61.7%	68.7%	75.6%	9.3%	65.5%	74.8%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 9: Junction

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
motorway	3	10	19	323	25	203	131	648	24	151	27	20	229	1355
no motorway	155	268	314	3331	654	1681	1272	2667	104	531	281	302	2801	8780
Total	158	278	333	3654	679	1884	1403	3315	128	682	308	322	3030	10135

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
motorway	1.9%	3.6%	5.7%	8.8%	3.7%	10.8%	9.3%	19.5%	18.8%	22.1%	8.8%	6.2%	7.6%	13.4%
no motorway	98.1%	96.4%	94.3%	91.2%	96.3%	89.2%	90.7%	80.5%	81.3%	77.9%	91.2%	93.8%	92.4%	86.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 10: Motorway

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
darkness, lights unlit or no lights	11	57	70	1246	178	683	327	0	22	141	89	95	697	2222
darkness, street lights lit	19	54	29	514	127	727	215	0	36	180	30	39	456	1514
daylight or twilight	124	225	234	2949	370	1876	851	0	70	691	188	298	1837	6039
unknown	4		0		4	21	10	5625	0	16	1	13	19	5675
Total	158	336	333	4709	679	3307	1403	5625	128	1028	308	445	3009	15450

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
darkness, lights unlit or no lights	7.0%	17.0%	21.0%	26.5%	26.2%	20.7%	23.3%	0.0%	17.2%	13.7%	28.9%	21.3%	23.2%	14.4%
darkness, street lights lit	12.0%	16.1%	8.7%	10.9%	18.7%	22.0%	15.3%	0.0%	28.1%	17.5%	9.7%	8.8%	15.2%	9.8%
daylight or twilight	78.5%	67.0%	70.3%	62.6%	54.5%	56.7%	60.7%	0.0%	54.7%	67.2%	61.0%	67.0%	61.1%	39.1%
unknown	2.5%	0.0%	0.0%	0.0%	0.6%	0.6%	0.7%	100.0%	0.0%	1.6%	0.3%	2.9%	0.6%	36.7%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 11: Lightning conditions

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
dry	85	304	242	4515	458	2795	1147	4067	100	909	194	338	2226	12928
fog or mist	0	1	0	102	0	28	0	56	0	16	0	16	0	219
other	0	0	0	85	0	60	0	804	5	0	0	0	5	949
rain	38	29	81	530	213	362	233	663	23	75	71	44	659	1703
snow, sleet, hail	35	30	7	64	0	23	4	23	0	8	41	15	87	163
strong wind	0	0	0	22	0	0	0	12	0	3	0	0	0	37
Total	158	364	330	5318	671	3268	1384	5625	128	1011	306	413	2977	15999

	FI		FR		UK		IT		NL		SE		Total	
	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE	WP5	CARE
dry	53.8%	83.5%	73.3%	84.9%	68.3%	85.5%	82.9%	72.3%	78.1%	89.9%	63.4%	81.8%	74.8%	80.8%
fog or mist	0.0%	0.3%	0.0%	1.9%	0.0%	0.9%	0.0%	1.0%	0.0%	1.6%	0.0%	3.9%	0.0%	1.4%
other	0.0%	0.0%	0.0%	1.6%	0.0%	1.8%	0.0%	14.3%	3.9%	0.0%	0.0%	0.0%	0.2%	5.9%
rain	24.1%	8.0%	24.5%	10.0%	31.7%	11.1%	16.8%	11.8%	18.0%	7.4%	23.2%	10.7%	22.1%	10.6%
snow, sleet, hail	22.2%	8.2%	2.1%	1.2%	0.0%	0.7%	0.3%	0.4%	0.0%	0.8%	13.4%	3.6%	2.9%	1.0%
strong wind	0.0%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	0.2%	0.0%	0.3%	0.0%	0.0%	0.0%	0.2%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 12: Weather conditions

Appendix 5.1 Aggregation of SARTRE variables

Accident causes

Sartre respondents rated several possible accident causes. Examining this list of causes, it appears that some of them are dependent upon the behaviour of the driver (e.g.: “Driving when tired”; “Drinking and driving”); whereas others are more of an external nature (e.g.: “Poorly maintained roads”; “Traffic congestion”). The rationale behind the choice and calculation of scores for this variable was that drivers who consider external factors as a relatively more important cause of accidents than their own driving behaviour might also be more reckless drivers, and thus be at higher risk to have accidents. Therefore, the variable <CauseExternalInternal> was generated to reflect the mean of all scores on the internal factors relative to the mean of all scores on the external factors.

Comparison to other drivers

Two questions of the Sartre questionnaire (questions 6 and 8) explicitly asked the drivers to rate their own driving behaviour as compared to the one of others. The first question assessed the dangerous nature of one’s way of driving as compared to others’ the other one focused on speeding. The aggregated variables <MoreDangerous> and <FasterThanOthers> corresponded to the percentage of respondents in each group who indicated that they behaved more dangerously or drove faster than others respectively.

Questions	***Answer range***	***Aggregated variable definition***	***Aggregated variable name***
ACCIDENT CAUSES			
4/ "How often do you think each of the following factors are the cause of road accidents?" 4a Driving when tired 4b Drinking and driving 4c Following too closely vehicle in front 4d Driving too fast 4e Taking medicines and driving 4f Taking drugs and driving" 4g Poorly maintained roads 4h Using a mobile phone (hand held) and driving 4i Using a mobile phone (hand free) and driving 4j Traffic congestion 4k Bad weather conditions 4l Poor brakes 4m Bald tyres 4n Faulty lights 4o Defective steering	"Never" (1) "Rarely" (2) "Sometimes" (3) "Often" (4) "Very often" (5) "Always" (6)	Ratio: Mean score on all external causes (4 g; j;k;l;m;n;o) to mean score on all internal factors (4a; b; c; d; e; f; h; I)	<CauseExternalInternal>
COMPARISON TO OTHER DRIVERS			
6/ "Compared to other drivers, do you think your driving is … dangerous?"	"Much more" (1) "A bit more" (2) "About the same" (3) "A bit less" (4) "A lot less" (5)	- Percentage of (1) and (2) in each group.	<MoreDangerous>
8/ "Compared to other drivers, do you generally drive … than average speed?"	"Much faster" (1) "A little faster" (2) "About average" (3) "A little slower" (4) "Much slower" (5)	- Percentage of (1) and (2) in each group.	<FasterThanOthers>

Speeding

Sartre participants indicated the extent to which they tend to exceed the speed limit on various types of roads. The aggregated variable <Speeding> computed corresponded to the average reported speeding rate for the various types of roads.

Perceived likelihood to be checked

Three aggregated variables were calculated on the basis of the original Sartre items measuring the perceived likelihood to be checked for speed and alcohol respectively. These two variables, labelled <LikelyCheckSpeed> and <LikelyCheckAlcohol> are defined as the percentage of respondents in each age group who answered that they thought on a typical journey they would be checked sometimes, often, or always.

Fined or Punished

Two questions from the Sartre study assessed the number of times the respondents had been fined or punished for speed, seatbelt, and drink-driving infringements. These questions have been used to compute three aggregated indicators of the extent to which the different groups had been the object of enforcement related to these infringements (<FinedPunishedSpeed>, <FinedPunishedSeatBelt>, and <FinedPunishedAlcohol>). These aggregated variables were computed as the percentage of people in each group reporting having been "only fined" or "fined and /or other penalty".

Questions	***Answer range***	***Aggregated variable name***	***Aggregated variable definition***
SPEEDING			
9/ "*In general, how often do you drive faster than the speed limit on the following types of road?*" 9a Motorways 9b Main roads between towns 9c Country roads 9d Built-up areas	"Never" (1) "Rarely" (2) "Sometimes" (3) "Often" (4) "Very often" (5) "Always" (6)	<Speeding>	Average score on 9a to 9d
PERCEIVED LIKELIHOOD TO BE CHECKED			
11/ "On a typical journey, how likely is it that your speed will be checked for?"	"Never" (1) "Rarely" (2) "Sometimes" (3) "Often" (4) "Very often" (5) "Always" (6)	<LikelyCheckSpeed>	Percentage equal or smaller (3) in each group.
25/ "On a typical journey, how likely is it that you will be checked for alcohol?"		<LikelyCheckAlcohol>	Percentage equal or smaller (3) in each group.
FINED OR PUNISHED			
12/ "In the last 3 years, have you been fined, or punished in any other way, for breaking the speed limit?"	"No" (1) "Yes, only fined" (2) "Yes, fined and/or other penalty" (3)	<FinedPunishedSpeed>	Percentage of (2) or (3) in each group
18/ "In the last 3 years, have you been fined, or punished in any other way, for not wearing a seat belt?"		<FinedPunishedSeatbelt>	Percentage of (2) or (3) in each group
24/ "In the last 3 years, have you been fined, or punished in any other way, for drink driving?"		<FinedPunishedAlcohol>	Percentage of (2) or (3) in each group

Dangerous Behaviour

This variable was generated on the basis of a series of questions measuring the perceived frequency of various dangerous driving behaviours The aggregated variable <OftenDangerousBehavour> was defined as the percentage of respondents in each group that had reported displaying any of these behaviour "often", "very often" , or "always".

No Seat Belt

The variable <NoSeatBelt> was generated from a series of questions assessing the frequency to which respondents wore their seatbelt on various types of roads. Scores to the items were reversed, so that high scores reflected infrequent seatbelt wear. The aggregated variable was computed as the average score on the set of questions (scores on the individual items had previously been reverse-scored, so that a high score express a low frequency of seatbelt wear).

DANGEROUS BEHAVIOUR			
13/ "How often do you….?"			
13a "Follow the vehicle in front too closely" 13b "Give way to a pedestrian at pedestrian crossings" *(Reverse-scored)* 13c "Drive through a traffic light that is on amber" 13d "Overtake when you think you can just make it" 13e "Signal other drivers to warn them of a police speed trap ahead"	"Never" (1) "Rarely" (2) "Sometimes" (3) "Often" (4) "Very often" (5) "Always" (6)	- Selection of maximum score on any of the question. - Percentage of (4), (5), (6) in maximum scores in each group	<OftenDangerousBehaviour>
SEATBELT			
"When driving this car, how often do you wear the seat belt when making a journey...?"			
15a On motorways 15b On main roads between towns 15c On country roads 15d On built-up areas	"Never" (1) "Rarely" (2) "Sometimes" (3) "Often" (4) "Very often" (5) "Always" (6)	Average reversed scores on questions 15a to 15e	<No seatbelt>

Drink Alcohol

Sartre respondents reported the number of days in a week they usually drank alcoholic beverages (question 19) as well as the number of days in a week they usually drove after having drunk (even a small amount of alcohol, question 20). The scores to each of the questions were reversed first, so that a high score indicate frequent alcohol consumption and drink driving in a week. The aggregated variable <DrinkAlcohol> was then generated so as to take answers to both questions into account. Indeed, answers to question 19 should form an upper boundary for answers to question 20 (if you never drink alcohol, you cannot drink drive). However, in a number of cases the scores to question 20 were – sometimes even substantially -- higher than those to question 19[5]. Therefore, the aggregated variable was defined as the mean of the maximum of the two scores.

Drink Drive

Another question from the Sartre questionnaire focused on drink driving, this time explicitly referring to infringements of the legal limit.("Over the last week,

[5] Possibly this might be due to different interpretations of "drinking alcoholic beverages" and "drinking even a small amount of alcohol".

Transport
Project co-financed by the European Commission, Directorate-General Transport and Energy

how many days did you drive, when you may have been over the legal limit for drinking and driving?”). The scores were reversed so that high scores indicate large infringements. The aggregated variable <DrinkDrive> corresponded to the mean score to this question within each group.

Checked for Alcohol
The variable <CheckedAlcohol> was generated on the basis of a variable assessing the number of times respondents had been checked for alcohol infringements in the past. The aggregated variable indicates the percentage of respondents in each group that have not been checked at all.

DRINK ALCOHOL			
19/ “In general, how many days per week do you drink alcoholic beverages?” *20/ “How many days per week do you drive after drinking even a small amount of alcohol?”*	“Most days” (1) “5 or 6”(2) “3 or 4”(3) “1or 2”(4) “Less than 1”(5) “Never”(6)	- Reverse scored - Selection of the maximum score on any of the two questions - Mean of the maximum scores for each group	<Drink alcohol>
DRINK DRIVE			
21/ “Over the last week, how many days did you drive, when you may have been over the legal limit for drinking and driving?”	“Most days” (1) “5 or 6”(2) “3 or 4”(3) “1or 2”(4) “Less than 1”(5) “Never”(6)	Reverse scored, Mean reverse score	<DrinkDrive>
CHECKED FOR ALCOHOL			
23/ “In the past 3 years, how many times were you checked for alcohol?”	“Never” (1) “Only once” (2) “More than once” (3)	Percentage of (1)	<CheckedAlcohol>

Driving Attitude
A series of questions assessed general driving attitudes. Of these items, two were considered to represent a rather impatient (and possibly not very safe) attitude towards driving. *Impatient drivers* were defined as those having a score equal to or lower than four on the sum of these two items. Two other items were interpreted as indicating the perception of a risk rather than of an enjoyment in driving. *Drivers who are not worried* were defined as those having a total score of 5 or more on the sum of these two questions. The aggregated variables <ImpatientDriver> and <NotWorried> respectively indicate the percentage of Impatient drivers and of Drivers who are not worried in each group.

Children
On the basis of the assumption that people used to carry children in their car might be characterized by different attitudes and driving behaviours than others. In order to examine this possibility, an aggregated indicator was computed on the basis of a question originally measuring the extent to which people use

appropriate child restraint systems was used[6]. This question indeed included the response option “I Never carry child(ren)”. Therefore, the variable <NeverTransportChild> was created to indicate the percentage of people who never carry children.

Aggressive Behaviour

The aggregated variable <Aggressive> gives the percentage of respondents in each group who admitted having directed aggressive behaviour towards other road users in the last 10 months.

Phone Calls

The average score to a question measuring the frequency with which participants used their phone while driving made up the aggregated variable <PhoneCalls>.

[6] It was assumed that the use of child restraint systems does not consist of a relevant indicator in the context of the present analysis, which consists of predicting the number of accidents which is based on the drivers’ data.

DANGEROUS BEHAVIOUR			
29/ "How much do you agree?"			
29a "I sometimes get very annoyed with other drivers" 29b "I enjoy driving fast" 29c "I worry when members of my family are out driving" 29d "I think a car is just a means of transport"	"Very" (1) "Fairley" (2) "Not much" (3) "Not at all" (4)	- Mean of 29a and 29b. - Mean of reversed 29c and 29d	<Impatient driver> <Not worried>
CHILDREN			
15/ "When you carry a child in your car, how often do you make them ear seat belt or use appropriate restraint?"			
15a On motorways 15b On main roads between towns 15c On country roads 15d On built-up areas	"Always" (1) "Usually" (2) "Sometimes" (3) "Never" (4) "Never carry child(ren)" (5)	Percentage (1)	<NeverTransportChildren>
AGGRESSIVE BEHAVIOUR			
35/ "In the last 12 months have you had an experience of aggressive behaviour on the road?"			
35a Directed towards you by another road user 35b By yourself towards another road user	"Yes" (1) "No" (2)	Percentage (1) in 35a	<Aggressive>
PHONE CALLS			
47/ "*How many times on an average day do you make or answer a telephone call while driving?*"	"I drive for my profession" (1) "I need to drive during my work" (2) "I drive to and from work" (3) "None of these" (4)	Percentage (1) and (2)	<Aggressive>
47a "You make a call:" 47b "You anser a call:"	"_ _ times a day"	Sum of 47a and 47b Mean	<PhoneCalls>

Education

To investigate whether there is a relation between the level of education in a particular goup and the occurreneces of accidents, the variable <OnlyPrimaryEducation> gives the percentage of people who had no education or primary schooling only.

Urban

Age groups might show systematic differences as to where they live (e.g., older people more often in the country and younger people more often in the city). To capture these possible differences the variable <Urban> gives the percentage of respondents that lived in urban or suburban areas.

Driving and Profession

Whether people drive for their profession or not makes a huge difference with respect to exposure and therefore risk of having an accident as well as driving experience. The variable <DriveProfession> marks all those persons who drive as part of their work as opposed to those who do not.

Driving and Profession

Whether people drive for their profession or not makes a huge difference with respect to exposure and therefore risk of having an accident as well as driving experience. The variable <DriveProfession> marks all those persons who drive as part of their work as opposed to those who do not.

EDUCATION			
44/ "What level of education did you achieve?"	"Primary School" (1) "Secondary school (2)" "Further education (3)" "None" (4)	- percentage of (1) and (4)	OnlyPrimaryEducation
AREA OF LIVING			
45/ "How would you describe the area in which you live?"	Rural/village (1) Small town (2) Suburban/city outskirts (3) Urban/city/large town (4)	Percentage of (3) or (4)	<Urban>
DRIVING AND PROFESSION			
48/ "What applies most to you?"	"I drive for my profession" (1) "I need to drive during my work" (2) "I drive to and from work" (3) "None of these" (4)	Percentage (1) and (2)	<DriveProfession>

Project co-financed by the European Commission, Directorate-General Transport and Energy

Appendix 5.2 Results Principal Components Analysis (PCA)

The PCA was conducted on 10 countries: Austria, Denmark, Finland, Greece, Ireland, Netherland, Spain, Sweden, and Great Britain.

Rotated Component Matrix(a)

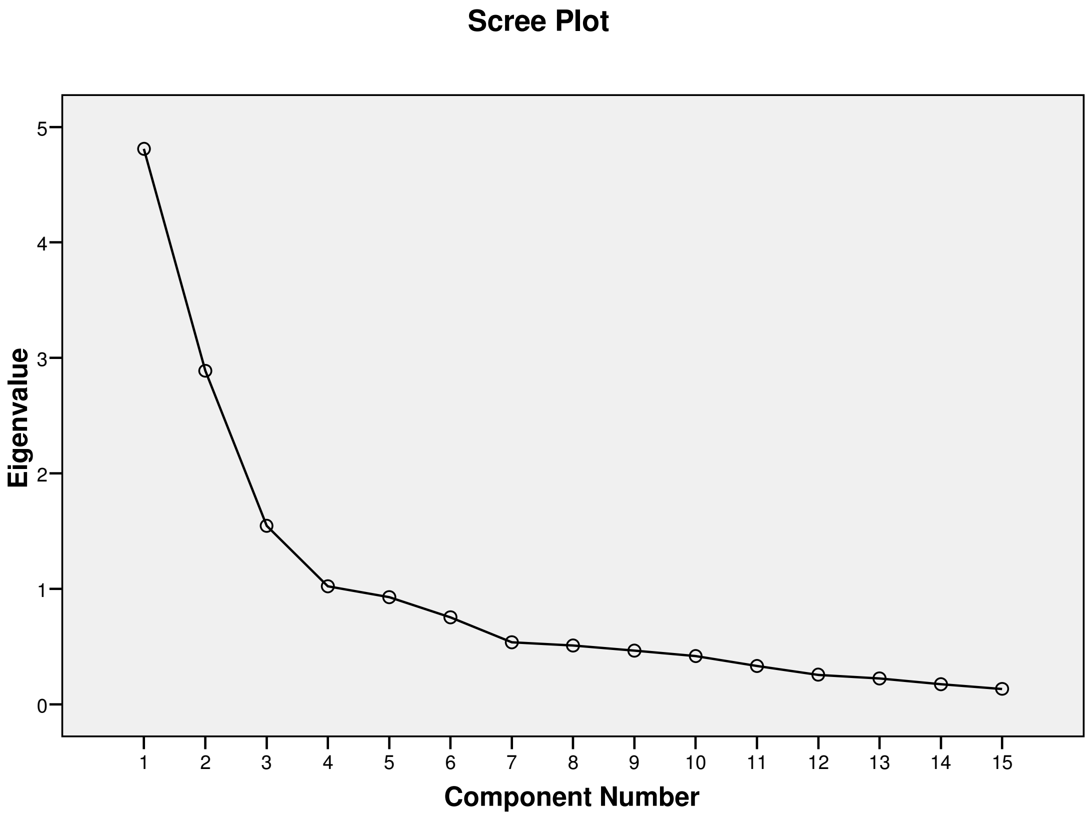
	Component		
	1	2	3
CauseExternalInternal		.745	
FasterThanOthers	.858		
Speeding	.817		
LikelyCheckSpeed			.759
FinedPunishedSpeed	.556		
SeatBelt		-.749	
NeverTransportChild		.638	
FinedPunishedSeatBelt	.495	.520	
DrinkDrive		.838	
CheckedAlcohol			.667
LikelyCheckedAlcohol			.883
ImpatientDriver	.819		
NotWorried	.643		
Aggressive	.712		
PhoneCalls	.682		

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 5 iterations.
Loadings <.35 suppressed.

Total Variance Explained

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.810	32.068	32.068	4.327	28.847	28.847
2	2.889	19.258	51.326	2.741	18.272	47.119
3	1.546	10.309	61.634	2.177	14.516	61.634

Extraction Method: Principal Component Analysis.


Scree Plot
Eigenvalue
0
1
2
3
4
5
Component Number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Appendix 5.3 Assumptions for modelling Accident Severity

Accident severity is defined as the logarithm of the ratio of fatalities (F) by accidents (A).

Accident severity = Log (F/ A) = Log (F) – Lob (A)

Variance Log(F)-Log(A) = Var(Log(F))+Var(Log(A))-2Cov(Log(F),Log(A))

Var(Log(F))= $Sum(f^2)/(Sum(f))^2$
(see Bijleveld, 2005)

$Sum(f^2)$ = n * g2;
with g2 being the mean of f^2.

Sum(f) = n * g

Var(Log(A))-2Cov(Log(F),Log(A)) ≅ - 1/n

Therefore:
Log(F)-Log(A) ≅ n * g2 / $(n * g)^2$ - 1/n ≅ (g2/g^2 - 1) / n

It thus seems possible to use normal linear regression with a weight 1/n under the assumption that g2/g^2 is approximately constant across observations. The condition applies when the number of killed people per accident is either 0 or 1. Consequently this approximation neglects the rare cases when there is more than one driver killed in the accidents.

Appendix 5.4 Detailed results for regression models

In the following, the detailed description and results of six regression analyses summarized in Table 5.5 are given. All models were estimated using the "R" software (package lme4). The Accident Severity models were estimated by maximising the restricted maximum likelihood (REML). For the log number of fatalities and log number of accidents a Laplacian approximation was used.

The formula used for all models was:

~ COMP1 +COMP2 +COMP3 + (1 +COMP1 +COMP2 +COMP3 | Country)

meaning that the components COMP1 to COMP3 were fixed factors. Moreover, there was a random factor Country, for which a random intercept was estimated and random slopes for all 3 components.

For each model the following information is given:

AIC	Akaaike Information Criterion
BIC	Bayesian Information Criterion
LogLik	Log Likelihood

Moreover, for each random effect the variance is estimated and for each fixed effect the coefficient, the standard errors (SE) and t-value.

As the restricted maximumlikelihood estimation is known to generate anti-conserative p-values, the significance of the fixed effects was established by MCMC (monte-carlo-markov-chain) estimation, using the MCMCsamp procedure in R. This procedure gives the following values:

Estimate:	The original estimation of for the coefficient
MCMCmean:	The mean coefficient according to the MCMC procedure. This is more reliable than the original coefficient.
HPD95 lower:	lower threshold of 95% confidence interval
HPD95upper:	Upper threshold of 95% confidence interval
pMCMC:	The significance according to the MCMC procedure (this is the value that was considered to determine a coefficients significance)
Pr(>/t/):	The significance according to the t-statistic.

For more explanation of these concepts, please refer to D7.4 (Dupont & Martensen, 2006).

Accident Severity: Men

```
Linear mixed-effects model fit by REML

    AIC     BIC logLik   Deviance
 -207.8 -164.8  117.9     -256.9

Random effects:
 Groups         Name         Variance
 Country       (Intercept)    0.04274017
               COMP1          0.00482651
               COMP2          0.00066499
               COMP3          0.00417345
 Residual                     0.00794903

Number of observations: 160, groups: Country, 10

Fixed effects:
             Estimate     SE          t value
(Intercept) - 0.74418     0.06685 -11.132
COMP1         0.09442     0.02406   3.924
COMP2         0.02258     0.02088   1.081
COMP3         0.04060     0.02859   1.420
```

MCMC estimation of fixed effects parameters and p-values

```
            Estimate   MCMCmean HPD95lower HPD95upper pMCMC  Pr(>|t|)
(Intercept)   -0.7442  -0.7274    -1.1045    -0.3476 0.0087   0.0000
COMP1          0.0944   0.0947    -0.0400     0.2256 0.1074   0.0001
COMP2          0.0226   0.0278    -0.0759     0.1349 0.4760   0.2812
COMP3          0.0406   0.0441    -0.1082     0.2009 0.4276   0.1576
```

Accident Severity: Women

```
Linear mixed-effects model fit by REML

    AIC     BIC logLik Deviance
 -208.4 -165.7  118.2   -258.5

Random effects:
 Groups         Name         Variance
 Country (Intercept)          3.7216e-02
               COMP1          2.5175e-03
               COMP2          5.0650e-04
               COMP3          4.0815e-12
 Residual                     8.1630e-03 9.0349e-02

Number of observations: 156, groups: Country, 10

Fixed effects:
             Estimate      SE          t value
(Intercept) - 0.474747    0.062812  -7.558
COMP1         0.042526    0.020719   2.053
COMP2        -0.004259    0.021679  -0.196
COMP3        -0.038760    0.016317  -2.375
```

MCMC estimation of fixed effects parameters and p-values

```
            Estimate   MCMCmean HPD95lower HPD95upper pMCMC  Pr(>|t|)
(Intercept)  - 0.4747  -0.4585    -0.7955    -0.1142 0.0230   0.0000
COMP1          0.0425   0.0383    -0.0695     0.1513 0.3550   0.0418
COMP2         -0.0043   0.0092    -0.1302     0.1511 0.8745   0.8445
COMP3         -0.0388  -0.0427    -0.1611     0.0737 0.3343   0.0188
```

Number of accidents: Men

```
Generalized linear mixed model fit using Laplace
Family: quasipoisson(log link)

   AIC   BIC logLik  Deviance
 34060 34103 -17016  34032

Random effects:
 Groups       Name          Variance
 Country (Intercept)        238.873
              COMP1           6.995
              COMP2          16.452
              COMP3          17.221
 Residual                   216.377
Number of observations: 160, groups: Country, 10

Fixed effects:
             Estimate    SE         t value
(Intercept)   7.24848    4.88969  1.4824
COMP1         0.44272    0.83845  0.5280
COMP2        -0.03255    1.29129 -0.0252
COMP3        -0.05635    1.31918 -0.0427
```

MCMC estimation of fixed effects parameters and p-values

```
            Estimate   MCMCmean HPD95lower HPD95upper pMCMC  Pr(>|t|)
(Intercept)    7.2485   7.2427     5.3047     9.2108 0.0013   0.1403
COMP1          0.4427   0.4429     0.1096     0.7676 0.0231   0.5982
COMP2         -0.0325  -0.0323    -0.5439     0.4767 0.8534   0.9799
COMP3         -0.0564  -0.0572    -0.5716     0.4717 0.7510   0.9660
```

Number of accidents: Women

```
Generalized linear mixed model fit using Laplace
Family: quasipoisson(log link)

   AIC   BIC logLik  Deviance
 22063 22106 -11017  22035

Random effects:
 Groups        Name         Variance
 Country (Intercept)         311.121
              COMP1           10.192
              COMP2           10.780
              COMP3            6.756
 Residual                    142.031

Number of observations: 156, groups: Country, 10

Fixed effects:
             Estimate    SE       t value
 (Intercept)  6.91708    5.58221  1.2391
COMP1         0.51575    1.01475  0.5082
COMP2        -0.09417    1.05644 -0.0891
COMP3        -0.12051    0.83675 -0.1440
```

MCMC estimation of fixed effects parameters and p-values

```
            Estimate   MCMCmean HPD95lower HPD95upper pMCMC  Pr(>|t|)
(Intercept)    6.9171   6.9167     4.1731     9.6165 0.0038   0.2172
COMP1          0.5157   0.5159     0.0174     1.0054 0.0454   0.6120
COMP2         -0.0942  -0.0985    -0.6342     0.4006 0.5886   0.9291
COMP3         -0.1205  -0.1232    -0.5175     0.2980 0.4063   0.8857
```

Number of fatalities: Men

```
Generalized linear mixed model fit using Laplace
Family: quasipoisson(log link)

   AIC   BIC logLik  Deviance
 24956 25000 -12464  24928

Random effects:
 Groups        Name         Variance
 Country (Intercept)         189.0054
              COMP1            8.3384
              COMP2           19.0201
              COMP3           16.5743
 Residual                    160.8144
Number of observations: 160, groups: Country, 10
```

```
Fixed effects:
             Estimate    SE          t value
(Intercept)   6.531541   4.351349  1.5010
COMP1         0.536893   0.916245  0.5860
COMP2        -0.017041   1.391682 -0.0122
COMP3        -0.009708   1.297700 -0.0075
```

MCMC estimation of fixed effects parameters and p-values

```
            Estimate   MCMCmean HPD95lower HPD95upper pMCMC  Pr(>|t|)
(Intercept)    6.5315   6.5355     4.5055     8.5027 0.0020   0.1354
COMP1          0.5369   0.5386     0.1108     0.9584 0.0279   0.5587
COMP2         -0.0170  -0.0157    -0.6477     0.6371 0.9450   0.9902
COMP3         -0.0097  -0.0142    -0.6176     0.5867 0.9434   0.9940
```

Number of fatalities: Women

```
Generalized linear mixed model fit using Laplace
Family: quasipoisson(log link)

   AIC   BIC logLik  Deviance
 15132 15175  -7552  15104

Random effects:
 Groups      Name         Variance
 Country (Intercept)       238.8056
              COMP1           8.9428
              COMP2          10.1214
              COMP3           4.3951
 Residual                    98.4323
Number of observations: 156, groups: Country, 10

Fixed effects:
             Estimate    SE        t value
(Intercept) 6.47890     4.89239  1.3243
COMP1        0.55082    0.95225  0.5784
COMP2       -0.06665    1.02861 -0.0648
COMP3       -0.16772    0.68340 -0.2454
```

MCMC estimation of fixed effects parameters and p-values

```
            Estimate   MCMCmean HPD95lower HPD95upper pMCMC  Pr(>|t|)
(Intercept)    6.4789   6.4754     3.6467     9.3425 0.0056   0.1874
COMP1          0.5508   0.5508    -0.0143     1.0953 0.0501   0.5638
COMP2         -0.0667  -0.0712    -0.6966     0.5376 0.7357   0.9484
COMP3         -0.1677  -0.1716    -0.5783     0.2206 0.2670   0.8065
```