



D7.9 – Analysing European in-depth data: Methodological framework and results

Please refer to this report as follows:

Dupont, E., and Martensen, H. (Eds.) (2008) Analysing European in-depth data: Methodological framework and results. Deliverable D7.9 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

Report Author(s):

Emmanuelle Dupont (IBSR); Heike Martensen (IBSR); Yvan Casteels (IBSR), Lars Akkermans (IBSR); Eleonora Papadimitriou (NTUA); Dimitris Katsochis (NTUA); George Yannis (NTUA); Christian Branstätter (KfV)

Due Date of Deliverable: 31/10/2008

Submission Date: 05/11/2008

Project Start Date: 1st May 2004

Duration: 4.5 years

Dissemination level

Public



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

Table of Contents

Executive summary	4
1. Introduction	7
1.1 Methodological Issues	9
1.2 Description of the Fatal Accident Investigation database	17
1.3 Comparison with fatal accident data in CARE	28
1.4 Overview of the deliverable	37
2. Chapter 2 – Misreporting injury severity	39
2.1 Introduction	39
2.2 Modelling injury severity misreporting in the first wave of FAI data	40
2.3 Modelling injury severity misreporting in the final FAI data	47
2.4 Conclusion	51
3. Chapter 3 – Predicting accident size – A contrast between single and multivehicle accidents	55
3.1 Objectives	55
3.2 Methodology	56
3.3 Results	58
4. Chapter 4 – Factors affecting the fatality risk of road-users involved in fatal accidents	73
4.1 Introduction	73
4.2 Modelling the fatality risk for all road-users	79
4.3 The car-car analysis	89
4.4 An opponent analysis for car occupants	101
4.5 Conclusions	109
5. Conclusions	112
5.1 Results summary	112
5.2 Future perspectives	114
References	117
Appendices	120
Appendix 1.1 - Raw counts and percentages for the CARE-FAI data comparison	120
Appendix 2.1 - Comparative injury severity distributions for other countries except Italy (1st FAI database)	124
Appendix 2.2 - Comparative injury severity distributions for other countries except Italy (2nd FAI database)	126
Appendix 2.3 - Exploratory analysis of WP 5 data with respect to accident severity scores (2nd FAI database)	128
Appendix 2.4 - Significance of each selected predictor examined alone (2nd FAI database)	132

Appendix 2.5 - Correlation testing for the selected variables (2nd FAI database)	134
Appendix 3.1 - Joint modelling of road condition variable	135
Appendix 3.2 - Joint modelling of time variable	136
Appendix 3.3 - Joint modelling of road user variables	137
Appendix 3.4 - Joint modelling of accident development variables	138
Appendix 4.1 - Bayesian estimation of multilevel logit models	140
Appendix 4.2 - Defining the “Opponent Type” variable	141
Appendix 4.3 - Proportions of fatalities, severe injuries, slight injuries and uninjured cases for remaining road-user-opponent combinations	143
Appendix 4.4 - Commando files for data preparation	145

Executive Summary

The work package 5 of SafetyNet has produced detailed road accident data under the form of the Fatal Accident Investigation Database (FAI database, or database 5.1) and of the Accident Causation Database (AC database, or database 5.2). In the present deliverable, a series of analytical approaches that are appropriate for the exploitation of the information contained in the FAI database are suggested. These suggestions are made taking into account the methodological difficulties associated with the fact that the data concern fatal accidents exclusively.

A thorough definition of the above-mentioned difficulties and of their implications for the interpretation of results based on FAI data is provided in the introduction of this document. The specific features of the data, and their distribution on various important variables is also provided in the introduction. Finally, the distribution of the FAI data is compared with the distribution of the fatal accidents recorded in CARE on a series of key factors (e.g., road-user age...). The rationale behind this comparison is that similar distributions for the two types of data would consist of an indirect indication that the FAI data can be considered representative of the fatal accidents occurring in the country where they had been collected. Overall, the two datasets appear reasonably comparable. Yet, as concluded in this section, one should ideally take care that the two datasets are compared for each specific feature that appears central to the research question investigated: Representativity can never be considered to be established in a general way.

The second chapter of this document addresses the specific question of knowing whether two types of injury severity records – namely, one provided by the police, the other by the SafetyNet team – can be considered to be in agreement or in disagreement. The factors affecting the degree of agreement / disagreement are explored. The analysis was performed twice. Initially, it was run on the first wave of the FAI data collection. The results of this first step indicated that the analysis was efficient in detecting mismatches between the two injury severity scores. The analysis was then performed on the second wave of data, leading to the conclusions that mismatches problems had been solved, and that the few remaining mismatched records were no subject to systematic variation. Assuming that the SafetyNet injury records can be considered « accurate standards », this result suggests that no systematic bias is at play in the recording of injury severity by the police forces for fatal accidents in the countries that took part in the FAI data collection.

Chapter 3 focuses on the factors that differentiate single and multi-vehicle fatal accidents. This investigation is based on the observation of the important proportion of single vehicle accidents in fatal accidents, and on the assumption that both accident types are qualitatively different. The results indeed show that single vehicle fatal accidents, as compared to multi-vehicle ones, tend to occur mainly on relatively empty roads and on road sections between junctions. Car drivers tend to be more frequently involved in single accidents than

motorcyclists or heavy good vehicle drivers are. The drivers involved in single-vehicle accidents are most often unfamiliar with the surroundings.

The outcomes of fatal accidents for each individual road-user involved are examined in Chapter 4. The risk and protection factors of road-users involved in fatal accidents were analysed, taking account of the fact that all accidents examined generated at least one fatality. The analyses there presented were performed in three successive steps, which each provided particular improvements in handling the methodological difficulties posed by the limitation of the data to fatal accidents. First the complete dataset, including the observations made on all road-user types (i.e.: car-drivers as well as pedestrians or heavy good vehicles) was analysed. The remaining analyses focused on car-occupants specifically. Some variables emerged as “risk factors” in a consistent way for all three analyses, such as the fact that the road-user him/herself could (or could not) be considered as “senior” (i.e.: as being more than 65 years), or that the driver did not reacted properly to the occurrence of the accident by braking. In the case of car occupants, seatbelt appeared to be an important protection factor, even if the accident was *a-priori* characterized by a high severity level. The risk of the road-user being fatally injured appears to increase with the age of the vehicle. Finally, the fact that the accident took place on a road junction also leaves the road-users with increased chances to survive. The effects of other factors, such as the protective effect offered by front damage appear to hold only for collisions between car-occupants with other cars or with light good vehicles.

The results from all sections are summarized in the conclusions of this deliverable. In so doing, the emphasis is put on the fact that the analyses presented here have to be placed in the broader set of possibilities offered by the FAI data, which were certainly not exhausted in the present work.

Chapter 1 - Introduction

Accident data collection is often described in terms of macroscopic data versus microscopic (or in-depth) data. As presented in the lower part of Figure 1.1, macroscopic data like the CARE database and those collected by most national authorities contain many cases but only a relatively low level of detail. Such data may provide interesting and useful results. However, their analysis is often limited by a lack of details on several key factors of the accident process. Moreover, macroscopic data are in most cases unavailable or not accessible in disaggregated format, making it difficult and often impossible to link the specific conditions under which an accident took place to its consequences. On the other hand, the need for more detailed information has been served by a number of projects, in which fewer accidents are described in greater amount of detail (see, for example, the PENDANT and STAIRS projects). The present deliverable focuses on the analysis of the Fatal Accident Investigation database (the FAI database). This database was created as part of the work package 5 of SafetyNet. Figure 1.1 indicates some characteristic numbers of the FAI data and situates it at an intermediate level between macroscopic data (such as those collected in CARE), and fully microscopic data, such as the Accident Causation database, also created in work package 5.

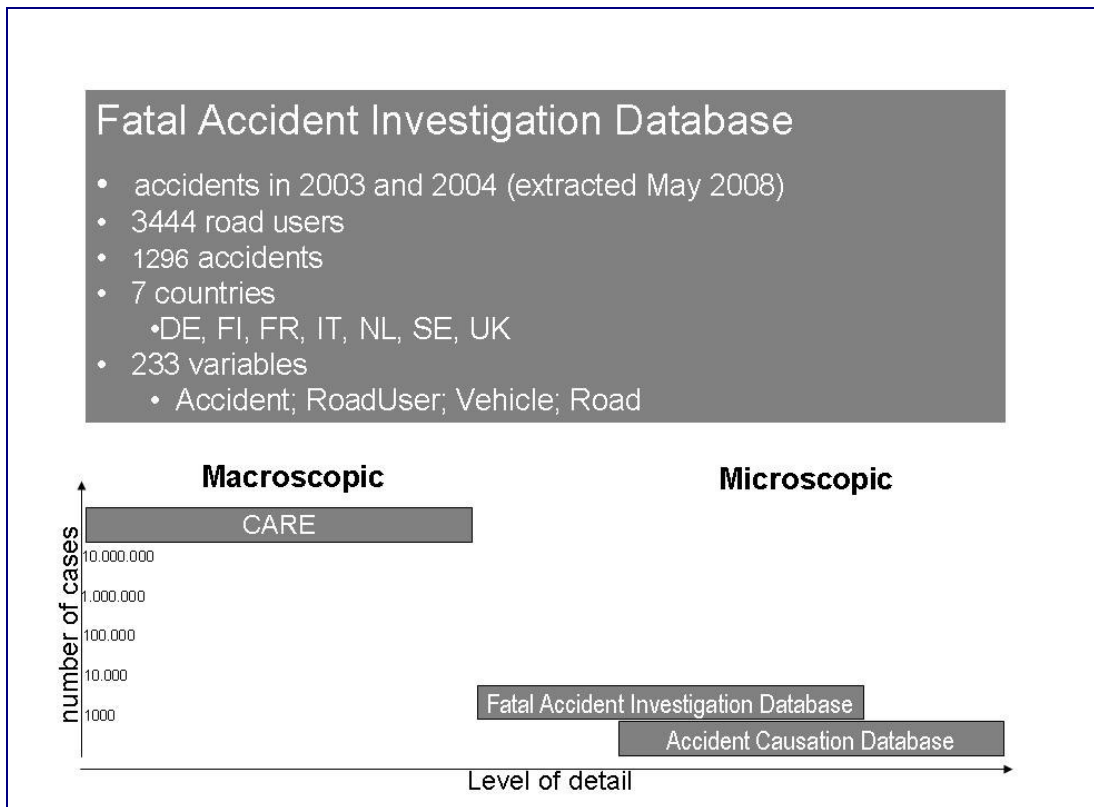


Figure 1.1: Description of the FAI database (Adapted from Morris, 2007)

The FAI database differs from macroscopic accident data as it includes a number of variables for which information is seldom available or reliable in

national databases. The FAI data also differs from macroscopic data in that they are available in a disaggregated format, as the information is recorded for each road user individually. More specifically, in the FAI database, the chain of events of each accident is identified and described in detail. Important variables related to the road user are also available, including the use of safety equipment, impairment, familiarity with the road network, etc. Finally, detailed additional information on the road and traffic environment is recorded, including speed limit, traffic volume, road design (gradient, curvature etc.), the presence of pedestrian facilities, etc.

The present deliverable is entirely devoted to the analysis of the FAI data. Three general questions were addressed in these analyses:

- How reliable are the injury severity scores assigned by the Police to road accident casualties? Are there any factors systematically affecting the misreporting of injury severity? (Chapter 2)
- What differentiates single from multiple vehicle fatal accidents? (Chapter 3)
- What is the probability of being killed, given that one is involved in a fatal accident? What factors affect this probability and thus the consequences of a severe accidents for the persons involved? (Chapter 4)

The first question can be examined on the basis of the FAI data, given that the initial Police scores of injury severity were checked, and, if necessary, corrected by the SafetyNet team. The second question emerged from the intuitive idea that single vehicle and multi-vehicle fatal accidents are two qualitatively different types of accidents. Single-vehicle accidents make up an important proportion of the total FAI data (see detailed explanation in section 1.1.1). The FAI data include the appropriate information to identify the factors upon which single and multi-vehicle accidents mostly differ. Finally, the third question concerns the fundamental issue of the consequences of accidents for the persons they involve, bearing in mind that all the accidents recorded here were fatal accidents. With respect to this question, the added value of the FAI data lies on the increased level of reliability and detail of the available variables, allowing the investigation of effects that cannot be examined with macroscopic data.

Before tackling any of the above basic questions, however, the particular nature of the FAI data, and, most notably, the fact that these data are limited to fatal accidents, requires addressing a number of conceptual and methodological issues. These are described in detail in the next section (Section 1.1). Section 1.2 then provides a thorough description of the distribution of the FAI data on several key factors (e.g.: the age and gender of the road-user, the type of road where the accident occurred...). The last section of this introduction provides a complete overview of the results of the comparison of the FAI data and the fatal accidents recorded in CARE. The rationale underlying this comparison is that the fatal data in CARE, because they correspond to all fatal accidents occurring in each European member state, can be considered to be the population from which the FAI data were extracted. To be considered “representative”, the FAI data should thus be distributed similarly to the fatal CARE data.

1.1 Methodological issues:

Emmanuelle Dupont & Heike Martensen, IBSR

1.1.1 Having data limited to fatal accidents: General implications

Obviously, observations that are limited to fatal accidents provide information that is restricted to the high-end of the accident-severity continuum. This is illustrated in Table 1.1, in which the FAI data are compared with (1) the CARE data, which contain information about accident from all severity level (left column), (2) the CARE data restricted to the fatal accidents observations (centre column).

Usually, the fatality risk is defined as *the ratio of people killed in road accidents to the total number of accidents (see Table 1.1, "Fatalities per accident")*. When the data include accidents of all severity levels, the proportion of fatalities per accident (the fatality risk) is small (0.3), reflecting the fact that only a few accidents result in fatalities. In a fatal accident database, however, every accident contains a fatality, so that the proportion of fatalities per accident goes up to 1 (it is in fact somewhat larger than one because one is the minimal number of fatalities in the accident, there can be more).

Table 1.1 also clearly shows that the structure of the FAI data and that of the CARE *fatal* accidents data are similar. The FAI and CARE data sharply differ, however, once the comparison is made with the complete CARE database. In this case the fatality risk in the FAI data ("Proportion of road-users killed" in Table 1.1) looks tremendously high in comparison to the one computed on the whole CARE data.

In the next sections, the types of analyses that are appropriate to apply to fatal accident data are described in more detail. For now it is important to note that because of the peculiarities of fatal-accident data, the term *fatality risk* will be used – throughout this document - in a way that differs from its common definition. In the specific context of the analysis of fatal accident data, the term "fatality risk" will be used to denote the *risk that a road-user runs of ending among the fatalities rather than among the survivors of an accident, given that the accident was fatal*. This risk is estimated here on the basis of the "proportion of road-user killed", or the *number of fatalities to the total number of road-users in the database*, as defined in the last row of Table 1.1. This specific way of defining the fatality risk will be of application throughout this deliverable

	CARE data (all accidents)	CARE data (fatal accidents only)	FAI data
Number of crashes	555558	14129	1296
Number of <i>fatal</i> crashes	14129	14129	1296
Percentage of fatal crashes in database	2%	100%	100%
Fatalities per accident (total fatalities/total accidents in database)	0.03	1.09	1.11
<i>Proportion road-user killed</i> (total fatalities/total road users in database)	0.01	0.45	0.42

Table 1.1 Characteristics of the Fatal Accident Investigation Database as compared to CARE

Asking appropriate questions of the data:

Below is a summary of percentages calculated on the raw FAI data for several variables:

- 87% of the drivers involved in fatal accidents were sober
- 83% of the drivers were men
- 60 % of the cars were damaged most at the front

How should these percentages be interpreted? Should we conclude that fatal accidents specifically involve men and sober drivers? Or that front damage is particularly likely to result in fatalities?

The problem with the interpretation of the above percentages lies in the absence of a meaningful reference point to which they can be compared. As previously noted, the data exclusively refer to fatal accidents. For this very reason, they will not provide any information about the features or characteristics that are *specific* to fatal accidents. As an example, front damage is the most frequent type of damage in *all* accidents in the FAI database, hence the high proportion of front damages reported above. This high proportion does not, however, allow the conclusion that front damage is particularly likely to result in an accident being *fatal*. To ascertain such a conclusion, one would need a reference point – such as the percentage of front-damaged cars in *non-fatal* accidents, indicating that front damage is significantly less frequent in non-fatal accidents as compared to fatal accidents. To state it alternatively, *in and by themselves*, observations from fatal accidents – whatever the level of detail they offer – do not deliver information on the causes of fatal accidents. To obtain such information, data from fatal accidents have to be combined with exposure data, or with similar information from non-fatal accidents.

Consequently, the restriction of the FAI data to fatal accidents rules out all investigations focusing on differences between fatal and non-fatal crashes.

However, contrary to the accidents themselves - which were all fatal - the severity of their consequences for the individual road-users differs. The fatality risk (here defined and estimated as the proportion of road users who died in the accident to the total number of road users in the database) is 0.45 for the whole FAI database. This indicates that, overall, a bit less than one person out of two died in the accident. The proportions of road users who died and who survived the accident are consequently quite even, and the question of knowing what differentiates the survivors of these crashes from the fatalities is interesting in its own right. The survivor group can therefore be used as a point of reference for the fatality group and enables us to identify person-, vehicle-, or accident-characteristics that make it more – or less -likely that an individual survived although he/she was involved in such a severe accident.

In this way, conclusions can be drawn about the fatality risk for those involved in fatal accidents. With some care, one can extrapolate the fatality risk results to nonfatal accidents by assuming that if the fatality in a fatal accident would have had the same characteristics as the survivor, this would not have become a fatal accident at all. One should, however, verify for each variable whether this assumption is reasonable, and keep in mind that the results say nothing about the general risk for road-users of dying in an accident.

Although this approach solves one basic problem - namely the one of the need for a reference group - several methodological pitfalls still apply to this analysis, which also relate to the fact that the data are restricted to fatal accidents. This lies at the origin of two general problems, which will be referred to here as the accident-size bias and the incomparability of the baseline risk. In the following, these two problems are described along with possible solutions.

The Accident-Size bias

The accident-size bias is best understood in reference to a particular type of observation, namely accidents in which one - and only one - road-user was involved. For this type of accidents, the percentage of road-users killed will inevitably be 100%. Accordingly, the probability for a road-user to end up among the fatalities of the accident will invariably be estimated to be 1, for the simple reason that this casualty would otherwise not have been part of the observations!

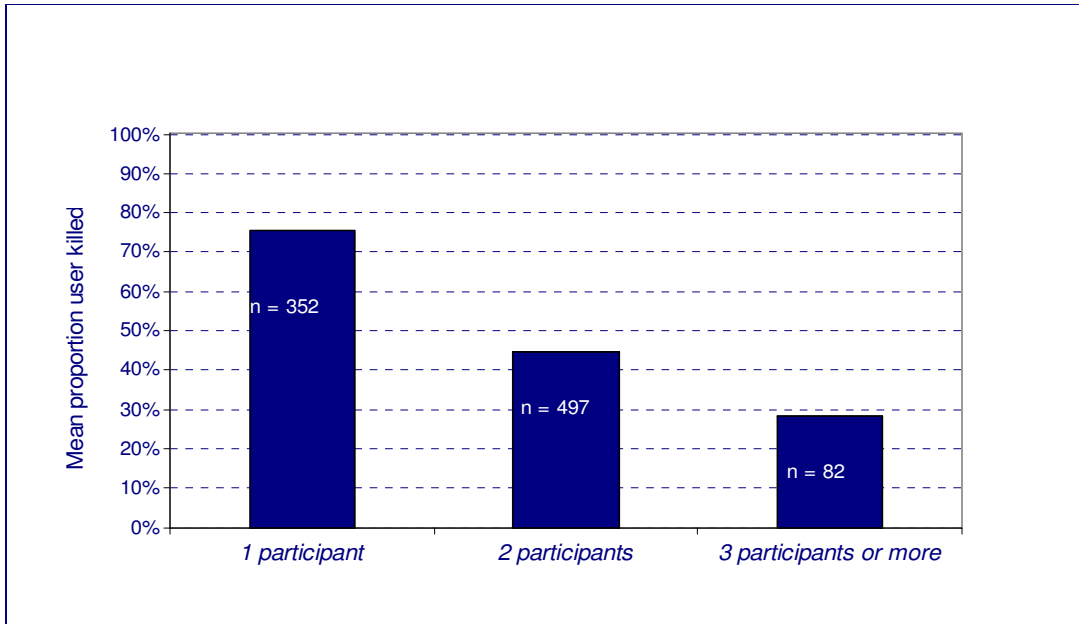


Figure 1.2: Mean proportion of road-user killed in an accident as a function of the number of participants in the accident

Because it is a priori given that each accident in the database has a least one fatality, whether there were also survivors or not most crucially depends on whether there was more than one person involved in the accident. There is, as a consequence, a strong relationship between the size of the accident (the total number of road-users that it involved) and the proportion of people killed in it. This is illustrated in Figure 1.2: The proportion of road-users killed is maximal in the case in which there was only one participant¹ in the accident (it is not exactly 100% though in this case because one “accident participant” often happens to be a car, with more than one road-user on board). This proportion then quickly and steadily decreases, as the number of accident participants increases.

This relationship between the size and the outcome of an accident is, in this context, qualified as *bias* because it actually does not tell much about the true relationship that may otherwise exist between the size of an accident and its consequences. In this case, it reflects mainly the selection criteria that were applied to the data collection: Only fatal accidents were included in the data.

In Chapter 4, specific ways to deal with this bias are developed on the basis of alternative models for the fatality risk.

¹ The term « accident participant » will be used whenever reference is made to one of the *protagonists* of the accident. Although accident protagonists sometimes happen to correspond to individual road-users (as this is the case for pedestrians or bicycles, for example), this is often not the case: all motor vehicles are protagonists in and by themselves, and they often “contain” several road users.

The (in)comparability of the baseline risk

The risk run by a road-user, whenever involved in an accident, always strongly depends on (1) his/her own travel mode (i.e., whether he./she is a car occupant, a bicyclist, or a truck driver), and (2) the travel mode of the road user he/she happened to collide with in the course of the accident. A car driver’s chances to die in an accident, for example, are *de facto* dramatically different depending on whether he/she collided with a bicycle, or with a heavy good vehicle. This is a general law in road accidents.

Consequently, when performing analyses focusing on the outcomes of the accidents for the individual road-users, care has to be taken that the cases selected offer sufficient baseline (i.e., a priori) comparability.

<i>Type of accident</i>	<i>Expected risk distribution between accident participants</i>
Car-Car	« 50-50 »
Car-Vulnerable	« 0-100 »
Car-Heavy	« 100-0 »

Table 1.2 *Distribution of risk in different accident types in a fatal accidents data base³*

This general problem is exacerbated when dealing with a fatal accident database: The fact that there was at least one fatality in the accident leads to the extreme incomparability situation depicted in Table 1.2. This table presents the distribution of the risk between the participants involved in three different types of accidents². The first row corresponds to car-car accidents (cf. Section 4.3). Knowing that one road-user at least died in each of these accidents, one can assume each car occupant to have a 50% chance to be the fatality. Matters are very different for the other two types of accidents. If the accident took place between a car and a pedestrian, one can be certain that the car occupant survived the accident. On the contrary, if the car collided with a heavy good vehicle, that car’s occupant has a 100% certainty to become the fatality. Generally speaking, in the case of fatal accidents data, incompatible accidents are accidents in which the risk to die is maximal for the more vulnerable of the road-users, while this risk is basically null for the relatively stronger participant.

In Chapter 4, the way to handle the incomparability of the baseline risk is specifically addressed, and different types of modelling frameworks are examined in order to identify the optimal approach for addressing this complex issue.

² These are theoretical numbers, simplified by the assumption that each car is occupied by one person only (Section 4.4 provides the real figures in the FAI-data).

1.1.2 Dealing with the hierarchical structure of accident data

Generally speaking, accident data are hierarchically organised. As presented in Figure 1.1, accident data have three different levels. At the first level individual road-users – who eventually sustain the consequences of accidents – are «nested» within vehicles, which form the second level. The vehicles in turn are nested within the accidents. This makes accidents the third level in the data. In our analyses, we want to know how characteristics at all three levels affect the fatality risk ran by each individual road-user. While the value of road-user variables (e.g. age, gender, seatbelt use, etc.) differ for each of them, all users in one vehicle have in common the values of the vehicle characteristics (e.g. they all occupied a 5 year-old car). Similarly, all road-users in a particular accident have in common the characteristics of that accident (e.g., the accident took place at a junction). This three-level structure is illustrated in Figure 1.3.

The importance of this hierarchical structure can be explained using the analysis of the factors affecting the fatality risk as an example. The question there asked is the one of knowing how the probability for each road-user to die in the accident is affected by characteristics of the road-users themselves (e.g., their age), but also by characteristics of the car (e.g., the car's age), or of the accident (e.g.: whether it occurred at a road junction or not). The hierarchical organisation of the data introduces correlation between the observations: As an example, Lenguerrand, Martin, & Laumon (2006) estimated³ that, “when a driver sustained no injury, the other driver had 12% risk to be severely injured or killed. When one of the drivers died, the other had a probability of 60% to be seriously injured or killed”. In other words, the probability of dying in an accident is more similar for two road-users involved in the same accident than for road-users involved in different accidents, and even more similar for two road-users that occupied the same vehicle than for individuals in different vehicles.

Because the FAI data are so hierarchically organised, the observations also tend to be correlated within vehicles and within accidents. However, because they include fatal accident data exclusively, the exact nature of these correlations differs from those described in the above example. The important point here is that – for the FAI data just as for any other type of accident data – observations coming from the same accident and/or vehicle are not independent from each other, and cannot be considered so. This is especially problematic when one wants to examine the effect of accident or vehicle-level characteristics on the probability of dying in the accident.

³ On the basis of 15525 two-car head-on crashes injury data

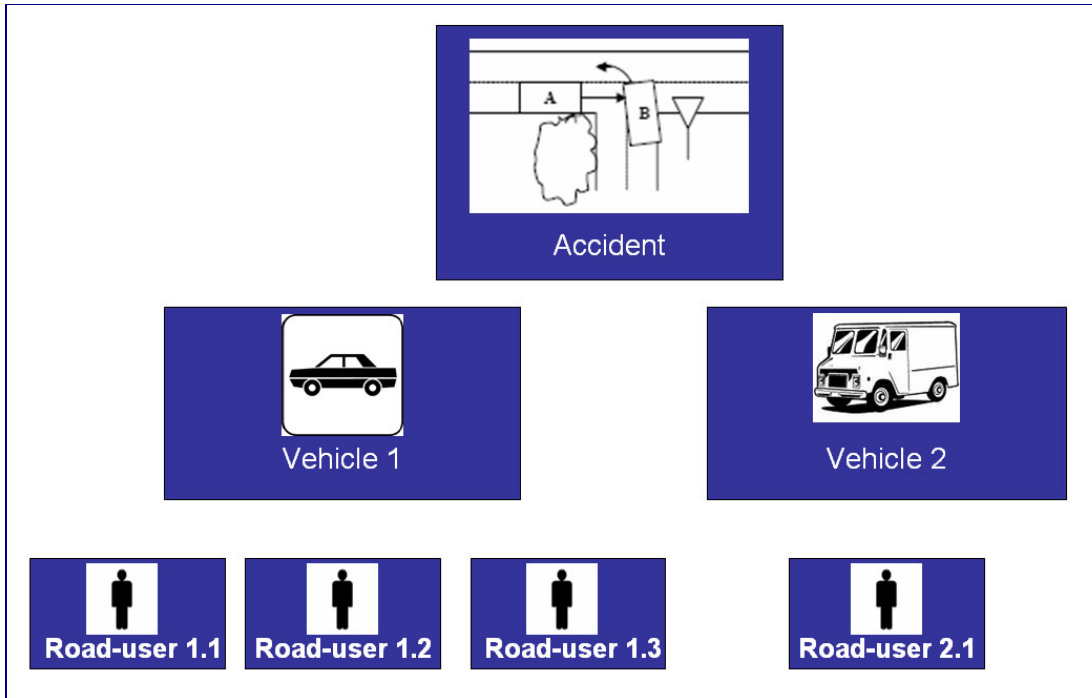


Figure 1.3: The hierarchical structure of accident data

Performing a standard analysis means ignoring the hierarchical nature of the data, and duplicating all the accident or vehicle characteristics of interest by the total number of individual road-users in each accident/vehicle. As an example: In a database containing five accidents with two vehicles involved in each and two road users per vehicle, we would end up with 20 observations, which is correct for the characteristics of the road-users (e.g. age). However, for the accident variables (e.g. junction) there were actually only five different observations (5 accidents). The danger with this way of proceeding is that it amounts to artificially augmenting the sample size for all observations that are situated at the accident or vehicle level. By so doing, one runs the risk of minimizing the variance in the observations that is related to the vehicle/accident level. Consequently, one may also come to decide that the accident and/or vehicle predictor(s) examined are significantly related to the dependent variable (the fatality risk), while in fact they are not.

For these reasons, hierarchical structures in accident data receive increasing attention from the road safety research community (Lenguerrand et al., 2006; Dupont & Martensen, 2007; Jones & Jorgensen, 2003) and endeavours are made to apply statistical models that allow taking the relation among the observations into account. These models are known as multilevel models, (e.g., Goldstein, 2003; Hox, 2002; Kreft and de Leeuw, 2002; Snijders and Bosker, 1999), random-coefficient regression models (Longford 1993) or mixed effects models (Pinheiro & Bates, 1995), and have been developed in other domains, predominantly biology and pedagogy, where hierarchical structures are even more prevalent.

However, compared to other data hierarchies, accident data present some peculiarities that prevent any straightforward application of multilevel models for their analysis. The most important is the fact that there are usually few road-users in one vehicle - most often is there only one - and it is also frequently the case that the road-user and its vehicle are one and the same thing (think about pedestrians, for example). For this reason, computational problems may be encountered and/or biased estimates can be obtained when multilevel models are applied to these data.

All the analyses here presented started on tests for the presence of hierarchies in the accident data, to check for the necessity of applying multilevel models. Whenever significant variation was identified at higher level(s), multilevel models were fitted. In most cases, however, the results yielded the conclusion that there was no significant variation in the observations at the accident or vehicle level. In principle, this means that no substantial correlation is introduced among the data by the hierarchical structure. However, precisely because of the particular nature of accident data (many accidents with few vehicles and people involved in each), it is difficult to tell whether there is indeed no substantial variation at the accident and/or vehicle level, or whether this *seems* to be the case simply because the number of observations per accident and/or vehicle is insufficient to allow the estimation of the different variance components.

In such cases, the choice was made to leave the model in its single-level state, and to perform a standard logistic regression analysis. This choice was motivated by the fact that the particular hierarchical structure of accident data (i.e.: a lot of accidents with few vehicles and people) also leads them to be “closer to that of independent observations than that of data traditionally processed by correlated data model” (Lenguerrand et al., 2006, p. 44). The artificial increase of the sample size mentioned earlier is indeed less important if there are few observations per higher-level unit than in the case of hierarchically organised school data, for example. In this case, ignoring the hierarchical structure of the data also means that the school or classes observations will be multiplied by the total amount of pupils, thus a far more considerable artificial increase.

1.2 Description of the FAI Database

George Yannis , Dimitris Katsochis, Eleonora Papadimitriou (NTUA)

Although limited to fatal accidents, the Fatal Accident Investigation (FAI) dataset of SafetyNet WP5 is considered likely to yield interesting conclusions with respect to injury severity and the identification of common injury causes. In absence of non-fatal accidents as a control group, it is still possible to identify suitable variables allowing for the formation of models. Keeping in mind the major methodological questions posed within the preliminary analysis of the data (absence of exposure data, accident size bias, and incomparability of risks), the objective of this section is to create an overall picture of the structure and main contents of the FAI database and to track those parameters that appear meaningful for the purposes of the analysis.

1.2.1 Structure of the FAI database

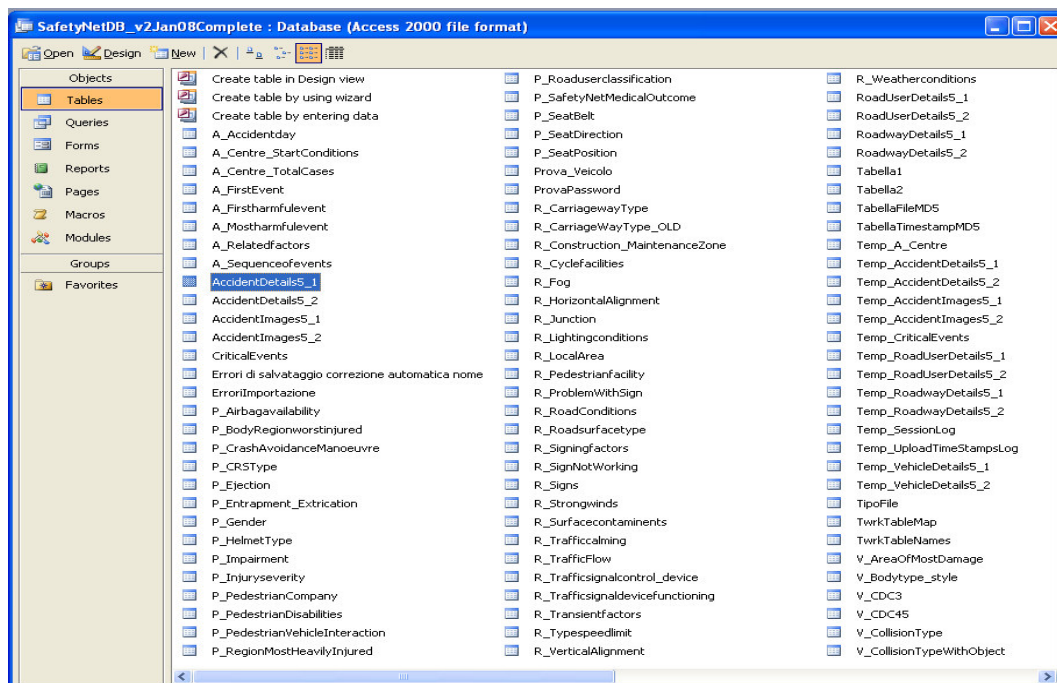


Figure 1.4: Structure and contents of the FAI database

As most accident databases, the FAI database consists of four separate yet linked main Tables: Accident details, Road details, Vehicle details and Road user details (Figure 1.4), in MS Access format. These Tables are linked by means of the following "keys" (common variables): StringDateTime⁴, Vehicle Number, Person Number. It is noted that several additional Tables are available

⁴ It is noted that this variable corresponds to the date and time that the researcher created the case when entering it into the database (and not to the actual date and time of the accident).

Analysing European in-depth data: Methodological framework and results

in the database, concerning e.g. variables coding information, pictures or sketches etc.

For instance, the Road User details table includes the date/time and vehicle number keys, allowing the linking of information between accident, road and vehicle tables (see Figure 1.5).

ID	Roadway5_1_ID	StringDateTime	CaseNumber	VehicleNumber	PersonNumber	RoadUserClass	Age	Gender	Impairment
19	17	01.02.2007 14.14.01	98569	6010	1	1 Driver	60	Male	None
276	195	01.06.2007 08.52.34	09953	4110	1	1 Driver	23	Female	None
277	196	01.06.2007 08.52.34	09953	4110	2	1 Driver	46	Male	None
278	197	01.06.2007 10.14.37	27935	4111	1	1 Driver	37	Male	None
94	68	01.08.2006 12.46.47	40620	4038	1	1 Driver	19	Female	None
187	130	01.11.2006 09.21.00	50016	4075	1	1 Driver	34	Male	None
606	93	01/03/2007 09.15.58	38232	7048	1	1 Pedestrian	999	Female	None
607	94	01/03/2007 09.15.58	38232	7048	2	1 Driver	999	Male	None
608	95	01/03/2007 10.00.19	61738	7049	1	1 Driver	21	Male	None
581	95	01/03/2007 10.00.19	61738	7049	1	2 Passenger	25	Male	
582	95	01/03/2007 10.00.19	61738	7049	1	3 Passenger	20	Male	
583	95	01/03/2007 10.00.19	61738	7049	1	4 Passenger	20	Male	
609	96	01/03/2007 10.00.19	61738	7049	2	1 Driver	23	Male	None
610	96	01/03/2007 10.00.19	61738	7049	2	2 Passenger	23	Male	
611	97	01/03/2007 12.40.05	11229	7050	1	1 Driver	38	Male	None
612	98	01/03/2007 12.40.05	11229	7050	2	1 Driver	27	Female	None
613	98	01/03/2007 12.40.05	11229	7050	2	2 Passenger	27	Male	
325	228	01/03/2007 14.01.48	92627	3136	1	1 Driver	22	Male	Alcohol
326	228	01/03/2007 14.01.48	92627	3136	1	2 Passenger	42	Female	
327	229	01/03/2007 16.02.56	95153	3137	1	1 Driver	14	Male	None
328	230	01/03/2007 16.02.56	95153	3137	2	1 Driver	66	Male	None
869	601	01/07/2007 09.30.59	71102	2424	1	1 Driver	10	Male	None
870	602	01/07/2007 09.30.59	71102	2424	2	1 Driver	42	Male	None
871	603	01/07/2007 10.30.49	01393	2425	1	1 Driver	32	Male	None
872	603	01/07/2007 10.30.49	01393	2425	1	2 Passenger	31	Male	
873	603	01/07/2007 10.30.49	01393	2425	1	3 Passenger	46	Male	
874	604	01/07/2007 17.08.52	67053	2426	1	1 Driver	19	Male	None
875	605	01/07/2007 17.08.52	67053	2426	2	1 Driver	20	Male	None
876	606	01/07/2007 18.19.49	28845	2427	1	1 Driver	56	Male	None
877	607	01/07/2007 18.19.49	28845	2427	2	1 Driver	57	Male	None
878	608	01/07/2007 22.13.23	46709	2428	1	1 Driver	30	Female	None
879	609	01/07/2007 22.13.23	46709	2428	2	1 Driver	52	Male	None
100	68	01/08/2006 10.44.04	48413	3043	1	1 Driver	23	Male	Other
101	69	01/08/2006 13.44.00	97115	3044	1	1 Driver	25	Male	None

Figure 1.5: View of the Road User details Table

By linking the "key" variables across the four tables and selecting all the fields of each table, it is possible to obtain a single Table containing all the information in a matched and sorted way (see Figure 1.6).

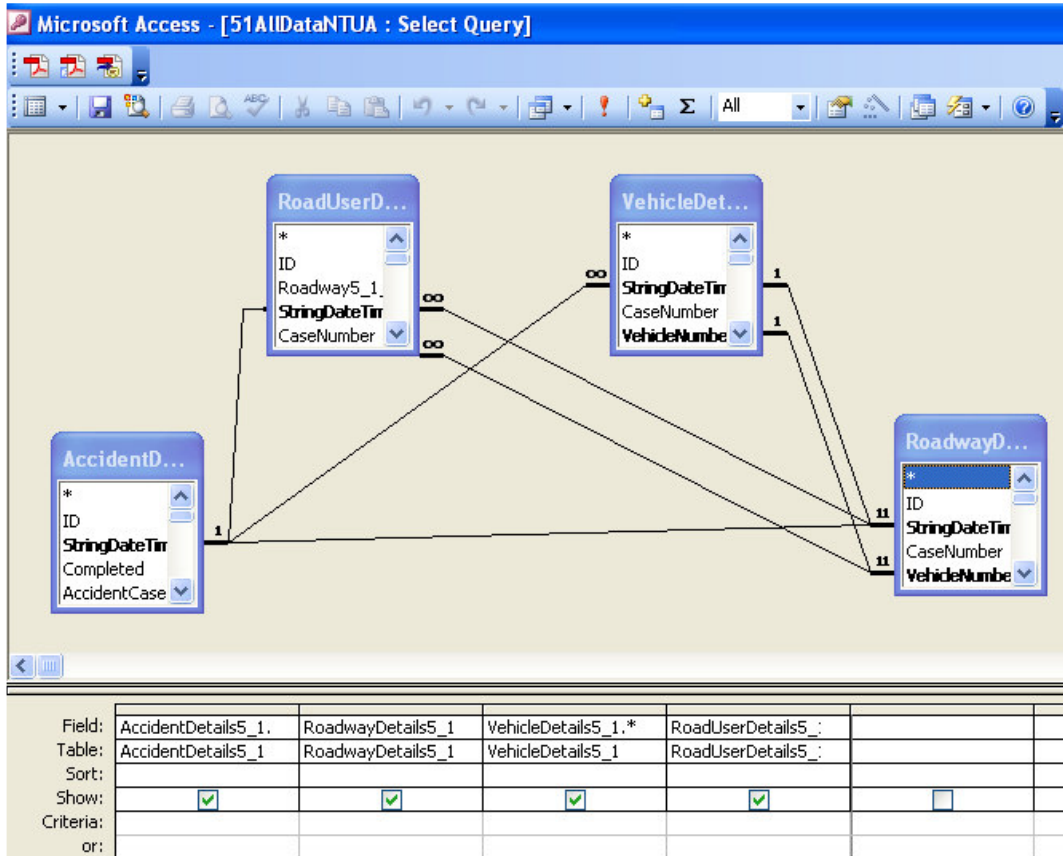


Figure 1.6: Linking and matching accident, road, vehicle and user information in the FAI database

A more detailed presentation of the data linkage and integration process is beyond the scope of this document. The above process was briefly demonstrated in order to provide an overall understanding of the initial and final structure of the database. As regards the contents of the database, the following list of variables is available (Table 1.2.1). More details on variables, values and definitions can be found in the WP5 Glossary of the data variables collected (SafetyNet D.5.5, 2008).

Analysing European in-depth data: Methodological framework and results

AccidentDetails5_1_ID	StrongWinds	HazardousCargo	IsFamiliar
AccidentDetails5_1_StringDateTime	Fog	CargoDischarged	Comments2
Completed	CommentCond	PreImpactSpeed	CrashAvoidMan
AccidentCase	SurfaceContaminants	NumberOfEvents	Comments3
AccidentDetails5_1_CaseNumber	SignRelated	MostHarmfulEvent	SeatPos
CentreName	TrafficCalming	AreaOfMostDamage	SeatDir
AccidentDate	WasTrafficCalm	EventType1	SeatBelt
AccidentDay	CommentsFact	EventDetail1	Comments4
TimeOfDay	NumOfSigns	InteractedWith1	AirbagAvail
HitAndRun	Sign1	CollisionType1	AirbagDeploy
AnimalInvolved	ProblemWithSign1	EventType2	CommentsAirbag
AccidentTypeClass	NotWorking1	EventDetail2	PoliceInjurySeverity
FirstEvent	Sign2	InteractedWith2	SafetyNetMedicalOutcome
RelatedFactors	ProblemWithSign2	CollisionType2	BodyRegionMostInjured
CrashParticipants	NotWorking2	EventType3	Comments5
CarMPV	Sign3	EventDetail3	Ejection
Van	ProblemWithSign3	InteractedWith3	EntrapmentExtraction
BusMinibus	NotWorking3	CollisionType3	TakenToHospital
Truck	Sign4	EventType4	HospDuration
AgriculturalVehicle	ProblemWithSign4	EventDetail4	DiedAtScene
MotorcycleMoped	NotWorking4	InteractedWith4	NDaysUntilDeath
Bicycle	Sign5	CollisionType4	Comments5b
TrainTram	ProblemWithSign5	EventType5	SuspicionAlcohol
ShoeVehiclePedestrian	NotWorking5	EventDetail5	PoliceRepOtherDrug
Other	RoadwayDetails5_1_OtherComments	InteractedWith5	FailureOfDriverRider
UnknownVehicle	RoadwayDetails5_1_FullFields	CollisionType5	WhatCausal
AccidentDetails5_1_Comments	RoadwayDetails5_1_EmptyFields	EventType6	Comments6
AccidentSummary	RoadwayDetails5_1_CreationDate	EventDetail6	ChildRestrFitted
AccidentDetails5_1_FullFields	RoadwayDetails5_1_LastUpdateDate	InteractedWith6	ChildRestrUsed
AccidentDetails5_1_EmptyFields	RoadwayDetails5_1_DeleteDate	CollisionType6	CRSType
TotalFiledPercentage	RoadwayDetails5_1_SessionID	EventsComments	Comments7
StateString	VehicleDetails5_1_ID	ABS	MCycleHelmetWorn
AccidentDetails5_1_CreationDate	VehicleDetails5_1_StringDateTime	BAS	MHelmetType
AccidentDetails5_1_LastUpdateDate	VehicleDetails5_1_CaseNumber	ACS	Comments8
AccidentDetails5_1_DeleteDate	VehicleDetails5_1_VehicleNumber	ESP	PartialLeathersProtJack
CaseCheck	NumOfOccupants	LDW	PartialLeathersProtJackTrou
AccidentDetails5_1_SessionID	VehicleType	CSS	MGloves
RoadwayDetails5_1_ID	VehicleMake	TCS	MBoots
RoadwayDetails5_1_StringDateTime	VehicleModel	ESafetyComments	MReflitemWorn
RoadwayDetails5_1_CaseNumber	CarBodyStyle	VehicleDetails5_1_OtherComments	Comments9
RoadwayDetails5_1_VehicleNumber	DrivenWheels	VehicleDetails5_1_FullFields	BHelmetWorn
CarriagewayType	DriveOfVehicle	VehicleDetails5_1_EmptyFields	BHelmetType
NumberOfLines	VehicleColour	VehicleDetails5_1_CreationDate	Comments9b
Motorway	VehicleLength	VehicleDetails5_1_LastUpdateDate	HighVisCloth
SpeedLimit	VehicleWidth	VehicleDetails5_1_DeleteDate	ThickCloth
TypeOfSpeedLimit	WasVehicleTowing	VehicleDetails5_1_SessionID	Comments10
Junction	EnginePower	RoadUserDetails5_1_ID	PedVehInteraction
LocalArea	YearOfManufacture	Roadway5_1_ID	PedCompany
VerticalAlignment	KerbWeight	RoadUserDetails5_1_StringDateTime	PedDisabilities
HorizontalAlignment	NumberOfAxles	RoadUserDetails5_1_CaseNumber	PReflectiveItemsWorn
ConstrMaintZone	SpecificSpeedLimit	RoadUserDetails5_1_VehicleNumber	Comments11
RoadwaySurfaceType	GeneralComments	PersonNumber	AnyOtherComment
PedestrianFacility	AreDefects	RoadUserClass	RoadUserDetails5_1_FullFields
CycleFacilities	VehicleDetails5_1_Comments	Age	RoadUserDetails5_1_EmptyFields
RoadConditions	PassedInspection	Gender	RoadUserDetails5_1_CreationDate
LightConditions	DriverManoeuvre	Impairment	RoadUserDetails5_1_LastUpdateDate
TrafficFlow	TransientFactors	Comments1	RoadUserDetails5_1_DeleteDate
WeatherConditions	VehicleHeading	IsAResident	RoadUserDetails5_1_SessionID

Table 1.3: Overview of accident, road, vehicle and user variables in the FAI data

1.2.2 Basic descriptive statistics of the FAI data

This section is intended to present the profile of accident severity in the examined dataset. This is achieved by describing how the severity scores vary across factors pertaining to the following main components of the road network:

- Accident (e.g. number of events, number of opponents)
- Road (e.g. speed limit, carriageway type, junction)
- Vehicle (e.g. type of vehicle, safety equipment)
- User (e.g. gender, age, seat belt use).

However, it is important to think about how non-fatally injured persons end up in this database. Only accidents with at least one fatality were included. Consequently the presence of non-fatally injured persons means that these persons were *additional* victims. Therefore, it would be reasonable to consider that proportionally more non-fatally injured people does not necessarily mean less severe accidents overall. Moreover, the degree of vulnerability of the road user certainly affects his or her fatality risk, once involved in a fatal accident. In the next chapters, specific conceptual and statistical considerations are proposed in order to address these issues. In this section, therefore, no conclusions on fatality risk are drawn and the figures presented mainly serve the purpose of data description.

The FAI database includes 3398 road users involved in 1296 fatal accidents. These road users were drivers or passengers of 2244 vehicles, out of which 257 are included as "shoe vehicle / pedestrians". Each road user's injury severity is recorded by two variables:

- The Police severity score, which is the score initially assigned by the Police in their records
- The SafetyNet medical outcome, which is the final score assigned by the SafetyNet team, on the basis of Police, hospital, insurance companies and other data sources.

Table 1.2.2 shows the distribution of the road users across injury severity score for the two variables. It can be seen that differences between the two scores exist even in the serious and fatal injuries, suggesting that Police records had to be corrected by the SafetyNet team in several cases. Mismatches between the two scores are even more striking in the slightly injured and uninjured road users. It is noted that such differences were observed (to a greater extent) in the first wave of FAI data analysed in Deliverable D.7.6 of SafetyNet. A separate chapter of this Report is devoted to a further analysis of these mismatches and to possible systematic factors behind them.

	<i>Fatal</i>	<i>Serious</i>	<i>Slight</i>	<i>Not Injured</i>	<i>Unknown</i>	<i>Grand Total</i>
<i>Police severity</i>	1351	507	455	834	251	3398
<i>SafetyNet severity</i>	1433	479	880	490	116	3398
<i>Difference</i>	-82	28	-425	344	135	

Table 1.4: Police severity score vs. SafetyNet medical outcome in the FAI data

In the remainder of this section, descriptive statistics are calculated on the basis of the SafetyNet medical outcome, which is considered to be far more accurate than the Police score. In particular, Figure 1.7 shows the distribution of casualties in the FAI data per injury severity and gender. It is shown that men are proportionally more often the fatality in fatal accidents. Moreover, women are slightly more often an additional casualty in fatal accidents than a fatality.

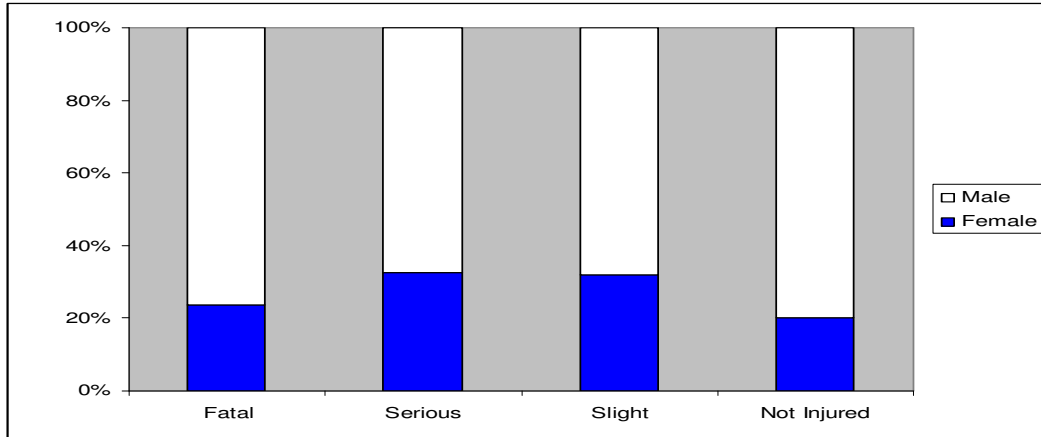


Figure 1.7: Injury severity per road user gender in the FAI data

The effect of road user age is shown in Figure 1.8, where persons aged more than 55 years are clearly overrepresented among the fatalities of the fatal accidents examined, obviously due to their increased vulnerability. The effect is even more apparent for persons older than 65 years. An additional interesting finding is that road users aged between 18 and 24 years are more often an additional serious casualty in fatal accidents.

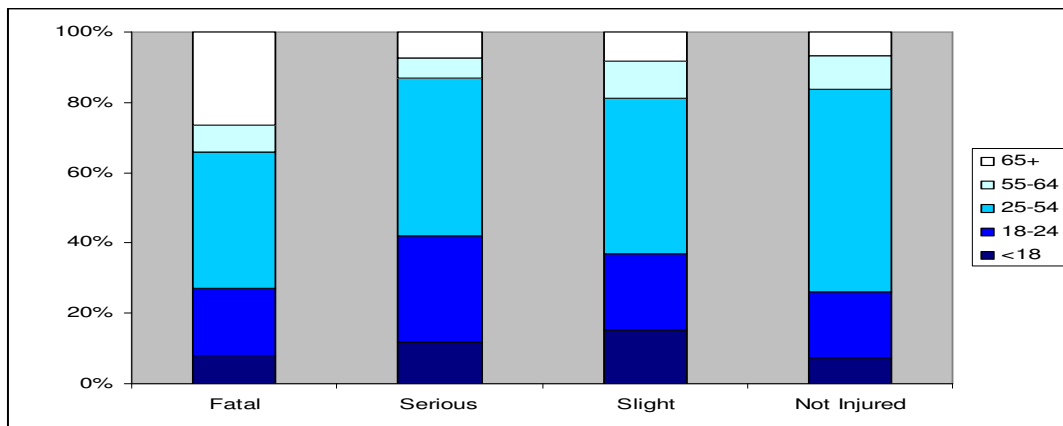


Figure 1.8: Injury severity per road user age group in the FAI data

Furthermore, Figure 1.9 concerning the distribution of different road-user types per injury severity in fatal accidents, shows that pedestrians are very unlikely to survive, once they are involved in a fatal accident. In fact, there were no uninjured pedestrians in the FAI data. Moreover, drivers appears to be more often among the fatalities of fatal accidents, at least when they are among the

casualties. More specifically, when excluding non-injured drivers, drivers are more often fatally injured than passengers, which is not surprising as the driver is more likely to be the only occupant in several cases.

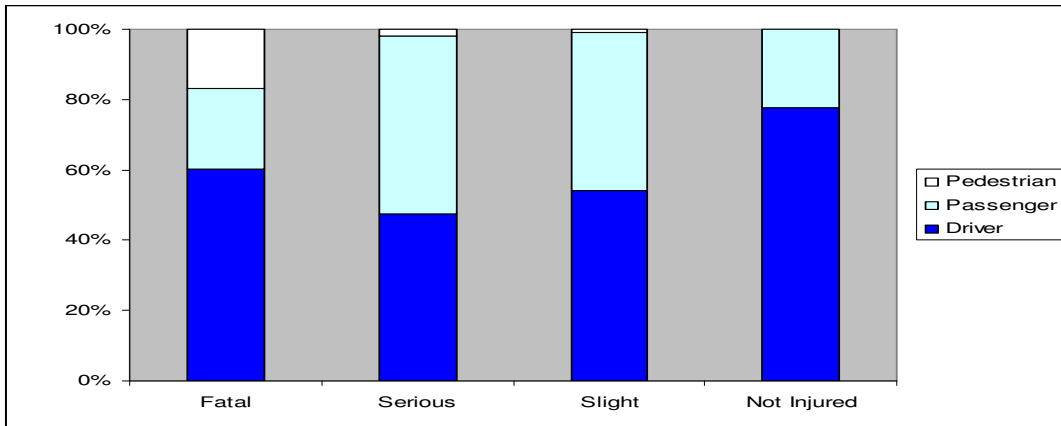


Figure 1.9: Injury severity per road user type in the FAI data

Figure 1.10 presents the effect of road user impairment (alcohol, drugs, fatigue etc.) on is or her injury severity outcome in fatal accidents. Although impaired road users were a very small proportion of all road users in the FAI data, a clear pattern is indicated, according to which impaired road users are more likely to be the fatality of a fatal accident. It is noted that this result only concerns those road users for which impairment was confirmed, and should therefore be considered with some caution.

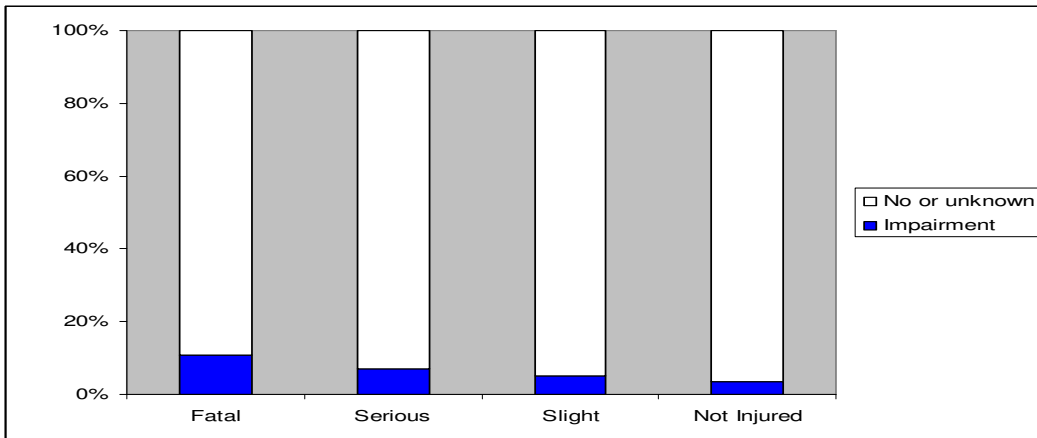


Figure 1.10: Injury severity per road user impairment in the FAI data

In Figure 1.11 it is shown that road users using their seat belt are less often the fatalities of fatal accidents, although they can still be among the casualties. These figures only concern vehicle occupants.

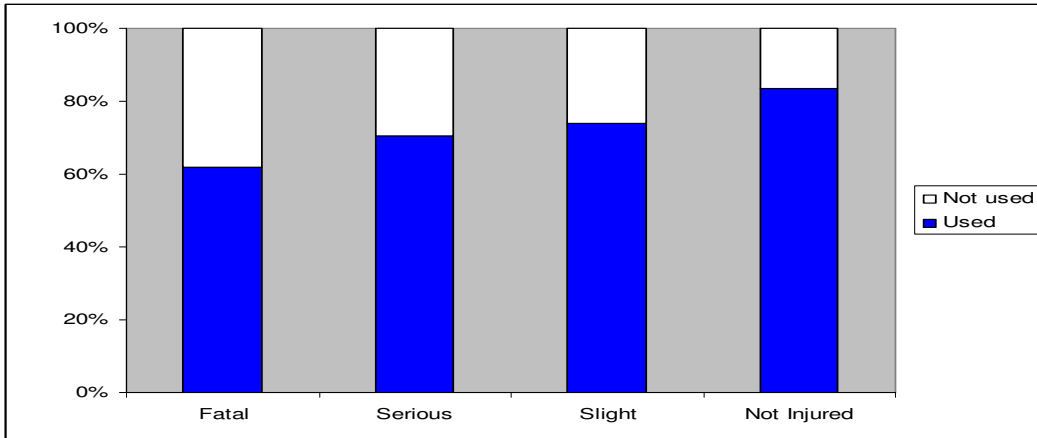


Figure 1.11: Injury severity per seat belt use (where applicable) in the FAI data

Proceeding to some vehicle related basic variables; Figure 1.12 shows that, when there is only one occupant in a vehicle involved in a fatal accident, that person is more often the fatality, compared to persons in vehicles with two or more occupants. This intuitive finding directly corresponds to the accident size effect, which is involved when analyzing fatal accidents alone. Another aspect of this effect is reflected in the number of accident participants.

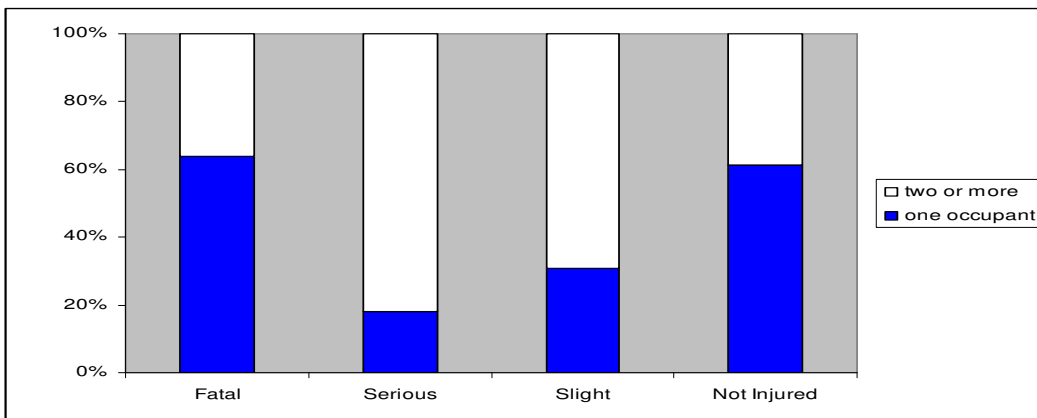


Figure 1.12: Injury severity per number of vehicle occupants in the FAI data

Figure 1.13 confirms the picture shown in Figure 1.9, according to which a pedestrian involved in a fatal accident will most probably be the fatality. Moreover, it shows a similar pattern for moped and motorcycle riders. These road users suffer an increased fatality risk by default, once involved in a fatal accident. The opposite is the case for heavy vehicle occupants (bus, truck etc.), who appear to be far more protected when involved in fatal accidents, whereas no clear pattern can be identified for passenger car occupants.

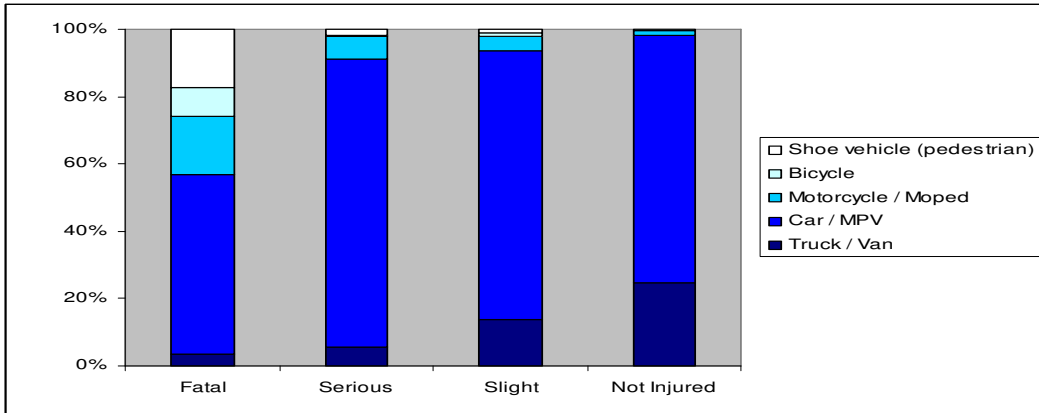


Figure 1.13: Injury severity per vehicle type in the FAI data

When looking at vehicle safety equipment (Figure 1.14), it can be seen that occupants of vehicles equipped with ABS are slightly less often the fatalities of fatal accidents.

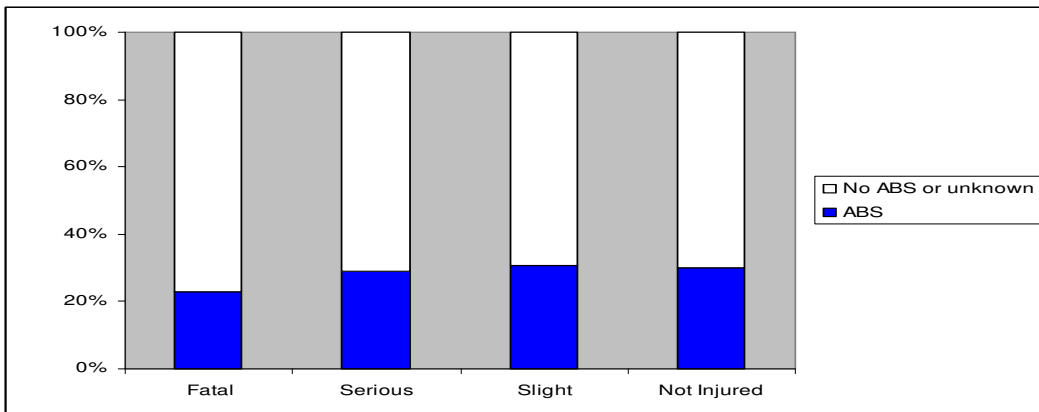


Figure 1.14: Injury severity per vehicle ABS (where applicable) in the FAI data

In Figure 1.15, the distribution of casualties per speed limit category is represented, showing that road users involved in fatal accidents are more likely to be killed when the speed limit is higher than 50 km/h. This effect may be partly due to increased impact speed in accidents within areas of higher speed limits, but may also be due to the occurrence of more single vehicle accidents within these (mainly rural) areas.

As regards the distribution of casualties on roads being physically divided or not in fatal accidents, Figure 1.16 shows no striking difference between fatalities and serious injuries in fatal accidents, which occur equally less often on physically divided roads.

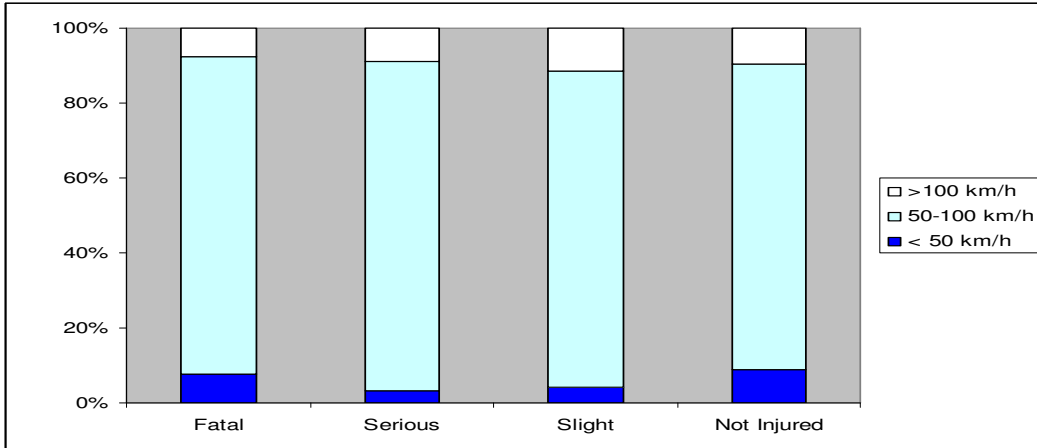


Figure 1.15: Injury severity per road speed limit in the FAI data



Figure 1.16: Injury severity per road separation in the FAI data

Fewer fatalities are associated with accidents at junctions, most probably due to the fact that more than one vehicle are involved in accidents at junctions. These results, presented in Figure 1.17, may be seen as another aspect of the accident size effect.

Finally, Figure 1.18 shows that, the higher the number of events in a fatal accident, the higher the probability of each road user involved in this accident of being the fatality. Moreover, when there is more than one event in a fatal accident, there is also increased probability of being a casualty of the fatal accident. In particular, in accidents of more than one event, the proportions of killed and seriously injured persons do not differ significantly.

The basic descriptive statistics presented above aimed at providing an overall picture of the main factors differentiating between fatalities and non fatalities in fatal accidents. Confirming what was discussed in section 1.1, the probability of being the fatality, given that one is involved in a fatal accident, appears to depend on specific issues: first, the accident size, whose effect can be identified not only in variables directly describing it (e.g. number of occupants), but also in

variables indirectly describing it (e.g. junction or not). Second, the vulnerability of the road user; elderly persons, impaired road users, pedestrians and motorcyclists are far more often the fatalities in fatal accidents. It is therefore shown that these factors need to be explicitly accounted for in statistical analysis, before proceeding in the investigation of other effects.



Figure 1.17: Injury severity at or not at junction in the FAI data



Figure 1.18: Injury severity per number of events in the accident in the FAI data

Other factors affecting the fatality risk were identified during the data exploration (e.g. speed limit, road separation, seat belt use), they did not, however, presented such strong patterns. These factors will be exhaustively tested in the statistical analysis.

Finally, it should be underlined that these descriptive statistics may reveal only part of the picture, given that each variable is examined separately. In the statistical analyses, where all the parameters will be examined jointly, the effects of some variables may eventually appear somewhat different.

1.3 Comparison with fatal accident data in Care:

Christian Brandstätter, KfV

As mentioned earlier, the CARE database is a community database in which every fatal or injury accident occurring in each European member state is stored. For this reason, the CARE data can be considered as the *recorded* population of fatal and injury accidents in Europe (and, by the same token, in each of the member states). Samples of injury or fatal accidents data that are collected in the member states for analytical purpose should consequently adequately “represent” the properties of this population.

In the following section, the distribution of the FAI data and the distribution of the fatal accidents in CARE are compared. The sampling methods applied to collect the FAI data are described in SafetyNet Deliverable 5.1: Fatal Data Methodology Development Report (Brace, 2005). The results of the present comparison are discussed with respect to representativeness of the FAI database.

1.3.1 Data

All data available from the FAI database is used for the following analysis. Table 1.5 gives an overview about the number of road users included in the Fatal Accident Investigation database (FAI data). Table 1.6 gives an overview of the availability of CARE data in the countries where the FAI data were collected, for the years during which these data were collected.

These tables show that it is not possible to compare the two dataset directly on the basis of year and country. Therefore, the following rules for aggregating the data for the comparison tables in the next section are defined:

- The FAI data were aggregated for each country irrespective of the years, (see. the “Totals” for each country).
- The CARE data from the years 2003 and 2004 were used for this comparison, to the exception of the Netherlands, where the data from 2003 only were used.
- Germany is excluded from the following analysis because there is no CARE data available.
- All variables used from the FAI dataset were recoded and re-categorized when necessary to make them compatible with the CARE definitions.

Once these rules are applied, the distributions of the two types of data can easily be compared, account taken of a series of basic characteristics of fatal accidents⁵. These characteristics are (exact definitions can be found in the glossary of the Annual Statistical Report):

⁵ These characteristics were chosen following the rules of 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 (http://www.erso.eu/data/content/accident_statistics.htm#_Accident_statistics).

Code	Country	2003	2004	Total
FR	France	333	0	333
DE	Germany	432	3	435
IT	Italy	814	589	1403
NL	The Netherlands	67	61	128
FI	Finland	59	99	158
SE	Sweden	154	154	308
UK	United Kingdom	357	285	642
	<i>Total</i>	2216	1191	3444

Table 1.5: Data available in the fatal accident investigation 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	United Kingdom	X	X	X	X
	North Ireland	X	X	X	

Table 1.6: Data availability in the CARE Database: X denotes data available.

- Person class (driver, passenger, pedestrian)
- Person killed (age group and/or gender)
- Area type (inside or outside urban area)
- Motorway (yes or no)
- Junction: (here only yes or no, because the counts for each and every type of junction 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)
- Day of week: Instead of the time-related variable in the ASR (hour), the variable “light conditions” was used, given that it is correlated with the hour, and presents higher counts in each of its categories.
- In addition to the variables listed above, the variable “security equipment” was selected.

1.3.2 Results

The comparison of the two datasets in this section is based on the most relevant variables listed above. These 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 Appendix 1.1, which includes the raw data in the upper part and the column percentages in the lower part.

Figure 1.19 shows the results for gender.

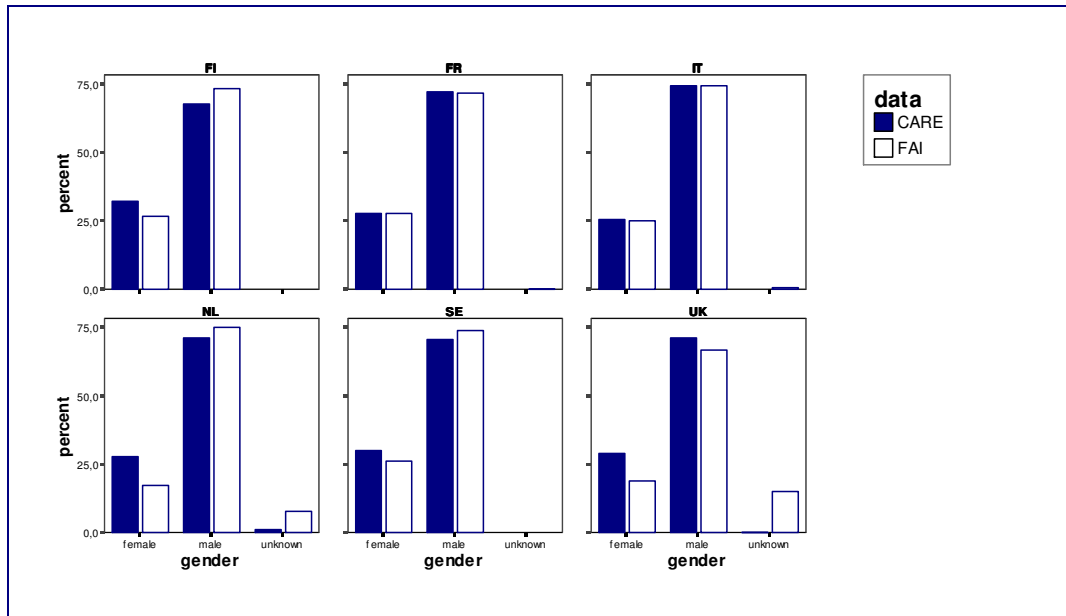


Figure 1.19: Gender

The only remarkable deviation can be found for the UK, but this results from the 15% unknown values in the FAI data. The percentages based only on the valid counts in male and female are 22% for females and 78% for males, which is close to the CARE proportions.

The result for the variable “age of the victims”, categorized in 10 year intervals, is similar, as shown in figure 1.20. The notable differences in the data for UK are, again, the consequence of the large number of unknown age values in this sample. The pattern based on the percentages of “known” age values only is very similar to the CARE distribution.

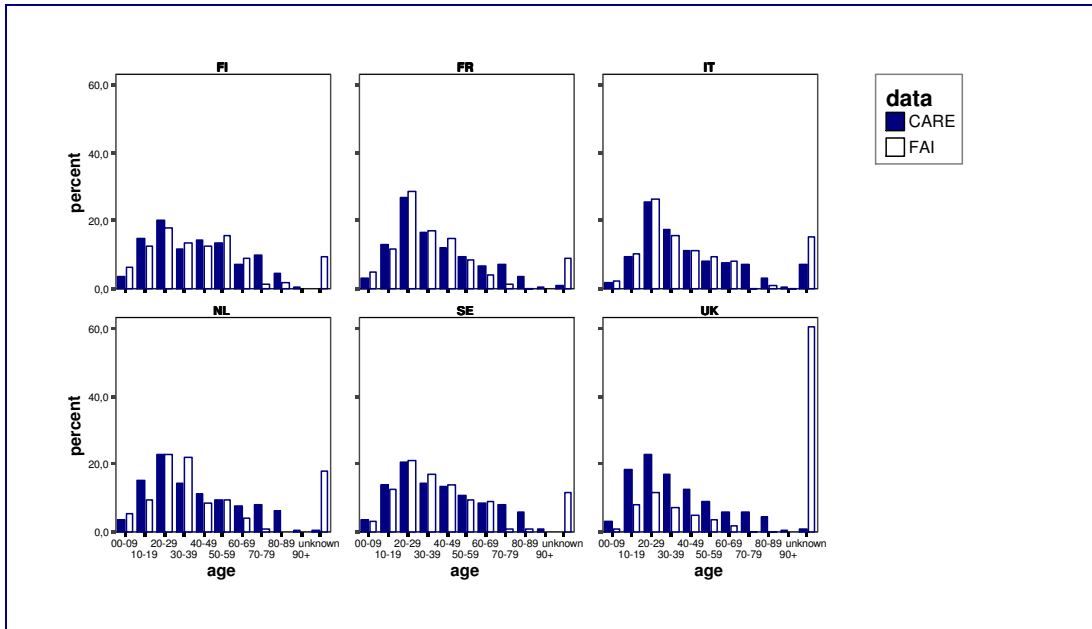


Figure 1.20: Age groups

Concerning the person class (Figure 1.21), the proportion of drivers in the FAI data is a good sample from CARE. Pedestrians, however, are under-represented in the FAI data as compared to CARE, and passengers show opposite patterns in one and the other database. This bias is maybe caused by different recording rules of or different accuracy levels in the information obtained.

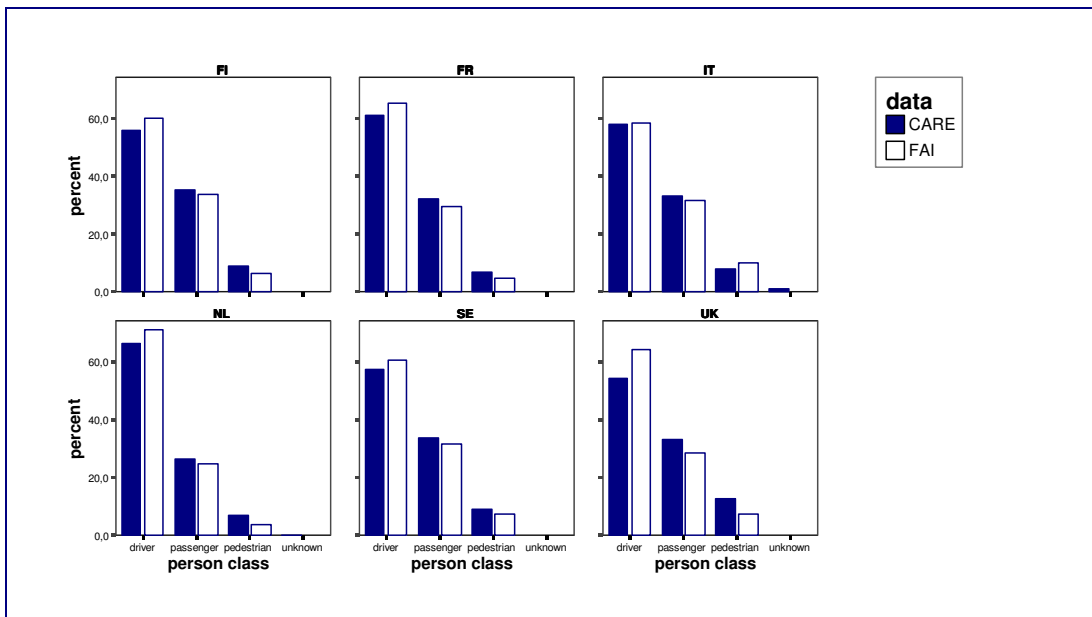


Figure 1.21: Person Class

The distribution for the categories in the variable 'vehicle group' is mostly similar in the two data sources. The percentage of cars involved is slightly higher in the

FAI data than in the CARE data, while the percentage of motor cycles and mopeds is slightly lower (see figure 1.22).

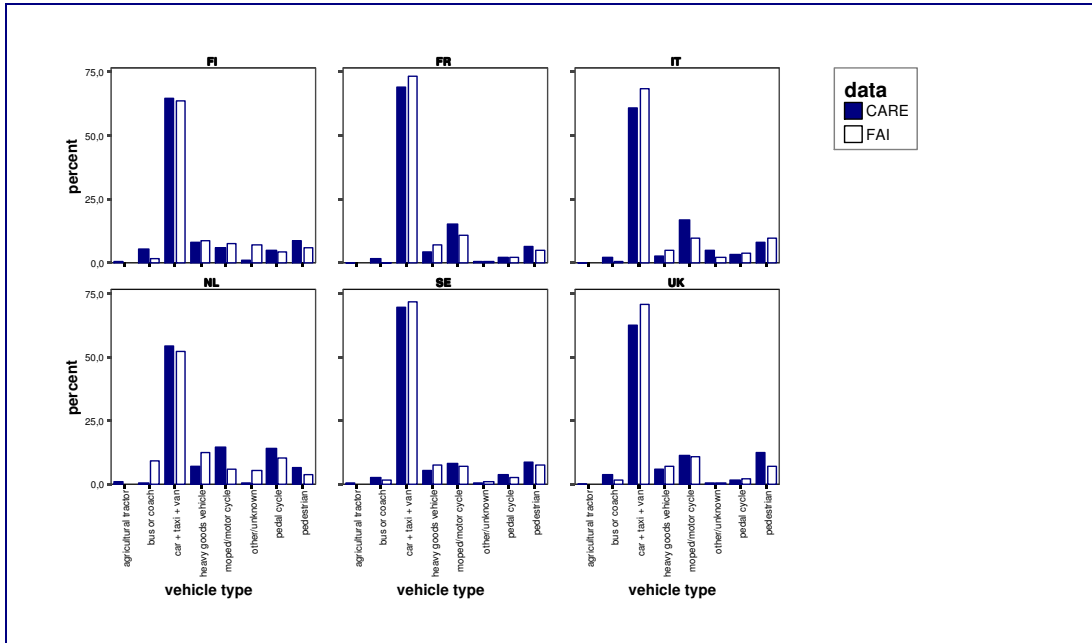


Figure 1.22: Vehicle group

The availability of information about security equipment differs between the countries in the CARE database as shown in Figure 1.23, making it difficult to compare the two dataset in terms of security equipment.

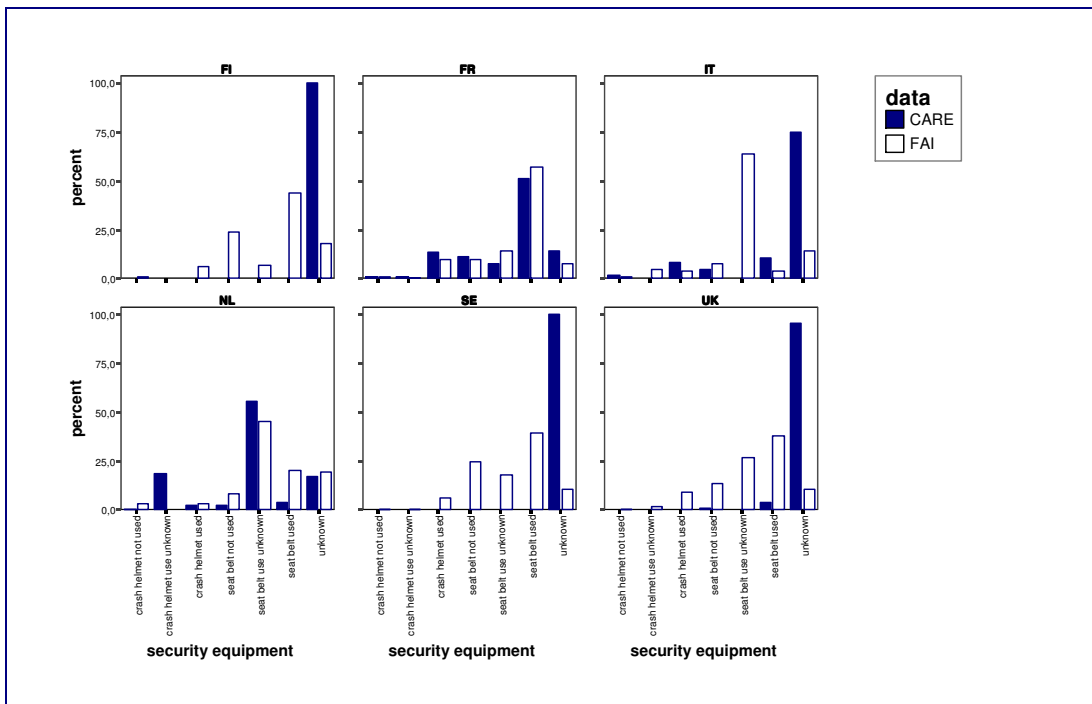


Figure 1.23: Security equipment

Figures 1.24 to 1.26 show variables describing the location of the accidents, figures 1.27 and 1.28 show the distribution of the CARE and FAI observations over the different lightning weather conditions at the time the accident happened. The results are similar: The main pattern of the CARE distribution is reproduced in the FAI sample (e.g.: for Italy, the proportion of fatal accident is higher outside than inside urban areas for both the CARE and the FAI data). There are, however, differences in the exact values of the proportions for the CARE and FAI data (e.g.: the proportion of fatal accidents outside urban areas is 75% for the CARE data, but only a bit more than 50% for the FAI data). The size of these differences also varies from one country to the other. The deviation of the FAI distribution from the CARE distribution for the variable “junction” is due to the large amount of missing values (74.6%) in the CARE data for Sweden for this variable.

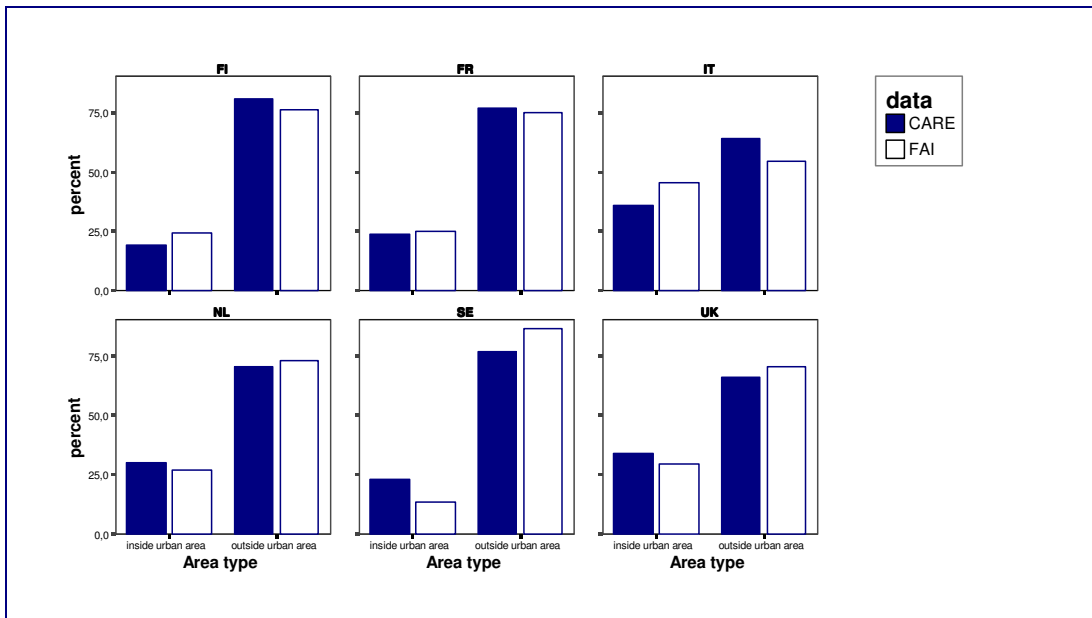


Figure 1.24: Area type

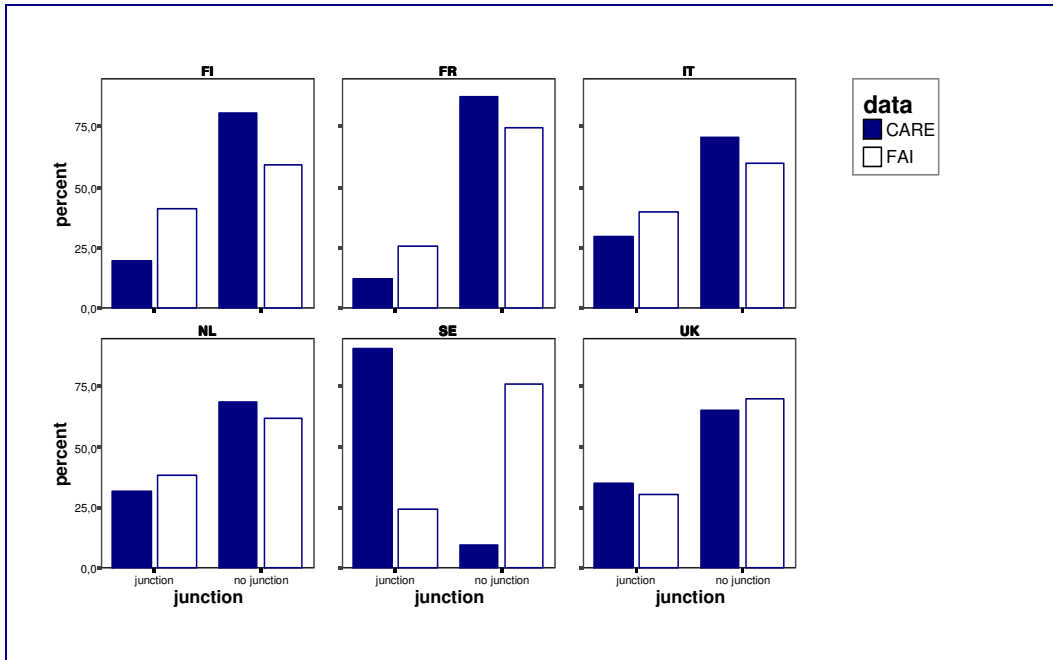


Figure 1.25: Junction

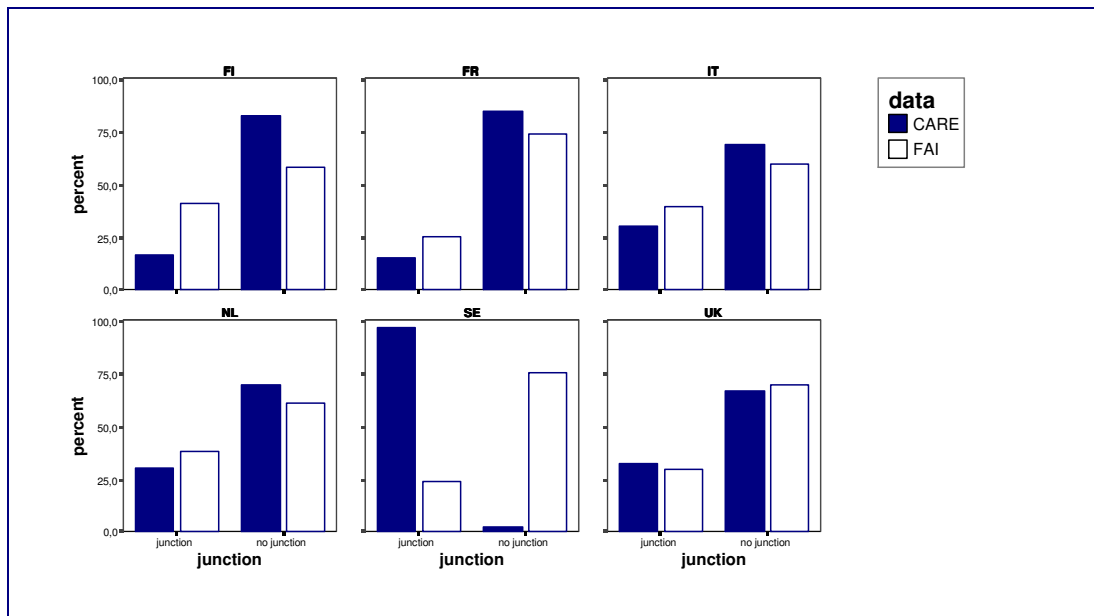


Figure 1.26: Motorway

Analysing European in-depth data: Methodological framework and results

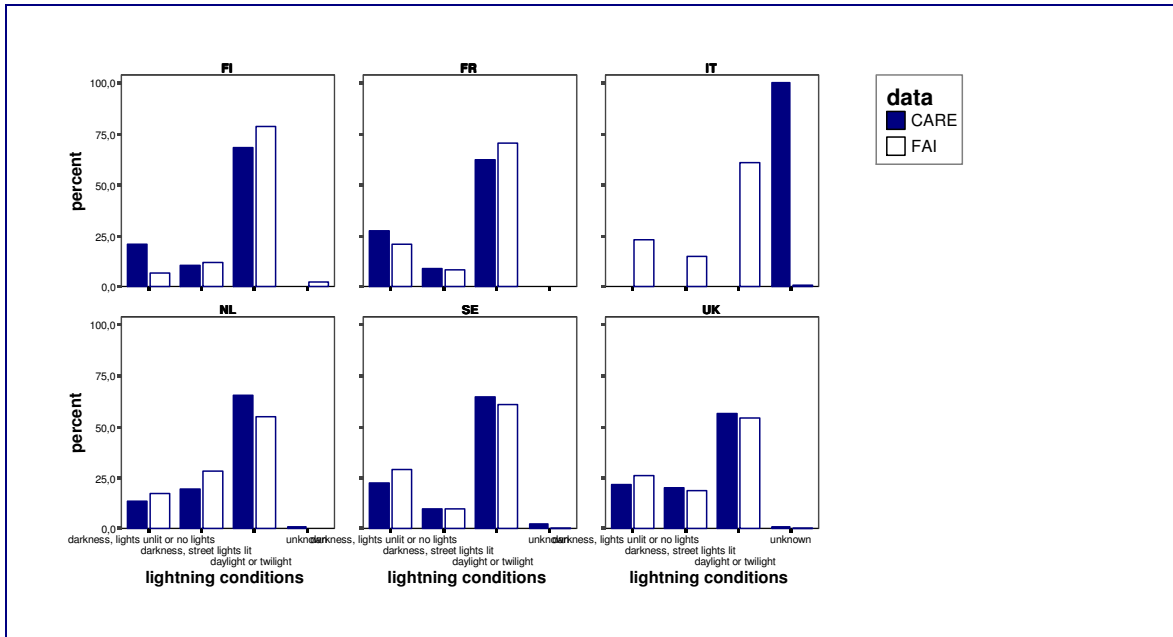


Figure 1.27: Lightning conditions

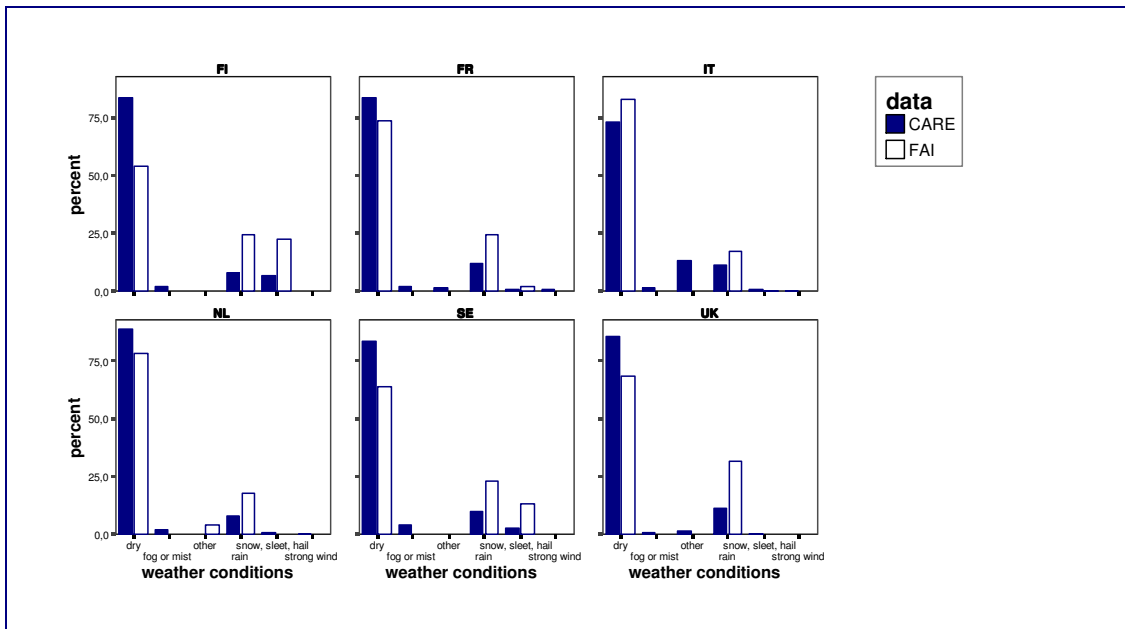


Figure 1.28: Weather conditions

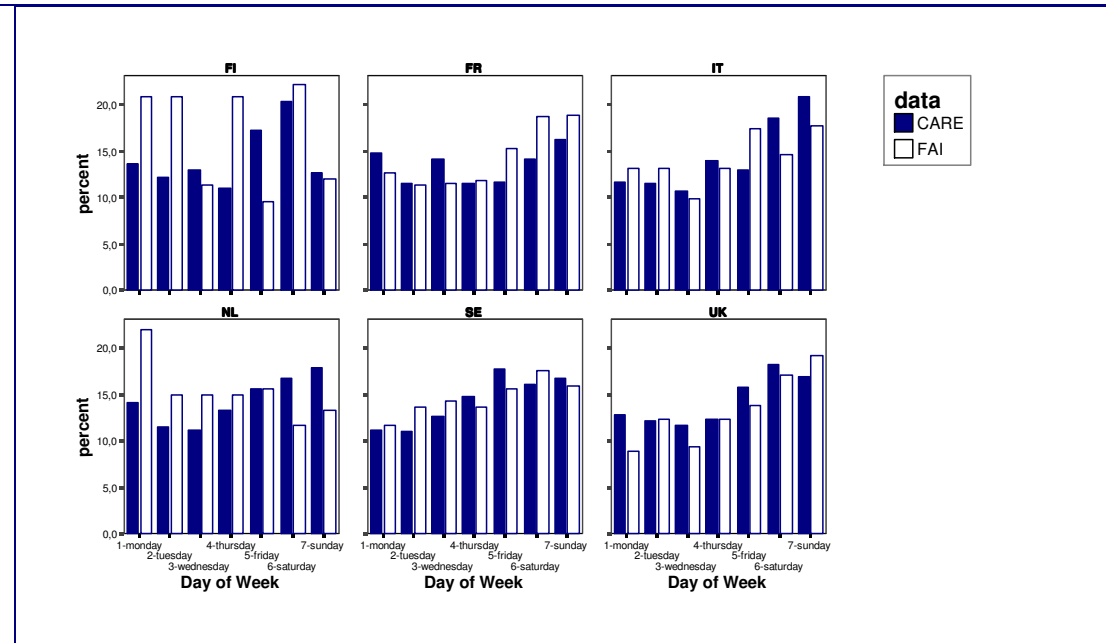


Figure 1.29: Day of the week

1.3.3 Conclusion

To conclude, 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 the CARE and FAI data with respect to the most important variables available in CARE indicates that their respective distributions match relatively well. Consequently, many of the results of the analyses of the FAI data can be generalized to the country in question.

Representativity, however, can never be considered a general characteristic of a given data set. Rather, it has to be re-investigated for each research question chosen 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 is tailored to the analysis to be conducted, and – if necessary – appropriate weights can be calculated to correct for deviations.

1.4 Overview of the deliverable:

This introduction offered a complete description of the the FAI database. It started on the methodological issues posed by the fact that the data are limited to fatal accidents. The structure of the database and the distribution of the observations on several key variables are then described. Finally, the FAI data are compared to the fatal accident data contained in the CARE database.

The next chapter (Chapter 2) illustrates a first possible use of the FAI data, which mainly lies on the enhanced accuracy with which the severity of injury is recorded in fatal accident investigations. The data indeed comprise two distinct severity scores for each road user in the sample: The “usual” one, assigned by the police, and the second, assigned later on by the SafetyNet research team. The analysis presented there consists of an attempt to model the (mis)matches between these two ratings as a function of several factors (e.g.: country, ...) This chapter illustrates how such a model allows to efficiently detect incorrect injury severity reporting.

In Chapter 3, the FAI data are used to investigate the variables that differentiate single from multiple vehicle accidents. Single- and multiple-vehicle accidents indeed appear to be two qualitatively distinct accident types, and single-vehicle accidents represent an important part (37%) of all car accidents in the database.

Chapter 4 is entirely devoted to the analysis of the fatality risk for each individual road-user involved in a fatal accident. This chapter comprises 3 sections, each focusing on the investigation of variations of the fatality risk while correctly accounting for accident size bias and incomparability of risk. In all three sections, the question asked is: What affects the probability for each road-user involved in a fatal accident to end up among the fatalities, rather than among the survivors of this accident.

Chapter 5 essentially aims at summarizing the main findings made on the basis of the different analytical attempts presented throughout the deliverable. The additional possibilities offered by the FAI data are also envisaged there, and recommendations are made in terms of desirable features and augmentation of the data.

Chapter 2 – Misreporting injury severity

George Yannis, Dimitris Katsochis and Eleonora Papadimitriou (NTUA)

2.1 Introduction

In most European countries, road accident data are collected by the Police, who are responsible for determining injury severity. As regards road fatalities, the common EU definition of fatalities within 30 days from the accident has contributed significantly to the reduction of fatality under-reporting, although a small proportion of fatalities may be still not reported. However, no such definitions are available with respect to injuries (serious or slight). Therefore, injury under-reporting or inappropriate reporting is a critical issue in working towards the full harmonization of EU road accident data.

The issue of injury *under-reporting* (i.e. of casualties who are unavailable in police records, but who can be possibly found in hospital records) is extensively treated within SafetyNet WP1, by means of comparisons between macroscopic police and hospital data. However, as regards *inappropriate reporting* (i.e. misclassification of injury severity), little or no information is available. In general, it is acknowledged that there exists a reporting inaccuracy problem (from now on “misreporting”), accounting in many countries for over 50% of all injuries (especially slight ones) (ETSC, 2007). This serves as a basis to proceed for further investigation within the framework of SafetyNet WP7.

In the Fatal Accident Investigation (FAI) database of WP5, two distinct classifications are available concerning accident severity at the level of individual road user. According to the glossary of terms of the database (SafetyNet D.5.5, 2008), there are:

- “Police injury severity”, i.e., injuries or complications directly due to the accident within 30 days of the crash, as recorded by the Police
- “SafetyNet medical outcome”, i.e. overall outcome of the crash, validated by the SafetyNet team on the basis of various data sources (Police, hospital, insurance companies records), as police only follows the situation of each individual’s health for a limited period of time.

In both recording systems, there are four possible outcomes, namely "killed", "seriously injured", "slightly injured" and "not injured". The SafetyNet medical outcome is considered more reliable, for all possible combinations of recorded severity extent.

The objective of this analysis is the identification of the degree of disagreement / mismatch between "Police" and "SafetyNet" scores. Moreover, the analysis aims to investigate whether any prevailing factors emerge that are related to such differences, making the initial "Police injury severity" change when corresponded to the "SafetyNet medical outcome". This research question is rather critical, especially since there have been several practical problems in the determination of the degree of injury severity across different countries, as a

result of the application of national police definitions. In this section, that type of inappropriate injury severity recording is referred to as "misreporting".

The results of such analyses may be a very first step towards the development of correction coefficients for inappropriate severity recording (misreporting), like those that are under development in SafetyNet WP1 for injury under-reporting.

In the present case, two analyses have been performed, on two sets of FAI data:

- The processing of the 1st wave of FAI data yielded important misreporting. The detailed results are included in D.7.6 of SafetyNet (Martensen, 2007).
- Similar modelling attempts with the 2nd (final) FAI dataset gave almost no misreporting.

A summary of the main results of the two distinct portions of work is provided in the following sections. In brief, it may be deduced that, during the transition from the first to the second "wave" of selected FAI datasets for analysis, the WP5 team thoroughly checked all cases and put effort in improving the quality of cases before including them into the database. This is particularly evident in the case of Italy, illustrating the selection of cases stored in a more reliable manner. By the time of the 1st analysis, however, such a careful examination of records was not feasible due to time pressure to ascertain a minimum FAI dataset.

Overall, SafetyNet WP5 has contributed to the collection of in-depth data that allow for detailed and disaggregated analysis of various injury-related aspects. It is therefore possible to examine the effect of several interesting parameters on the probability to eventually record accurate descriptions of the injury severity in a road accident. These factors involve all major components of road systems, namely network users, vehicles and roads. Moreover, some of those are seldom adequately stored in national databases, a feature that adds certain value to this effort.

It is noted that a careful examination of partial cases was required in the present analysis, as there were several rather incompatible cases of misreporting, i.e. combinations of initial outcomes changing category in the end (e.g. cases switching from "slightly injured" to "not injured", as opposed to "seriously injured" changing to "killed"). Such an approach provided insight on the most promising field (data subsets) for further analysis.

2.2 Modelling injury severity misreporting in the first wave of FAI data

In this section, the models developed and the results obtained by utilising the 1st Fatal Accident Investigation database (or FAI data) are presented. The detailed results are available in SafetyNet / WP7 – Data Analysis & Synthesis / D.7.6. – Modelling severity recording reliability through in-depth data. The 1st wave of the FAI data (referred to as FAI 1 from this point onwards) included 943 casualties in fatal accidents. Table 2.1 presents the classification of these casualties per injury severity, according to the two recording systems, the Police score and the

SafetyNet score. The results suggest that important misreporting was present in the 1st wave of the FAI data, especially as regards those casualties that were initially recorded by the Police as "seriously injured". A more detailed exploratory analysis of the data was required to identify the subsets of cases upon which the analysis should focus.

Source	Killed	Seriously Injured	Slightly Injured	Not Injured	Grand Total
Police	328	247	163	205	943
SafetyNet	404	98	165	243	910
Difference	+76	-149	+2	+38	-33

Table 2.1: Distribution of casualties recorded in different severity scores (FAI 1 data)

2.2.1 Exploratory analysis

According to the examination of possible injury severity combinations between the two aforementioned recording systems, there are quite a few cases in which the severity score changes from the Police to the SafetyNet records (see Table 2.2). These represent a proportion in the range of 20-35% of police severity scores as far as "fatal/not injured/slightly injured" scores are concerned. "Seriously injured" constitutes a notably different case, with almost 60% of initial scores changing to some other severity category. Parts of the road users initially rated as "serious" by the police are afterwards categorized either as "slightly" injured (20%), as "fatal" (30%) or as "unknown" (10%) by the SafetyNet team.

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	328					328
Not Injured		201		4		205
Serious	75	2	95	50	25	247
Slight		40	3	111	9	163
Unknown	1				10	11
Grand Total	404	243	98	165	44	954

Table 2.2: Corresponding injury severity for the two severity scores in the FAI 1 data

Note: Figures in the diagonal (grey) present the cases where the original reporting was correct; off-diagonal cells (white) present misreporting.

Further analysis showed that the validity of an analysis that would examine cases from all participating countries in a single subset of the available database is influenced by data from Italy. In fact, further exploration of the data showed that this country, which has contributed to more than 40% of all accidents in the database, reveals some striking differences when compared to the other countries (see Tables 2.3 - 2.4).

SafetyNet Medical Outcome						
Police Injury Severity	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
Fatal	241	0	0	0	0	241
Not Injured	0	102	0	4	0	106
Serious	5	0	80	5	6	96
Slight	0	2	3	87	4	96
Unknown	1	0	0	0	1	2
Grand Total	247	104	83	96	11	541

Table 2.3: Corresponding injury severity for the two severity scores (FAI 1 data): all countries except from Italy (541 road users)

The values of Table 2.3 include Sweden, Germany, France, Finland and the United Kingdom in the FAI 1 data⁶. It is acceptable to treat those road users in a uniform way, since the distributions of injury severity for each of these countries are most similar (see Appendix 2.1). On the other hand, Table 2.4 confirms that Italy follows a very different pattern with respect to injury severity misreporting in the FAI 1 data. The detailed Tables for each country are presented in Appendix 2.1.

SafetyNet Medical Outcome						
Police Injury Severity	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
Fatal	87					87
Not Injured		99				99
Serious	70	2	15	45	19	151
Slight		38		24	5	67
Unknown					9	9
Grand Total	157	139	15	69	33	413

Table 2.4: Corresponding injury severity for the two recording systems (FAI 1 data): Italy (413 road users)

Three main conclusions may be drawn for the FAI 1 data, based on tables 2.3 and 2.4:

- In all countries except Italy, the large majority of road users are on the diagonal, i.e. there are proportionately few differences between the injury severity recorded by the police and by the SafetyNet team.
- In the group of all countries except Italy, cases of misreporting are relatively more frequent for the entries that have been initially rated as “serious” by the police; approximately 11% of these cases ended up in a different severity category for the SafetyNet score.

⁶ The Netherlands contributed cases only in the 2nd wave of FAI data.

- Italy constitutes a striking exception to the generally satisfactory picture, as a large proportion of Police scores were initially incorrect. Especially with respect to seriously injured road users according to the Police, SafetyNet concluded that only about 10% of these were appointed this score correctly. As mentioned, changes in score can take place in both directions (i.e. to a status of either heavier or less severe type of casualty).

The dependent variable considered in the analysis of the FAI 1 data was at first a binary one, indicating whether the two severity scores (Police and SafetyNet) are the same or different. As explained previously, this exploratory analysis is intended to suggest a subset of available data on which analyses would be meaningful. After attempts involving broader subsets, it was concluded that a model only offers some insight for the “seriously injured” cases of Police scores.

Dealing with a not so common dependent variable calls for careful examination of available candidate predictors: Independent variables should be selected on the basis of their potential explanatory value with respect to the specific research question of injury severity misreporting. Once the main promising predictors are identified, more elaborate multinomial models may be fitted to the 3-category response variable, i.e.:

- changing to less severe injury,
- remaining the same,
- changing to more severe injury.

As mentioned in section 1.1, quite a few variables in the FAI database exhibit high overlap in terms of variance, implying multicollinearity if included simultaneously in a model. In order to address such uncertainties, some descriptive analysis of various variables precedes the building of models, so that the latter is performed in an appropriate manner. Relative frequencies (i.e. distributions of variable scores across possible injury severity scores) have been calculated for most variables appearing in the database, often adopting various groupings until some striking difference is observed. The detailed results of this process are included in D.7.6 of SafetyNet.

In the following subsections, logistic regression models are developed to compare alternative combinations of injury severity scores between the police and SafetyNet team records. Binomial regression models are presented, for the probability of the occurrence of misreporting between the police and SafetyNet severity scores. In each case, separate models are developed for Italy and all the other examined countries⁷.

The dependent variable is a binary one (1: same outcome, 0: different outcome) indicating the record of injury severity for each road user between the two

⁷ It is noted that, as a first step of investigation, a model involving individuals classified as “Seriously injured” by the Police had also been produced for the total of all countries. Normally, this would facilitate the detection of any patterns characterising the whole database - e.g. systematic errors, or impacts of general nature. Due to the inconsistency between Italy and the other countries, though, the results had been misleading as the results identified peculiarities of the Italian data set rather than characteristics of the cases where the reporting changed.

recording systems. All explanatory variables have been defined as categorical (see Table 2.5). Categories were formed by means of independent contrasts for the sake of simplicity. It is noted that, for each categorical variable, the first category is used as reference group of the parameter estimates (i.e. the related parameter is set equal to zero). Different coding schemes have been considered, especially for those variables that were first found to be non-significant.

Variable	Values
Same/Different	0: Different Recording, 1: Same recording
Body Region Most Injured	0: Head/Thorax/Multiple, 1: All other (known) cases
Crash Participants	0: 1, 1: >=2
Road User Class	0: Driver / Passenger, 1: Pedestrian
Age	0: 15 - 54, 1: 0 - 14 / >=55
Gender	0: Male, 1: Female
Impairment	0: No, 1: Yes
Resident	0: No, 1: Yes
Familiar	0: No, 1: Yes
Avoidance	0: No, 1: Yes
Motorway	0: No, 1: Yes
Speed Limit	0: <50, 1: >50
Weather Conditions	0: Dry, 1: Wet
Light Conditions	0: Daylight/Dazzling sunlight, 1: Other (known) cases
Carriageway Type	0: Dual divided, 1: Other cases (uniform)
Number Of Lanes	0: 1/direction, 1: >=2/direction
Junction	0: No, 1: Yes
Area	0: Rural, 1: Urban / Mixed
Traffic	0: Light, 1: Normal / Heavy
Vertical Align.	0: Flat, 1: Uphill / Downhill
Horiz. Align.	0: Straight, 1: Bend / Junction / Other
Most harmful event	0: 1 st event, 1: 2 nd -plus event
Vehicle Type	0: 4wheelers, 1: 2wheelers & pedestrian / shoe vehicle
Crash Participants	0: 1, 1: >=2
Road Conditions	0: Dry, 1: Other
Event Type 1	0: Non-collision, 1: Collision
Accident Day	0: Weekdays, 1: Weekend

Table 2.5: Variables and values used in the analysis

2.2.2 A binomial model for Italy

Italy contributed 132 known cases in the FAI 1 database (19 more were unknown according to SafetyNet and therefore excluded from the analysis).

The model yields the probability of observing the same severity score from the Police and the SafetyNet team in relation to the explanatory variables. As a part of the process, all variables were initially tested alone, in order to see whether they are significant when no other effect is present.

The best performing model is presented in Table 2.6.

There is certain improvement from the "empty" model in terms of fit, as indicated by the likelihood ratio. Its value is reduced from 93.470 to 74.288, an important reduction: Since 4 degrees of freedom are introduced into the model, the expected reduction would be 4. Another useful indicator of the quality of the model can be obtained by means of the percentage of correctly classified cases; more than 91% of the outcomes are correctly predicted by the model (98.2% of different and 40% of matching scores)⁸. These results are quite satisfactory. The 'empty' model predicted correctly only non-matching cases (around 88% of total cases).

Variables	Parameter estimates			
	B	S.E.	Sign.	Exp(B)
Traffic(normal/heavy)	-1.791	0.628	0.004	0.167
Traffic(light)
Vehicle Type(pedestrians-riders)	-1.550	0.830	0.062	0.212
Vehicle Type(occupants)
Junction(yes)	-1.103	0.670	0.100	0.332
Junction(no)
Gender(female)	-1.643	0.850	0.053	0.193
Gender(male)
Constant	0.150	0.563	0.790	1.161

Table 2.6: Parameter estimates of the best fitting binary logit model (probability of misreporting - Italy)

Some interesting conclusions may be drawn:

- The heavier the traffic, the more likely it becomes to observe misreporting of injury severity between Police and SafetyNet.
- The same appears to hold for the presence of a junction.
- Non-matching scores are also more frequent for female road users. No straightforward interpretation may be applied, at least not before further investigation is carried out by means of a multinomial model
- Two-wheelers riders and pedestrians are much more likely to have their injury severity changed than vehicle occupants. This appears reasonable as far as the change from serious injuries to fatalities is concerned. A multinomial model would be useful in verifying that. The corresponding model is analytically presented in section 6.4.1 of D.7.6.

These results may be considered to suggest that, the more complex the conditions of the accident, the higher the probability of different severity scores between the police and SafetyNet. It may be the case that higher traffic volumes (and consequently more accident participants), the presence of junctions etc. make data collection and classification a more complex task for the police, increasing the probability of errors in recording. This may be better understood

⁸ The low prediction of cases with not-changing score may be attributed to the fact that matching entries are very few (10% of the total). Subsequently, a model cannot capture very well individuals with injury severity status remaining "seriously injured".

when considering that all FAI cases concern fatal accidents, which are already more complex cases for the Police carrying out the accident investigation.

2.2.3 A binomial model for all other countries

All other countries except Italy contributed 90 known cases (6 more were excluded from the analysis as unknown). The results of the analysis are presented in Table 2.7.

Variables	Parameter estimates			
	B	S.E.	Sign.	Exp(B)
Age(0-14 / 55+)	-1.689	0.776	0.030	0.185
Age(15-54)
Light Conditions(dusk/night)	2.087	1.129	0.065	8.064
Light Conditions(daylight)
Area(urban/mixed)	-2.062	0.980	0.035	0.127
Area(rural)
Constant	2.666	0.632	0.000	14.378

Table 2.7: Parameter estimates of the best fitting binary logit model (probability of misreporting - other countries)

Since the standard likelihood ratio (as a statistic test) equals 62.790 for the empty model and the likelihood statistic for this model equals 49.45, a chi-square test verifies that the three selected predictors produce a really improved model (reduction by 13.341 with three degrees of freedom).

About 91% of the cases are correctly predicted by the model (100% of matching and 20% of non-matching scores). Since the model is developed around the prediction of the majority of cases - i.e. the matching ones -, there is room for improvement concerning the prediction of the non-matching cases.

Similarly to the preceding analysis for Italy, in the case of all other countries the main findings are the following:

- The absence of daylight appears to enhance matching scores between the two recording systems. It is likely that Police recording is more careful during the night, in an attempt to compensate for the more complicated conditions during the night.
- It appears that there is increased probability to obtain different score eventually for individuals who are either very young or rather old (reference age group: 15-54). This is a significant and rather strong effect. Some justification could be provided by the fact that children and aged people are often more vulnerable to deteriorate when injured; this addresses the shift from the state of injured to that of killed.
- The same observation holds for individuals participating in collisions occurring in urban or mixed areas. In the following section it will be investigated whether this finding is working towards both directions (heavier or lighter severity) in a multinomial model.

Although the significant predictors are rather few, it may also be suggested that, the more complex the conditions of the accident (e.g. urban environment), or the more vulnerable the road user groups (children/elderly), the higher the probability of different severity scores between Police and SafetyNet.

In addition to the models presented above, multinomial models have been produced for Italy and the remaining countries respectively, focusing again on misreporting of serious injuries only, with a multinomial dependent variable (0: change to fatality, 1: same outcome, 2: change to slight or no injury). In this case, the effect for each of the independent variables was tested for two contrasts between each of the “change” categories and the “same” category (0 vs. 1 and 2 vs. 1). This process serves as a means to investigate whether the parameters identified in the binomial models work on one or both directions (final injury severity status heavier or lighter than “seriously injured”). This is a meaningful order to follow, as some factors may only work towards one direction (misreporting implying over- or under-estimation of the injury severity), while others may work towards both over- and underestimation. A complete presentation of corresponding analyses may be found in the Deliverable D.7.6.

2.3 Modelling injury severity misreporting in the final FAI data

The results obtained, with the subsequent discussion presented in this section, refer to the 2nd (final) Fatal Accident Investigation database (referred to as FAI 2 data). This database includes around 3,400 individuals involved in fatal accidents. Table 2.8 present the classification of the injury severity of these individuals according to the two examined scores.

Source	Killed	Seriously Injured	Slightly Injured	Not Injured	Grand Total
Police	1,367	511	460	855	3,193
SafetyNet	1,449	494	481	903	3,327
Difference	+82	-17	+21	+48	+134

Table 2.8: Distribution of casualties recorded in different severity scores (FAI 2 data)

The figures of Table 2.8 exclude cases stored as unknown. SafetyNet involves somewhat more road users than Police, which is reasonable given the SafetyNet records are assigned after a longer period of observation after each accident occurred. There are notable differences, at all levels of injury severity. However, comparing the two set of records with the “unknown” observations included may be slightly misleading, as it does ensure that each pair of records relate to the same observation. If the identity of each participant in the examined casualty-involving accidents is followed, meaning that injury severity is compared for the same individuals in the two recording systems, the precise picture of cases resulting in a different status is obtained. Leaving out cases recorded as unknown in either of the two recording systems, yields the results of Table 2.9.

The four columns with the possible severity levels indicate the final outcome for those specific individuals listed in the Police records whose injury severity could be tracked all the way, i.e. where not registered as unknown, either by the Police, or by SafetyNet. In other words, this table describes the different distribution of a common set of recorded individuals' injury severity (i.e. leaving out unknown cases) across the two systems. The main macroscopic finding concerns the shift of some cases characterised as "seriously or slightly injured" to "killed" or "not injured". A more detailed picture is obtained by Table 2.10, in which a cross-classification of the two severity scores is presented.

Source	Killed	Seriously Injured	Slightly Injured	Not Injured	Grand Total
Police	1,367	504	452	853	3,176
SafetyNet	1,383	480	450	863	3,176
Difference	+16	-24	-2	+10	0

Table 2.9: Status evolution of casualties with initially known result – Police data

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	1,366		1			1,367
Not Injured		848		5	2	855
Serious	16	2	473	13	7	511
Slight	1	13	6	432	8	460
Unknown	66	40	14	31	100	251
Grand Total	1,449	903	494	481	117	3,444

Table 2.10: Corresponding injury severity for the two severity scores in the WP 5.2 database (FAI 2) - Note: Cells in blue italics represent the two injury categories (Police) of higher interest to the performed analysis

It appears that there are a number of changes of the injury level recorded, depending on whether the Police or the SafetyNet recording system is considered. This does not hold for entries initially stored as "fatal/not injured" (negligible changes, in the order of 0-0.5%). "Seriously/slightly injured" constitute two categories of higher interest, as almost 6% and 5% of the initial scores change to some other severity category. In fact, this change is working towards both directions, distributed as follows:

- **Serious injuries:** 3% of Police entries change to "fatal", with another 3% resulting to "slightly/not injured".
- **Slight injuries:** 1.5% of Police entries change to "fatal" or "severely injured" and 3% to "not injured".

Unlike the analysis presented in the previous sections, the validity of an analysis that would examine cases from all participating countries in a single subset of the available database is moderately questioned, since less striking differences are observed among countries. To verify that, Tables 2.11 - 2.12 present injury severity distribution for all countries (except Italy) and Italy alone respectively. Given that the distributions of injury severity are relatively homogeneous across all countries, the analysis was performed here on all countries including Italy.

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	818					818
Not Injured		446		5	2	453
Serious	11	2	237	8	7	265
Slight	1	9	6	258	8	282
Unknown	66	40	14	31	72	223
Grand Total	896	497	257	302	89	2,041

Table 2.11: Corresponding injury severity for the two severity scores (FAI 2 data): all countries except from Italy

Table 2.11 depicts cases from Sweden, Germany, France, Finland, Netherlands and the United Kingdom. Similarly to the 1st FAI data those records are analysed together, exhibiting similar distributions of injury severity (see Appendix 2.2). At the same time, Italy (Table 2.12) follows a rather similar pattern in terms of accuracy (with even better matching, proportionally fewer unknown cases). The detailed Tables for each country are presented in Appendix 2.2.

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	548		1			549
Not Injured		402				402
Serious	5		236	5		246
Slight		4		174		178
Unknown					28	28
Grand Total	553	406	237	179	28	1,403

Table 2.12: Corresponding injury severity for the two recording systems (FAI 1 data): Italy

Three main conclusions may be drawn, based on tables 2.11 and 2.12:

- In all countries the table is almost diagonal, i.e. there are proportionately few instances of different injury severity between the two scores.
- This is more notable in the case of Italy, which has contributed alone 2/3 of all other cases together (40% of total). In fact, Italy's matrix is extremely coherent, only showing 28 cases with "unknown" outcome (common in both databases as "unknown" from the beginning).
- The most frequently changing status is the Police entry of "severe injury", with quite a few cases changing to another score (8.1% and 4.2% respectively in Tables 2.11 & 2.12). This is followed by initial entries of slightly injured road users according to the Police, with changing scores of 5.8% and 2.2% respectively. As mentioned, changes in score involve both directions (i.e. to a status of either heavier or less severe type of casualty).

This time, Italy has contributed much more uniformly to the data collection, with proportionately fewer cases (as compared to the other countries) switching to some another injury severity level, according to SafetyNet. This was the result of an extensive effort of the SafetyNet team to enhance the quality of the cases included in the FAI 2 data, allowing in the end to delete from the database a large part of the FAI 1 cases presenting injury severity misreporting.

Obviously, the significantly reduced degree of misreporting in the FAI 2 data does not call for very detailed analysis. Nevertheless, modelling efforts were carried out; since the primary objective is the validation of the effects identified through the analysis of the FAI 1 data (see Appendix 2.4). A binomial model, limited to entries initially rated as “seriously injured” from all countries including Italy, was eventually examined.

As it was the case in FAI 1 data, several variables in the database are likely to be correlated. Similarly to the analysis of section 2.2, this problem was addressed through a preliminary descriptive analysis of various variables, to enhance the robustness of the models built. The distributions of possible outcomes across possible injury severity scores have been calculated for several variables, with varying grouping until some striking difference is observed, as done in the FAI 1 data analysis. A representative subset of variables is presented in Appendix 2.3.

The results confirmed the initial suspicion that such a small degree of misreporting is most likely to be random. Thorough exploratory analysis of the FAI 2 data (presented in Appendix 2.3) resulted in the selection of some possible predictors, after having checked for correlations among variables (see Appendix 2.5). However, only the effect of the “Taken to hospital” variable has been confirmed, as expected beforehand. In particular, it was found that individuals taken to hospital exhibit a higher probability to keep the injury severity score they had originally been assigned by the Police. A possible interpretation is that failure to be transferred to the hospital may cause avoidable loss of lives - although misreporting may involve changes to lighter injury severity extent as well.

Some other predictors with significant impact were identified, including age group (confirming the FAI 1 analysis results), number of occupants in vehicle, road conditions (dry/wet) and vertical alignment. However, the model’s fit is not adequate to consider it a predicting tool. In fact, it can hardly be perceived as an explanatory one. Consequently, the interpretation of these effects might not be meaningful and is not presented here. The tests of individual variables effects in the FAI 2 data are presented in Appendix 2.4.

The modelling attempt described above is treated with great reserve, as the influential predictor (“Taken to hospital”) resembles a constant term, given that only very few individuals were not transferred to the hospital. Hence, the large coefficient of this variable may be due mostly to the prediction of matching cases - as they constitute the vast majority of the observations. It would be interesting to have a comparable number of cases of individuals that were not

taken to the hospital. This would probably allow for a much more complete approach on different severity scores between Police and SafetyNet.

In spite of several efforts to work with all available potential predictors, it has not been possible to produce a multinomial model worth presenting in this report, due to the very small degree of misreporting in the FAI 2 data. Moreover, the aforementioned explanatory variables (that participate in a binomial model somewhat fitting the considered dataset) do not behave equally well when incorporated in a two-directional multinomial model.

2.4 Conclusion

The analyses performed on the 1st FAI data mainly revealed two issues with respect to the investigation of accurate injury severity reporting:

- The selection of an appropriate subset of variables
- The interpretation of conclusions/implications provided by such an analysis

From the results obtained by that time, a general trend could be identified, according to which, the more complex the accident and the accident site, and the more vulnerable the road user, the higher the probability of injury severity score to be different between the police and SafetyNet. An additional issue that needs to be addressed is whether score differences are mainly due to recording bias (e.g. the Police may tend to record severity incorrectly under some conditions), or to the lack of a sound definition of injury severity (making it difficult to identify the correct severity score).

Summarizing the results of the binomial models, it is interesting to note that the "person's age" variable was most significant for all countries but non-significant in the model for Italy. It may be that, in the other countries, serious injuries' reporting problems come from the type of injury and not from reporting errors as such. It is reasonable to assume that the scores obtained by the Police in the other countries are mainly influenced by special injury features alone, as non-matching scores only represent a very small proportion of total cases. On the other hand, additional parameters related to the type of accident are dominant in the Italy model, suggesting presence of recording bias (partly verified at the later stage).

Multinomial models were built as well, aimed at further analysing the major conclusions drawn through the respective binomial models. All predictors that appeared to be significant in the respective binomial models have been tested in the multinomial ones. In most cases, it was found that the impact of the selected variables on the dependent variable was practically of the same kind (i.e. towards matching or non-matching of Police and SafetyNet records), but not always of the same magnitude and direction.

However, additional predictors that were not found significant in the binomial models could be significant in the multinomial models. For example, if "females" have an increased probability of "overestimation" of their injury severity by the Police in relation to "males", and at the same time a decreased probability of an

“underestimation” of their injury severity by the Police in relation to “males”, then because the two types of misreporting are pooled together in the binomial model, the effect of “gender” would not appear significant in the case of the binomial model.

Overall, the 1st wave of FAI data included several cases presenting injury severity misreporting, especially as regards Italy. The analysis of these cases allowed for some interesting indications to be obtained, as regards the presence of a few non-random factors in misreporting. However, the final FAI data presented a very small degree of misreporting, as a result of the extensive effort put in by the SafetyNet WP5 team, and especially the Italian team, to enhance the quality of the data, by collecting much more improved cases, eventually allowing deleting from the database most of the initial cases presenting misreporting.

Modelling attempts on the 2nd and final FAI dataset showed that the prevailing factor affecting misreporting is the “taken to hospital” variable, which reduces the probability of misreporting by the Police. There is one additional independent variable whose effect was found consistent with the one identified in the FAI 1 analysis; in particular, it was confirmed that the more vulnerable the road user, the higher the probability of injury severity score to be different between Police and SafetyNet. A general trend on the complexity of the accident and the accident site, identified in the FAI 1 analysis, cannot be identified or validated this time. Despite several attempts to come up with some explanatory causes, it should be stressed that the degree of misreporting in the FAI 2 database is very limited overall. At least in the context of the examined cases, this phenomenon seems to be probably almost random, not calling for further statistical analysis.

It is noted that, given the particularity of the Italian data, as far as “Police injury severity” is concerned, the authors have asked for some clarifications on the precise definition of the term and the data collection process in Italy. Compared to the previous analysis performed in terms of D7.6, Italian data were extremely improved in terms of accuracy. The Italian team confirmed that the FAI 1 cases included several cases dated before year 2003, a time where the common 30-days definition of fatality has not been adopted in Italy. An interesting observation on the particular data sub-set, also reflecting the thorough data selection process, is that all cases of serious injuries involve individuals transferred to a hospital.

Overall, however, the analysis of the FAI 1 data illustrated that differences in severity scores are largely observed due to recording bias (e.g. the Police tends to record severity incorrectly under some conditions). The lack of a sound international definition of injury severity also hinders the identification of the correct severity score. At least partly, the examined observation may simply be attributed to the fact that in some cases severe injuries have an increased probability of becoming fatal ones - a hint that also deserves further investigation.

Concluding the above, it may be stated that the whole sequence of reporting reliability analyses carried out within SafetyNet WP7 constitutes a success story, since the homogeneity of stored FAI datasets has been enhanced, leading to uniform level of accuracy. This is important because it allows for more reliable comparative analyses on the basis of in-depth data. In spite of the fact that the 2nd analysis was practically hindered from yielding some notable, measurable finding, due to this improvement of data consistency, the selected topic is still considered most interesting and is worth attracting more attention in the future because data quality is crucial when trying to interpret repeated occurrences and other patterns in the road safety context.

In any case, it would be most interesting to have a larger data subset of individuals initially stored as seriously injured. Furthermore, it would open many opportunities to explore the issue in a more complete manner if the “Taken to hospital” variable were utilized. For this to happen, data should include at least some 15-20% of total seriously injured individuals not transferred to a hospital according to traffic Police. Such a dataset might provide more insight on the degree of Police reporting errors in relation to other factors making individuals eventually change injury severity scores.

Chapter 3 – Contrasting single and multivehicle fatal accidents

Heike Martensen, Emmanuelle Dupont, & Lars Akkermans (IBSR)

3.1 Objective

In the introduction, it has been described how in a fatal accident database the size of an accident affects the fatality risk, i.e. the proportion of people killed among those involved in the fatal accidents. The next chapter will be dedicated to identifying the factors that affect the fatality risk while controlling for the size of the accident. In Sections 4.2 and 4.4, accident size will be controlled for statistically and in Section 4.3 by considering only accidents with two cars. By doing so, all information that is somehow related to the size of the accident is filtered out of the analysis.

The question which factors characterize accidents of different size is however, interesting in its own right. In particular, fatal accidents involving only a single vehicle differ in many variables from accidents involving two or more participants. Roughly speaking suggested explanations can be sorted into two categories: (1) conditions of the road and (2) characteristics of the road users involved.

With respect to *road conditions*, Ivan, Paupathy, and Ossenbruggen (1999) demonstrated that the traffic flow (Level of Services) was an important predictor for the number of single vehicle injury accidents. For more busy roads there were less single vehicle accidents than for empty roads. The authors also demonstrated that this relation does not hold for multivehicle injury accidents. Single vehicle injury accidents have also been associated with narrow lanes and narrow road-shoulders (Persaud & Mucsi, cited in Ivan et al., 1999) and they are reported to take place outside of rural areas on the road sections between junctions (in short road sections), while multivehicle injury accidents tend to happen at junctions in urban areas (Wegman & Aarts, 2006). Another difference between single and multi vehicle accidents concerns the time of the accident. Single vehicle accidents are more likely to happen at night and during the weekends (Casteels & Scheers, 2008, Öström & Eriksson, 1993, Persaud & Mucsi, cited after Ivan et al., 1999). All these variables can be deduced from one principle: Single vehicle accidents, rather than multivehicle ones, occur when vehicles are less likely to encounter each other and the error of one driver is less likely to involve another driver.

With respect to *road user characteristics*, Öström and Eriksson (1993) found a whole range of differences between drivers that had been severely injured or killed in single vehicle (SV) accidents as opposed to those in multivehicle accidents. Young, male, and intoxicated drivers were found to be more likely involved in SV accidents. Moreover the drivers from SV accidents more often had liver diseases that are associated with permanent alcohol abuse, did not wear a seatbelt, and had no driver's license. These variables sketch the picture of irresponsible drivers being overrepresented in single vehicle injury accidents.

This agrees with the idea that drivers in single-vehicle accidents were (on average) more responsible for the accident than drivers in multi-vehicle accidents. The logic behind this is the assumption that in single-vehicle accidents all errors leading to the accident have been made by the one driver involved. In contrast, in multi-vehicle accidents, there are drivers involved who have not made mistakes (or at least the mistakes were more evenly distributed between the various drivers involved). On average, the drivers in single-vehicle accidents have made more (severe) mistakes than the drivers in multivehicle accidents. The variables that differentiate the two types of accidents could, therefore, be seen to characterize drivers that are more prone to making errors than others.⁹

While the reasoning about responsibility in single-vehicle accidents should principally apply to fatal and non-fatal accidents in the same way, it remains to be explored whether the factors described above that have been found to play a role in injury accidents (i.e. fatal and nonfatal ones) have a similar results in the FAI database, which consists of fatal accidents only. In the present study we will model variables of different types jointly. This allows us to weigh different accounts of single vehicle as opposed to multivehicle accidents. In particular it is interesting to determine whether variables concerning the road conditions and variables concerning the road users are equally important in understanding the differences between single- and multivehicle accidents.

3.2. Methodology

3.2.1 Selection criteria

In a preliminary screening of the data, it turned out that accidents involving non-motor vehicles or pedestrians all show the same pattern: Two participants, of which one (the vulnerable road user) is killed. While the motor vehicles all have a substantial proportion of single-vehicle accidents, the pedestrians and pedal cyclists do not. To separate factors that differentiate between single multivehicle accidents from those that differentiate between accidents involving vulnerable road users and those that do not, all accidents involving non-motorised vehicles have been excluded from the following analyses. Moreover, all accidents for which there was no information on the time when the accident took place¹⁰ or on the traffic flow or whether the drivers had been drinking, were otherwise impaired or were unfamiliar with the area were excluded, leaving 772 cases of which 275 were single vehicle accidents and 497 multiple vehicle accidents.

3.2.2 The dependent variable

The number of crash participants (either vehicles or pedestrians) ranges from one to five. The number of accidents as a function of the number of crash-participants is presented in Figure 3.1.

⁹ One should, however, keep in mind that this is a very indirect way of reasoning, as the drivers in multi-vehicle accidents that were included in the fatal accident investigation database have been involved in a very severe accident as well and should not be seen as “exemplary”.

¹⁰ This forced us to exclude all cases that had been collected in Sweden.

As can be seen in Figure 3.1, the majority of the accidents involved either one or two participants. There are only a few accidents with 3 or more participants. It was therefore decided to model only the difference between accidents with one participant and those with two or more participants. We consequently have a dichotomous dependent variable “single vehicle accident” that takes a zero value for all accidents with more than one participant and a one for all single-vehicle accidents. This variable is used as the dependent variable in a logistic regression model.

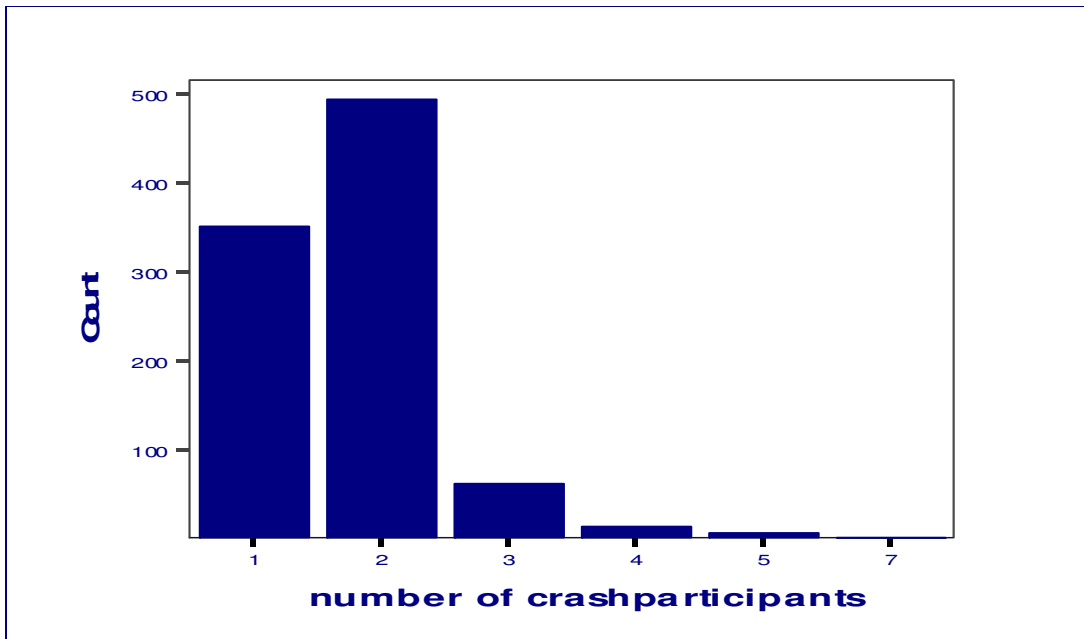


Figure 3.1 The number of accidents with 1, 2, 3... crash participants

3.2.3 The model

The dependent variable <single-vehicle accident> was entered in a logistic regression model. Variables that might affect the number of vehicles involved were entered as independent variables. The predictors were selected with two stepwise selection procedures (see below). All independent variables selected this way explained a significant part of the variance in the dependent variable over and above what the other variables in the model already can explain. That means that the effect of each variable is corrected for the effects of all the others.

The independent variables considered often explain the difference between single – and multi-vehicle accidents in a non-unique way. For example, the variable “Daytime” differentiates between single and multi-vehicle accidents (single vehicle accidents are more frequent during the night), but so does the variable light (more single vehicle accidents when it is dark or twilight). Of course, these variables are not independent of each other: It is more often dark during the hours defined as “night” than during those hours defined as “day”. When entering both variables simultaneously, this can lead to collinearity

problems (Dupont & Martensen, 2004). To tackle these problems a three-step procedure was followed to select the variables in the model. (1) For all variables that were initially considered it was tested whether they were a significant predictor of single-vehicle accidents by themselves. The significant variables were entered into step (2): They were grouped in 4 thematic groups (road condition, time of the accident, road-user characteristics, and accident development). The grouping was conducted at the discretion of the authors with the aim of forming groups of conceptually (and therefore often statistically) related variables. Within these groups the variables were entered in a stepwise procedure¹¹ into the model to select those that had a contribution over and above the other variables in the thematic group. The “group winners” were entered into step (3) where variables from all thematic groups were entered jointly and again a stepwise procedure was used to select those with a contribution over and above that of other variables.

The stepwise procedure was applied in two steps (2 and 3) to investigate which variables had the strongest predictive power but at the same time keep an overview of which variables excluded the other. While the knowledge gained in the intermediate step of modelling each thematic group separately is presented together with the results for each variable in section 3.3.2, the resulting four regression models are not presented in the text, but in Appendix 3.1 – 3.4.

3.3 Results

In the first place it is described how fatal multi and single vehicle accidents differ in their consequences. Subsequently, the explanatory variables considered for the difference between multivehicle and single vehicle accidents are presented in thematic groups with the results of modelling them by themselves and together with the other variables. Finally the joint model for the four different themes will be presented.

3.3.1 Consequences of multi and single vehicle accidents

This study aims to find factors that *predict* fatal single vehicle accidents as opposed to multivehicle ones. The consequence variables have conceptually a different relation with the dependent variable than the accident and road user characteristics analysed below. It is nevertheless important to realize that single and multivehicle accidents do not only differ in their origin but also in their consequences. These differences will be presented here. Because the unit of analysis in this chapter is the accident, we cannot analyse the factors contributing to the fatality or injury severity for each victim. We can, however, compute variables that summarize the severity level of each accident, namely the number of people with a particular consequence in each accident and the proportion of people suffering a particular consequence. The idea is to indicate whether single- and multivehicle accidents differ in the severity of outcomes.

¹¹ The forward stepwise procedure starts with an empty model and includes step by step the variables that will be significant when entered into the model. The backward stepwise procedure starts with the full model (all variables entered) and removes step by step those variables that are not significant. Both procedures lead to the same result in all cases.

Note, that the figures here apply to the dataset analysed in the present study (see 3.2.1) rather than to the complete database¹².

	Single Vehicle Accidents	Multivehicle Accidents
Average number killed	1.11	1.17
Average number seriously injured	0.36	0.59
Average number slightly injured	0.31	0.58
Average proportion killed	0.74	0.40
Average proportion seriously injured	0.12	0.31
Average proportion slightly injured	0.09	0.36

Table 3.1 Outcome variables for single and multivehicle accidents

It is important to keep in mind that only fatal accidents are included in this analysis. If either single- or multivehicle accidents end up as non-fatal accidents more often, these cases would simply not be included in this analysis. Moreover, the figures presented in Table 3.1 have to be interpreted in this context. In multivehicle accidents the total number of people involved is (on average) higher, simply because there are more participants. Consequently, the total number of victims – killed, seriously injured, or slightly injured – is usually higher for multivehicle accidents. As opposed to that, the proportion of people killed is lower for multivehicle accidents. Given that there is always at least one fatality – and in most cases exactly one fatality – the proportion of people killed depends predominantly of the presence of other people in the accident, which is higher in multivehicle accidents (this is the accident size bias described in Section. 1.1.1 in more detail). Finally the proportion of injured people is lower for single vehicle accidents, simply because in many cases the one person involved died in the accident while in multivehicle accidents there are always at least two persons involved. Consequently, in a fatal accident database, multivehicle accidents offer a greater chance to observe injured victims next to the fatality.

3.3.2 Modelling per thematic group

Road Condition Variables

Traffic Flow

The variable traffic flow describes how congested the road is. “Normal” stands for a situation with free flowing traffic, “light” signifying a more or less empty road and “heavy” a congested road. On more congested roads the proportion of fatal single vehicle accidents is lower than on free flowing roads (B=-.961, SE=.136, p<.001). The reasons can be twofold: On empty roads the chance to encounter other road users is low. An error would therefore be less likely to

¹² Note that these figures differ from those presented in Chapter 4 for another reason: The overall proportion of a consequence (e.g. the overall proportion of people killed) is not necessarily the same as the mean of the proportions of people killed per accidents. The main reason for the difference is that accidents involving many people weigh much higher in the overall proportion than in the mean of the proportion per accident.

involve another road user. Moreover, the travelling speed on empty roads is higher, making it more likely to lose control of the car and run off the road.

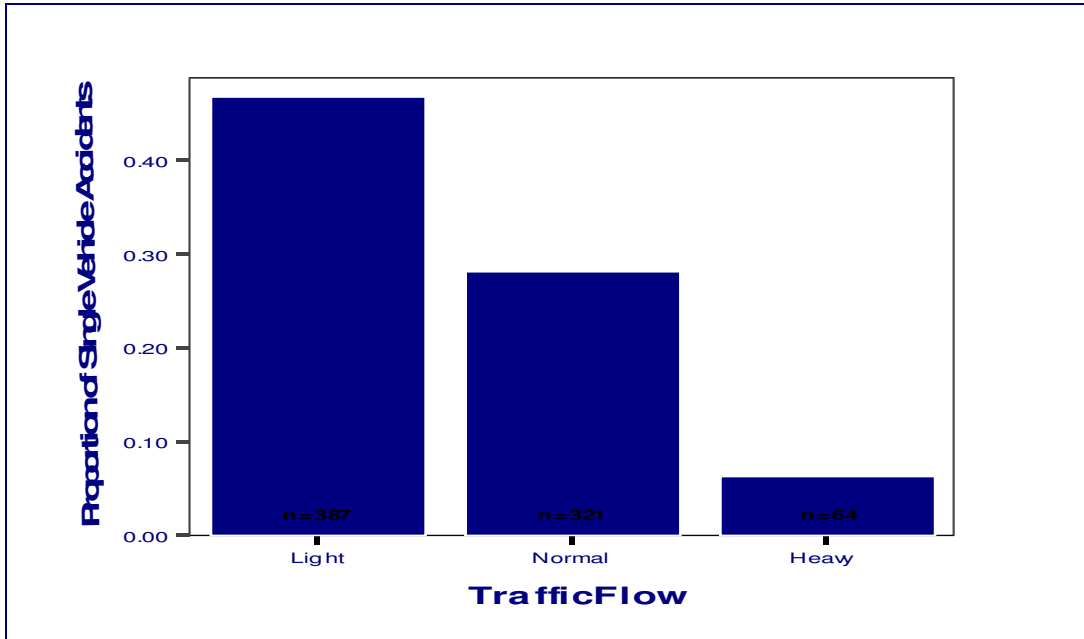


Figure 3.2 Proportion of single vehicle accidents as a function of traffic flow

Junction

At junctions there are fewer single-vehicle accidents than at road-sections ($B=1.778$, $SE=.206$, $p<.001$). This is not an unexpected result as junctions are the place per definition where vehicles interact and therefore have the largest chance to collide.

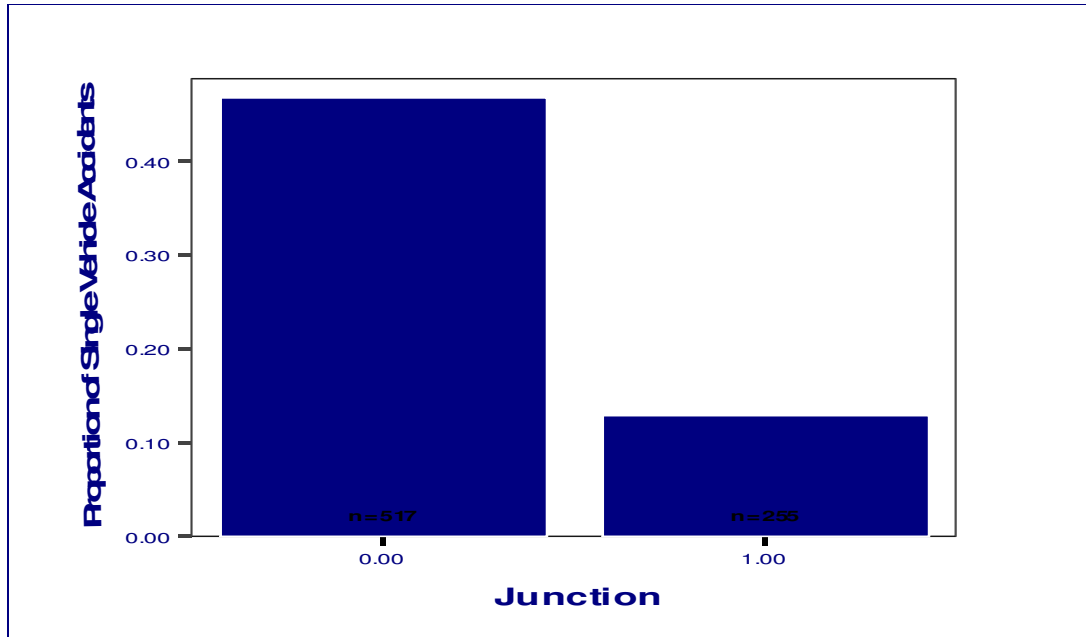


Figure 3.3 Proportion of single vehicle accidents as a function of Junction (0 road-section, 1 junction)

Carriageway Physically Divided

Generally speaking only a small percentage (14%) of the fatal accidents reported in the FAI data took place on roads with physically divided carriageways. Although, detailed exposure data for different road types in the sampling areas would be necessary to draw firm conclusions, this percentage seems very low and probably indicates that roads with physically divided carriage ways see disproportionately few severe accidents. Among the few accidents that did happen on these roads, the proportion of single vehicle accidents (50%) is higher than on other roads (.34%, $B=.572$, $SE=.206$, $p<.01$). This suggests that the physical divisions, which are meant to protect road-users against each other, indeed prevented the error of one road user from involving others'. As a consequence roads with divided carriageways have a low overall risk for fatal accidents, and an even lower one for fatal accidents involving more than one vehicle.

Motorway

As most roads that have physically divided carriageways are motorways, the variable motorway shows a strong overlap with the previous variable. By itself, motorway is a significant predictor of single vehicle accidents ($B=.651$, $SE=.259$, $p<.05$) as there are a higher proportion of single vehicle accidents on motorways as compared to other road types. However, when the variable is considered jointly with the variable Carriageway Physically Divided, motorway becomes nonsignificant while Carriageway remains significant. This suggests that it is the fact that motorways have physically divided carriageways is responsible for the low number of multivehicle accidents.

Speed limit



As can be seen in Figure 3.4, there is a tendency for roads with higher speed limits to have a higher proportion of single vehicle accidents. As can be seen there is a break between speed limits of 100 and above and speed limits of 90 and below. Consequently this division (91 and above vs. lower) is the best predictor of single vehicle accidents ($B=.720$, $SE=.155$, $p<.000$). However, this predictor becomes non-significant when taken up jointly with Physically Divided Carriageway. This is logical given that it is mostly high-speed roads that have physically divided carriageways.

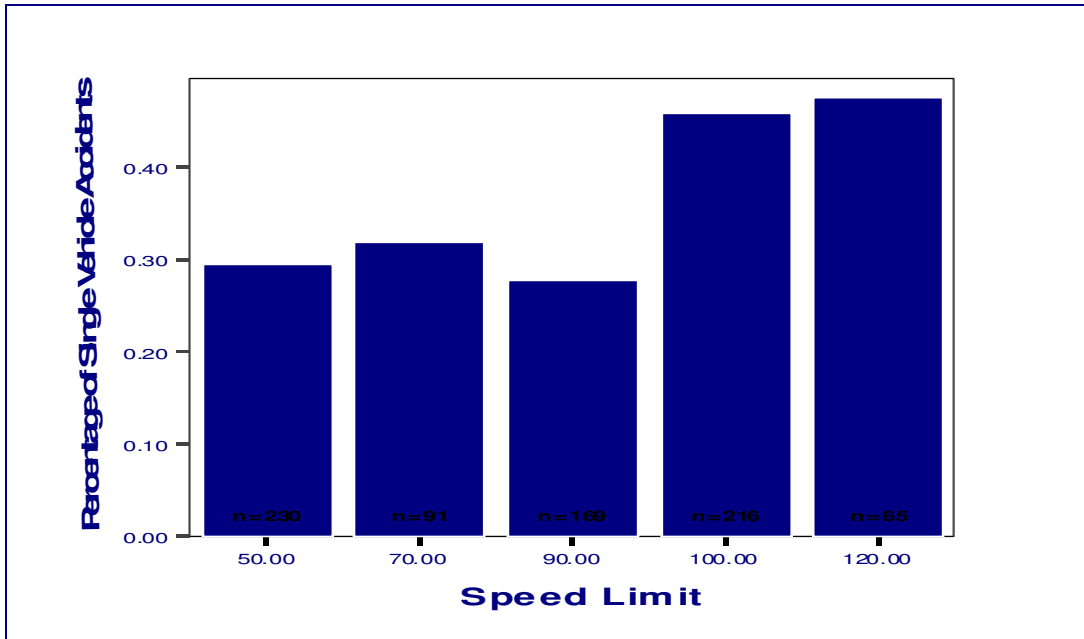


Figure 3.4 Proportion of single vehicle accidents as a function of Speed Limit. The bars represent categories ranging from the number indicated below each bar (e.g. 50) to 1 minus the number indicated below the following bar (e.g. 69).

Rural Area

In rural areas as opposed to urban or mixed areas, there are more single vehicle accidents ($B=.329$, $SE=.168$, $p=.05$). When analysing this variable jointly with junction and physically divided carriageways it turns out that the latter two variables are the more important ones in predicting single vehicle accidents. In rural areas there are comparatively few junctions and more roads with divided carriageways as compared to urban areas.

For the joint regression model of the road-condition variables see Appendix A3.1.

Time of the accident

The time variables were based on two variables: weekday and hour. In Figure 3.5, the proportion of single vehicle accidents is plotted for each hour. It can be seen that the proportion of single vehicle accidents is particularly high in the late night- early morning hours.

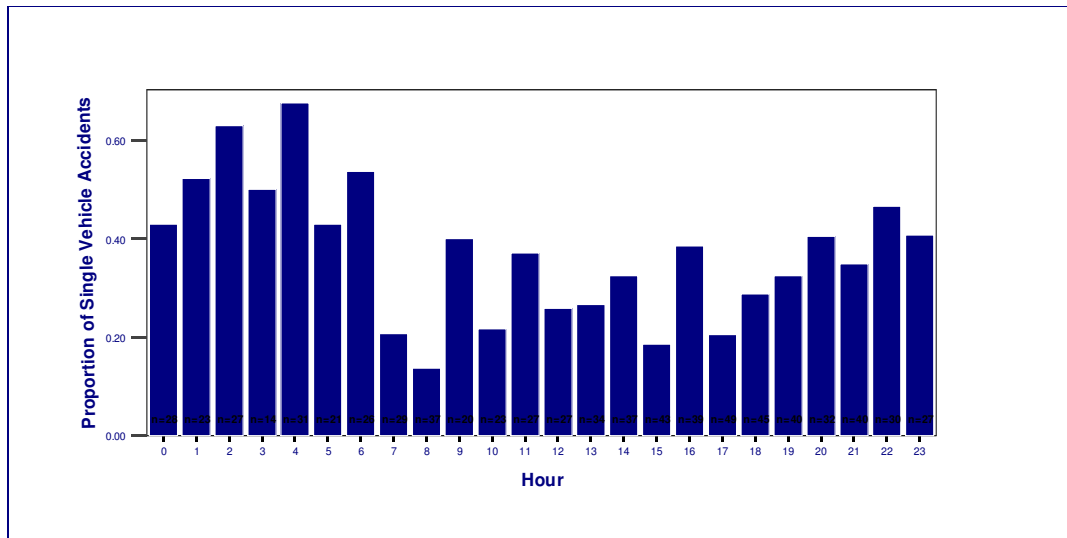


Figure 3.5 Proportion of single vehicle accidents at different hours.

Daytime

The variation across the hours was captured in a dichotomous variable *daytime*, which was defined as the hours between 6 a.m. and 8:59 p.m. During day time the proportion of single vehicle accidents was lower than at night time ($B=-.886$, $SE=.163$, $p<.001$). Together with traffic flow however, this variable becomes nonsignificant. Apparently, it is the empty roads rather than the night-time per se that is responsible for a higher proportion of single vehicle accidents.

Weekend

The weekend was defined as lasting from Friday night 9:00 p.m. to Sunday 11:59 p.m. During the weekend the proportion of single vehicle accidents was lower than during the week ($B=.390$, $SE=.152$, $p<.05$). Like the daytime variable, this variable is no longer significant when modelled jointly with traffic flow. Again it is the emptier roads during the weekends that promote single vehicle as opposed to multivehicle accidents.

Weekend Night

This variable is the conjunction of the variables Weekend and Day. It was generated to test the hypothesis that it is especially the weekend nights that have a high proportion of single vehicle accidents. Although this is the case ($B=.906$, $SE=.220$, $p<.001$), this variable does not remain significant, when analysed jointly with the main effects of weekend and daytime. As can be seen in Figure 3.6, the difference between day and night at the weekend is not larger than the difference between day and night during the week. Moreover, just like daytime and weekend, this variable is not significant when considered jointly with traffic flow.

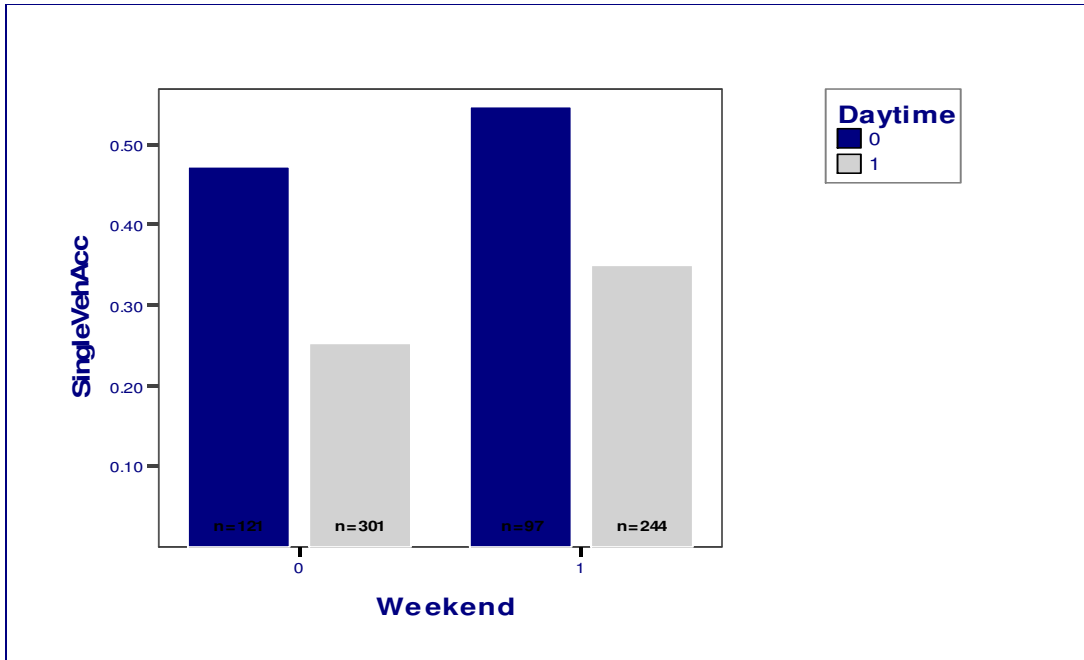


Figure 3.6 Proportion of single vehicle accidents as a function of weekend (1=weekend, 0=week) and daytime (1=day, 0=night).

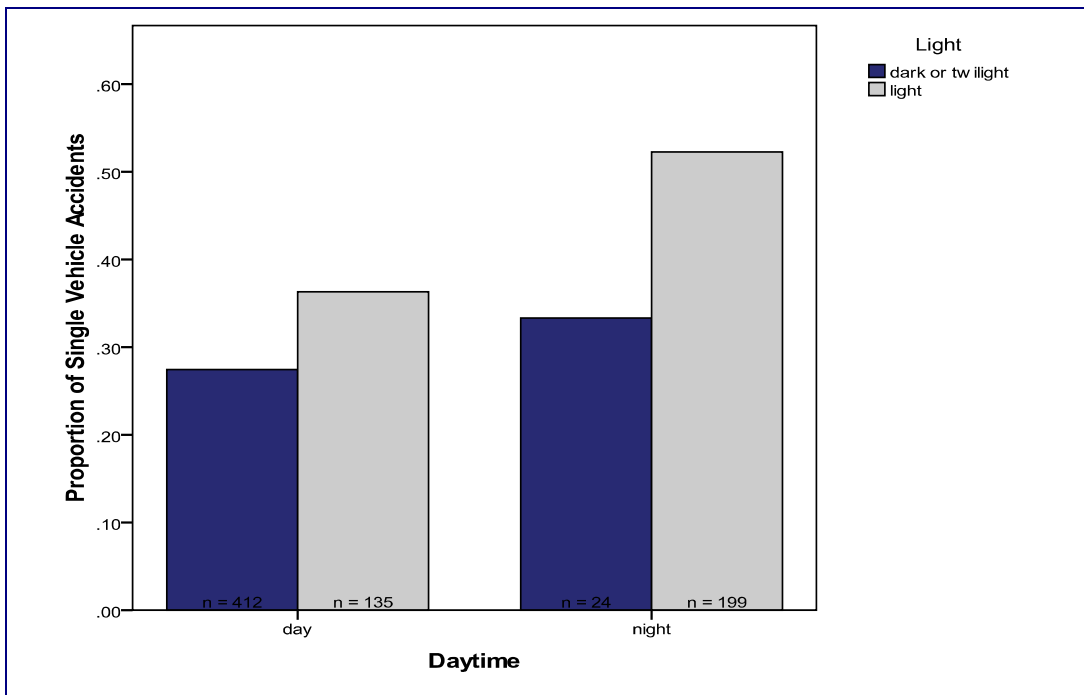


Figure 3.7 Proportion of single vehicle accidents as a function of light and daytime.

Light

Although light conditions is not a time variable, it was decided to analyse it together with the time variables, because of the obvious relation between *light* (daylight as opposed to all other light conditions like darkness or twilight) and *daytime*. By itself, the variable light had an effect comparable to that of daytime:

The proportion of single vehicle accidents was lower in daylight. Interestingly, both variables, light and daytime, remain significant when modelled jointly (see Appendix A3.1 for the joint model of all time variables). There is a substantial number of cases, where it was night, but not dark or where it was daytime, but not daylight. Each variable has its own effect. However, in the final model including the significant variables from all thematic groups, neither light nor daytime became significant. Again this is due to the variable traffic flow.

As a summary for the time variables, we can see large differences between different times of the day and of the week with respect to the occurrence of single vehicle accidents. All these differences however can best be explained in terms of the amount of traffic on the roads. See Appendix 3.2 for the model including all time variables to predict single vehicle accidents.

Road User Variables

As the data for the present analysis were aggregated at the accident level, road-user characteristics cannot be considered in their pure form. It is nevertheless interesting to investigate whether single-vehicle and multivehicle accidents differ with respect to the road-users involved. This was done by including road-user variables that gave a summary of the accident, rather than describing one individual user. In the following, the variables that were initially considered because they were by themselves significant predictors of single-vehicle crashes are described.

Road user type

The involvement of different road user types was coded in two variables. PTW involved signified whether a powered two-wheeler was involved in the accident ($B=-.924$, $SE=187$, $p<.001$) and HGV involved identified those accidents in which a heavy goods vehicle was involved ($B=-2.171$, $SE=291$, $p<.001$). As can be seen in Figure 3.7 fatal accidents involving cars only have the highest proportion of single vehicle accidents, followed by fatal accidents involving powered two-wheelers. The fewest single vehicle accidents are among those involving heavy goods vehicles.

Mean Driver Age

For single-vehicle accidents this is simply the age of the one driver involved. For the other accidents the mean of the ages of all drivers involved was taken. The variable has a marginally significant negative effect, suggesting that on average the drivers involved in single vehicle accidents are younger than those in multivehicle accidents ($B=-.011$, $SE=.006$, $p=.057$). In the joint model, this variable is not significant anymore. Again we see a relation between this variable and traffic flow: Young driver tend to drive in the evenings and in the weekends, which is also the time when the roads are relatively empty, while middle-aged and older drivers use their cars predominantly for commuting when the roads are crowded.

Gender of the driver.

The percentage of women among all drivers involved was calculated for each accident. For the single vehicle accident this was either 100 (in the cases in which the driver was a woman) or 1, for multivehicle accident it could take different values. Although in the literature (Öström & Eriksson, 1993) it has been observed that male drivers are proportionally overrepresented in single vehicle accidents, in the present dataset no effect of the gender of the driver was observed ($B=-.139$, $SE=.295$, $p=.638$).

Average number of occupants

On average, vehicles involved in single vehicle accidents had more occupants than those in multivehicle accidents ($B=.435$, $SE=.081$, $p<.001$).

Alcohol and other impairments

In the literature single vehicle injury accidents have been linked to alcohol consumption and other impairments (e.g., Öström & Eriksson, 1993). Two dichotomic variables were generated indicating whether any of the drivers involved in the fatal accident of the FAI data were suspected or proven to have been under the influence of alcohol ($B=.226$, $SE=.213$, $p=.289$) or other impairments ($B=.310$, $SE=.195$, $p=.112$). In the present analysis, these two variables were not significant¹³.

Familiarity with the area

This variable, which is not available in macroscopic databases, contrasts accident in which all drivers are unfamiliar with those in which at least one driver was familiar. In the first place, it is interesting to note that only in 24% of the fatal accidents at least one driver was familiar with the area. This is surprising in itself, because generally speaking people drive much more on roads that they are familiar with than on roads they are unfamiliar with. So purely on the basis of risk exposure, one would expect a high percentage of the accidents to have taken place in area's familiar to one or more of the drivers. And indeed, insurance figures (Strillacci, 2002) tell us that 83% of all accidents (fatal and nonfatal) happen within a range of 10 miles of the drivers' homes. This close-to-home effect is sometimes interpreted in terms of driving style: When getting close to home, drivers start to relax and feel more confident, because they know the area well and they become less careful.

The much lower percentage of accidents within familiar area in the FAI database suggests, however, that being *unfamiliar* with the area increases the risk to have a *fatal* accident. So *if* the close-to-home effect is due to overconfidence, this does at least not apply to the fatal accidents in the FAI data. (Moreover, the higher risk exposure also applies to the close-to-home effect and might be sufficient to explain that effect).

¹³ This variable was significant when including those cases that had missing values on the traffic flow variable or the time variables. However, in that analysis the effect became non-significant when the variables were entered together with the other road-user variables (most importantly the seatbelt variable).

Among the fatal accidents the percentage of accidents with at least one familiar driver is lower for single vehicle accidents (13%) than for multivehicle accidents (30%). In other words, drivers who were unfamiliar with the area had a higher chance of having single vehicle accidents than those who are familiar ($B=-.759$, $SE=.207$, $p<.001$). For drivers in fatal single vehicle accidents we know for sure that they made a fatal error, while in fatal multivehicle accidents there are also drivers involved who became the victim of another driver's error. The results therefore suggest that unfamiliarity with the area increases the likelihood of making a fatal error

The joint model for all road-user variables predicting single vehicle accidents is presented in Appendix 3.3.

Accident Development Variables

To the difference of macroscopic accident databases, the FAI database includes a number of variables that describe the development of the accident. In particular each accident was described in a chain of events. Moreover, the manoeuvres that the drivers were executing when the accident happened and possible manoeuvres that were taken to prevent the accident were reported. The results below show that single and vehicle accident differ strongly on these variables.

First Event

In the fatal accident database the development of the accident was described in events. The First Event variable in this analysis indicated whether the first event of an accident was a non-collision (1), a collision with another road user (2), a collision with a non-fixed "object" (e.g., an animal) (3), or a fixed object (4). As can be seen in Figure 3.8, Categories 1 and 3 occurred approximately equally often for single- and multivehicle accidents. Category 2, a collision between road-users, was practically never the first event in single-vehicle accidents, while category 3 (collision with a fixed object) was the first event for single vehicle accidents much more often than for multivehicle accidents. Although this variable is highly significant (joint Wald test: $t(3)=72$, $p<.001$) it was decided not to include it in the following analyses because its strong relation with the dependent variable is in some sense trivial: Obviously, single vehicle accidents cannot start with collisions between two vehicles.

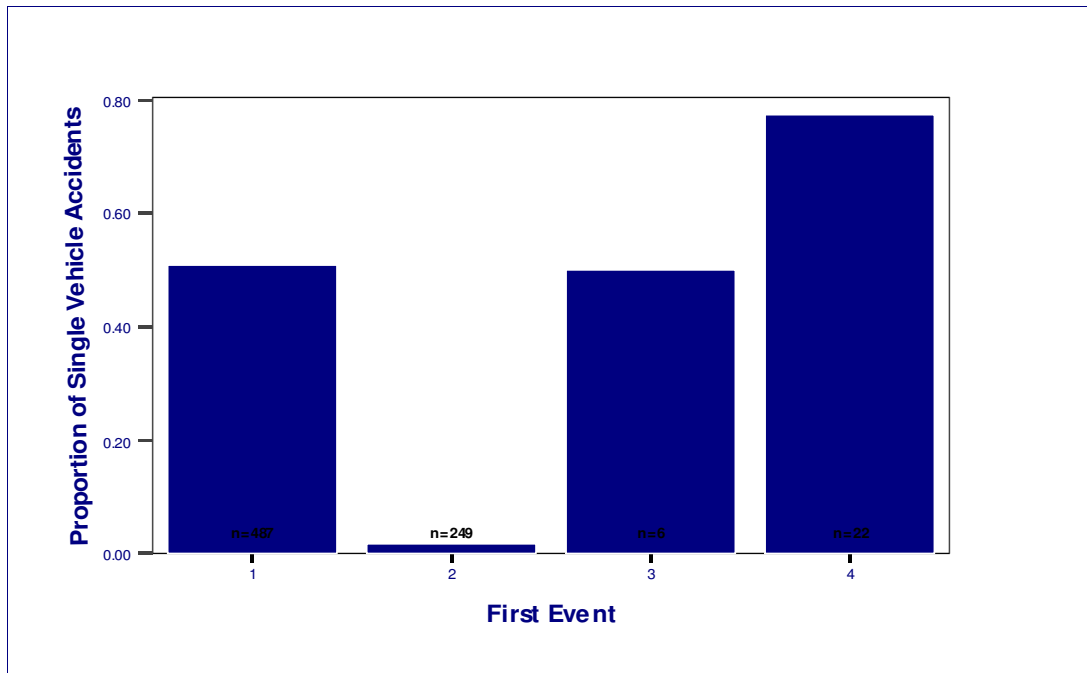


Figure 3.8 Proportion of single vehicle accidents as a function of First Event: 1=non-collision, 2= collision with another road-user, 3= collision with non-fixed object, 4=collision with fixed object.

Avoidance Manoeuvre

This variable described the type of avoidance manoeuvre that road users (possibly) took. In the present accident level analysis, it is differentiated whether all drivers attempted to avoid the accident by steering or braking (No Avoidance Manoeuvre=0) or whether there was a driver involved for whom no avoidance manoeuvre is reported (No Avoidance Manoeuvre=1). When there was one driver who did conduct a manoeuvre, the accident was more likely a single vehicle accident as compared to accidents where all drivers had attempted to avoid the accident by braking or steering ($B=-.502$, $SE=.217$, $p<.05$).

Driver Manoeuvre

The variable Driver Manoeuvre described the type of manoeuvre the driver was about to execute when the accident happened. It contains many categories and was recoded into two dichotomic variables: Executed Manoeuvre and Loss of Control. *Executed manoeuvre* was 0 when none of the drivers had executed any kind of manoeuvre (i.e. they were all driving straight or had lost control) and 1 if at least one of them had. *Loss of Control* was 1 when any of the drivers had lost control and 0 otherwise.

Executed Manoeuvre had a negative effect ($B=-.355$, $SE=.154$, $p<.05$) indicating that accidents in which none of the participants had executed a manoeuvre were more likely to be single vehicle accidents than accidents in which at least one of the drivers had executed a manoeuvre. This variable shows however, a strong overlap with *junction*, as it is on junctions that most manoeuvres take place at. When entered together with *junction*, the executed manoeuvre variable became nonsignificant, suggesting that the fact that a road-

user passes a junction is more important than whether he conducted a manoeuvre or not.

Loss of Control was not significant when entered into the analysis by itself (B=-.015, SE=.367, p=.985) and consequently excluded from the further analyses.

See Appendix A3.4 for the model including all accident development variables to predict single vehicle accidents.

3.3.2 Joint modelling of thematic groups

All variables that were significant in the sub models conducted for each thematic group (see Appendices 3.1 – 3.4) were entered into the final model to predict single vehicle as opposed to multivehicle accidents. The results of this stepwise analysis are presented in table 3.2.

	B	S.E.	Wald	Sig.	Exp(B)
Junction	-1.609	.229	49.480	.000	.200
TrafficFlow	-.831	.165	25.357	.000	.436
Carriageway Physically Divided	1.199	.274	19.198	.000	3.317
HGV Involved	-2.364	.327	52.196	.000	.094
PTW Involved	-.718	.230	9.747	.002	.488
Occupants	.266	.098	7.340	.007	1.305
Familiar	-.925	.250	13.651	.000	.397
No Avoidance Manoeuvre	-.502	.217	5.349	.021	.605
Constant	1.530	.363	17.816	.000	4.619

Table 3.2. Joint model of predictors for single vehicle accidents from different thematic groups. Variables entered: junction, trafficflow, carriageway physically divided, daytime, daylight, weekend, ptw involved, hgv involved, mean driver age, familiar, impairment, avoidance manoeuvre, manoeuvre executed.

The variables that remained significant in the joint model indicate that multivehicle accidents are more likely at junctions while single vehicle accidents have an increased chance to occur at road-sections between junctions. Moreover, the chance of single vehicle accidents to happen increases as the traffic volume decreases. On roads with divided carriageways single vehicle accidents are overrepresented in comparison to multivehicle accidents. Powered two-wheelers or heavy goods vehicles are less likely to be involved in a single vehicle accident than car drivers. Cars in single vehicle accidents tend to carry more passengers. Drivers unfamiliar with the area are more prone to have single vehicle accidents than those for whom the area is familiar and drivers in single vehicle accidents fail more often to execute an avoidance manoeuvre.

A number of variables that were significant in the thematic models, did not become significant in the global model: More specifically the time variables, daytime, light, and weekend are no longer significant once they are considered jointly with traffic flow. At night and at the weekends the streets are relatively empty allowing for speeding, which might lead to single-vehicle accidents. Moreover, a driver who makes a severe mistake on an empty road harms only himself (and his/her passengers) while on crowded roads, a severe error by one

driver is much more likely to involve other vehicles and therefore result in a multivehicle accident.

The mean driver-age is also strongly related to traffic flow. So strongly in fact, that together with traffic flow, driver age becomes insignificant. Young people drive often at the weekend and at night, when the roads are empty. This is also the time at which more single vehicle accidents happen. It is therefore difficult to say whether it is the young drivers or the fact that the roads are empty that cause accidents to be single- rather than multi-vehicle accidents. Our analyses, however, suggest that the emptiness of the road is the more important variable.

The results presented here, demonstrate the importance of modelling the difference between single and multivehicle accidents in a multivariate way. Earlier studies have attributed single vehicle accidents to rural areas (Wegman & Aarts, 2006; Scheers & Casteels, 2008) to young drivers (Öström & Eriksson, 1993) and to weekend nights (Öström & Eriksson, 1993, Scheers & Casteels, 2008). And although these results were generally replicated in the sense that single and multivehicle accidents do indeed differ on these variables, a more complete analyses suggested that these effects are in fact mediated by other variables (in particular traffic flow, junction, and divided carriageway).

3.3 Conclusion

Single and multivehicle accidents differ on a large number of variables. These variables and their relation with each other have been systematically investigated using stepwise logistic regression analyses.

It is important to keep in mind that this study is conducted on an accident data base that contains exclusively fatal accidents. Further analysis will have to show whether the findings reported here also hold for non-fatal accidents.

Conditions of the road have been found to be the most important predictors of whether an accident involves one or several vehicles. Multivehicle accidents take place on junctions and busy roads, while single vehicle accidents tend to take place on empty road-sections between junctions. Roads with physically divided carriageways see generally few accidents. Those that do happen are more often single-vehicle accidents than on other road types. This suggests that such physical divisions do in fact succeed in preventing drivers who make a mistake involving other drivers in the accident. The effect of these variables can be summarized under the principle that road-conditions that make it less likely that two vehicles encounter each other prevent the error of one driver involving another one and thus reduces the likelihood of multivehicle accidents as compared to single vehicle accidents.

Single vehicle accidents mostly involve cars. Motorbikes are less likely to be involved in single than in multivehicle accidents and heavy good vehicles are involved almost exclusively in multivehicle accidents. The cars in single vehicle accidents have more passengers on average and the drivers in single vehicle accidents tend to be unfamiliar with the area more often than the drivers in

multivehicle accidents. Moreover, these drivers failed disproportionately often to even attempt to avoid the accident by braking or steering.

It has also been found that single vehicle accidents take place especially at night and during the weekends and involve young drivers more often than multivehicle accidents. However, in the fatal database analysed here, these effects seem to be mediated by the fact that at those times the roads are much emptier than during the week at daytime, when middle-aged and older people tend to drive. Earlier reports of driver-effects (Öström and Eriksson, 1993), have however been based on fatal *and* non-fatal accidents. Further research is necessary, to investigate, whether in non-fatal accidents these effects can be explained by road-condition factors as well.

Three variables that are not commonly found in macroscopic databases have proven to be important in differentiating between single and multivehicle accidents. A low Traffic Flow is associated with a higher proportion of single vehicle accidents. Accidents where all drivers attempted to avoid the accident with a Manoeuvre like braking or steering and accidents in which all drivers were familiar with the area are more likely to be multivehicle accidents. In particular Traffic Flow turned out to be a very important variable. Not only is it highly significant, its inclusion into the model also gave the important insight that variables like the age of the driver or the time of the accident (weekend and night) do not in themselves differentiate between single- and multivehicle accidents, but rather happen to coincide with times when the roads are relatively empty.

Returning to the starting questions we can conclude that road conditions that prevent the error of one driver to spread to others are more important to understand the difference between single and multivehicle accidents than characteristics of the drivers involved.

Chapter 4 – Factors affecting the fatality risk of road-users involved in fatal accidents

4.1 Introduction

Emmanuelle Dupont & Heike Martensen, IBSR

The goal in collecting accident data is to learn from the past and gain information that can help preventing future accidents. As stated in the introduction, an important problem in interpreting *fatal* accident data however, is the lack of comparable data from more desirable situations, as for example nonfatal accidents.

As a solution, in the present analysis the survivors *in fatal accidents* served as a control group for the fatalities. The question asked is: What makes the difference between the road-users who ended up among the fatalities of the accidents and those who survived it? This information may help designing measures that – if they do not help preventing the occurrence of future accidents – may well contribute to mitigating the severity of their consequences for the individual road-users.

The consequences of a fatal accident for the road-users are thus the focal point of this chapter. The consequences will be most often defined as the *fatality risk* of road users involved in fatal accidents. It is important to recall that the fatality risk has to be understood, in this specific context, as the *risk that a road-user runs of ending up among the fatalities rather than among the survivors of an accident, given that the accident was fatal*. This risk is estimated here on the basis of the proportion of road-users killed, the survivors being defined as the sum of the persons injured (severely or slightly) and not injured. The SafetyNet injury severity score was used as indicator for the calculation of the fatality risk for all the analyses presented in this chapter. In some cases (Section 4.2), the analysis is extended so that the full range of the accident's consequences is examined (in other words, the probability to be killed, to be seriously or slightly injured, and the probability to be left uninjured). As thoroughly explained in this section, the interpretation of the results obtained for the “severely” and “slightly injured” categories requires extra-care in the context of fatal accident data. This is the main reason why the remainder of the analyses presented here is limited to the investigation of the fatality risk.

The following three sections correspond to three analytical attempts at identifying significant sources of variations of the fatality risk of road users involved in fatal accidents: In the first one, this is done for all types of road-users altogether, in the second, the analyses focuses on car-car accidents, by including only the data from accidents involving two cars. Finally, the third section continues to focus on the fatality risk for car occupants, but the scope is broadened to other types of car-accidents. The type of opponent with whom each car collided with is therefore included as a predictor in the model. Below, we describe with more detail the considerations that guided the evolution of these three “fatality risk models”. These considerations all relate to evolving

conceptions concerning the most appropriate way to handle the accident-size bias and the issue of the (in)comparability of the baseline risk .

A first definition and description of the two issues has been provided in the general introduction. Below is a reminder of the exact nature of each of these problems as well as a summary of the way they have been handled in the three models presented in this section.

○ 4.1.1 Controlling for the accident-size bias:

As already explained in Chapter 1, the probability for a road-user to die in a fatal accident is greatly affected by whether or not this person was the only one involved in the accident, and if not, by how many others there were.

Accident size was therefore taken into account in all three of the analyses presented in the following sections. In the first section, variations in the fatality risk for all the road users were examined while controlling for the number of participants in the accident as well as the number of occupants per vehicle. In the second analysis, the data were selected so that the number of participants was maintained constant: Only data from accidents involving two cars were analysed. In addition to this, the number of car occupants was controlled for and included as predictor in the model. The last model, also investigates variations of the fatality risk for car-occupants, but this time several types of accident are taken into account (and thus not only car-car accident). In this last model, the accident-size bias is controlled for by including both the number of accident participants and the number of car occupants as predictors in the model.

○ 4.1.2 Obtaining comparable baseline risk estimates:

The importance of ensuring comparable “baseline risk” for survivors and fatalities has already been underlined earlier. Although incomparability of the baseline risk is a general problem of accident data, it is exacerbated when these are circumscribed to fatal accidents. This is illustrated by Figures 4.1 to 4.3, which show, each for a different type of road-users, the variation of the proportion of fatalities, of severe or slight injuries, and of uninjured people associated with different opponent types¹⁴. Two things can easily be noticed when looking at these graphs: Firstly, there are road-user groups for which the outcome of the accident hardly varies, namely HGV occupants (Figure 4.1), and pedestrians (Figure 4.2).

¹⁴ Similar figures for other road-user types are provided in Appendix 4.3

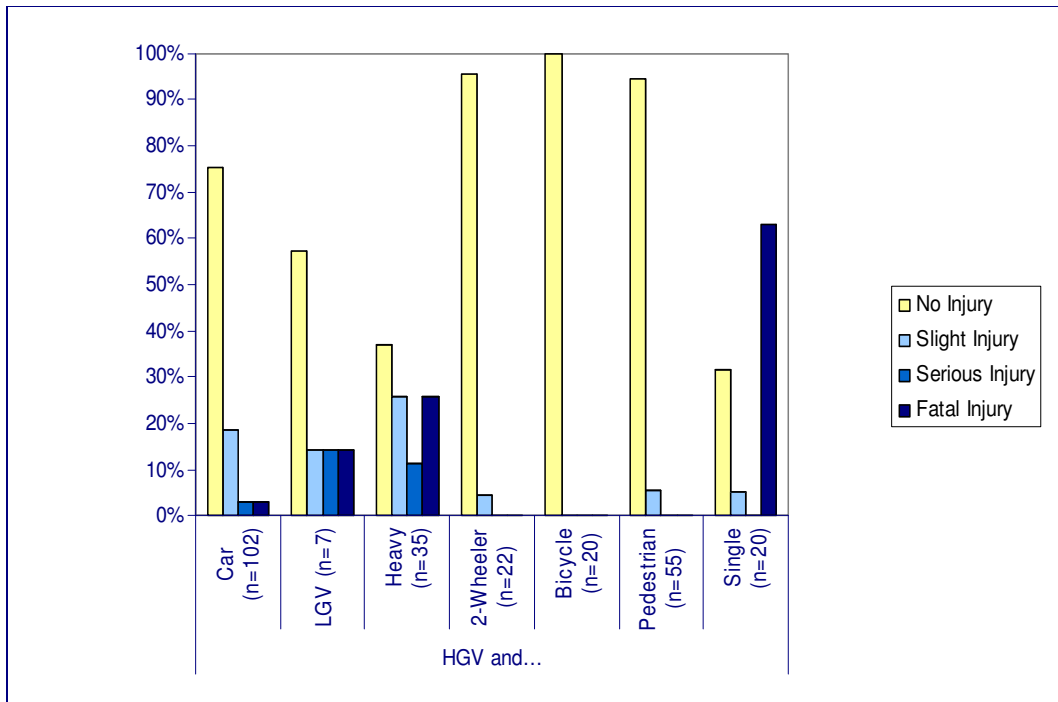


Figure 4.1: Percentage of survivors, fatalities, slightly and severely injured road-users among occupants of HGV depending on opponent type

HGV occupants tend to survive the accident (often without even being injured), whatever the type of opponent they collided with (with the exception of single collisions, or collisions with other heavy good vehicles, where the proportion of fatalities tends to be higher). The figure for pedestrian (Figure 4.2) is mainly composed of long dark blue bars, indicating that they almost inevitably end up among the accident fatalities, whatever their opponent type. The same pattern is observed for other vulnerable road-users, such as cyclists. The conclusion to be derived from this first observation, is that the transport mode of the road-users themselves has a dramatic influence on their fatality risk, once involved in a fatal accident.

The second important observation that can be made is that, for other types of road-users – such as the car occupants (Figure 4.3) – the type of *opponent* appears to exert a strong influence on the fatality risk. Indeed, while car-occupants survived the vast majority of collisions with vulnerable road-users, they most frequently deceased as a result of collisions with heavy and light good vehicles. The picture is different for car occupants who collided with another car. Among these, the different types of consequences are more evenly distributed.

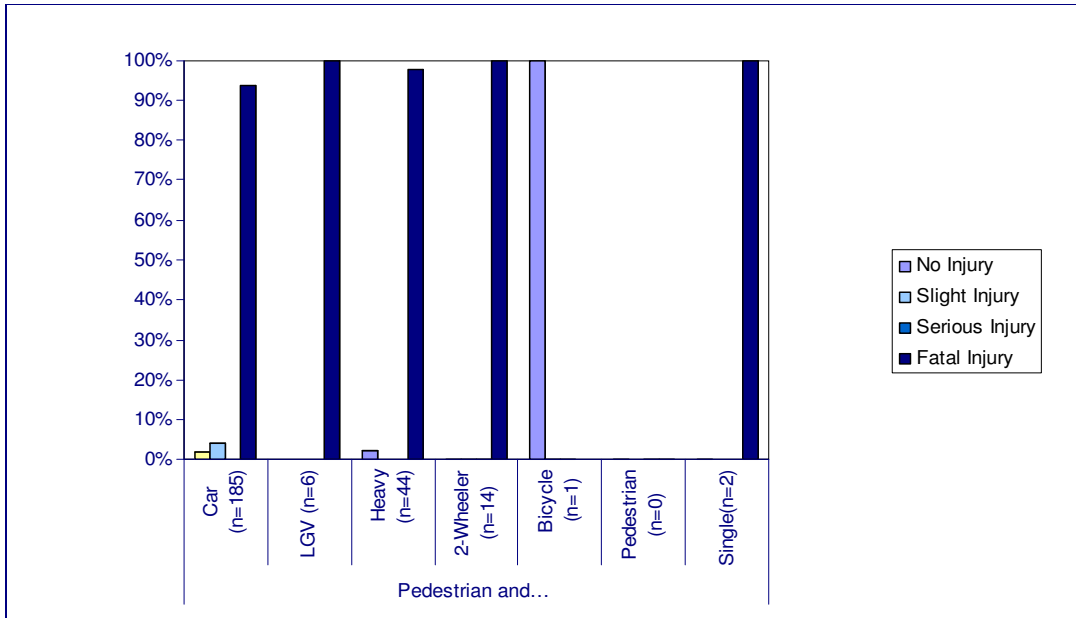


Figure 4.2: Percentage of survivors, fatalities, slightly and severely injured road-users among pedestrians depending on opponent type¹⁵

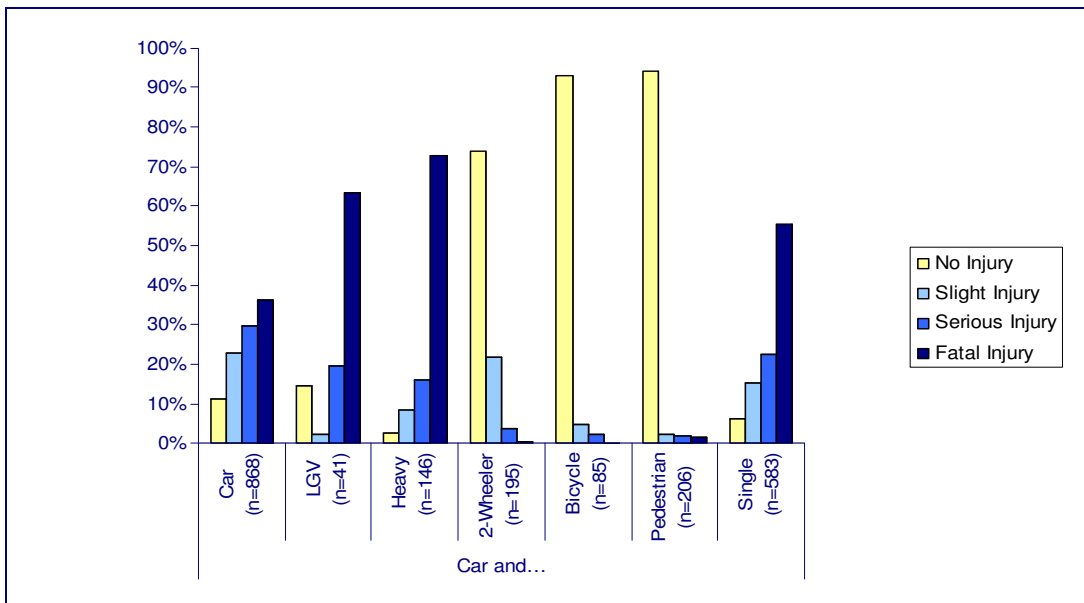


Figure 4.3: Percentage of survivors, fatalities, slightly and severely injured road-users among car occupants depending on opponent type

¹⁵ The two “single pedestrian” instances can best be understood when reading the corresponding accident summaries: “The accident took place in an urban area between a pedestrian and a Bus. The pedestrian was going out from a bus, when, unexpectedly, when he was with a foot on ground and keeping a support with the hand for to help his movement, the driver of the bus closed the door and the passenger remained blocked with his hand between the doors and was dragged for any metres then fell to the ground” – “The accident took place in a country road involving a car and a pedestrian. The pedestrian was on a ladder, under a tree, while his friend was inside the car. Unexpectedly the driver began a reversing manoeuvre and maybe, having the back view obstructed (maybe by a cloth on the windshield) hit the ladder and the pedestrian fell to the ground”.

Ensuring the comparability of the baseline risk is consequently of crucial importance when examining the factors that are associated with variations of the fatality risk, as estimated on the basis of the FAI database. The three models presented here also propose alternative ways - more or less efficient – to handle that problem.

The model presented in the first section includes a predictor representing the road-user's own transport mode, namely: pedestrian, car, motorcyclist, or HGV-occupant. In this way, it is possible to examine variations of the fatality risk associated with several key factors, *taking into account* the influence that the road-user's own transport mode intrinsically has on the fatality risk.

Yet, as we noticed on the basis of the above figures, the fatality risk of each individual road-user is in some cases also strongly affected by the transport mode of the one he/she collided with in the course of the accident (as an example: car-occupants opposed to pedestrians versus car-occupants opposed to HGV's). This influence of the type of opponent on the fatality risk should also be accounted for, if one is to investigate the relation of other predictors with the fatality risk. The last two models presented in this chapter constitute two distinct attempts at simultaneously controlling for the influence of the road user's own transport mode and that of his/her opponent on the fatality risk. To do so, the model presented in section 4.3 was circumscribed to accidents between two cars, so that the road-user and opponent types, as well as the number of vehicle in the accident were maintained constant. The third and last model presented in this chapter offers the most efficient way of handling the incomparability of risk issue: It also focused on car occupants, but is extended to car occupants involved in several collision types (and not only accidents with other cars). It was also not limited in terms of the number of accident participants. The specificity of this analysis lies in the integration of the type of collision opponent into the model (where car-car accident is taken as the reference category to which car-HGV; car-LGV; car-motorcycle and single car accidents are compared). In other words, the relationship between the fatality risk and other variable of interest was there examined among car opponents, while fully controlling for the influence of the type of opponent on the baseline risk.

It is important to note that the realisation of this "Opponent analysis" is rendered possible by the high level of details offered by the FAI data, which allows each road-user to be linked to his/her collision opponent (see Appendix 4.2 for details on how this opponent type variable was defined and computed). By the same token, the opponent analysis is also the one that best acknowledges and exploits that level of detail in the data.

The different solutions adopted in the three models to handle the accident size and incomparability of risk are summarized in Table 4.1. In the following section, we describe the general approach adopted in building the three models and in selecting the predictors of interest.

	Accident Size	(In)comparability of the baseline risk	N
<i>Section 4.2</i>	-Inclusion of the number of participants	-Inclusion of the road-user's transport mode (car-occupant vs vulnerable, motorcycle, or HGV occupant)	3072
<i>Section 4.3</i>	- Selection of 2-participants accidents only -Inclusion of the number of occupants	-Analysis limited to comparable collision types: Only accidents between car-occupants	584
<i>Section 4.4</i>	-Inclusion of the number of participants -Inclusion of the number of occupants in the vehicle	-Analysis limited to car occupants -Vulnerable road-users discarded from the analysis -Inclusion of opponent type in the analysis	1638

Table 4.1: Treatment of the accident size and incomparability of the baseline risk in the three models

○ 4.1.3 Model building and predictors of interest

Predictors from four main groups were investigated:

- (1) *Predictors corresponding to all sources of variation of the fatality risk that are necessary to control for:* the accident-size bias (number of participants in the accident, vehicle occupancy), the incomparability of the baseline risk (the road-user's transport mode, the type of opponent)
- (2) *Indicators of the personal (physical) vulnerability of the road-users* (gender, age, etc.)
- (3) *Passive-safety related predictors* (e.g., seatbelt)
- (4) *Predictors related to the circumstances of the accident* (e.g., time of day, or road section versus junction)

All these predictors were naturally not independent from each other. For example, the variables "Motorway" and "Carriageway" show strong overlap, as it is mostly the motorways that have physically divided carriageways. Moreover, these variables are related to "Junction" - as there are no junctions on motorways - and to "RuralArea" - as motorways are more often inside rural areas than outside. Strong interrelations between the predictors are likely to blur conclusions about their respective impact on the fatality risk. To solve these confounding problems, the individual relation of each predictor with the fatality risk was tested first. Other predictors that were expected to be related to the one under scrutiny were then added: Changes in the value of the first predictor subsequent to the introduction of the second one in the model indicate that some confounding is at hand.

Two criteria were taken into account to build the final model: The evolution of the DIC criterion, on the one hand, and the Wald-test for each predictor, on the other hand. The DIC criterion allows comparing two models in terms of their

efficiency in accounting for the data. The lower the DIC criterion, the better the model is. The significance of each single predictor was also evaluated on the basis of a Wald test¹⁶.

4.2 Modelling the fatality risk for all road users

George Yannis (NTUA), Eleonora Papadimitriou (NTUA)

○ 4.2.1 Objectives and main questions

In this section, a first approach for modelling fatality risk in fatal accidents is tested, by examining the whole Fatal Accident Investigation (FAI) database. The analysis aims at answering questions such as: “Given their involvement in a fatal accident, does the risk of ending up among the fatalities increase with the road-user’s age?”; “Can this risk be considered more elevated for road-users involved in rural accidents as compared to those in urban accidents?”, and so on.

An additional objective of the analysis is to examine whether the fundamental issues (the accident-size, the comparability of risk) involved in analyzing a set of only fatal accidents can be efficiently handled in an analysis of the whole FAI data, and whether this leads to useful and reliable conclusions. This preliminary analysis, if successful, may provide guidelines for dealing with the particularities of fatal data. If not fully successful, the modelling attempted here will nevertheless provide useful information for the next steps, by revealing which categories of accidents or road users can not be handled appropriately by such an approach.

The main interest of this analysis lies in the possibility it offers to exploit the entire dataset. However, each of the questions presented in section 4.1 need to be addressed, in order to eliminate the obvious bias that underlies this type of data as much as possible. In particular, the major issues that need to be addressed are:

- The accident size effect: Given that every accident resulted in at least one fatality, the fewer the vehicles involved in the accident, and the fewer the occupants in these vehicles, the higher the fatality risk for the road users

¹⁶ The DIC criterion is a value that can be calculated for a whole model (i.e., for all the predictors it includes). It favours models with a good fit but it penalizes complexity. Of two models, the one with the lowest DIC is to be preferred because it explains the data better with fewer variables.

The Wald test compares the size of the coefficient estimated for one predictor to the size of its standard error. A significant result indicates that one can be 95% sure that the true value of the coefficient differs from 0.

For each variable, the DIC of the model with and without this variable were compared. All variables for which the DIC increased when removing them from the model and for which the Wald-test was significant were kept. At the end of each section, a table is provided summarizing the final model obtained with the evolution of the DIC criterion corresponding to the inclusion of each predictor along with the p-value for this predictor (result of the Wald test performed on each predictor while all other predictors kept in the final model are included).

involved. In order to correct for this effect, it is necessary to initially include in the analysis variables indicating the number of accident participants (the variable chosen indicates whether the accidents involved one participants only, or two participants or more) and the number of vehicle occupants.

- The comparability of risk: Given that the data contained only fatal accidents, vulnerable road users in the sample (e.g. motorcyclists and pedestrians) by default suffer an increased fatality risk, whereas heavy vehicles occupants suffer a reduced fatality risk. Furthermore, the heavier the collision opponent, the higher the fatality risk of car occupants. It is likely that this effect can not be fully accounted for when examining the whole set of FAI data. It will be tested to which extent this incomparability of risks of fatal accidents participants can be dealt with by accounting for the road user vulnerability. On that purpose, variables showing whether the road user is a pedestrian, a motorcyclist or a heavy vehicle occupant will be examined, in order to at least account for the most obvious source of risk incomparability.

Having accounted for these basic effects, it will be tested whether other variables add explanatory effect on the fatality risk of road users involved in fatal accidents; in this framework, particular care is taken for eliminating any confounding variables effects from the model (i.e. multicollinearity). At the same time, the presence of hierarchies in the data will be tested.

The analysis first tackles the question of fatality risk, given that one is involved in a fatal accident, which is the main focus of this section. In addition, the question of serious and slight injury risk in fatal accidents is examined. This is a more complicated question, aiming to identify the parameters affecting the probability of a road user being an additional accident casualty, given that there is a fatality. This eventually translates into determining the factors that make a fatal accident have even more consequences than the ones expected by default (i.e. one or more fatalities).

Naturally, the accident size effects also apply in this case. The more vehicles involved in an accident that will be fatal, or the more the vehicle occupants, the lower the probability of serious or slight injuries among the non-fatalities. As regards the comparability of risks, additional injuries are less likely to be found among heavy vehicles occupants, and more likely to be found in light vehicle occupants. However, it is unlikely that in a fatal accident involving a pedestrian it is another road user who ends up as the fatality. Finally, among car occupants, the probability of being an additional non-fatal casualty in a fatal accident obviously depends on the (size of the) accident opponent. It will be examined whether the proposed ways to account for these effects in fatality risk analysis are efficient in the extended injury risk analysis.

○ 4.2.2 Modelling the fatality risk

The dependent variable initially selected is a binary variable (1: killed, 0: not killed) of the injury severity of each road user in the FAI data. A binary logistic regression model was fitted to the data. Initially, an "empty" single-level model (i.e. including a constant term only) is created. This model will be considered as

a baseline for comparing more analytical models in terms of performance. It is noted that standard estimation methods for logistic regression models can not provide reliable estimates of models fit (Martensen and Dupont, 2007). For this reason, Bayesian modelling is exploited, which is based on Markov Chain Monte Carlo (MCMC) simulation techniques (see details in Appendix 4.1).

The first step is to examine the presence of hierarchical dependences in the data. These dependences are examined in "empty" models. The results are presented in Table 4.2. Model 1, which is obtained after 15,000 MCMC iterations, is the baseline empty model with a constant term only. In Model 2, the possible hierarchies in the FAI data are tested in terms of geographical dependencies, through a two-level model in which road users are nested into countries. The results show that the random variation at the country level is not significant after 80,000 MCMC iterations, whereas only a marginal improvement in the likelihood is obtained.

Another type of dependencies examined in Model 3 is the dependencies due to the accident process. In particular, a three-level structure is considered, in which persons are nested into vehicles and vehicles are nested into accidents. In this case, the random variation at the accident level (level 3) is not significant; however the variation at the vehicle level (level 2) is very significant. The improvement of the likelihood statistic compared to Model 1 is also very important. It is thereby indicated that the probability of a person being killed in a fatal accident does vary systematically across different vehicles and that the consideration of such a hierarchical structure improves the fit of the model.

	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>
<i>Fixed effects</i>			
<i>Constant</i>	-0.260 (0.037)	-0.255 (0.071)	-0.219 (0.046)
<i>Random effects</i>			
σ^2_{u0} (country level)		0.022 (0.030)	
σ^2_{u0} (vehicle level)			0.934 (0.243)
σ^2_{v0} (accident level)			0.004 (0.006)
<i>-2*loglikelihood</i>	4208.6	4198.7	3678.0
<i>Number of iterations</i>	15,000	80,000	300,000

Table 4.2: Testing hierarchies in the data ("empty" multilevel models)

It is noted that it is possible that this vehicle dependence may become more or less significant once explanatory variables are introduced in the fatality risk model. Nevertheless, the results from the "empty" models suggest that it is necessary to examine this type of dependence. The building of a model with explanatory variables for fatality risk, given that one is involved in a fatal accident is presented in Table 4.3.

Model 4 only includes the variables meant to capture the accident size bias and correct for the incomparability of risks. The parameter estimates correspond to what could intuitively be expected, suggesting that, compared to car occupants, vulnerable road users have an impressively increased probability of being the fatalities in fatal accidents, whereas motorcycle riders also have increased

probability of being the fatalities in fatal accidents. On the other hand, HGV occupants have a significantly lower baseline probability than car-occupants of being the fatalities in fatal accidents. As regards accident size bias, the negative coefficients for the variables "number of participants" and "number of occupants" indicated that the baseline fatality risk is lower for road-users involved in accidents involving two or more vehicles (as compared to single-vehicle accidents) and for road-users in cars containing two occupants or more (as those being the sole car occupant).

In Model 5, the additional explanatory variables that were found to be significant are presented. These were selected after careful consideration of all possible correlations and were initially tested individually (i.e. in the "empty" model), in order to make sure that their parameter estimates are stable. The main explanatory effects identified can be summarized as follows:

- Senior road users (i.e. with more than 65 years of age) have a higher probability of being a fatality in a fatal accident, obviously due to their physical vulnerability.
- The probability of being killed, given that one is involved in a fatal accident, increases when there is more than one event for the road user's vehicle in the accident. Given that this effect is obtained in a model accounting for the accident size bias, it may suggest that there is a higher probability for more than one fatality in a fatal accident including more than one event.
- A speed limit higher than 50 km/h appears to increase the probability of being killed when involved in a fatal accident. This variable may reflect increased travel speeds and rural road environments, resulting in either increased probability of single vehicle accidents (e.g. run off-road) or higher impact speeds in multi-vehicle accidents.
- Road users in a vehicle that braked before the collision have a lower probability of being killed, once involved in a fatal collision. Probably, a vehicle braking results in lower impact speed for its occupants.
- Road users with confirmed impairment have a marginally significantly higher probability of being the fatalities in fatal accidents. It is likely that this is due to the limited physical capabilities of impaired road users. However, given that the vast majority of impaired road users are drivers, it is also possible that these drivers travel at higher speeds or carry out inappropriate manoeuvres, resulting in more severe accident consequences for themselves.

It is also noted that the vehicle-level random variation examined was found to be non significant in the last two model, not confirming the initially important difference in the fatality risk of road users in different vehicles. It appears that most of the random variation identified in the "empty" models of Table 4.2 is captured by the explanatory effects, and probably especially those of Model 4, which account for the different vehicle types occupied by the road users (e.g. motorcycle, heavy vehicle etc.).

		Model 4	Model 5
Constant	1.000	0.931 (0.155)	-1.343 (0.201)
Number of participants	One Vehicle	.	.
	Two or More	-1.623 (0.171)	-0.909 (0.122)
Number of occupants	One occupant	.	.
	Two or more	-0.634 (0.106)	-0.820 (0.108)
Motorcycle	Yes	2.128 (0.198)	1.827 (0.174)
	No	.	.
Heavy vehicle	Yes	-1.461 (0.195)	-1.354 (0.187)
	No	.	.
Vulnerable	Yes	3.693 (0.285)	3.746 (0.269)
	No	.	.
Senior	> 65 years old	.	1.430 (0.156)
	Younger	.	.
Number of Events	More than one	.	1.621 (0.123)
	One	.	.
Speed limit	> 50 Km/h	.	0.347 (0.105)
	=< 50 Km/h	.	.
Vehicle Braked	Yes	.	.
	No	.	0.455 (0.116)
Impairment	Yes	.	0.280 (0.170)
	No or unknown	.	.
σ^2_{u0} (vehicle level)		0.353 (0.229)	0.031 (0.036)
-2*loglikelihood		3018.36	2816.15
Number of iterations		300,000	250,000

Table 4.3: Model building for fatality risk in all accidents (binomial two-level models)

The reduction of the likelihood statistic of Model 5 is important compared to Model 4, confirming the additional explanatory effect offered by the new variables. Useful information on the model's performance is also obtained through the number of correctly predicted outcomes. In particular, Model 5 correctly predicts 82% of fatalities and 69% of non-fatalities in the FAI data, resulting in overall 76% of all road users' outcome being correctly predicted. It can be seen that the model's performance with respect to non fatalities is much less satisfactory, suggesting that the model includes sufficient variables for predicting fatalities, but lacks some variable(s) that increase the probability of non fatality.

○ 4.2.3 Modelling fatality and injury risk

The next stage of the analysis concerns the development of a multinomial model, in which the accident severity in fatal accidents can be considered in more detail. In this case, the dependent variable is a multinomial one, with the following categories: fatality, serious injury, slight injury, no injury. What in this case is modelled is the variation of the probability of occurrence of each injury

category associated with variations in the predictor variables (e.g., person age, vehicle types, speed limits etc.). Instead of speaking of the fatality risk, in the present analysis it will be recurrently referred to as the “slight, serious, or fatal injury risk”. The main objective of this analysis is to test whether explanatory variables have a different effect on these different casualty risks in fatal accidents. However, the interpretation is expected to be less straightforward in this case. Indeed, given that each and every accident is *bound to* contain at least one fatal injury, serious and slight injuries in the FAI data cannot but be additional casualties in fatal accidents. Variation in the respective injury risk will therefore be explained in this particular context.

An unordered multinomial model is opted for, assuming thus no "meaningful" order of the four categories (Martensen and Dupont, 2007); although there is an ascending order of accident consequences, the categories are not necessarily equidistant, i.e. a fatality compared to a serious injury can not considered as equivalent to a serious injury compared to a slight injury, and therefore an unordered model is considered to be a safest option.

Before testing the explanatory variables, a basic modelling structure was created, including the necessary variables to account for accident size and risk comparability bias. Moreover, a multilevel structure of road users nested into vehicles was defined. All additional explanatory variables that were found significant in the binomial model were initially included in the multinomial model. The results (Model 6) are presented in Table 4.4.

One can notice the positive intercept for fatalities, and the negative intercept for serious injuries. They indicate respectively that, overall, the probability of being killed in a fatal accident is larger than the probability being uninjured, while the probability of being seriously injured is lower than the one of being uninjured. This is reasonable when considering that there was at least one fatality in all the accidents in the dataset, but not necessarily a serious injury. The intercept for slight injuries was not significant.

As regards the effects of accident size variables, a common negative effect of the number of participants was estimated, given that the three separate effects for fatality, serious and slight injury were practically equal. This suggests that, in multi-vehicle fatal accidents, the fatality risk of all road users involved - but also the non-fatal casualty risk - is lower than in single-vehicle accidents. This is reasonable, given that when there is only one vehicle in a fatal accident, all fatalities and additional casualties are bound to be found in this vehicle.

Having corrected for the effect of the number of vehicles in a fatal accident, the effect of the number of occupants is less straightforward. The effect on fatality risk in particular is not significant in Model 6. Given that a significant negative effect is obtained when testing the two accident size variables alone, it appears that other explanatory variables account for this initial effect in the final model. Moreover, a positive effect of the number of occupants is obtained for serious and slight injuries. This suggests that, the more persons in a vehicle involved in an accident, the higher the probability of each one of them being injured, which

is intuitive and may be considered to reflect the accident size effect in *any* accident, and not just in a fatal one. Therefore, in fatal accidents, the higher the number of occupants, the lower the baseline fatality risk for each one of them (i.e. fatal accidents size bias), but the higher the baseline injury risk for each one of them not being the fatality (i.e. general accident size effect).

With regard to the issue of the incompatibility of risk, the results confirmed that motorcycle riders have increased probability of being the fatality in fatal accidents, and that they also have a higher probability of being additional non-fatal casualties. On the other hand, HGV occupants are much less likely to be part of the fatalities in a fatal accident, and to be seriously injured if there are additional casualties in the fatal accident. On the other hand, they only have a faintly lower probability of being slightly injured when involved in a fatal accident. It is thereby indicated that the protection offered by a heavy vehicle may not fully prevent slight injuries once involved in a fatal accident. As regards vulnerable road users, they have an increased serious or slight injury risk in fatal accidents and even more increased fatality risk. Given that only a couple of vulnerable road users were uninjured in the FAI data, not including this variable in the model would have resulted in important bias in the remaining parameter estimates.

Looking at the parameter estimates of the additional explanatory variables, the following can be identified:

- The variable "senior" was found to be significant only for fatalities and slight injuries. Senior road users have a significantly higher probability of being the fatality in a fatal road accident, a marginally higher probability of being an additional serious casualty, and also a higher probability of being slightly injured, obviously because of their vulnerability. It may seem somewhat surprising that the mean effect of "senior" appears to be higher on the slight casualty than on the serious casualty probability, but when examining the standard errors of these parameter estimates, one can see that they overlap, and therefore one can not conclude that these probabilities differ significantly.. Besides, in the FAI data, the numbers of senior road users that were seriously and slightly injured are practically equal.
- The fact that the vehicle braked before the collision does not appear to affect serious or slight injury risk in fatal accidents.
- Impairment appears to increase the probability of any type of casualty in a fatal accident. Impaired road users have increased probability of being additional non-fatal casualties once involved in a fatal accident. It is reminded that very few impaired road users were non drivers. It is also noted that in the FAI data there were more slightly injured than seriously injured impaired road users, however the parameter estimates appear to contradict this fact.
- Finally, speed limits higher than 50 km/h increases the probability of all injury scores, which is reasonable, for the reasons mentioned in the binomial model. It is also underlined that the standard errors of the estimated effects for serious and slight injury risk present some overlap, and therefore it is not possible to conclude which one is the highest effect.

Analysing European in-depth data: Methodological framework and results

		<i>Model 6</i>		
		<i>Fatality</i>	<i>Serious injury</i>	<i>Slight injury</i>
<i>Constant</i>	1.000	0.992 (0.124)	0.029 (0.174)	0.548(0.166)
<i>Number of participants</i>	One Vehicle		.	
	Two or More		-2.496 (0.027)	
<i>Number of occupants</i>	One occupant	.	.	.
	Two or more	0.120 (0.087)	1.518 (0.130)	0.792 (0.120)
<i>Motorcycle</i>	Yes	3.410 (0.158)	2.292 (0.209)	1.672 (0.246)
	No	.	.	.
<i>Heavy vehicle</i>	Yes	-1.548 (0.167)	-1.199 (0.222)	-0.144 (0.161)
	No	.	.	.
<i>Vulnerable</i>	Yes	5.527 (0.246)	3.794 (0.339)	3.221 (0.360)
	No	.	.	.
<i>Senior</i>	> 65 years old	1,490 (0,131)	0.353 (0.194)	0.575 (0.191)
	Younger	.	.	.
<i>Speed limit</i>	> 50 Km/h	1.128 (0.085)	0.830 (0.116)	0.941 (0.121)
	=< 50 Km/h	.	.	.
<i>Vehicle Braked</i>	Yes	.	.	.
	No	0.463 (0.096)	0.126 (0.127)	-0.033 (0.126)
<i>Impairment</i>	Yes	1.151 (0.142)	0.957 (0.172)	0.450 (0.206)
	No or unknown	.	.	.
σ^2_{u0} (vehicle level variances)		0.000 (0.000)	0.616 (0.132)	0.527 (0.105)
σ_{u0} (vehicle level covariances)			-0.376 (0.102)	
-2*loglikelihood				7964.74

Table 4.4 Model for fatality, serious and slight injury risk in all accidents (multinomial two-level model)

Although some of the parameter estimates of Model 6 do not bring straightforwardly interpretable effects, no other variables were found to be significant, at least from those that did not present significant correlations with the predictors already used. As regards hierarchical dependences, a significant variation of serious and slight injuries across different vehicles in fatal accidents was found, but no respective random variation of fatalities.

It is noted that Bayesian estimation was proved to be less efficient in the multinomial multilevel modelling, providing substantially higher and unstable estimates of the random parameters compared to the standard estimation, which is a known problem that may occur in this case (Browne, 2003). For these reasons, the standard estimation method's results were considered to be more reliable and are the only ones presented here. The likelihood statistic estimation, however, is quite approximate in this case and can only be taken as a rough measure of model fit.

Overall, the performance of Model 6 is less satisfactory compared to Model 5, although the likelihood statistic of Model 6 is significantly reduced compared to

the likelihood statistic of the respective "empty" model. In particular, Model 6 correctly predicts 78% of fatalities, but only 39% and 47% of serious and slight injuries respectively, resulting in 63% correctly predicted outcomes in total.

Hence, an important part of the variation in injury risk when involved in fatal accidents remains unexplained, especially as regards serious injuries. There is some suspicion that serious and slight injuries can not be distinguished by the specific model. However, none of the available additional variables could improve the model's ability to correctly predict the different types of injury risk. It is likely that injury risk in fatal accidents is to a large extent random, at least in this sample. Furthermore, it is likely that the consequences of the comparability of risk bias are even more complex when examining injury risk in fatal accidents (and not only the fatality risk). The interpretation becomes more complicated when considering that, in the FAI data, although at the road user level injuries indicate a less severe consequence than that of a fatality, at the accident level they are associated with a more severe accident, because there is an additional casualty.

Consequently, although the multinomial model confirms to a satisfactory degree the findings of the binomial model for fatality risk in fatal accidents, and it also reveals a few interesting effects on injury risk, the existing results can not be used for drawing conclusions.

○ 4.2.4 Conclusion

In this section, a first attempt at modelling the fatality risk of road users involved in fatal accidents is presented, by investigating the potential for analyzing the whole set of road users in the FAI data. Several known difficulties are involved in this first attempt, concerning both the accident size bias and the incomparability of risk, and a lot of effort was put in addressing these issues while at the same time working with a large and heterogeneous set of road users. The analysis focused on the development of a binomial model for fatality risk, followed by a multinomial model for fatality, serious and slight injury risk.

Although there are several encouraging findings in the proposed fatality risk model for all road users, such as the estimation of interesting and intuitive variables effects, and the satisfactory models fit, there are several weaknesses as well. First of all, non fatalities are less adequately handled by the model. This suggests that, despite the effort put in correcting for the incomparability of risks, there are additional factors affecting fatality risk in fatal accidents, which are not included in the model. One such factor may concern variables which are specific to car occupants (the use of safety equipment, the position in the vehicle etc.). These could not be included in the model, as they are not applicable to all road users.

Another type of missing factor, and maybe a more important one, concerns the collision opponent. In particular, the incomparability of risks among different road users involved in fatal accidents is not only due to the vehicle of each road user, but also to the type of opponent each one of them has collided with.

Especially for passenger cars, the opponent may have an important and heterogeneous effect on the fatality risk.

Consequently, it is likely that the absence of these variables in the present models results in unsatisfactory performance with respect to non-fatalities in fatal accidents, although fatalities are predicted quite successfully.

The fact that non-fatalities among road users involved in fatal accidents are more difficult to distinguish in statistical analysis is confirmed by the multinomial modelling results. In this case, the effects observed on fatality risk are consistent with those obtained on the basis of the binomial model. However, serious and slight injuries are less correctly predicted, and the parameter estimates for this type of casualty present a couple of strange effects. Consequently, and also due to the fact that non fatal injury risk factors in fatal accidents are more difficult to interpret as compared to fatality risk factors, the multinomial model is not appropriate for drawing conclusions.

Departing from the binomial model, however, it is possible to seek ways for improving the conceptual framework of the analysis, and consequently obtain better results in terms of fatality risk of road users involved in fatal accidents. A specific analysis devoted to car occupants and excluding single vehicle accidents may limit both the accident size and the incomparability of risks effects, and possibly reveal some additional effects. Furthermore, it would be worth coding a variable concerning the collision opponent vehicle, which is probably the main source of risk incomparability bias. The results of these alternative modelling processes are presented in the following sections.

4.3 The “Car-Car Analysis”¹⁷

Emmanuelle Dupont, Yvan Casteels & Heike Martensen, IBSR

The model presented in the previous section illustrated its ability to handle the accident size bias and incomparability of risk issues. It also, generally speaking, provided indications on the factors that are associated with variations of the fatality risk for road-users involved in fatal accidents, generally speaking.

Yet, given that all road-user types were pooled together in this analysis, the predictors that may be tested in the context of such a model are also *de facto* limited to those that are of applications for all road-users, whatever their transport mode. Of course, this rules out the test of a series of other predictors that may otherwise be suspected to have an important influence on the outcome that a fatal accident will have for each individual road user. Seatbelt use, for example, could not be tested in the present case because it is rated as “not applicable” for several road-users categories. Helmet use is also likely to be of importance in the case of motorbike riders, but it is not of application for all other road-users.

The analysis presented here aims at fully controlling for the incomparability of risk by focusing exclusively on accidents that took place between two cars. In so doing, both the type of road user and the type of collision opponent are kept constant. Focusing on car occupants additionally allows the selection of the predictors that are most relevant to this type of crash, and to provide a more detailed view of the factors that matter in differentiating between the survivors and fatalities among car occupants involved in fatal accidents.

Below, the results associated with the predictors relevant to controlling for the accident size and the incomparability of the baseline risk are presented. Then, the results for the road-user, vehicle, and accident predictors are described.

○ 4.3.1 Variables to be controlled for:

Accident Size

Vehicle occupancy appears to be significantly and negatively related to the fatality risk, as expected ($B=-0.33$, $S.E= 0.07$, $p<.001$). The probability for a road-user to end up among the fatalities of the accident decreases as the number of occupants in the vehicle increases. Figure 4.4 indicates that the percentage of road users killed is the highest in cars with one and only one occupant and is close to 50% (there were always two cars – and thus at least 2 road users – involved in the accident).

¹⁷ The model fitted here is a binary logistic model (logit link function) estimated by means of Bayesian methods (Markov Chain Monte Carlo). A preliminary step was taken to test the hierarchical structure of the data. The random variation at the country and accident levels was estimated as null, while the one at the vehicle level was not significant ($XX= 0.15$, $S.E = 0.17$, *n.s.*).

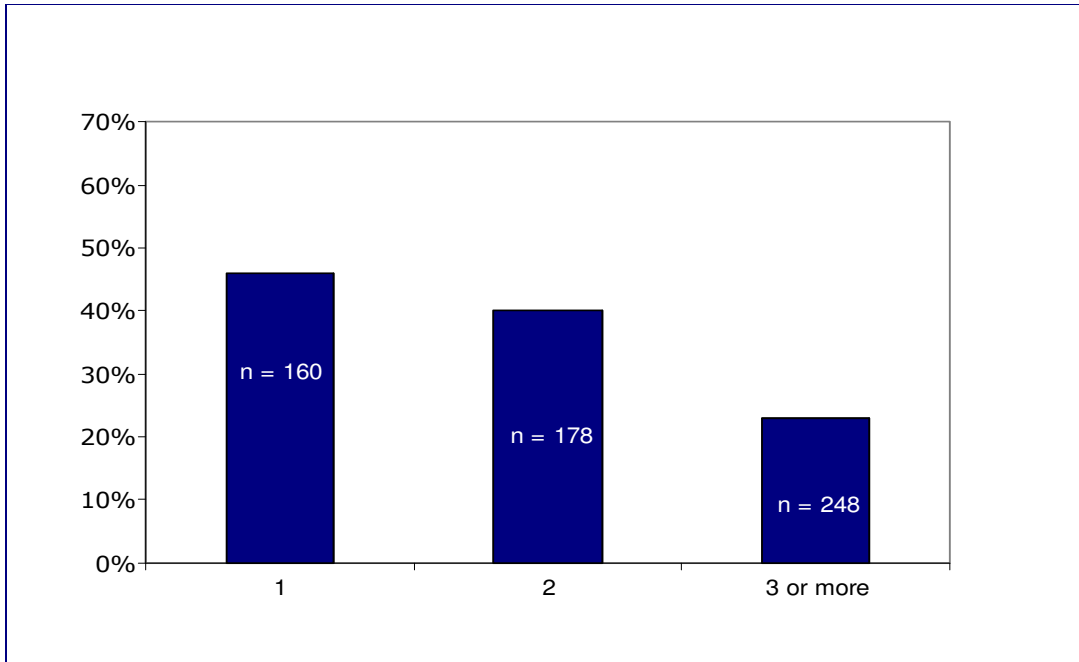


Figure 4.4: Percentage of fatalities as a function of vehicle occupancy

○ 4.3.2 Road-User vulnerability factors:

Age

Figure 4.5 shows the percentage of road-users killed in various age categories. The global relationship between age and the road-user fatality risk was not significant. However, the fatality risk is strongly affected by whether or not the road-user can be considered to be “senior”, i.e., to be aged 65 or more. As indicated in Table 4.5, the odds of dying in a fatal accident indeed increase for “senior” road-users as compared to all others ($B=1.22$, $S.E.=0.31$, $p<.001$).

Gender

When gender is entered alone as a predictor of the fatality risk, it does appear that the odds of dying in a fatal accidents are lower for women as compared to men ($B=-0.482$, $S.E.= 0.198$, $p=0.15$). This is, however, quite a counter-intuitive result (women were expected to be more vulnerable than men).

Several variables were thus examined, variables for which gender differences could also be found, and that could themselves better explain the gender difference identified concerning the fatality risk.

The inclusion of seatbelt in the model does not affect the findings for gender, suggesting that the lower fatality risk for women cannot be attributed to gender differences in terms of seatbelt usage, although the proportion of women wearing a seatbelt (53%) is indeed larger than that of men (38%).

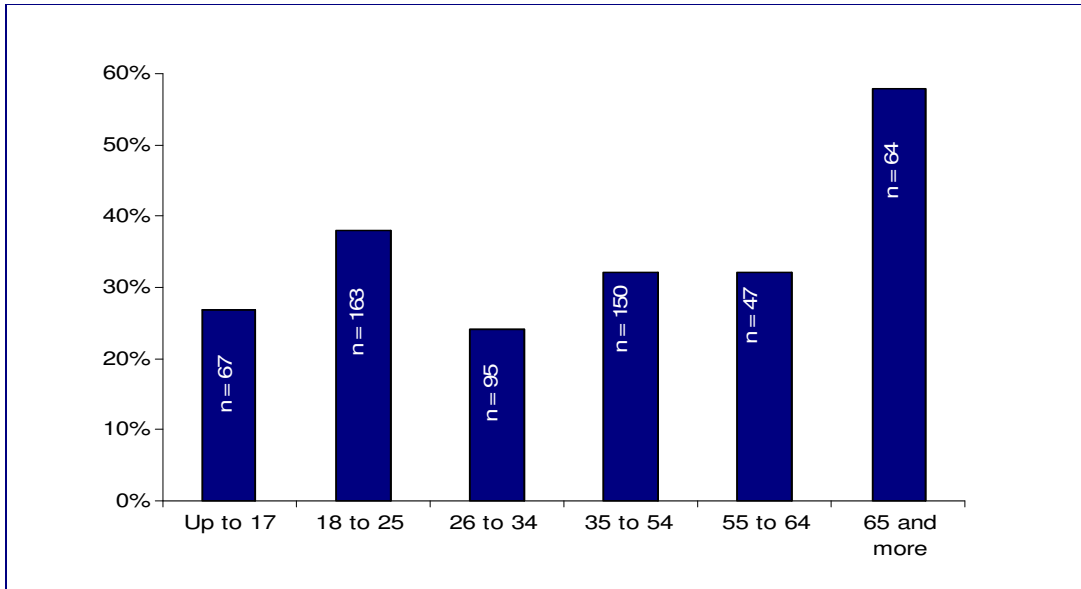


Figure 4.5: Percentage of fatalities as a function of road-user age

There also appears to be an important difference between men and women with respect to whether they occupy the car as driver or as passenger: The proportions of drivers and of passengers are completely reverted among men and women in the sample. While most of the men were drivers, the majority of the women were passengers.

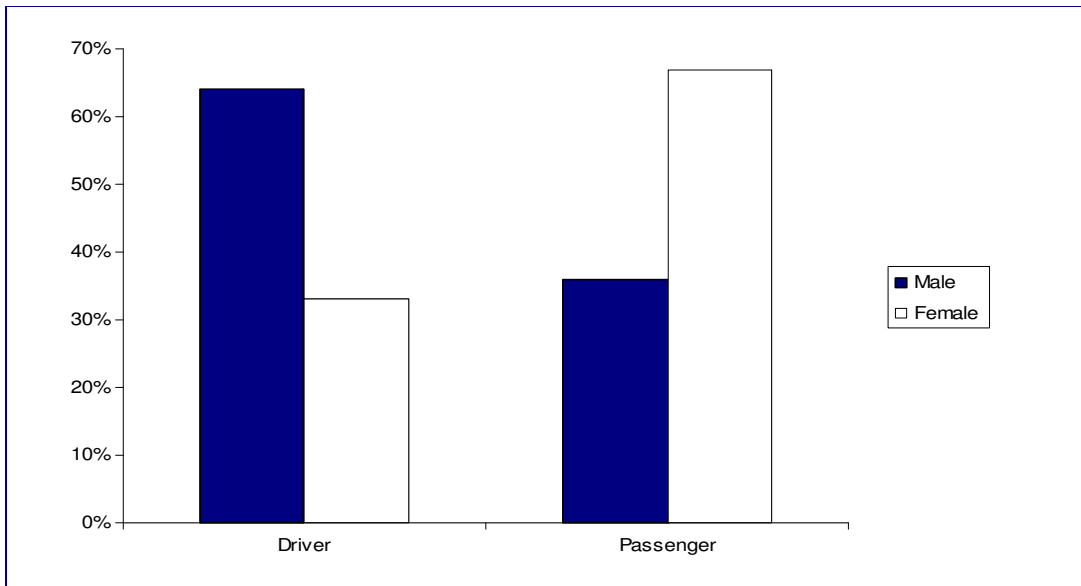


Figure 4.6: Percentage of fatalities among passengers and drivers as a function of gender

When entered alone in the model, the variable “UserClass” (i.e., whether the road-user is a driver or a passenger) does not appear to be significantly related to the fatality risk ($B=0.17$, $S.E.=0.21$, $p=0.44$). However, if the “UserClass-Gender” interaction term is entered in the model, the main effect of gender becomes non-significant ($B=0.356$, $SE=0.66$, $p=0.59$). Although this interaction

term was not significant itself ($B=-0.588$, $S.E=0.41$, $p=0.15$), the pattern of results for this interaction suggests that there is, indeed, a lower fatality risk for women as compared to men, but that this holds mostly among for passengers.

Because the gender effect does not appear to be reliable, it was not included in the final fatality risk model.

Position within the vehicle

87 percent of the road-users were seated at the front of the car at the time of the accident (53% as drivers and 24% as passengers). Although the proportion of road users killed is slightly higher among the occupants at the front (38%) rather than at the back of the car (22%), this difference is not reliable ($B= -0.344$, $S.E.=0.365$, $p= 0.35$)¹⁸.

The discussion of the results for the gender variable already indicated that whether the road-user was a driver or a passenger did not significantly affect their fatality risk.

○ 4.3.3 Passive Safety

Seatbelt

Seatbelt use was expected to be associated with a decreased fatality risk as compared to non-use of seatbelt. In its original format, the seatbelt variable was made up of four categories: “Used” (40% of the observations); “Not Used” (12% of the observations), as well as “Used Claimed” (2% of the observations) and “Unknown” (45% of the observations). Given the very small number of “Used Claimed” observations, these were merged with the observations in the “Used” category¹⁹. The “Unknown” category was maintained as a separate category and introduced as such in the model. Figure 4.7 shows the percentage of road-users killed in each of those three categories.

It was tested whether people not wearing their seatbelts had greater odds of dying as people wearing it. The results indicated that the odds of dying are higher for people not wearing their seatbelt than for those who did wear it ($B=0.89$, $S.E=0.29$). As can be seen in Figure 4.7, this was indeed the case.

¹⁸ More than the mere position of each road user in the car, their position *relatively to the area of the car that was damaged mostly* during the crash may strongly affect the fatality risk. To investigate this, a new variable was computed that indicates whether the road-user was seated in the car’s most damaged area or not. However, it turned out that most positions in the car can be related to more than one damaged-area category (for example: the driver’s seat can be considered to correspond to “front”; “right”, “roof” and “multiple” damaged areas altogether). As a consequence, the vast majority of the road users (94%) could eventually be considered as having occupied the most damaged area of the car and the variable was not significant.

¹⁹ It might be the case that, among people having claimed to have used their seatbelt, some actually did not do so. This should not be a major problem, however, given the few number of observations in the “Use Claimed” category. Besides, should it be the case that the used claimed observations correspond in fact to people that did not wear their seatbelt, then the inclusion of these observations in the “Used” category should only weaken the chances that the “Used” – “Not Used” comparison meets the expectation that seatbelt use is associated with decreased odds of dying in the accident.

The odds of dying in the accident were much higher for people not wearing their seatbelt as compared to those wearing it ($B=0.89$, $S.E.=0.29$)

The comparison between the seatbelt used category with the “unknown” one was not significant ($B=0.13$, $S.E. = 0,125$). As can be seen in Figure 4.7, the percentages of fatalities for those two groups are nearly similar. The result of this last comparison is probably affected by the fact that the police arriving at the accident scene are more often unable to tell whether a person had worn a seatbelt if that person survived than if she died in the accident.

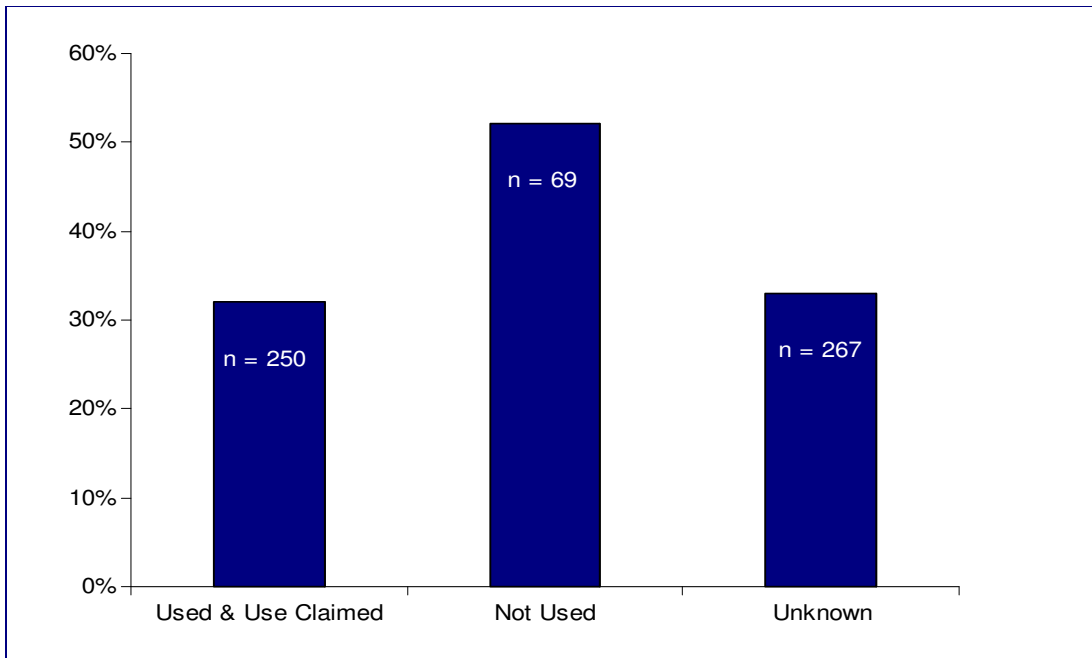


Figure 4.7: Percentage of fatalities among passengers and drivers as a function of seatbelt use

Airbag

For 67% of the observations, airbag was rated as “not applicable” because the place occupied by the road-user was simply not equipped with an airbag (this proportion strongly co-varies with the age of the vehicle). An airbag was available and deployed at the moment of the accident in 21% of the cases and did not – although present - deploy at the moment of the accident in only 5% of the cases. The functioning of the airbag was unknown in 13% of the cases.

Although one would naturally expect the odds of dying to decrease with airbag deployment, the nature of the airbag variable renders the test of this assumption difficult. Figure 4.8 shows the percentage of fatalities among the road-users depending on airbag availability and deployment status. The instances in which airbag was “Not Applicable” have to be distinguished from the “No Airbag” case: In the latter case, although the vehicle was indeed equipped with airbags, these failed to deploy at the crucial moment. Contrary to expectations, the percentage of fatalities appears to be the lowest among the road users in vehicle which were equipped with airbag, but in which the latter failed to deploy. Actually, this

can simply indicate that the crash impact was not strong enough to trigger the airbag, and has – by the same token – resulted in fewer casualties. The percentage of fatalities was higher in the case in which no airbag was available at all.

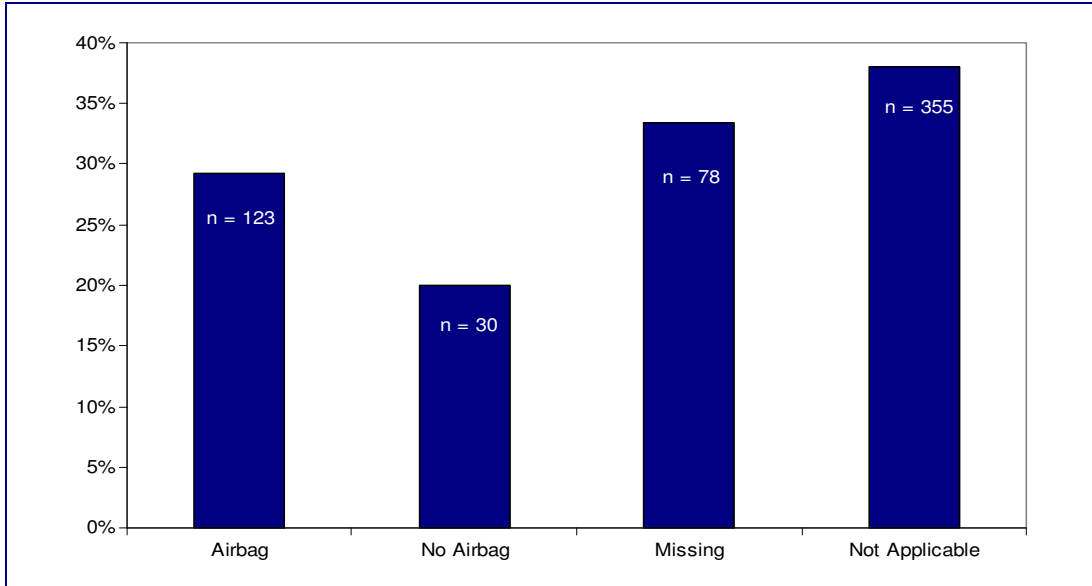


Figure 4.8: Percentage of fatalities among passengers and drivers as a function of airbag deployment

Driving Aids (ABS, ESP, CSS, TCS)

About one-third of the road users were travelling in a car that was equipped with ABS (30%). The car had no ABS in 40% of the case, while this was unknown for the remaining 26%. When entered alone in the model, ABS appears to be significantly related to the fatality risk (0 “ABS”, 1 “No ABS”, 2 “Unknown”). The risk is more elevated for the occupants of cars without ABS as compared to the occupants of cars with ABS ($B=0.76$, $S.E.= 0.22$, $p<0.000$). No difference was associated with the “Unknown” category ($B= 0.40$, $S.E= 0.25$).

However, once the variable “Vehicle Age” is entered in the model, the effect observed for the presence of ABS ceases to be significant. Obviously, this variable co-varies with vehicle age (and this is the case for all driving aids devices). The relationship between vehicle age and the fatality risk is discussed below. For now it is important to note that vehicle age, as a predictor, has the property to “summarize” the evolution of the car in terms of driving aids taken *altogether*. As such, it is likely to form a more powerful predictor than separate indicators for the presence of the various aid devices.

It was impossible to further test the effect of other driving aid devices, given the very small proportion of cars that were equipped with them (4% for ESP; 7% for CSS; and 5% for TCS).

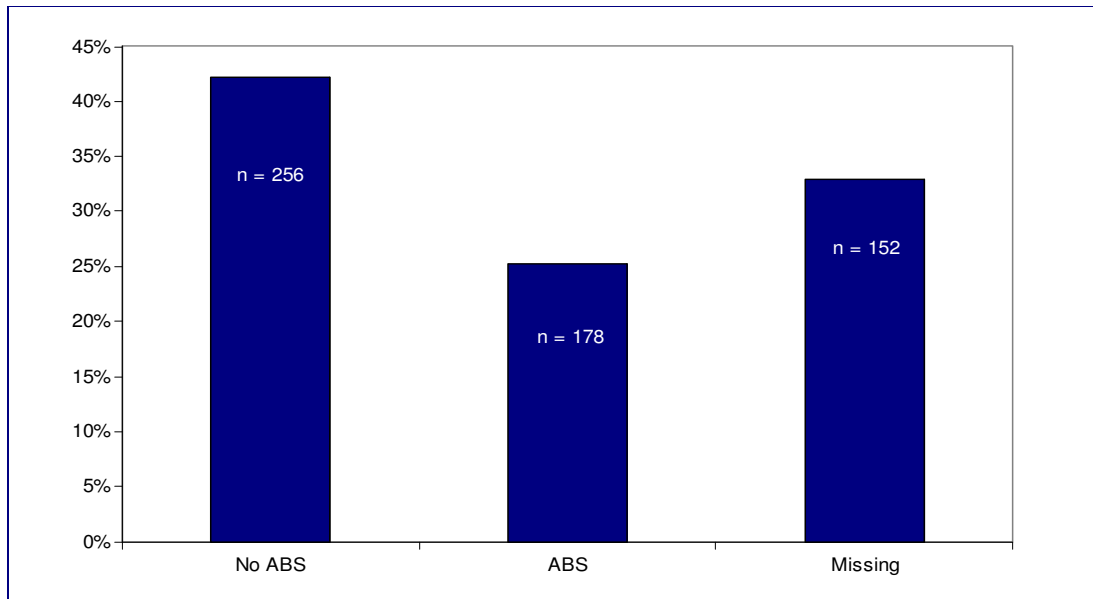


Figure 4.9: Percentage of fatalities among passengers and drivers as a function of whether the vehicle is equipped with ABS

Vehicle Age

The effect of vehicle age was strongly significant ($B=0.08$, $S.E=0.02$, $p<.000$). As expected, the results indicated that the fatality risk for the car occupants increased with vehicle age.

Vehicle Size and Mass

The number of missing values (117) for the variable “Kerbweight” precluded the inclusion of this predictor in the model. The mass of the car was, however, highly correlated with its length ($r=.83$) and with its width ($r=.84$), both variables for which the number of missing values was much lower (34). These were used as a proxy for car mass. When entered alone in the model, vehicle width appears to be significantly related to the fatality risk ($B=-4.67$, $S.E=1.05$, $p<.002$): The broader the car, the lower the risk for its occupants to die in the accident. When integrated in the built-up model width also appears to affect the fatality risk to a significant extent ($B=-3.64$, $S.E=1.16$, $p<.002$), but the inclusion of this predictor in the model is not associated with any decrease of the DIC criterion. Besides, once vehicle length is included in the model in turn, the standard error of the car width coefficient increases, and none of these car-size indicators is significant any more. Consequently, these predictors were not included in the final model.

○ 4.3.4 Accident Circumstances

Overall description

The sample sizes in Figure 4.10 show that the observations are quite evenly distributed across the different periods of the day and of the week. Similarly, the share of observations for accidents that took place under day light conditions and other light conditions (mostly darkness) is homogeneous (52 and 48%

respectively). The vast majority of the observations that were made came from accidents that took place on dry roads (75%) and a small number of them correspond to accidents that took place on wet roads (22%).

Most of the observations (88%) came from accidents on two-way roads divided by a painted line. The observations are quite evenly distributed as far as speed limit is concerned: 33% of the road-users were involved in accidents on roads with a speed limit lower or equal to 50 km/h, 14% in accidents with a speed limit between 80 and 100 km/h, and 53% on roads with a speed limit higher than 100 kmh. The traffic flow was light (43%) or normal (35%) in most cases. Instances of accidents that happened under heavy traffic flow conditions are far more scarce (5%). The number of missing values for the traffic flow variable was substantial (16%).

Very few observations were made for accidents that occurred on motorways (2%). 61% of the observations came from accidents that took place on road sections. When the accident occurred at a road junction, this was either a crossroad (54%) or a T-shaped junction (35%) in most cases. 75% of the observations were made in rural areas as compared to 25% in urban or mixed areas.

When the observations were very unevenly distributed across the categories of one predictor (e.g.: motorway, type of road), the relation between the corresponding predictor and the fatality risk was not investigated. The same is true for variables in which the number of missing values was substantial (e.g.: traffic flow).

Moment of the accident

The fatality risk for road users involved in fatal accidents does not significantly differ between week and weekend (B=0.21, S.E=0.18) or between night and day (B=-0.04, S.E=0.18).

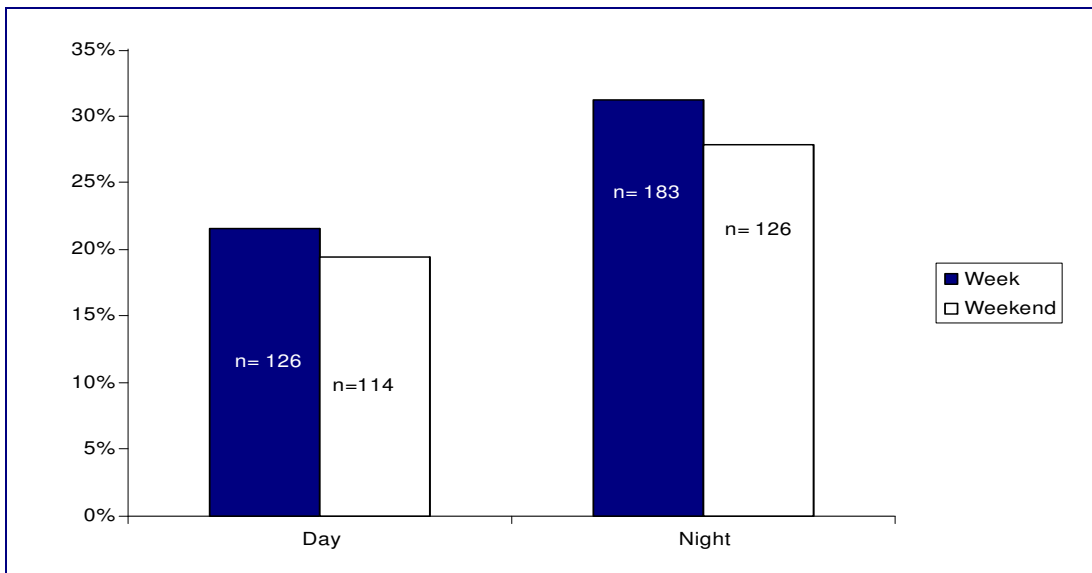


Figure 4.10: Percentage of fatalities among passengers and drivers as a function of moment of the week



Environmental conditions

Given the lack of variation in the weather conditions for the accident, this variable was not taken into account in the analysis.

For light conditions, the results revealed that they do not affect the fatality risk ($B=-0.19$, $S.E=0.18$).

Road

Speed limit can be considered an indication (although indirect and imprecise) of the speed at which the vehicle occupants were travelling at the moment of the accident, but this variable did not appear to be significantly related to the fatality risk.

The odds of dying are more elevated for the road-users who were involved in road-sections accidents as compared to road-users in road-junction accidents ($B=-0.70$, $S.E=0.21$, $p<.001$). Whether the accident took place in a rural or in an urban area did not affect the fatality risk.

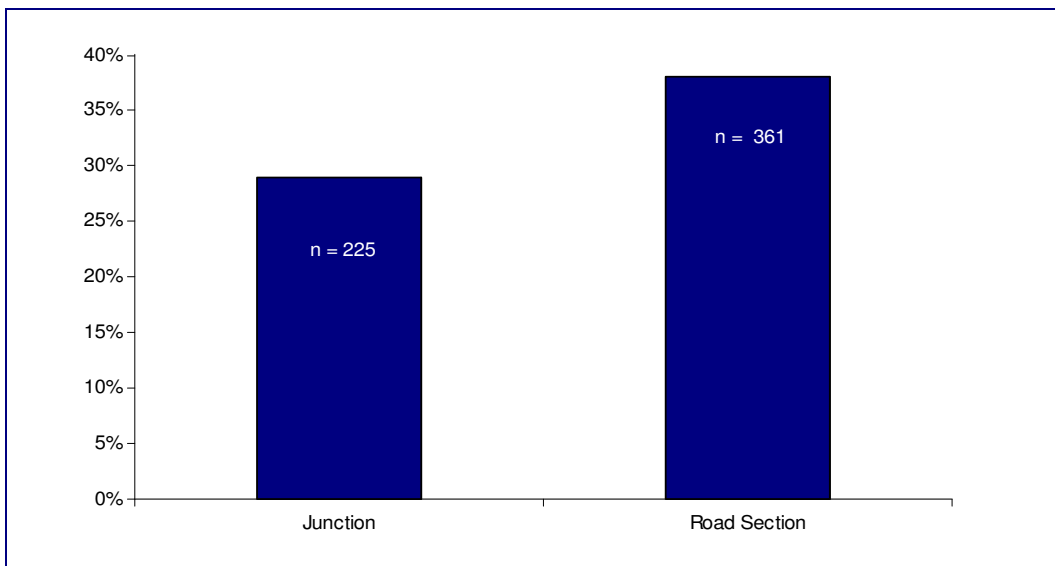


Figure 4.11: Percentage of fatalities among passengers and drivers as a function of whether the accident took place on a road section or on a road junction

Development of the accident

The area of the car that was most damaged provides an indication of the type of impact characterizing the accident. For the large majority of the cases the area of main damage was the front (67%). The impact area played an important role with respect to the fatality risk: The odds of dying were significantly lower for the road users in cars that had been damaged most at the front as compared to any other damaged area²⁰ ($B=0.50$, $S.E=0.24$, $p<.05$).

²⁰ Front damage: 67% ; left damage :10%, right damage : 12%, back, roof, and « multiple » damages : 2, 3, and 4% respectively.

Another way to characterize the development of the accident is to take account of the different avoidance manoeuvres that have or have not been conducted. Initially, it was differentiated whether the driver had braked to avoid the accident, steered, or braked and steered. It turned out however, that only whether the driver braked or did not brake make a significant contribution to the fatality risk of the occupants. As shown in Figure 4.12, cars where the driver had braked contained a fatality in 25% of the cases, while for cars where the driver had not brake this was 38%. This difference was marginally significant ($B=0.48$, $S.E=0.26$, $p<.10$)

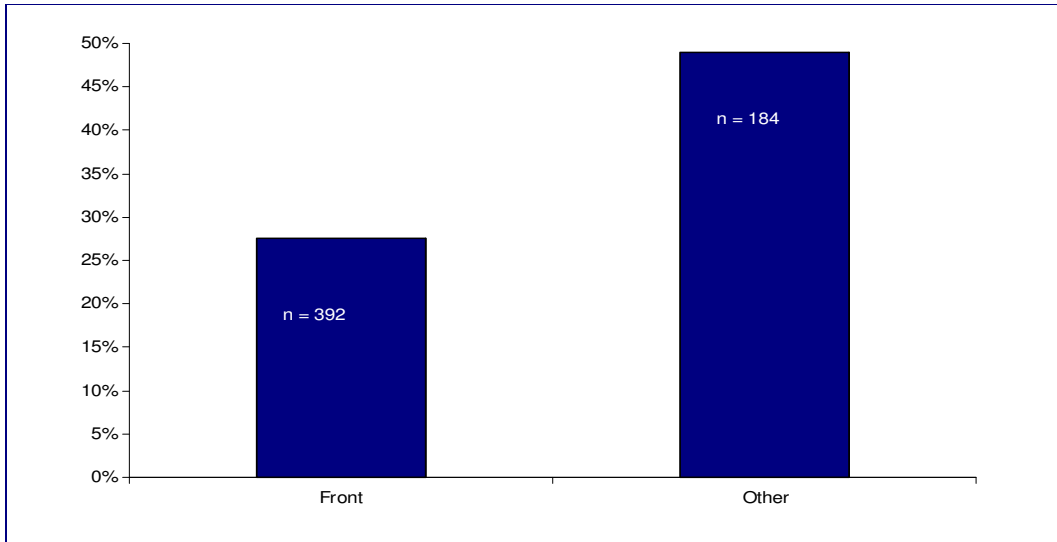


Figure 4.12: Percentage of fatalities among passengers and drivers as a function of area of most damage

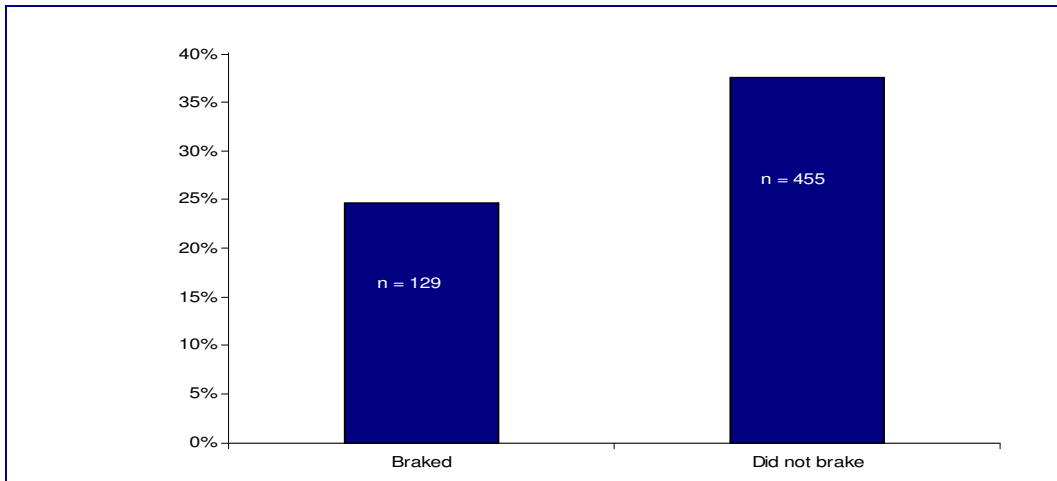


Figure 4.13: Percentage of fatalities among passengers and drivers as a function of moment of whether the accident driver braked or not at the moment of the accident

○ 4.3.5 Final model and conclusions

In the present analysis a large number of variables were tested as to their effect on the fatality risk for persons involved in fatal car-car accidents. Those variables were indicators of the personal vulnerability of the road-users, indicators related to passive safety devices and other characteristics of the vehicle, and predictors related to the circumstances of the accident.

For the interpretation of the results it is important to keep in mind that all cases analysed have been “worst cases”, in the sense that only fatal accidents were analysed. What can be derived from this analysis is a series of factors that appear to make a difference, *once one is involved in a fatal accident*.

By carefully selecting those cases where two accident partners were similar in their original fatality risk (i.e. only accidents involving two cars), the comparison of the characteristics of the fatalities to those of the survivors gives an indication about risk and protection factors in severe accidents. We can interpret the characteristics of the survivors as aspects that have the potential to prevent a severe accident from becoming a fatal accident.

The final model containing those risk and protection factors is presented in Table 4.5.

Parameters	Coefficients (logit)	Coefficients (Exponential)	Standard Error	p- value	DIC Evolution
Intercept	-0,55	0,577	0,31		754,21
Occupants	-0,36	0,698	0,08	0.000	732,22
Seatbelt					
Not Used vs Used	1,15	3,158	0,32	0.001	
Unknown vs Used	0,24	1,271	0,21	0.257	724,47
Senior	1,22	3,387	0,31	0.000	711,48
VehAge	0,09	1,094	0,02	0.000	694,72
Junction vs No Junction	-0,83	0,436	0,22	0.000	683,89
NO Braking vs Braking	0,48	1,616	0,26	0.070	680,36
FrontDamage vs other	-1,02	0,361	0,2	0.000	640,57

Table 4.5: The final car-car model

The factors that together determine whether a person involved in a fatal accident is among the fatalities or the survivors are the following:

- As most accidents have only one fatality, the effect of the number of occupants has to be understood as a spreading of risk: if there are many people involved in the accident, the chance for each of them to be the fatality is much lower than if there are only few individuals.

Analysing European in-depth data: Methodological framework and results

- Wearing a seatbelt increases the chances of surviving a fatal accident
- Older people, when involved in fatal accidents, suffer a larger chance of being the fatality.
- Occupants of older cars, when involved in fatal accidents, suffer a larger chance of becoming the fatality.
- Accidents on junctions have a lower fatality risk for all persons involved compared to accidents on road sections.
- Drivers who did not try to avoid the accident by braking were more often driving the car that contained the fatality than those who did brake.
- Cars that were damaged mostly on the side, at the back, or at the roof were more likely to contain the fatality than cars that were damaged at the front.

4.4 An opponent analysis for car-occupants²¹

Compared to the model run on all road-user type altogether, restricting the analysis to car-car accidents offers much more detailed information, but on a very restricted part of the sample (see Table 4.1). Not only does the analysis focus on car occupants, it also restricts the latter sample to one single opponent type. In other words, the choice of precision entailed a good deal of losses in data.

The model presented in this section consists of some sort of compromise between the models presented in the previous sections. Here, the incomparability of risk is also controlled for by focusing on car occupants, but this time the data are extended to several types of collision opponents: other cars, but also light and heavy good vehicles, motorcycle, or no opponent at all (single car accidents).

Running this analysis required linking each road-user to his/her collision opponent. It is important to adequately choose the opponent that will be used as a basis to define the “collision opponent” value for this road-user. One single road-user can indeed happen to have collided with several opponents in the course of the accident. Consequently, the identification of each road-user’s opponent was done using the chronological unfolding of the accident as a reference: The earlier opponents (those with whom the road-user interacted first) were considered as the most “significant” ones, and consequently designated as opponents over those that the road-user “encountered” in the later stages of the accident. More details over the procedure adopted to identify the opponent of each road-user are provided in Appendix 4.2.

One possible analysis, once this opponent variable is available, is one that would be run on the entire dataset, using one single predictor to represent all possible road-user-opponent combinations (i.e.: some 42 categories!). Such an analysis would be difficult to handle, both statistically and theoretically. An alternative solution consists of focusing on each road-user category at a time. Then, *within that category*, the relationship between the fatality risk and other key factors can be investigated, accounting for the important influence of the type of opponent on the fatality risk. This way of functioning (one road-user category at a time) offers the additional advantage that the predictors to be entered in the model can be chosen on the basis of their relevance for the road-user category examined (e.g., helmet wear), which is not possible when all road-user types are included in the analysis. This approach was adopted to run a model on car occupants, the results of which are described below.

²¹ The model fitted here is a binary logistic model (logit link function) estimated by means of Bayesian methods (Markov Chain Monte Carlo). A preliminary step was taken to test the hierarchical structure of the data. The random variation at the country level was not significant (XX= 0.03, S.E = 0.03, *n.s.*), while the one at the accident and vehicle level were estimated as null.

○ 4.4.1 Variables controlled for:

Collision opponent

The type of collision opponent is expected to strongly affect the fatality risk for all car occupants involved in fatal accidents. To examine the influence of other predictors on the fatality risk while taking this opponent-induced variation into account, the type of collision opponent was introduced as a predictor in all versions of the model. 4 collision categories were so taken into account: Car-car (as reference category), car-LGV, car-HGV, and single car. Car occupants who collided with bicycles and pedestrians were excluded from the analysis because no car occupant died in such collisions, which means that the fatality risk equalled 0 exactly for all these road-users.

The fatality risk for car occupants involved in car-car accidents appears to be the lowest. The odds of dying are significantly more elevated for car occupants who collided with a light goods vehicle or with a heavy good vehicle as compared to car occupants that collided with another car. The odds of dying were also more elevated among car occupants involved in single car accidents as compared to those involved in car-car accidents. This difference, however ceases to be significant once the number of accident participants is entered as a predictor in the model. In other words, the reason why single car accidents are associated with a higher proportion of fatalities than car-car accidents, is that they involve fewer participants.

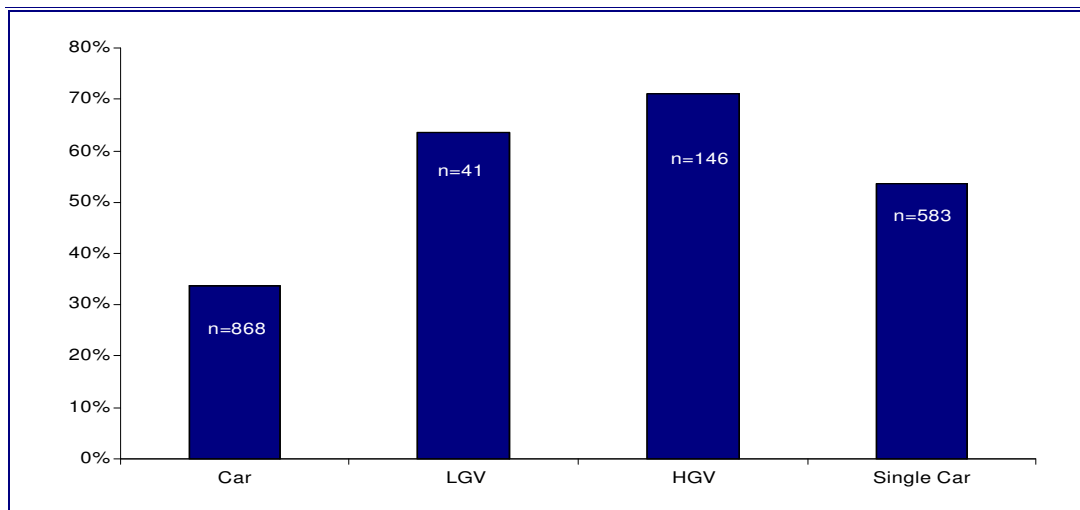


Figure 4.14: Percentage car-occupants killed as a function of collision opponent type

Number of participants in the accident

The more the participants (vehicles, pedestrians...) involved in the accident, the lower is the fatality risk for each road-user involved in the accident is ($B=-0.48$, $S.E= 0.1$, $p<.000$).

Vehicle Occupancy

Independently of the number of participants in the accident, vehicle occupancy - the second component of accident size in this case – still appears to be

significantly and negatively related to the fatality risk ($B=-0.53$, $S.E= 0.05$, $p<.000$). The odds of being among the fatalities of the accident decrease with the number of occupants in the same vehicle.

○ 4.4.2 Road-User vulnerability factors:

Age

The global relationship between age and the road-user fatality risk was positive and significant, indicating that the odds of dying for car occupants in a fatal accident increase with age ($B=0.001$, $S.E=0.003$, $p<.0001$).

Gender

Although the percentage of fatalities was higher among male (48.18%) than among female car occupants (40.17%), being male does not appear to be associated with a significant increase of the fatality risk ($B=-0.497$, $p=0.70$).

Position within the vehicle

As was the case in the car-car model, the fatality risk did not differ among car passengers and car drivers ($B=0.179$, $S.E=0.129$, $p=0.17$), nor as a function of the road user's position in the car (front or back) ($B= 0.096$, $S.E.= 0.189$, $p= 0.61$)

○ 4.4.3 Passive Safety

Seatbelt

As for the car-car model in the previous section, the results indicated that the odds of dying are higher among people not wearing their seatbelt as compared to those who did wear their seatbelt ($B=0.836$ $S.E=0.120$)²².

Vehicle Age

The effect of vehicle age was significant ($B=0.037$, $S.E=0.011$, $p<.001$). As expected, the results indicated that the fatality risk for the car occupants increased with vehicle age. As an illustration, the percentage of fatalities among the occupants of vehicles of various age categories is displayed in Figure 4.15.

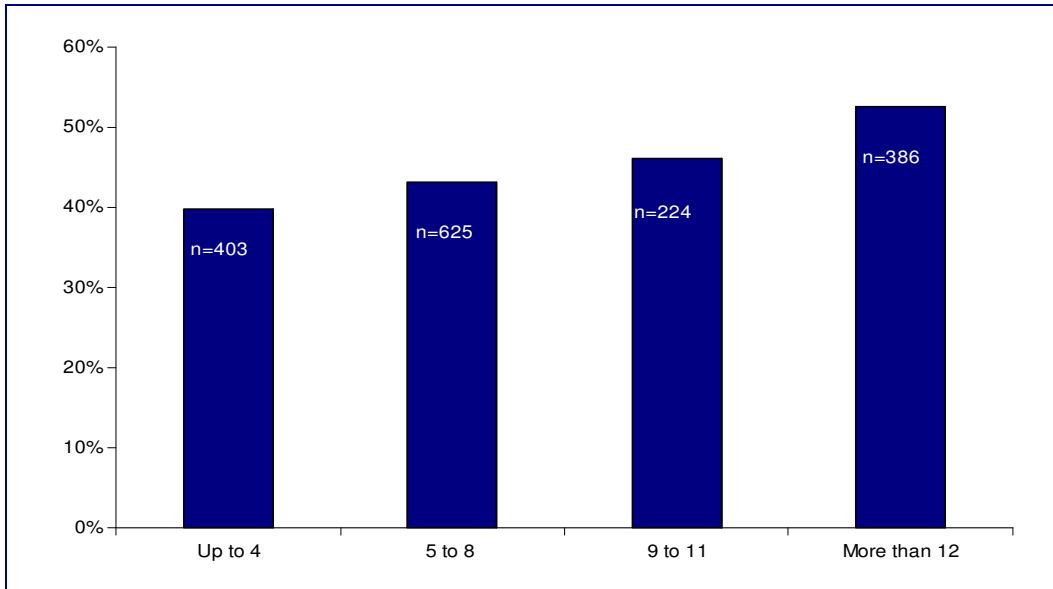


Figure 4.15: Percentage of fatalities as a function of vehicle age

○ 4.4.4 Accident Circumstances

Overall description

The observations are quite evenly distributed across the different periods of the day and of the week. Similarly, the share of observations for accidents that took place under daylight condition and other light conditions (mostly darkness) is homogeneous (49 and 51% respectively). The vast majority came from accidents that took place on dry roads (75%) and a small number of them took place on wet road (25%).

Most of the observations (84.59%) came from accidents on two-way roads divided by a painted line. The observations are quite evenly distributed as far as speed limit is concerned: 28.89% of the road-users were involved in accidents on roads with a speed limit lower or equal to 50 km/h, 15.4% in accidents with a speed limit between 80 and 100 km/h, and 57.7% on roads with a speed limit higher than 100 km/h. The traffic flow was light (45.42%) or normal (31.07%) in most cases. Instances of accidents that happened under high traffic flow conditions are far more scarce (6.90%). The number of missing values for the traffic flow variable was substantial (16.61%).

Very few observations were made for accidents that occurred on motorways (10.56%). 73.08% of the observations came from accidents that took place on road sections. When the accident occurred at a road junction, this was either a crossroad (42.73%) or a T-shaped junction (43.86%) in most cases. 76.07% of the observations were made in rural areas.

When the observations were very unevenly distributed across the categories of one predictor (e.g. = motorway, type of road), the relation between the corresponding predictor and the fatality risk was not investigated. The same is

true for variables in which the number of missing values was substantial (e.g.: traffic flow).

Moment of the accident

The fatality risk for road users involved in fatal accidents does not significantly differ between week and weekend ($B=0.014$, $S.E=1.014$, $p=0.90$) or between night and day ($B=-0.101$, $S.E=0.904$, $p=0.37$).

Road

The odds of dying are more elevated for the road-users who were involved in road-sections accidents as compared to road-users in road-junction accidents ($B=-0.436$, $S.E=0.140$, $p<.002$).

Development of the accident

The area of the car that was damaged most provides an indication of the type of impact characterizing the accident. For a small majority of the cases the area of most damage was the front (54.12%). The impact area was significantly related to the fatality risk: Generally speaking, the odds of dying were lower for the road users in cars that had been damaged most at the front, compared to cars damaged anywhere else²³ ($B=-0.97$, $S.E=0.162$, $p<.000$).

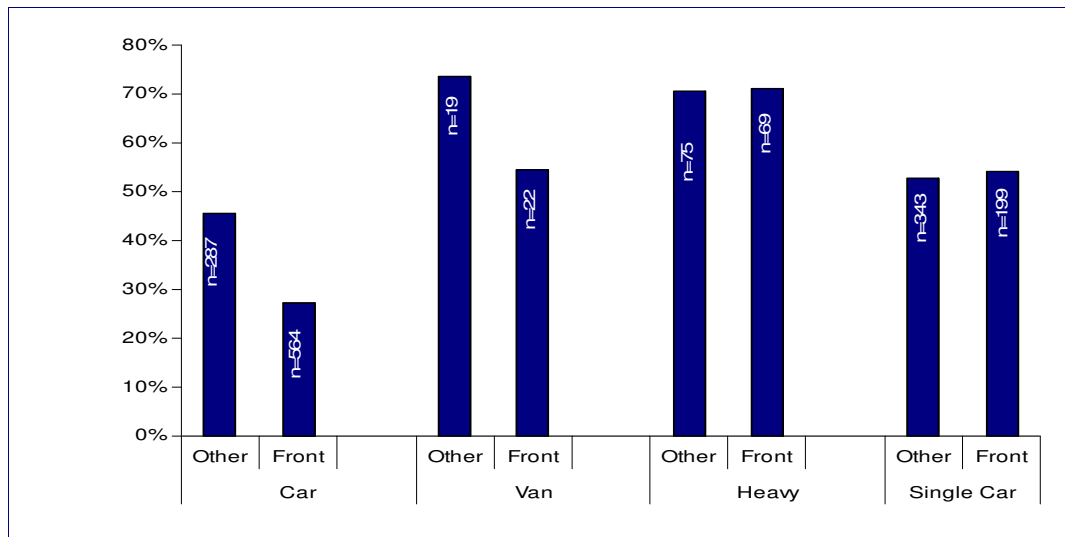


Figure 4.16: Percentage of car-occupants fatalities as a function of collision opponent and damaged area of the car

Yet, there also appears to be a significant interaction between front damage and collision opponent: Compared to car-car accidents, the effect of front damage was positive and significant for single car accidents ($B=0.886$, $S.E=0.252$, $p<.002$). The same tendency was observed in the case of car accidents with heavy good vehicles, although not significantly so ($B= 0.912$, $S.E= 0.417$, $p<.003$). The effect of front damage did not differ, however for car-LGV accidents and for car-car accidents. This result indicates that car-occupants are better-off in cars that are damaged most at the front if such damages result from

²³ Front Damage: 54.15% ; left damage :14.20%, right damage : 12.62%, back, roof, and « multiple » damages : 2.54, 7.61, and 8.81% respectively.

a collision with another car or with some light good vehicle. The same is not true if the car collided with a truck or with some fixed object (single car accidents): In such cases, the fatality risk does not vary as a function of the area of the car that was mostly damaged.

Another way to characterize the development of the accident is to distinguish between different avoidance manoeuvres that have or have not been conducted by the driver. In their original state, the data allow determining whether the driver had braked to avoid the accident, steered, or braked *and* steered. It turned out however, that only whether the driver braked or did not brake made a significant contribution to the fatality risk for the car occupants. As shown in Figure 4.17, cars of which the driver had braked contained a fatality in 29.77% of the cases while for cars of which the driver had not braked this was 45.50%. This difference was marginally significant ($B=0.269$, $S.E=0.136$, $p<.05$).

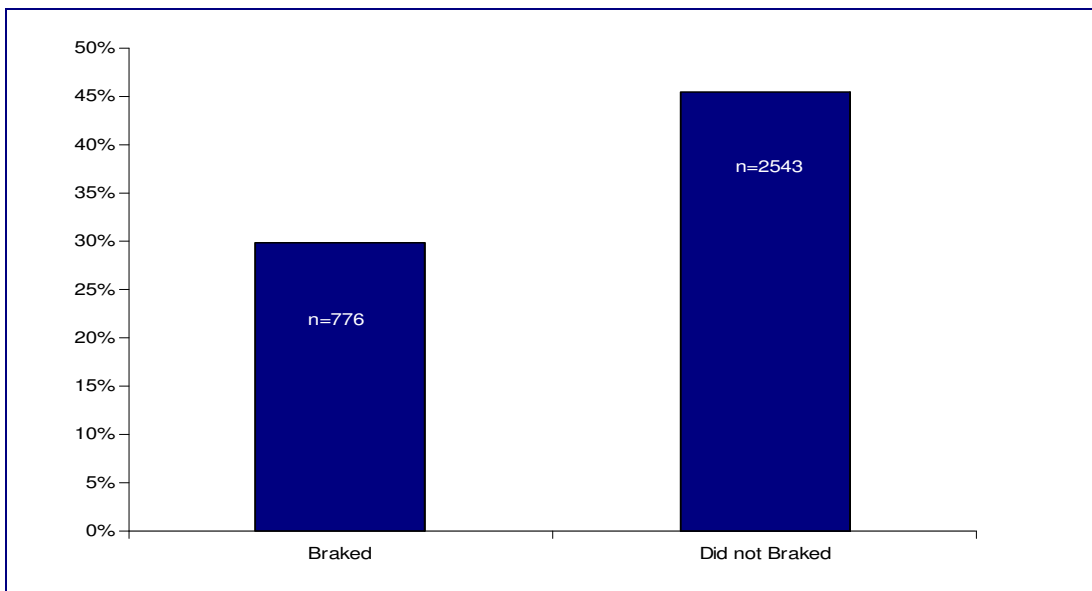


Figure 4.17: Percentage of fatalities among car occupants depending on whether the car driver braked or not at the moment of the accident.

Final model

The aim of this analysis was to extend the findings previously obtained by means of an analysis of two-car accidents to car accidents of different sorts. In order to account for the fact that different collision opponents are associated with different fatality risks for car occupants, the type of collision opponent was included and controlled for in the model.

The final model containing the risk and protection factors that have been identified for car occupants, *which takes their collision opponent into account* is presented in Table 4.6. The results confirmed the important influence of the type of collision opponent on the road-users' fatality risk. Overall, the remaining conclusions offered by this model are very similar to those of the car-car model. Yet, in the present case, the possibility exists to test interactions between the effects of the different predictors and the effect of opponent type, as this was done in the case of the "front damage" variable and the type of opponent, the

result of which indicated that the “protective effect” of front damage holds only in the case of car accidents with other cars or with light good vehicles. Testing such interactions thus really amounts to examining whether the effect of one predictor generalises to all accident types, differs depending on the type of opponent, or is circumscribed to one particular combination of road user – opponent.

Analysing European in-depth data: Methodological framework and results

<i>Parameters</i>		<i>Coefficients (logit)</i>	<i>Coefficients (Exponential)</i>	<i>Standard Error</i>	<i>p- value</i>	<i>DIC Evolution</i>
Intercept		-0.21	0.81	0.19		2255
Collision Opponent						
	LGV	0.69	1.99	0.56	0.46	
	HGV	0.85	2.34	0.30	0.02	
	Single	-0.21	0.81	0.21	0.61	2148
Number of Participants		-0.48	0.62	0.10	0.00	2138
Occupants		-0.55	0.58	0.05	0.00	1977
Age		0.01	1.01	0.00	0.00	1964
Seatbelt	Used vs :					
	Not Used	0.84	2.32	0.17	0.00	
	Unknown	0.17	1.19	0.13	0.41	1938
VehAge		0.03	1.03	0.01	0.01	1932
Junction		-0.49	0.61	0.14	0.00	1923
FrontDamage		-1.06	0.35	0.16	0.00	1895
FrontDamagex CollisionOpponent						
	LGV	0.18	1.20	0.72	0.99	
	HGV	0.80	2.23	0.43	0.32	
	Single	1.03	2.80	0.26	0.00	1880
Braking		0.29	1.34	0.14	0.03	1877

Table 4.6: The final car-opponent model

4.5 Conclusions²⁴:

In this section, much attention has been devoted to the problems posed – in terms of data analysis and results interpretation – by the fact that the data are limited to fatal accidents. Several solutions have been outlined, which are not mutually exclusive. The general model presented at the beginning of the chapter can, for example, be used to explore the data and to identify the factors associated with the fatality risk across all collision types. More focused models, such as the opponent analysis can be run afterwards.

In all cases, the importance of accounting for the accident size bias and the incompatibility of risk was illustrated. The evolution of the analytical solution chosen also clearly indicates that appropriate and full handling of the incomparability issue is feasible only at some cost, namely, the exclusion of a number of data. The result, however, seems to be worth the cost. The generality of the approach initially adopted indeed imposes important restrictions on the types of predictors that may be tested. But running a model on all road-users and opponents types also raises problems in terms of results interpretation. Indeed, and notably in the case of car occupants, the type of opponent strongly affects the fatality risk (see Figure 4.3). Although the different road-user types are distinguished in the “all road-users model”, the fatality risk as it is estimated for car opponent by this model is a risk that holds both for the case in which the car collided with a pedestrian and the case in which it collided with a heavy good vehicle. Overlooking the importance of the type of collision opponent may also result in spurious effects for other predictors. As an example, consider the speed limit effect that was observed on the basis of this model. The results indicate that the fatality risk is higher on locations where the speed limit exceeds 50kmh. Is this result to be attributed to speed limit itself, or to the fact that pedestrians are more likely (and harmless) collision opponents in area where the speed limit is lower than 50kmh, while cars and heavier vehicles are more likely (and dangerous) collision opponents in areas where the speed limit is higher?

Up to now, the last solution presented, or the opponent analysis, is the one that offers the best compromise between data loss and adequate control of the baseline risk incomparability. It also has the additional advantage of offering a general framework that can be applied to the different road-user groups represented in the data.

Ideally, one should be able to run such an analysis for each road-user type. Yet, as explained in the introduction of this chapter, for some collision types there is basically no variation in the consequence of the accident. Including this category in the model would consequently be technically impossible (most statistical techniques do not even give results when the outcome probability is 0 or 1) and conceptually questionable. Figure 4.3, for example, clearly showed

²⁴ All the commandos that have been used to define and recode the variables, as well as to define the opponent variable can be found in Appendix 4.4.

that all car occupants survived collisions with pedestrians. How can the fatality risk be estimated for car drivers who hit pedestrians if none of the car drivers was killed? There is thus one important restriction to the application of the “opponent analysis” to the different road-users group: There must be sufficient variation in the outcome of the accident for each combination of this road-user type with the others.

A second factor that prevents the application of the “opponent model” to all types of road-users is the number of observations. For some collision types, the number of observations is simply too small. At this stage of the development of the FAI database, there are sufficient numbers of observations to perform the analysis on the car occupants, but not on the other road-user types (there are, for example only 8 powered 2-wheeler drivers who collided with a bicycle). The possibility is there, however, to run this detailed and focused analysis on the other groups. This, however, would require the collection of additional data.²⁵ The possibilities offered by such an analysis – among others: a detailed screening of the factors affecting the fatality risk of each road-user category, and the possibility to examine whether the effects identified remain stable across different categories of collision opponent – justify, in our view, the cost that such an additional data collection would entail.

²⁵ Attempts were made at running the opponent analysis on other road-user categories, such as powered 2-wheeler. The effects observed in this case cannot fully be trusted given that the number of motorcycles in the total sample is much lower than that of car occupants, and even the more so that several opponent types are considered. Should the size of the database be augmented motorcycles and heavy good vehicles are the two road-user categories for which the opponent analysis could be most straightforwardly applied.

5 Conclusions

This document has demonstrated several possible uses of the Fatal Accident Investigation database. Having reached the end of this document, it is useful to remind the reader of the most important conclusions derived from the analyses performed, and also to remind that all the possibilities offered by the data have certainly not been exhausted here.

5.1 Results summary

Chapter 2 showed that - although they were not initially meant for this - the FAI data could be very useful in detecting inaccurate reporting of injury severity under the form of misreporting (e.g.: slight injuries recorded as serious, fatalities recorded as serious injuries, and so on). Generally speaking, the analysis performed offered encouraging conclusions with respect to the quality of the reporting of injury severity in fatal accidents, at least in the member states that took part in the FAI data-collection. Indeed, several sources of misreporting had been identified on the basis of this analysis among the first wave of the FAI data, suggesting a general pattern according to which, the more complex the road accident conditions were (e.g. more accident participants, higher traffic flow, night time), the higher was the probability of misreporting injury severity. However, running the same analysis on the second wave of data revealed a very limited number of inaccurate reported cases, indicating that most of the inaccuracy previously identified had been solved.

The remaining inaccuracies mostly concerned serious injuries, and very few systematic sources could be identified for these inaccuracies. Notably, inaccuracies were less probable when the victim was taken to the hospital. It is not fully clear, however, whether this was due to the fact that injuries necessitating transportation to the hospital are intrinsically likely to evolve in the time span taking place between the rating of the injury by the police and its recording by the SafetyNet team; or whether this indicates that the injury status of those cases who were *not* taken to the hospital is especially likely to deteriorate.

Independently of these considerations, it must be acknowledged that the FAI data offer the great advantage of providing some “standard”, namely the SafetyNet injury severity score, against which the accuracy of the police records for injury severity can be evaluated. The fact that no notable inaccuracy could be identified in the second wave of the FAI data, and that there is no systematic country variation in the few inaccuracies identified suggests that no major problems should be expected in terms of country comparison on the basis of different level of injury severity on the basis of the FAI data.

Chapter 3 showed that the FAI data can also be used to investigate the occurrence of different *accident types of fatal accidents*. Single vehicle fatal

accidents were contrasted with multivehicle ones. The results showed that these two types of accidents differ on a large number of variables. As compared to multivehicle accidents, single vehicle fatal accidents were found to be more likely on road sections (rather than junctions) with low traffic volume. Overall, few fatal accidents occur on roads with physically divided carriageways. Among these few, there are a comparatively high proportion of single vehicle accidents. Moreover, single vehicle accidents have been found to involve a different type of road-user as compared to multivehicle ones: Car drivers have disproportionately many single-vehicle fatal accidents as compared to motorcyclists and especially heavy good vehicle drivers. Moreover, drivers who were unfamiliar with the area were more frequently involved in a single vehicle fatal accident, and there also appeared to be more occupants in the car in that case. Finally, the drivers in fatal vehicle accidents seem to have made fewer attempts at avoiding the accident in the case of a single as compared to a multivehicle accident.

Throughout this document, the important influences of the size of the accident and of the incomparability of the baseline risk for road users involved in fatal accidents have been stressed. Yet, it is for the analysis of the outcome of accidents that the problem appears to be most acute. The analyses presented in Chapter 4, which all focus on the issue of the accident outcomes, have illustrated these two effects on the estimation of the chances of survival of the road-users who happen to be involved in a fatal accident. Three different approaches to modeling the fatality risk of road users involved in fatal accidents were presented: an "all users" analysis, a "car-car" analysis and a "car-opponent" analysis, each one testing different ways of correcting for these two effects, by using different subsets of the FAI data and different variables. The results showed that, over and above the basic – and sometimes dramatic – influence of accident size and risk incomparability, a variety of factors appeared to make a difference for each of the road-users involved.

Some variables emerged as "risk factors" in a consistent way for all three analyses, such as the fact that the road-user him/herself could (or could not) be considered as "senior" (i.e.: as being more than 65 years), or that the driver did not react properly to the occurrence of the accident by braking. Apart from that, the set of predictors identified was rather different depending on whether the whole dataset was analyzed, or whether the analysis specifically focused on car occupants. Using the whole dataset prevents the examination of predictors that are relevant – and also maybe very important – for one particular road-user type only. In a related vein, should one predictor appear to have opposite effects on two different road-user groups, then this effect is not likely to show up in an analysis where all road-user types are pooled together. For instance, in the case of car occupants, seatbelt appeared as a protection factor, an effect which could not be identified when examining all road users together. The risk to be killed in the accident increases with the age of the vehicle. The analysis, however did not allow further insight on which feature(s) of the vehicle exactly are responsible for this vehicle age effect. Finally, the fact that the fatal accident took place on a road junction also leaves the road-users involved with increased chances to survive. The effects of other factors, such as the

protective effect offered by front damage appear to hold only for collisions between car-occupants with other cars or with light good vehicles (not for single-car accidents or collisions between cars and heavy good vehicles)

The importance of the type of opponent has, hopefully, been demonstrated. The opponent analysis presented here can be improved in many ways, and should at least be applied to other road-user categories. Within one such category, examining the interaction between the so-called risk or protection factors can also prove particularly informative. Yet, the basic condition to be met in order to be able to perform these more advanced forms of the opponent analysis is an augmentation of the size of the dataset. This is especially true for these road-users that are less frequently encountered on the roads than car drivers. For other road-user types, such as “extremely strong” or “extremely vulnerable” road-users, increasing the number of observations will not suffice, however, given that these *invariably* tend to end-up as the survivors, or as the victims, in severe accidents. For these road-users categories, then, the joint analysis of fatal and non-fatal data is an additional requirement.

5.2 Future perspectives

The reader must be aware that, although the models presented throughout this deliverable differ, they all rest on the same broad analytical approach, namely the regression model. Other, very different analytical strategies could have been chosen, and should be envisaged to exploit the rich information contained in the FAI data. The choice of this global approach, despite its several advantages, has also imposed some constraints on the results obtained. Regression analysis requires, for example, that one variable (or several of them, although this has not been the case here) be assigned the status of “dependent variable”, while others must be designated as “predictors”. Constraints related to the number and distribution of missing values, or to the number of observations per category of each predictor also have to be taken into account. This has more than once precluded the investigation of some newly available variables, and somewhat impaired the full exploitation of the level of detail offered by the FAI data. Other techniques, such as classification techniques (e.g.: multiple correspondence analysis, clustering techniques...) are available and do not have the same requirements as regression analyses. They would thus consist of an interesting complementary approach to the one adopted here.

As noted in the conclusions of Chapter 4, the possibilities offered by the FAI data also depend on the size of the dataset. The more observations are included in the data, the higher the possibility to tailor the analyses to their detailed character. The protection and risk factors for different road-user-opponent combinations, of whom we have seen that they really constitute different risk groups, should therefore also be better known. For some of these groups, disquieting trends in terms of the number of fatalities have been observed in the recent years (think of motorcyclists, for example), stressing the need for relevant information on the road-user category considered, an information that would not be blurred by information on other categories – such

as car occupants, which happen to be predominant on roads everywhere in Europe.

Finally, given all the attention devoted here to having data limited to fatal accidents and the implications thereof, it is also important to stress that the possibility exists to run analyses similar to those presented here on both fatal and non-fatal accident data. The Accident Causation Database, also created in the framework of the work package 5 of SafetyNet indeed includes non-fatal accident data. Although this database contains even more detailed information than the FAI data, there is a set of variables that is common to both. Analyses focusing on these variables would thus nicely complete the ones presented in this deliverable: It would indeed allow the occurrence of fatal accidents and fatal injuries to be explained better by using non-fatal accidents and non-fatal injuries as control groups.

References

- 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.
- ETSC (European Transport Safety Council). “*Social and economic Consequences of Road Traffic Injury in Europe*”, Brussels, 2007.
- Goldstein H. (2003). *Multilevel statistical models*, London: Arnold.
- Hox, J., (2002). *Multilevel Analysis, Techniques and Applications*. Lawrence Erlbaum Associates, Publishers, London.
- Ivan, J.N., Raghubhushan, K.P., Ossenbruggen, P.J. (1999). Differences in causality factors for single and multi-vehicle crashes on two-lane roads. *Accident Analysis & Prevention*, 31, 695-704.
- Jones A. P., & Jørgensen, S. H. (2003). The use of multilevel models for the prediction of road accident outcomes. *Accident Analysis and Prevention*, Vol. 35, pp. 59-70.
- Kreft, I., & De Leeuw, J., (1999). *Introducing multilevel modelling*. SAGE Publications.
- Lenguerrand, E., Martin J.L., & Laumon, B. (2006) Modelling the hierarchical structure of road crash data – Application to severity analysis. *Accident Analysis and Prevention*, 38, 43-53.
- Longford, N. (1993) *Random coefficient models*, Oxford: Clarendon Press.
- Martensen, H. (Ed.) (2007) *Analysis of Fatal Accident Investigation Database*. Deliverable D7.6 of the EU FP6 project SafetyNet.
- Öström, M. & Eriksson, A. (1993). Single-vehicle crashes and alcohol: a retrospective study of passenger car fatalities in Northern Sweden. *Accident Analysis & Prevention*, 25, 171-176.
- PENDANT - Pan-European Coordinated Accident and Injury databases
- Pinheiro, J.C., & Bates, D. M. (1995) Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model *Journal of Computational and Graphics Statistics* 4, 12-35.
- Reed, S., & Morris, A. (Eds.) (2006) *Review of fatal accident pilot study: Fatal data methodology development report*.
- SafetyNet WP5, "Work Package 5 Database Glossary", version of 29 November 2006.

Scheers, M. & Casteels, Y. (2008). Evolutie van de verkeersveiligheid in België 2000-2006. IBSR, Belgian Road Safety Institute.

Snijders, T., & Bosker, R. (1999). Multilevel analysis. An introduction to basic and advanced multilevel modeling, London: Sage Publications.

STAIRS consortium (1999) – Stairs: Standardisation of Accident and Injury Registration systems – Final Report.

Strillacci, L. (2002). Car Accidents Tend to Occur Close to Home. www.chase.com/cm/cig/advice-and-planning/page/june02caraccidents.html (retrieved, October 2008).

Wegman & Aarts (eds.) (2006). Advancing Sustainable Safety: National Road Safety Outlook for 2005-2020. SWOV Institute for Road Safety Research, Leidschendam.

Appendices

Appendix 1.1: Raw counts and percentages for the CARE-FAI data 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%
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 A.1.1: 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%

Table A.1.2: Age groups

Analysing European in-depth data: Methodological framework and results

	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 1.1.3: 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 A.1.4: Vehicle group

Analysing European in-depth data: Methodological framework and results

	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 A.1.5: 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 A.1.6: 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 A.1.7: 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%



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

Analysing European in-depth data: Methodological framework and results

Table A.1.8: 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 A.1.9: 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 A.1.10: Weather conditions

APPENDIX 2.1: Comparative injury severity distributions for other countries except Italy (1st FAI database)

Tables for the five (5) remaining examined countries, similar to Table 2.4 for Italy

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	45					45
Not Injured		7		2		9
Serious	3		11	1	3	18
Slight			1	25	3	29
Unknown	1					1
Grand Total	49	7	12	28	6	102

Table A.2.1.1: Corresponding injury severity for the two recording systems in the WP5.1 (1st FAI) database: Sweden

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	66					66
Not Injured		24				24
Serious	1		29	4	2	36
Slight		2	1	15	1	19
Unknown						
Grand Total	67	26	30	19	3	145

Table A.2.1.2: Corresponding injury severity for the two recording systems in the 1st FAI database: Germany

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	108					108
Not Injured		59		2		61
Serious			35		1	36
Slight			1	37		38
Unknown					1	1
Grand Total	108	59	36	39	2	244

Table A.2.1.3: Corresponding injury severity for the two recording systems in the 1st FAI database: France

Analysing European in-depth data: Methodological framework and results

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	3					3
Not Injured						
Serious			1			1
Slight				1		1
Unknown						
Grand Total	3		1	1		5

Table A.2.1.4: Corresponding injury severity for the two recording systems in the 1st FAI database: Finland

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	19					19
Not Injured		12				12
Serious	1		4			5
Slight				9		9
Unknown						
Grand Total	20	12	4	9		45

Table A.2.1.5: Corresponding injury severity for the two recording systems in the 1st FAI database: United Kingdom

APPENDIX 2.2: Comparative injury severity distributions for other countries except Italy (2nd FAI database)

Tables for the remaining six (6) examined countries, similar to Table 2.12 for Italy

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	140					140
Not Injured		40		2		42
Serious	3		53	1	3	60
Slight	1	3	1	57	3	65
Unknown	1					1
Grand Total	145	43	54	60	6	308

Table A.2.2.1: Corresponding injury severity for the two recording systems in the WP 5.2 (2nd FAI) database: Sweden

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	153					153
Not Injured		85		2		87
Serious			48		1	49
Slight			1	42		43
Unknown					1	1
Grand Total	153	85	49	44	2	333

Table A.2.2.2: Corresponding injury severity for the two recording systems in the 2nd FAI database: France

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	188					188
Not Injured		116				116
Serious	2	2	52	6	3	65
Slight		6	1	54	3	64
Unknown					2	2
Grand Total	190	124	53	60	8	435

Table A.2.2.3: Corresponding injury severity for the two recording systems in the 2nd FAI database: Germany

Analysing European in-depth data: Methodological framework and results

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	52					52
Not Injured		44		1	2	47
Serious	3		6			9
Slight				10	2	12
Unknown					8	8
Grand Total	55	44	6	11	12	128

Table A.2.2.4: Corresponding injury severity for the two recording systems in the 2nd FAI database: Netherlands

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	2					2
Not Injured		4				4
Serious			1			1
Slight				1		1
Unknown	65	40	14	31		150
Grand Total	67	44	15	32		158

Table A.2.2.5: Corresponding injury severity for the two recording systems in the 2nd FAI database: Finland

Police Injury Severity	SafetyNet Medical Outcome					Grand Total
	Fatal	Not Injured	Serious	Slight	Unknown	
Fatal	283					283
Not Injured		157				157
Serious	3		77	1		81
Slight			3	94		97
Unknown					61	61
Grand Total	286	157	80	95	61	679

Table A.2.2.6: Corresponding injury severity for the two recording systems in the 2nd FAI database: United Kingdom

APPENDIX 2.3: Exploratory analysis of WP 5 data with respect to accident severity scores (2nd FAI database)

This analysis is carried out prior to the development of models, intending to present the profile of accident severity as expressed by the two severity scores at the road user level, namely:

- "Police injury severity",
- "SafetyNet medical outcome".

This is achieved by means of distributions of these two variables across characteristics related to user, vehicle, accident and road characteristics.

The selection of the parameters in question was limited to present a few of the most influential factors that intervene in the context of the present analysis. These factors were selected from an extensive list of candidate variables for the models developed in the following sections (see Table 2.5 of main text). The list included all factors that theoretically may be involved in the reporting problem.

- Age

The grouping adopted in this analysis distinguishes between persons that are either very young (0-14) or rather old (55+), on the one hand, and all the others (aged 15-54). According to SafetyNet results, the balance between "heavy" casualties (fatalities - severe injuries) and "light" ones (slight / no injuries) is similar for both groups (in the order of 63%-37%).

Police injury severity

Age	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
15-54	37.7%	18.6%	15.2%	28.5%	100.0%
0-14 / 55+	51.3%	11.7%	13.1%	23.9%	100.0%
Total	42.8%	16.0%	14.4%	26.8%	100.0%

SafetyNet medical outcome

Age	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
15-54	38.1%	17.5%	15.2%	29.2%	100.0%
0-14 / 55+	52.6%	10.4%	13.3%	23.7%	100.0%
Total	43.6%	14.8%	14.5%	27.1%	100.0%

% Difference

Age	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
15-54	0.4%	-1.1%	0.0%	0.7%	100.0%
0-14 / 55+	1.3%	-1.3%	0.2%	-0.2%	100.0%
Total	0.8%	-1.2%	0.1%	0.3%	100.0%

Table A.2.3.1: Distribution of casualties by age group

Interestingly, individuals from the second age group switch to a larger extent their status from “seriously injured” to “killed” (and “slightly injured”, as well). This is in accordance to what would be expected, given all other conditions were identical, since children and elderly should be more vulnerable when seriously injured. The observed pattern is not very strong, though; this may also have to do with the travelling environment of people in the [15-54] group (more frequently on interurban roads, i.e. at high speed, more often impaired due to fatigue or alcohol, etc), affecting the type of their injuries.

- *Transfer to hospital*

The grouping adopted in this analysis distinguishes between individuals taken to the hospital after an accident or not. According to SafetyNet results, the allocation between "heavy" casualties (fatalities - severe injuries) and “light” ones (slight / no injuries) is similar for both groups (in the order of 73%-27%).

Police injury severity

Hospitalisation	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
No	45.4%	0.2%	3.6%	50.8%	100.0%
Yes	40.0%	33.0%	25.6%	1.3%	100.0%

SafetyNet medical outcome

Hospitalisation	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
No	45.7%	0.1%	3.4%	50.8%	100.0%
Yes	41.7%	31.1%	25.8%	1.4%	100.0%

% Difference

Hospitalisation	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
No	0.3%	-0.1%	-0.2%	0.0%	100.0%
Yes	1.7%	-1.9%	0.2%	0.1%	100.0%

Table A.2.3.2: Distribution of casualties by hospitalisation

The patterns exhibited by both groups are (reasonably) rather different. Seriously injured entries according to the Police are slightly reduced in SafetyNet records. This small share is mainly added to fatalities.

- *Number of occupants*

The selected grouping detects the presence of passengers apart from the driver of a vehicle. It is interesting that the second group with respect to this variable exhibits a much more balanced distribution over severity extent outcomes. According to the following Table, over 50% of cases where there was only one occupant were killed, in both Police records and SafetyNet database. This group shows a notable shift from seriously or slightly injured to killed or not injured (presumably respectively, to be examined in section 5.3). The other group contains all other casualties and reveals a (proportionately) similar shift from the status of severely injured to all other outcomes.

Police injury severity

Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
One	53.1%	5.9%	9.4%	31.6%	100.0%
More than one	31.8%	26.8%	19.8%	21.6%	100.0%

SafetyNet medical outcome

Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
One	54.1%	5.2%	8.7%	32.0%	100.0%
More than one	32.3%	25.1%	20.6%	22.0%	100.0%

% Difference

Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
One	1.0%	-0.7%	-0.7%	0.4%	100.0%
More than one	0.5%	-1.7%	0.8%	0.4%	100.0%

Table A.2.3.3: Distribution of casualties by number of occupants in vehicle

- Road conditions

The grouping adopted in terms of this analysis distinguishes casualties recorded in accidents occurring on dry pavement from those involving other, more adverse road conditions (wet or frozen rolling surface due to rain/snow).

Police injury severity

Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Dry	41.9%	14.9%	15.1%	28.1%	100.0%
Other	45.3%	19.5%	12.6%	22.5%	100.0%

SafetyNet medical outcome

Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Dry	42.9%	13.8%	15.1%	28.2%	100.0%
Other	45.1%	18.1%	12.9%	23.8%	100.0%

% Difference

Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Dry	1.0%	-1.1%	0.0%	0.1%	100.0%
Other	-0.2%	-1.4%	0.3%	1.3%	100.0%

Table A.2.3.4: Distribution of casualties by road conditions

From a mere injury severity point of view, it seems that accidents occurring when the pavement is not dry involve somewhat more severe casualties (in the order of 10-15% in both Police and SafetyNet records). With respect to recording reliability (matching/non matching scores), the first group reveals some shift from seriously injured to killed, while the second one shows some reduction of "heavy" casualties (fatalities - severe injuries) along with a respective increase of "light" ones (slight / no injuries).

- Vertical alignment

The selected grouping distinguishes flat roads from stretches driving uphill/downhill. It is generally acceptable that large slopes place a negative impact in drivers' interaction, depending of course on many other parameters.

Police injury severity

Traffic	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Flat	42.5%	14.7%	14.4%	28.4%	100.0%
Other	43.0%	19.7%	14.6%	22.7%	100.0%

SafetyNet medical outcome

Traffic	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Flat	43.4%	13.6%	14.3%	28.6%	100.0%
Other	43.1%	18.2%	15.1%	23.6%	100.0%

% Difference

Traffic	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Flat	0.9%	-1.1%	-0.1%	0.2%	100.0%
Other	0.1%	-1.5%	0.5%	0.9%	100.0%

Table A.2.3.5: Distribution of casualties by road vertical alignment

In terms of severity, this is the variable with the lowest deviation from the distribution of total cases across severity extent, implying that it is not so critical in the examined dataset. In both records, flat surface (first group) reveals a notably smaller share for seriously injured and a similarly larger share for not injured individuals. As far as reporting is concerned, the first group is described by a reduction in severe injures' share –mostly added to killed individuals. The second group shows a shift of the same order from seriously injured to the opposite direction (slightly/not injured).

APPENDIX 2.4: Significance of each selected predictor examined alone (2nd FAI database)

SPSS output for each independent variable examined alone

Constant

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	145,597(a)	,000	,000

a Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 Constant	2,424	,228	113,305	1	,000	11,286

Road Conditions

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	136,101(a)	,036	,082

a Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a) RoadConditions(1)	1,856	,755	6,043	1	,014	6,401
Constant	2,004	,244	67,265	1	,000	7,421

a Variable(s) entered on step 1: RoadConditions.

Age

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	140,239(a)	,021	,048

a Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a) Age(1)	-1,067	,469	5,166	1	,023	,344
Constant	2,944	,363	65,890	1	,000	19,000

a Variable(s) entered on step 1: Age.



Vertical Alignment

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	139,608(a)	,021	,048

a Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	VerticalAlignment(1)	1,197	,571	4,389	1	,036	3,309
	Constant	2,042	,258	62,743	1	,000	7,706

a Variable(s) entered on step 1: VerticalAlignment.

Horizontal Alignment

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	136,474(a)	,033	,077

a Estimation terminated at iteration number 6 because parameter estimates changed by less than ,001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	HorizontalAlignment(1)	1,367	,483	8,008	1	,005	3,924
	Constant	1,730	,290	35,609	1	,000	5,643

a Variable(s) entered on step 1: HorizontalAlignment.

Accident Day

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	144,719(a)	,003	,008

a Estimation terminated at iteration number 5 because parameter estimates changed by less than ,001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1(a)	AccidentDay(1)	-,427	,456	,877	1	,349	,652
	Constant	2,625	,327	64,234	1	,000	13,800

a Variable(s) entered on step 1: AccidentDay.

APPENDIX 2.5: Correlation testing for the selected variables (2nd FAI database)

Correlations	Variables in the model					
	Road Conditions	Road Conditions	Age	Vertical Alignment	Horizontal Alignment	Accident Day
Road Conditions	Pearson Correlation	1	0,115	-0,052	0,005	-0,090
	Sig. (2-tailed)		0,063	0,400	0,935	0,146
	N	264	264	260	261	264
Age	Pearson Correlation	0,115	1	0,113	-0,057	-0,024
	Sig. (2-tailed)	0,063		0,067	0,359	0,699
	N	264	265	261	262	265
Vertical Alignment	Pearson Correlation	-0,052	0,113	1	0,140	0,158
	Sig. (2-tailed)	0,400	0,067		0,024	0,010
	N	260	261	261	261	261
Horizontal Alignment	Pearson Correlation	0,005	-0,057	0,140	1	0,127
	Sig. (2-tailed)	0,935	0,359	0,024		0,040
	N	261	262	261	262	262
Accident Day	Pearson Correlation	-0,090	-0,024	0,158	0,127	1
	Sig. (2-tailed)	0,146	0,699	0,010	0,040	
	N	264	265	261	262	265
Grand Total	N	264	265	261	262	265

Table A.2.5.1: Correlation testing for the variables selected for the binomial model developed in the 2nd FAI database

APPENDIX 3.1: Joint modelling of road condition variable

Road condition variables

In Table A3.1 the proportion of single vehicle accidents is predicted by variables describing the road condition.

	B	S.E.	Wald	df	Sig.	Exp(B)
Carriageway Phys. Div.	.843	.235	12.880	1	.000	2.323
TrafficFlow	-.997	.146	46.311	1	.000	.369
Junction	-1.696	.212	63.935	1	.000	.183
Constant	1.228	.229	28.896	1	.000	3.416

Table A3.1 Joint modelling of road condition variables. Variables entered: Junction, Carriageway Physically Divided, Traffic Flow, Speed limit greater 50, Speed limit greater 90, Speed limit greater 100, Motorway, Area

APPENDIX 3.2: Joint modelling of time variable

Road condition variables

In Table A3.2 the proportion of single vehicle accidents is predicted by variables characterizing the time of the accident.

	B	S.E.	Wald	df	Sig.	Exp(B)
Light	.520	.193	7.282	1	.007	1.682
Daytime	-.554	.206	7.256	1	.007	.575
Weekend	.416	.156	7.056	1	.008	1.515
Constant	-.645	.234	7.573	1	.006	.525

Table A3.1 Joint modelling of time variables. Variables entered: Light, Daytime, Weekend, WeekendNights

APPENDIX 3.3: Joint modelling of road user variables

In Table A3.3 the proportion of single vehicle accidents is predicted by the road user variables.

	B	S.E.	Wald	df	Sig.	Exp(B)
MeanDriverAge	-.013	.007	3.922	1	.048	.987
Familiar	-1.149	.238	23.394	1	.000	.317
Occupants	.287	.090	10.100	1	.001	1.332
ptwInvolved	-1.045	.208	25.218	1	.000	.352
hgvInvolved	-2.278	.299	57.965	1	.000	.103
Constant	.248	.339	.534	1	.465	1.281

Table A3.3 Joint modelling of road user variables. Variables entered: MeanDriverAge, Familiar, Occupants, ptwInvolved, hgvInvolved, percentageFemaleDriver, Alcohol, Impairment

APPENDIX 3.4: Joint modelling of accident development variables

In Table A3.4 the proportion of single vehicle accidents is predicted by the accident development variables.

	B	S.E.	Wald	df	Sig.	Exp(B)
AvoidMan01	-.591	.176	11.284	1	.001	.554
Constant	-.140	.153	.836	1	.361	.870

Table A3.4 Joint modelling of accident development variables. Variables entered: AvoidanceManoeuvre, ExecutedManoeuvre.

Appendix 4.1. Bayesian estimation of multilevel logit models

After the standard estimation procedure of a binomial logit model, a MCMC (Markov Chain Monte Carlo) estimation method is applied by means of sampling from prior parameter distributions in order to obtain more accurate (interval) estimates for the parameters ("posterior" distribution) and the likelihood statistic (Rasbash et al. 2000). The starting values of the process are those obtained through the conventional estimation method, whereas specific criteria are used to determine the number of iterations required for the convergence of posterior parameter estimates to a given confidence level (Browne, 2003).

In this framework, the possible hierarchies in the dataset can be first tested in terms of geographical dependencies. A two-level "empty" model is considered, in which road users are nested into countries.

$$\text{logit}(\pi_{ij}) = \beta_{0j} \text{ cons}$$

$$\beta_{0jk} = \beta_0 + u_{0j}$$

$$u_{0j} \sim N(0, \sigma_{u0}^2)$$

Apart from the geographical dependencies, another type of dependencies that needs to be examined is the dependencies due to the accident process. In particular, a three-level "empty" structure is considered, according to which persons are nested into vehicles and vehicles are nested into accidents.

$$\text{logit}(\pi_{ijk}) = \beta_{0jk} \text{ cons}$$

$$\beta_{0jk} = \beta_0 + u_{0jk} + v_{0k}$$

$$u_{0jk} \sim N(0, \sigma_{u0}^2)$$

$$v_{0k} \sim N(0, \sigma_{v0}^2)$$

On the basis of the statistical significance of the random effects and the models fit, it can be determined whether one of the examined hierarchical structures is meaningful. Furthermore, explanatory variables can be tested.

It is noted that Bayesian estimation was proved to be less efficient in the multinomial multilevel modelling, providing substantially higher estimates of the random parameters compared to the standard RIGLS estimation, which is a known problem that may occur in this case (Browne, 2003). Moreover, the posterior distributions of the random parameters were not stabilised even after several hundreds thousands iterations. For these reasons, the standard estimation method's results were considered to be more reliable and are the only ones presented here. The likelihood statistic estimation, however, is quite approximate in this case and can only be taken as a rough measure of model fit.

Appendix 4.2 - Defining the “Opponent Type” variable

Assigning each Road-User an opponent:

In the FAI database, the information about each accident includes the different “events” that make up the accident (up to 6 events per accident). As a consequence, the road-user involved in one given accident, can be described as having interacted with several others (e.g.: a car that, after having collided with another car, is pushed back and hits a pedestrian standing next to the accident scene). It is important to adequately choose the opponent that will be used to define the “collision opponent” for each road-user. This was done using the chronological unfolding of the accident as a reference: The earlier opponent (those with whom the R-U interacted first) was considered as the most “significant” one, and consequently designated as opponent over those that the road-user “encountered” in the later stages of the accident (so in the case of the example above, the vehicle that was hit first will be designated as opponent, so that the all the occupants of this car will have the value “car” as opponent). The data indicate that, in 91.5% of the cases, the event that was chosen as basis to assign the road users their opponent corresponded to the event that had been rated as “most harmful” in the database (see Table 2). In the majority of the remaining cases, the most harmful event followed the first interaction with the opponent taken into account (turquoise fields), and only in 42 cases the most harmful event preceded the event that had been considered critical for the determination of the opponent (grey fields).

		Most Harmful Event						
		1	2	3	4	5	6	Unknown
Event taken into account to assign opponent	1	1913	70	33	28	5	1	74
	2	31	375	19	8	0	0	27
	3	5	6	54	4	5	0	4
	4	0	1	0	16	1	3	3
	5	0	0	0	0	2	0	5
	6	0	0	0	0	0	0	5

Table A 4.2: Basis for the opponent assignment and most harmful event in the accident

This way of proceeding also implies that there are as many “collision opponent” values for as there are road users involved in the accident. Hence, the variable that is here defined should not be understood as a variable that characterises the accident itself (to keep on with the car-car-pedestrian example: collision type will be “car-car” for both of the cars in the accident, while it will be “pedestrian-car” for the pedestrian), but rather each individual road-user.

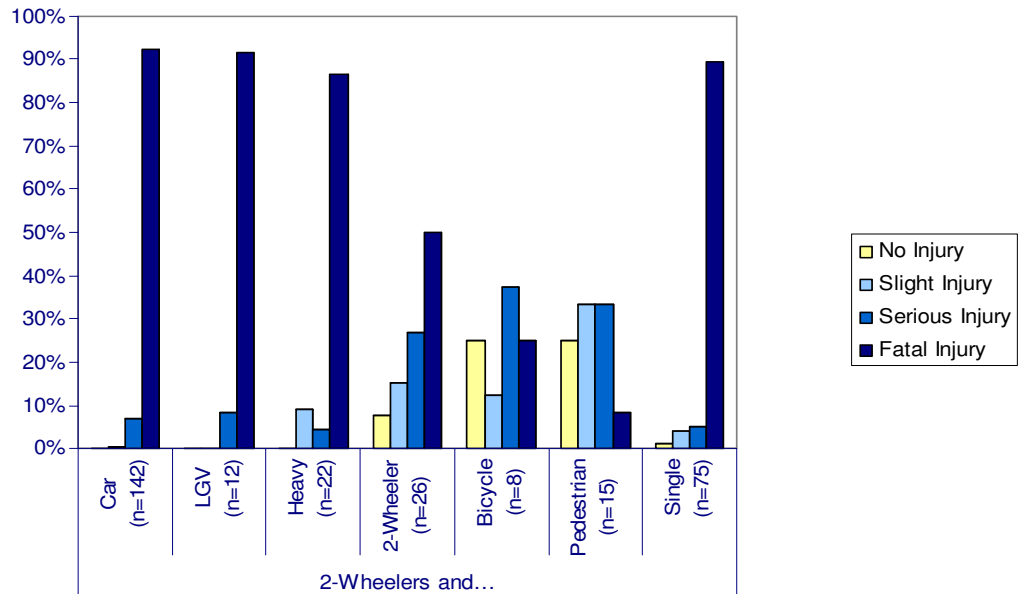
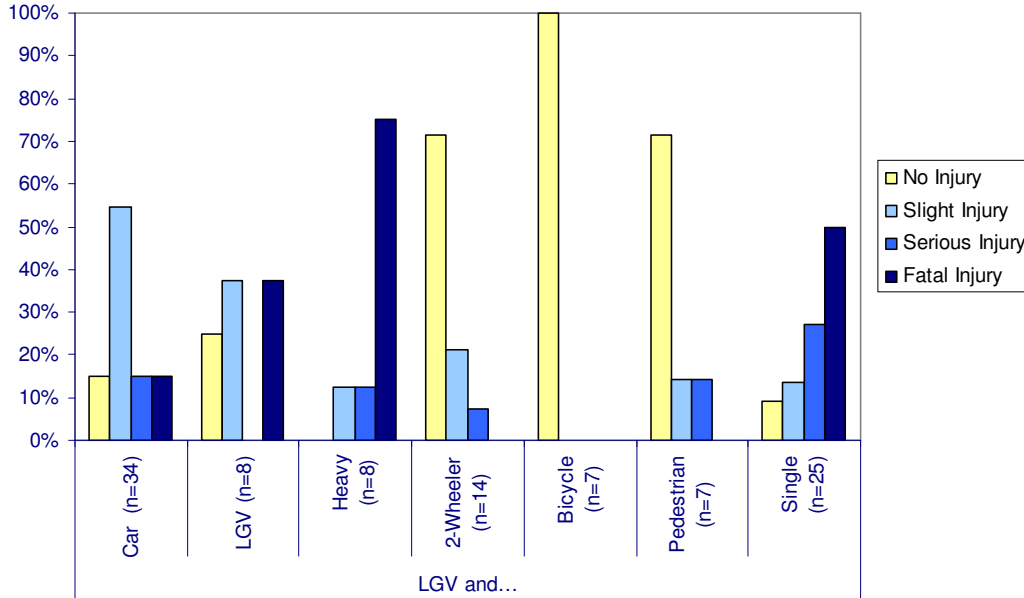
2. Ambiguous accidents/accidents involving parked vehicles

In certain cases, however, the accident pattern was complicated and could even prove ambiguous. This was typically the case for accidents involving parked vehicles. Parked vehicles can be both vehicles that are parked along the road at some place where they are to be expected, but also vehicles stopped for one reason or another at some place where they are unexpected (e.g.: defective vehicle...). The status of one road-user can change in the course of the accident (the driver who steps out of the car to check the motor and gets hit by

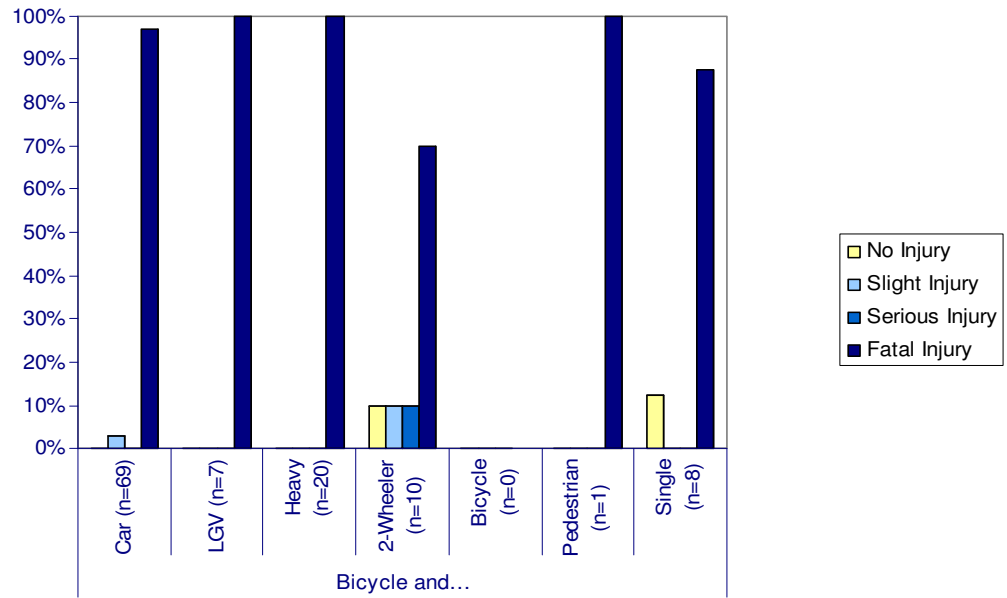
a truck is no car driver anymore, but a pedestrian). Extra-attention was consequently devoted to accidents involving parked vehicles, so as to ensure that the collision type value had been rightly coded for all road-users in these accidents. As an example, for the accident described below, the opponent type was originally coded as “car” for all occupants of the first vehicle (there was no occupant in the second vehicle): “The accident took place in an urban area between two cars. The first car was travelling at high speed when the driver lost control of his vehicle and crossing the centre line hit on back a parked vehicle and then had a rollover-overturn.” After verification, and on basis of the accident summary, the opponent type was as “none” in this case.

When there was too much ambiguity about the exact nature of the interaction, or when the circumstances appeared to be too exceptional, the accident case was discarded of the analysis. As an example, the information related to the occupant of vehicle 2 was excluded from the analysis in the following case: The accident took place on a motorway on a viaduct and two vehicles were involved. The first truck was parked on the emergency lane and the driver was outside the vehicle while his cousin was sitting inside. At that moment was arriving a truck (tractor + semitrailer) out of control (for unknown causes): it hit the parked vehicle pushed it outside the viaduct in an escarpment falling for 80 metres. No brake marks have been found by police. The driver of vehicle 2 stated to the police that he saw the parked vehicle immediately before the collision”.

APPENDIX 4.3: Proportions of fatalities, severe injuries, slight injuries and uninjured cases for remaining road-user-opponent combinations.



Analysing European in-depth data: Methodological framework and results



Appendix 4.4 Commando files for data preparation

4.4.1 Variable recoding

```
// String variables will be transformed into numerical variables
Missing values will here be coded as "." or ".a" (when it is
necessary, for example to distinguish unknown from mere missing
data). The presence of values such as "-999" or 999 is checked for
every variable (see summaries above), and in most cases transformed
into missing values (unless clear indications that this is indeed a
value within the range of the variable) -777 values will also be
recoded as missing values.//

// The variables "Carmpv", "busminibus", "bicycle"... indicate the
number of participants of each type involved in the accident. It is
thus an accident-level variable. Another variable "vehicletype"
contains the same information at the vehicle level. It is a 9-
categories variable indicating which class of road user the vehicle
(participant) described belongs to. The numbers of participants
contains missing values instead of 0 when the participants involved in
the accident are from another type than the variable's name
(e.g.:van), and will be replaced by 0 so that sum can be calculated on
the basis of these participants type. //

//variables.//

recode carmpv (.=0)
recode van (.=0)
recode busminibus (.=0)
recode truck (.=0)
recode agriculturalvehicle (.=0)
recode motorcyclemoped (.=0)
recode bicycle (.=0)
recode traintram (.=0)
recode shoevehiclepedestrian (.=0)
recode unknownvehicle (.=0)
recode unknownvehicle (.=0)

label variable carmpv "Total number of cars in the accident"
label variable busminibus "Total number of buses/minibuses in the
accident"
label variable van "Total number of vans in the accident"
label variable truck "Total number of trucks in the accident"
label variable agriculturalvehicle "Total number of agricultural
vehicles in the accident"
label variable motorcyclemoped "Total number of motorcycles/mopeds in
the accident"
label variable bicycle "Total number of bicycles in the accident"
label variable traintram "Total number of trains/trams in the
accident"
label variable shoevehiclepedestrian "Total number of pedestrians in
the accident"

// Construction of 2 broad categories, one for Heavy and the other for
vulnerable road users//
```



Analysing European in-depth data: Methodological framework and results

```
generate SumHeavy=(van+busminibus+truck)
label variable SumHeavy "Total nb of heavy vehicles involved"
generate SumVulnerable=(bicycle+shoevehiclepedestrian)
label variable SumVulnerable "Total nb of vulnerables RU involved"

// In case MlwiN should be used with this datafile, the following
variables are created:

generate Cons=1
generate AccidentID=casenumbr
generate VehicleID=AccidentID*10+vehiclenunder
generate UserID=VehicleID*10+personnumber

// Construction of a numerical variable for the Safetynet medical
outcome//

generate Consequence=4 if (safetynetmedicaloutcome=="Fatal")
replace Consequence=3 if (safetynetmedicaloutcome=="Serious")
replace Consequence=2 if (safetynetmedicaloutcome=="Slight")
replace Consequence=1 if (safetynetmedicaloutcome== "Not Injured")
replace Consequence=.a if (safetynetmedicaloutcome=="Unknown")
label variable Consequence "Consequence"
label define Consequence 1 "Fatal" 2 "Serious" 3 "Slight" 4
"Uninjured"
label values Consequence Consequence
describe Consequence

// Creation of indicator variables for each consequence level://

recode Consequence (1=1) (else=0), generate (Uninjured)
recode Consequence (2=1) (else=0), generate (Slightly)
recode Consequence (3=1) (else=0), generate (Seriously)
recode Consequence (4=1) (else=0), generate (Killed)
recode Consequence (4=1) (3=1) (else=0), generate (KSI)
inspect (Uninjured Slightly Seriously Killed KSI)

// Creation of categorical variable for the number of crash
participants//

generate OccInVeh= numofoccupants

// Creation of categorical variable for the number of crash
participants://

recode crashparticipants (1=1) (2=2) (3/max=3), generate
(CrashPartCat)
inspect CrashPartCat

// Creation of an indicator variable for single vs multiple car
accidents://

recode crashparticipants (1=0) (2/max=1), generate (MultipleCarAcc)
inspect MultipleCarAcc
label define MultipleCarAcc 0 "Single-vehicle" 1 "Multi-vehicle"
label values MultipleCarAcc MultipleCarAcc

// Creation of a numerical categorical variable from vehicletype://

generate VehicleCat1=1 if (vehicletype=="Van")
```



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

Analysing European in-depth data: Methodological framework and results

```
replace VehicleCat1=1 if (vehicletype=="Truck")
replace VehicleCat1=4 if (vehicletype== "Bicycle")
replace VehicleCat1=1 if (vehicletype== "Bus / Minibus")
replace VehicleCat1=2 if (vehicletype== "Car / MPV")
replace VehicleCat1=3 if (vehicletype=="Motorcycle / Moped")
replace VehicleCat1=5 if (vehicletype=="Shoe vehicle (pedestrian)")
replace VehicleCat1=6 if ((vehicletype=="Other") |(vehicletype=="Train
/ Tram")|(vehicletype=="Agricultural vehicle"))
label variable VehicleCat1 "Recode from vehicletype rare vehicles as
6th category"
label define VehicleCat1 1 "Heavy" 2 "Cars" 3 "2-wheelers" 4
"Bicycles" 5 "Pedestrians" 6 "Other"
label values VehicleCat1 VehicleCat1
tabulate VehicleCat1

// Creation of indicator variable for the road user class to which the
accident participant belongs://

recode VehicleCat1 (1/2=0) (3=1) (4/6=0) (-999=.), gen (Moto)
recode VehicleCat1 (1/3=0) (4/5=1) (6=0)(-999=.), gen (Vulnerable)
recode VehicleCat1 (1=1) (2/6=0) (-999=.), gen (Heavy)

// Creation of numeric variable from roaduserclass:

generate UserClass=1 if (roaduserclass=="Driver")
replace UserClass=2 if (roaduserclass== "Passenger")
replace UserClass=3 if (roaduserclass== "Pedestrian")
label define UserClass 1 "Driver" 2 "Passenger" 3 "Pedestrian"
label values UserClass UserClass

//Replacement of missing age values (coded as 999) by mean age: 38.73
(based on D7.6 syntax)//

replace age=39 if(age==999)
summarize age

//Creating a categorical variable for age://

recode age (0/17=1) (18/25=2) (26/34=3) (35/54=4) (55/64=5)
(65/100=6), generate (AgeCat)
tabulate AgeCat
label define AgeCat 1 "00 to 17" 2 "18 to 25" 3 "26 to 34" 4 "35 to
54" 5 "55 to 64" 6 "65+"
label values AgeCat AgeCat

// Creating an indicator variable for "Senior":

generate Senior=0 if
((AgeCat==1)|(AgeCat==2)|(AgeCat==3)|(AgeCat==4)|(AgeCat==5))
replace Senior=1 if (AgeCat==6)
label define Senior 1 "Yes" 0 "No"
label values Senior Senior

// Creation of categorical variable for gender

generate Gender=1 if (gender=="Female")
replace Gender =2 if (gender=="Male")
replace Gender=. if (gender=="Unknown")
label define Gender 1 "Female" 2 "Male"
label values Gender Gender
```

Analysing European in-depth data: Methodological framework and results

```
recode Gender (1=1) (2=0), generate (Female)
tabulate Gender
tabulate Female
label define Female 1"Female" 2"Male"
label values Female Female

// Creating a numeric variable for Familiarity:
tabulate isfamiliar
generate IsFamiliar=1 if (isfamiliar=="No")
replace IsFamiliar=0 if (isfamiliar=="Yes")
replace IsFamiliar=. if ((isfamiliar=="Unknown")|(isfamiliar=="."))
tabulate IsFamiliar
inspect IsFamiliar
label define IsFamiliar 0"Yes" 1"No"
label values IsFamiliar IsFamiliar

// Creating a numeric variable for alcohol suspicion:
generate Alcohol=1 if (suspicionalcohol=="Yes")
replace Alcohol=0 if (suspicionalcohol=="No")
replace Alcohol=.a if (suspicion alcohol=="Not applicable")
inspect Alcohol
label define Alcohol 0 "No" 1 "Yes"
label values Alcohol Alcohol

// Creating numeric variable for impairment:
generate Impairment01=1 if
((impairment=="Alcohol")|(impairment=="Combination of the
above")|(impairment=="Drugs")|(impairment=="Drugs and
alcohol")|(impairment=="Fatigue")|(impairment=="Medication")|(impairme
nt=="Other"))
replace Impairment01=0 if (impairment=="None")
replace Impairment01=. if (impairment=="Unknown")
inspect impairment
inspect Impairment01
tabulate impairment
tabulate Impairment01

//-----
//-----
//*****FOR CAR-CAR ANALYSIS ONLY: ("Not applicable" are coded ".a"
and "Unknown" are coded as ".")
//-----
//-----

*Creating a numeric variable for SeatPosition:

generate BackSeat=0 if
((seatpos=="1.1")|(seatpos=="1.2")|(seatpos=="1.3"))
replace BackSeat=1 if
((seatpos=="2.1")|(seatpos=="2.2")|(seatpos=="2.3")|(seatpos=="2.4")|(
seatpos=="3.3"))
replace BackSeat=.a if (seatpos=="Not Applicable")|(seatpos=="Not
applicable")
replace BackSeat=. if (seatpos=="Unknown")
inspect BackSeat
inspect seatpos
tabulate BackSeat
```



Analysing European in-depth data: Methodological framework and results

```
tabulate seatpos
label define BackSeat 0 "Front" 1 "Back"
label values BackSeat BackSeat

*Creating numeric variable for Seatbelt

tabulate seatbelt
generate NoSeatbelt=1 if (seatbelt=="Not used")
replace NoSeatbelt=0 if (seatbelt=="Used")|(seatbelt=="Use claimed")
replace NoSeatbelt=.a if (seatbelt=="Not Applicable")
replace NoSeatbelt=.a if (seatbelt=="Not applicable")
replace NoSeatbelt=. if (seatbelt=="Unknown")
tabulate NoSeatbelt
label define NoSeatbelt 0 "Seatbelt on" 1 "No seatbelt"

* Creating numeric variable for Airbag Deployment

generate NoAirbagDeployed=1 if (airbagdeploy=="No")
replace NoAirbagDeployed=0 if (airbagdeploy=="Yes")
replace NoAirbagDeployed=.a if (airbagdeploy=="Not
Applicable")|(airbagdeploy=="Not applicable")
replace NoAirbagDeployed=. if (airbagdeploy=="Unknown")
tabulate airbagdeploy
tabulate NoAirbagDeployed
describe NoAirbagDeployed
label define NoAirbagDeployed 0 "Airbag" 1 "No Airbag"
label values NoAirbagDeployed NoAirbagDeployed

*Creating categorical variables to summarize avoidance manoeuvre:
**1/ Binary indicator: Manoeuvre (0) versus no-manoevre (1) -
Caution: like the 2 previous ones, this variable's coding is the
opposite as the one defined by H!!!

generate CrashAvoidMan=1 if (crashavoidman=="Braking (no skid marks
evident)")
replace CrashAvoidMan=2 if (crashavoidman=="Braking (skid marks
evident)")
replace CrashAvoidMan=3 if (crashavoidman=="No avoidance manoeuvre
reported")
replace CrashAvoidMan=4 if (crashavoidman=="Not reported/inconclusive
(by police)")
replace CrashAvoidMan=5 if (crashavoidman=="Other")
replace CrashAvoidMan=6 if (crashavoidman=="Other avoidance manoeuvre")
replace CrashAvoidMan=7 if (crashavoidman=="Steering (evidence or
stated)")
replace CrashAvoidMan=8 if (crashavoidman=="Steering and braking
(evidence or state)")
replace CrashAvoidMan=8 if (crashavoidman=="Steering and braking
(evidence or stated)")
replace CrashAvoidMan=9 if (crashavoidman=="Unknown")

label define CrashAvoidMan 1 "Braking (no skid marks)" 2 "Braking
(skid marks)" 3 "None reported" 4 "Not reported by
police/inconclusive" 5 "Other" 6 "Other manoeuvre" 7 "Steering" 8
"Steering and braking" 9 "Unknown"
label values CrashAvoidMan CrashAvoidMan

* Creation of indicator variable describing whether any manoeuvre took
place before the crash (or not): 0 (the "normal case") is defined as
```



Analysing European in-depth data: Methodological framework and results

the occurrence of a manoeuvre, 1 as the absence of any type of manoeuvre

```
generate NOCrashAvoidMan=0 if
((CrashAvoidMan==1)|(CrashAvoidMan==2)|(CrashAvoidMan==6)|(CrashAvoidMan==7)|(CrashAvoidMan==8))
replace NOCrashAvoidMan=1 if (CrashAvoidMan==3)
replace NOCrashAvoidMan=. if (CrashAvoidMan==9)
replace NOCrashAvoidMan=.a if ((CrashAvoidMan==5)|(CrashAvoidMan==4))
label define NOCrashAvoidMan 0 "Manoeuvre" 1 "No Manoeuvre"
label values NOCrashAvoidMan NOCrashAvoidMan
```

*Creation of indicator variable forNObraking:

```
generate NOBraking=0 if
((CrashAvoidMan==1)|(CrashAvoidMan==2)|(CrashAvoidMan==8))
replace NOBraking=1 if
((CrashAvoidMan==3)|(CrashAvoidMan==5)|(CrashAvoidMan==6)|(CrashAvoidMan==7))
replace NOBraking=. if
((CrashAvoidMan==4)|(CrashAvoidMan==9)|(CrashAvoidMan==5))
label define NOBraking 0 "Braked" 1"Did NOT braked"
label values NOBraking NOBraking
tabulate NOBraking
```

*Creation of indicator variable for steering (coefficient indicates change from having steered to not having done so)

```
generate NOSteering=0 if ((CrashAvoidMan==7)|(CrashAvoidMan==8))
replace NOSteering=1 if
((CrashAvoidMan==1)|(CrashAvoidMan==2)|(CrashAvoidMan==3)|(CrashAvoidMan==4)|(CrashAvoidMan==5))
replace NOSteering=. if ((CrashAvoidMan==4)|(CrashAvoidMan==9))
label define NOSteering 0 "Steered" 1"Did NOT steered"
label values NOSteering NOSteering
```

```
tabulate CrashAvoidMan
tabulate NOCrashAvoidMan
tabulate NOBraking
tabulate NOSteering
inspect CrashAvoidMan
inspect crashavoidman
tabulate crashavoidman CrashAvoidMan
```

***Recode of the "First Event" variable in order to get identical values than for "Event Detail"

```
tabulate firstevent
tabulate eventdetail
generate FirstEvent= "Cross median/centreline" if (firstevent=="Cross median/centreline")
replace FirstEvent= "Embankment - material type unknown" if
(firstevent=="Embankment - material type unknown")
replace FirstEvent="Equipment failure (blown tyre, brake failure, etc.)" if (firstevent=="Equipment failure (blown tire, brake failure, etc.)")
replace FirstEvent="Fell/jumped from vehicle" if
(firstevent=="Fell/jumped from vehicle")
replace FirstEvent="Vehicle struck by falling/shifting cargo or anything set in motion by another motor vehicle transport" if
```



Analysing European in-depth data: Methodological framework and results

```
(firstevent=="Motor vehicle struck by falling/shifting cargo or
anything set in motion by another motor vehicle in transport")
replace FirstEvent="Vehicle travelling on other roadway" if
(firstevent=="Motor vehicle travelling on other roadway")
replace FirstEvent="Vehicle travelling on same roadway" if
(firstevent=="Motor vehicle travelling on sameroadway")
replace FirstEvent="Parked vehicle (not travelling)" if
(firstevent=="Parked motor vehicle (not travelling)")
replace FirstEvent="Ran off road - Nearside" if (firstevent=="Ran off
road - near side")
replace FirstEvent="Ran off road - Offside" if (firstevent=="Ran off
road - off side")
```

*Checking whether Road Users Event1 match the first event in the accident - Value is 1 whenever first event in accident is first event user.

```
generate FirstEventIsEvent1=1 if (FirstEvent==eventdetail1)
replace FirstEventIsEvent1=0 if (FirstEvent!=eventdetail1)
```

* A numerical variable is first computed on basis of mostharmfulevent: What is the position of most harmful event in the chain of events?

```
generate MostHarmfulEvent=1 if (mostharmfulevent=="Event 1")
replace MostHarmfulEvent=2 if (mostharmfulevent=="Event 2")
replace MostHarmfulEvent=3 if (mostharmfulevent=="Event 3")
replace MostHarmfulEvent=4 if (mostharmfulevent=="Event 4")
replace MostHarmfulEvent=5 if (mostharmfulevent=="Event 5")
replace MostHarmfulEvent=6 if (mostharmfulevent=="Event 6")
replace MostHarmfulEvent=. if (mostharmful event=="Unknown")
tabulate mostharmfulevent MostHarmfulEvent
```

*Now, indicator variable:

```
generate MostHarmfulNOTFirst=1 if
((MostHarmfulEvent==2)|(MostHarmfulEvent==3)|(MostHarmfulEvent==4)|(MostHarmfulEvent==5))
replace MostHarmfulNOTFirst=0 if (MostHarmfulEvent==1)
replace MostHarmfulNOTFirst=. if (MostHarmfulEvent==6)
label define MostHarmfulNOTfirst 0 "First most harmful" 1 "Later most harmful"
label values MostHarmfulNOTFirst MostHarmfulNOTFirst
```

** Creation of a variable indicating of what type the most harmful event is:

```
generate HarmEventType=eventdetail1 if (MostHarmfulEvent==1)
replace HarmEventType=eventdetail2 if (MostHarmfulEvent==2)
replace HarmEventType=eventdetail3 if (MostHarmfulEvent==3)
replace HarmEventType=eventdetail4 if (MostHarmfulEvent==4)
replace HarmEventType=eventdetail5 if (MostHarmfulEvent==5)
replace HarmEventType=eventdetail6 if (MostHarmfulEvent==6)
replace HarmEventType="." if (MostHarmfulEvent== .)
tabulate HarmEventType
tabulate MostHarmfulEvent MostHarmfulNOTFirst
```



Analysing European in-depth data: Methodological framework and results

```
** Recoding all Event variables into Collision categories

generate FirstEventCat=1 if (firstevent=="Cross
median/centreline")|(firstevent=="Equipment failure (blown tire, brake
failure, etc.)")|(firstevent=="Fell/jumped from
vehicle")|(firstevent=="Other non-
collision")|(firstevent=="Overturn/rollover")|(firstevent=="Pavement/r
oad surface irregularity (pothole, grooved,
grates)")|(firstevent=="Ran off road - near side")|(firstevent=="Ran
off road - off side")|(firstevent=="Ran off road -
Nearside")|(firstevent=="Ran off road -
Offside")|(firstevent=="Separation of
units")|(firstevent=="Ejection")|(firstevent=="Fire/explosion")|(first
event=="Injured in vehicle")|(firstevent=="Jack-
knife")|(firstevent=="Vehicle went
airborne")|(firstevent=="Immersion")
replace FirstEventCat=2 if (firstevent=="Motor vehicle travelling on
other roadway")|(firstevent=="Motor vehicle travelling on same
roadway")|(firstevent=="Parked motor vehicle (not
travelling)")|(firstevent=="Parked vehicle (not
travelling)")|(firstevent=="Vehicle travelling on other
roadway"=2)|(firstevent=="Vehicle travelling on same roadway")
replace FirstEventCat=3 if (firstevent=="Motor vehicle struck by
falling/shifting cargo or anything set in motion by another motor
vehicle in transport")|(firstevent=="Other object (not
fixed)")|(firstevent=="Pedal
cycle")|(firstevent=="Pedestrian")|(firstevent=="Vehicle struck by
falling/shifting cargo or anything set in motion by another vehicle")
replace FirstEventCat=2 if (firstevent=="Motor vehicle travelling on
other roadway")|(firstevent=="Motor vehicle travelling on same
roadway")|(firstevent=="Parked motor vehicle (not
travelling)")|(firstevent=="Parked vehicle (not
travelling)")|(firstevent=="Vehicle travelling on other
roadway")|(firstevent=="Vehicle travelling on same roadway")
replace FirstEventCat=4 if (firstevent=="Embankment -
earth")|(firstevent=="Embankment - material type
unknown")|(firstevent=="Guardrail face")|(firstevent=="Guardrail
end")|(firstevent=="Kerb")|(firstevent=="Other fixed
object")|(firstevent=="Other traffic barrier")|(firstevent=="Traffic
signal support/signal")|(firstevent=="Tree (standing tree
only)")|(firstevent=="Bridge parapet end")|(firstevent=="Bridge pier
or abutment")|(firstevent=="Building")|(firstevent=="Culvert")|
(firstevent=="Ditch")|(firstevent=="Luminary/light
support")|(firstevent=="Utility pole")|(firstevent=="Wall")|
(firstevent=="Other post, other pole, or other supports")

replace FirstEventCat=.a if (firstevent== "Not Applicable")
tabulate firstevent
tabulate FirstEventCat
inspect FirstEventCat

generate HarmEventCat=1 if (HarmEventType=="Cross median/centreline")|
(HarmEventType=="Equipment failure (blown tyre, brake failure, etc.)")
|(HarmEventType=="Fell/jumped from vehicle")|(HarmEventType=="Other
non-collision")|(HarmEventType=="Overturn/rollover")|
(HarmEventType=="Pavement/road surface irregularity (pothole, grooved,
grates)")|(HarmEventType=="Ran off road - near side")|
(HarmEventType=="Ran off road - off side")|(HarmEventType=="Ran off
road - Nearside")|(HarmEventType=="Ran off road - Offside")|
```



Analysing European in-depth data: Methodological framework and results

```
(HarmEventType=="Separation of units")|(HarmEventType=="Ejection")|
(HarmEventType=="Fire/explosion")|(HarmEventType=="Injured in
vehicle")|(HarmEventType=="Jack-knife")|(HarmEventType=="Vehicle went
airborne")|(HarmEventType=="Immersion")

replace HarmEventCat=2 if (HarmEventType=="Motor vehicle travelling on
other roadway")|(HarmEventType=="Motor vehicle travelling on same
roadway")|(HarmEventType=="Parked motor vehicle (not
travelling)")|(HarmEventType=="Parked vehicle (not
travelling)")|(HarmEventType=="Vehicle travelling on other
roadway"=2)|(HarmEventType=="Vehicle travelling on same roadway")
replace HarmEventCat=3 if (HarmEventType=="Motor vehicle struck by
falling/shifting cargo or anything set in motion by another motor
vehicle in transport")|(HarmEventType=="Other object (not
fixed)")|(HarmEventType=="Pedal
cycle")|(HarmEventType=="Pedestrian")|(HarmEventType=="Vehicle struck
by falling/shifting cargo or anything set in motion by another
vehicle")

replace HarmEventCat=2 if (HarmEventType=="Motor vehicle travelling on
other roadway")|(HarmEventType=="Motor vehicle travelling on same
roadway")|(HarmEventType=="Parked motor vehicle (not travelling)")|
(HarmEventType=="Parked vehicle (not travelling)")|
(HarmEventType=="Vehicle travelling on other roadway")|
(HarmEventType=="Vehicle travelling on same roadway")
replace HarmEventCat=4 if (HarmEventType=="Embankment - earth")|
(HarmEventType=="Embankment - material type unknown")|
(HarmEventType=="Guardrail face")|(HarmEventType=="Guardrail
end")|(HarmEventType=="Kerb")|(HarmEventType=="Other fixed
object")|(HarmEventType=="Other traffic barrier")|
(HarmEventType=="Traffic signal support/signal")|(HarmEventType=="Tree
(standing tree only)")|(HarmEventType=="Bridge parapet end")|
(HarmEventType=="Bridge pier or abutment")|
(HarmEventType=="Building")|(HarmEventType=="Culvert")|
(HarmEventType=="Ditch")|(HarmEventType=="Luminary/light
support")|(HarmEventType=="Utility pole")|(HarmEventType=="Wall")|
(HarmEventType=="Other post, other pole, or other supports")
tabulate HarmEventType
tabulate HarmEventCat

tabulate firstevent FirstEvent

label define CollisionType 1 "No Collision" 2 "Collision with other
vehicle" 3 "Collision with object nonfixed" 4 "Collision with fixed
object"
label values FirstEventCat CollisionType
label values HarmEventCat CollisionType
label values Event1Cat CollisionType
label values Event2Cat CollisionType
label values Event3Cat CollisionType
label values Event4Cat CollisionType
label values Event5Cat CollisionType
label values Event6Cat CollisionType

**Recoding traffic flow into numerical variable

generate TrafficFlow3=1 if (trafficflow=="Light traffic flow")
replace TrafficFlow3=2 if (trafficflow=="Normal traffic flow")
replace TrafficFlow3=3 if (trafficflow=="Heavy traffic flow")
```



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

Analysing European in-depth data: Methodological framework and results

```
replace TrafficFlow3=. if (trafficflow=="Unknown")
label define TrafficFlow3 1 "Light" 2 "Normal" 3 "Heavy"
label values TrafficFlow3 TrafficFlow3
tabulate TrafficFlow3 trafficflow

***Cleaning up "Year of Manufacture" - Generating the VehicleAge
variable:

destring yearofmanufacture, replace force
replace yearofmanufacture=1998 if (yearofmanufacture==1198)
replace yearofmanufacture=. if (yearofmanufacture<0)

gen acdate= date(accidentdate, "dmy")
format acdate %d
gen AccYear= year(acdate)
gen VehicleAge2=AccYear-yearofmanufacture
summarize VehicleAge2
generate vehagecent=(VehicleAge2-7.611858)

**Replacing missing values for vehicleage2 by mean value for
corresponding vehicle category:

by VehicleCat, sort: summarize VehicleAge VehicleAge2
replace VehicleAge2=6.1 if ((VehicleCat==1)&(VehicleAge2==.))
replace VehicleAge2=8.12 if ((VehicleCat==2)&(VehicleAge2==.))
replace VehicleAge2=5.65 if ((VehicleCat==3)&(VehicleAge2==.))
tabulate VehicleAge2

*Creation of indicator variables for the presence of technical
systems:

generate abs01=0 if (abs=="No")
replace abs01=1 if (abs=="Yes")
replace abs01=. if (abs=="Unknown")
replace abs01=.a if (abs=="Not Applicable")

generate ESP01=0 if (esp=="No")
replace ESP01=1 if (esp=="Yes")
replace ESP01=. if (esp=="Unknown")
replace ESP01=.a if (esp=="Not Applicable")

generate CSS01=0 if (esp=="No")
replace CSS01=1 if (esp=="Yes")
replace CSS01=. if (esp=="Unknown")
replace CSS01=.a if (esp=="Not Applicable")

generate TCS01=0 if (esp=="No")
replace TCS01=1 if (esp=="Yes")
replace TCS01=. if (esp=="Unknown")
replace TCS01=.a if (esp=="Not Applicable")

*Creation of different categorical variables for speedlimit:

*first replace 999 by ".":
replace speedlimit=. if (speedlimit==999)|(speedlimit==888)

replace speedlimit=70 if (speedlimit==7)
```



Analysing European in-depth data: Methodological framework and results

```
recode speedlimit (min/50=50) (51/80=80) (81/100=100) (101/200=130)
(200/max=.), gen (SpeedLimitCat)
tabulate SpeedLimitCat
recode speedlimit (min/50=0) (50/max=1), gen (SLfaster50)
```

*** Creation of a numerical variables for Area:

*Step 1: Keeping the three categories

```
generate LocalArea=1 if (localarea=="Rural")
replace LocalArea=2 if (localarea=="Urban")
replace LocalArea=3 if (localarea=="Mixed")
label define LocalArea 1 "Rural" 2 "Urban" 3 "Mixed"
label values LocalArea LocalArea
```

*Step 2: Differentiating between rural and all (Urban/Mixed):

```
recode LocalArea (1=1) (2=0) (3=0), gen (RuralArea)
```

* Creation of an indicator variable for junction: 0 means "no junction", 1 means "junction" (whatever the type)

```
generate junction01=1 if (junction == "Crossroads (+ junction)")
replace junction01=0 if (junction == "No junction")
replace junction01=1 if (junction == "Other")
replace junction01=1 if (junction == "Roundabout")
replace junction01=1 if (junction == "Other")
replace junction01=1 if (junction == "Slip road")
replace junction01=1 if (junction == "T junction")
replace junction01=. if (junction=="Unknown")
replace junction01=1 if (junction == "Y junction")
label define junction01 0 "No junction" 1 "Junction"
label values junction01 junction01
tabulate junction junction01
```

* Creation, on basis of carriagewaytype, of an indicator variable describing the presence of a physical barrier (0) or not (1)

```
generate CarrPhysDivided=0 if (carriagewaytype=="Junction")
replace CarrPhysDivided=1 if (carriagewaytype=="One way traffic")
replace CarrPhysDivided=0 if (carriagewaytype=="Other")
replace CarrPhysDivided=1 if (carriagewaytype=="Two way physically
divided roadway with traffic barrier")
replace CarrPhysDivided=1 if (carriagewaytype=="Two way physically
divided roadway without traffic barrier")
replace CarrPhysDivided=0 if (carriagewaytype=="Two way traffic
divided by painted line")
replace CarrPhysDivided=0 if (carriagewaytype=="Two way traffic with
no division markings")
label variable CarrPhysDivided "Presence of a physical barrier"
label define CarrPhysDivided 0 "Physical separation" 1 "No physical
separation"
label values CarrPhysDivided CarrPhysDivided
```

* Creation of indicator variable for motorway:

```
generate Motorway01=0 if (motorway=="No")
replace Motorway01=1 if (motorway=="Yes")
label define Motorway01 0 "Not on motorway" 1 "Motorway"
label values Motorway01 Motorway
```



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

Analysing European in-depth data: Methodological framework and results

```
*** Transformation of TimeOfDay into indicator for day vs night:

* Selecting the first two values of each TimeOfDay observation (21:59
becomes 21 just as 21:15!)

generate HourString=substr(timeofday,1,2)
replace HourString= "." if (HourString== "Un")

*Making a numerical variable out of the new HourString variable

destring (HourString), gen (Hour)

*Indicator for Day/Night

recode Hour (6/21=1) (else=0), generate (Day)
tabulate Hour Day

* Recoding days into "Week" (0) and "Weekend" (1) categories:

generate Weekend=0 if
(accidentday=="Monday")|(accidentday=="Tuesday")|(accidentday=="Wednes
day")|(accidentday=="Thursday")|(accidentday=="Friday")
replace Weekend=1 if (accidentday=="Saturday")|(accidentday=="Sunday")
tabulate accidentday Weekend
label define Weekend 0 "Week" 1"Weekend"
label values Weekend Weekend

* Newweekend variable: Friday nights included in weekends: "Week" (0)
and "Weekend" (1) categories:

tabulate accidentday
generate Newweekend=0 if
(accidentday=="Monday")|(accidentday=="Tuesday")|(accidentday=="Wednes
day")|(accidentday=="Thursday")|((accidentday=="Friday")&(Day==0))
replace Newweekend=1 if
(accidentday=="Saturday")|(accidentday=="Sunday")|((accidentday=="Frid
ay")&(Day==1))
tabulate Newweekend accidentday
tabulate Weekend Newweekend
label define Newweekend 0 "Week" 1 "Weekend"
label values Newweekend Newweekend
label variable Newweekend "Weekend including Friday Night"

*Creation of indicator variable for light conditions: Daylight is
coded 0, all other conditions are coded 1

generate Light=1 if (lightconditions=="Darkness")
replace Light=1 if (lightconditions== "Darkness with artificial
light")
replace Light= 0 if (lightconditions=="Daylight")
replace Light=0 if (lightconditions== "Dazzling sunlight")
replace Light=1 if (lightconditions=="Other")
replace Light=1 if (lightconditions=="Partial light")
label define Light 0 "Daylight" 1 "Other than daylight"
label values Light Light
tabulate lightconditions Light

save "C:\DATA\D7.6\WP5 2d Data\WP5 data preparation V2\Emmanuelle WP5
data V25.dta"
```



4.4.2 Aggregation and disaggregation of vehicle and accident information

**Make of vehicle summary file, which will be used to copy the driver info to the passengers in the same vehicles later on:

```
keep if (UserClass==1)|(UserClass==3)

collapse (mean) AgeDriver=Age SeniorDriver=Senior AgeCatDriver=AgeCat
GenderDriver=Female AlcoholDriver=Alcohol
ImpairmentDriver=Impairment01 FamiliarDriver=IsFamiliar
VehNOAvoidMan=NOCrashAvoidMan VehNOBraking=NOBraking
vehNOSTeering=NOSteering vehMostHarmfulEvent=MostHarmfulEvent
VehHarmEventCat=HarmEventCat VehFirstEventIsEvent1=FirstEventIsEvent1
VehNumberOfEvents=numberofevents numofoccupants kerbweight2
AccidentID, by (VehicleID)

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Vehicle Summary
File.dta"

use "C:\DATA\D7.6\WP5 2d Data\WP5 Accident Summary File\Accident
Summary File.dta"

keep (AccCrashPartCat HeaviestVehicle LightestVehicle VictimSum
distVeh LossControl_sum ExMan_Sum AvoidMan_Sum SumAlcohol
SumImpairment SumFamiliar KSISum MultiKilled MultiKSI AccidentID
VehicleID)

sort AccidentID

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Accident Summary
Variables for Road-User filemerge3.dta"

use "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Vehicle Summary
File.dta"

keep (VehNOAvoidMan VehNOBraking vehNOSTeering VehNumberOfEvents
VehicleID AccidentID numofoccupants AgeDriver SeniorDriver
AgeCatDriver GenderDriver AlcoholDriver ImpairmentDriver
FamiliarDriver)

sort VehicleID

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Vehicle Summary
Variables for Road-User filemerge3.dta"

use "C:\DATA\D7.6\WP5 2d Data\WP5 data preparation V2\Emmanuelle WP5
data V26.dta", clear

sort AccidentID

merge AccidentID using "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User
File\Accident Summary Variables for Road-User filemerge3.dta"

drop _merge

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Road User File with
Accident summarymerge3.dta"
```

Analysing European in-depth data: Methodological framework and results

```
sort VehicleID
```

```
merge VehicleID using "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User  
File\Vehicle Summary Variables for Road-User filemerge3.dta"
```

```
save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Road-User  
Filemerge3.dta"
```



A 4.4.3 Deselection of cases with missing values

```
use "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Road-User
Filemerge3.dta"

keep if
(!missing(FirstEventCat))&(!missing(VehicleCat))&(!missing(Consequence
))&(!missing(Female))&(!missing(Weekend))&(!missing(VehNOAvoidMan))

tabulate FirstEventCat
tabulate Consequence
tabulate Female
tabulate AlcoholDriver
tabulate ImpairmentDriver
tabulate Weekend
tabulate NOCrashAvoidMan
tabulate CrashAvoidMan
tabulate VehNOAvoidMan
tabulate TrafficFlow3
tabulate FamiliarDriver
tabulate Day

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Road-
Usermissinglistwise File.dta"

log close
```

A4.4.4 Data Preparation for car-car analysis

```
use "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Road-
Usermissinglistwise File.dta"

keep if
(((VehicleCat==2)&(CrashPartCat==2))&((HeaviestVehicle>1)&(LightestVeh
icle<3)))

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Car-Car
File.dta"

collapse (count) carcarsum=UserID (mean) VehicleID (min)
MinWeight=kerbweight2 (max) MaxWeight=kerbweight2, by (AccidentID)

generate CarcarCompatibility=(MaxWeight-MinWeight)

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User
File\carcarsumvehicle.dta"

use "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Car-Car
File.dta", clear

drop _merge

sort AccidentID VehicleID

merge AccidentID using "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User
File\carcarsumvehicle.dta"

edit AccidentID VehicleID VictimSum carcarsum

generate CarCarSelect=0
replace CarCarSelect=1 if (carcarsum==VictimSum)
tabulate CarCarSelect

keep if (CarCarSelect==1)

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Car-Car
File.dta", replace
```

A4.4.5 Data preparation for Car-Opponent Analysis

```
use "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\WP5 Road-User
Filemerge3.dta"

drop _merge

sort UserID

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\WP5 Road-
User Filemerge3.dta"

insheet using "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User
File\Opponent\Interacted With.txt", clear

generate AccidentID=casenumbr

generate VehicleID=AccidentID*10+vehiclenunder

generate UserID=VehicleID*10+personnumber

sort UserID

merge UserID using "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User
File\Opponent\WP5 Road-User Filemerge3.dta"

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\WP5 Road-
User Filemerge3.dta", replace

***Creation of a numeric variable based on "vehicletype", so as to be
able to distinguish between Trucks/Van/Buses in the "Heavy" category
(VehicleCat1):

generate VehicleType2=1 if (vehicletype=="Car / MPV")
replace VehicleType2=2 if (vehicletype=="Van")
replace VehicleType2=3 if (vehicletype=="Bus / Minibus")
replace VehicleType2=4 if (vehicletype=="Truck")
replace VehicleType2=5 if (vehicletype=="Motorcycle / Moped")
replace VehicleType2=6 if (vehicletype=="Bicycle")
replace VehicleType2=7 if (vehicletype=="Shoe vehicle (pedestrian)")
replace VehicleType2=8 if ((vehicletype=="Other") |(vehicletype=="Train
/ Tram")|(vehicletype=="Agricultural vehicle"))

tabulate VehicleType2 VehicleCat1

mean kerbweight2, over(VehicleType2)
by VehicleType2, sort: summarize kerbweight2

***LGV are defined as vehicles weighing < 3500kgs, HGV as all vehicles
exceeding the 3500kg: therefore, buses and minibuses are merged with
the trucks:

recode VehicleType2 (1=1) (2=2) (3/4=3) (5=4) (6=5) (7=6) (8=7)

label define VehicleType2 1 "Car" 2 "Van" 3"Truck/Bus" 4
"Motorcycle/Moped" 5 "Bicycle" 6 "Pedestrian" 7 "Other"
label values VehicleType2 VehicleType2
```



Analysing European in-depth data: Methodological framework and results

```
***Assigning each road user an opponent:

recode vehiclenu mber (1=1) (else=0), generate (user1)
recode vehiclenu mber (2=1) (else=0), generate (user2)
recode vehiclenu mber (3=1) (else=0), generate (user3)
recode vehiclenu mber (4=1) (else=0), generate (user4)
recode vehiclenu mber (5=1) (else=0), generate (user5)
recode vehiclenu mber (6=1) (else=0), generate (user6)
recode vehiclenu mber (7=1) (else=0), generate (user7)

generate typeuser1=VehicleType2*user1
generate typeuser2=VehicleType2*user2
generate typeuser3=VehicleType2*user3
generate typeuser4=VehicleType2*user4
generate typeuser5=VehicleType2*user5
generate typeuser6=VehicleType2*user6
generate typeuser7=VehicleType2*user7

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\WP5 Road-
User Filemerge3.dta", replace

** One observation for which User3 corresponded to a parked van and
was not further included in the database-Typeuser was re-defined as
below so that the opponent and accident type are correctly identified
replace typeuser3=1 if (UserID==416611)

collapse (max)TypeUser1=typeuser1 TypeUser2=typeuser2
TypeUser3=typeuser3 TypeUser4=typeuser4 TypeUser5=typeuser5
TypeUser6=typeuser6 TypeUser7=typeuser7, by (AccidentID)

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\Accident-
level user type.dta"

use "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\WP5 Road-
User Filemerge3.dta", clear

sort AccidentID

drop _merge

merge AccidentID using "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User
File\Opponent\Accident-level user type.dta"

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\WP5 Road-
User Filemerge3.dta", replace

generate IW1=substr(interactedwith1,-1,1)
generate IW2=substr(interactedwith2,-1,1)
generate IW3=substr(interactedwith3,-1,1)
generate IW4=substr(interactedwith4,-1,1)
generate IW5=substr(interactedwith5,-1,1)
generate IW6=substr(interactedwith6,-1,1)

generate Opponent=IW6
replace Opponent=IW5 if ((IW5!="e")&(IW5!="n"))
replace Opponent=IW4 if ((IW4!="e")&(IW4!="n"))
replace Opponent=IW3 if ((IW3!="e")&(IW3!="n"))
```



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

Analysing European in-depth data: Methodological framework and results

```
replace Opponent=IW2 if ((IW2!="e")&(IW2!="n"))
replace Opponent=IW1 if ((IW1!="e")&(IW1!="n"))

generate TypeOpponent=TypeUser1 if (Opponent=="1")
replace TypeOpponent=TypeUser2 if (Opponent=="2")
replace TypeOpponent=TypeUser3 if (Opponent=="3")
replace TypeOpponent=TypeUser4 if (Opponent=="4")
replace TypeOpponent=TypeUser5 if (Opponent=="5")
replace TypeOpponent=TypeUser6 if (Opponent=="6")
replace TypeOpponent=TypeUser7 if (Opponent=="7")
replace TypeOpponent=8 if (Opponent=="e")
replace TypeOpponent=9 if (Opponent=="n")

label define TypeOpponent 1 "Car" 2 "Van" 3"Truck/Bus" 4
"Motorcycle/Moped" 5 "Bicycle" 6 "Pedestrian" 7 "Other" 8 "Single" 9
"Unknown"
label values TypeOpponent TypeOpponent

generate TypeAccident=1 if (VehicleType2==1) & (TypeOpponent==1)
replace TypeAccident=2 if (VehicleType2==1) & (TypeOpponent==2)
replace TypeAccident=3 if (VehicleType2==1) & (TypeOpponent==3)
replace TypeAccident=4 if (VehicleType2==1) & (TypeOpponent==4)
replace TypeAccident=5 if (VehicleType2==1) & (TypeOpponent==5)
replace TypeAccident=6 if (VehicleType2==1) & (TypeOpponent==6)
replace TypeAccident=7 if (VehicleType2==1) & (TypeOpponent==8)
replace TypeAccident=. if (VehicleType2==1) & (TypeOpponent==9)

replace TypeAccident=8 if (VehicleType2==2) & (TypeOpponent==1)
replace TypeAccident=9 if (VehicleType2==2) & (TypeOpponent==2)
replace TypeAccident=10 if (VehicleType2==2) & (TypeOpponent==3)
replace TypeAccident=11 if (VehicleType2==2) & (TypeOpponent==4)
replace TypeAccident=12 if (VehicleType2==2) & (TypeOpponent==5)
replace TypeAccident=13 if (VehicleType2==2) & (TypeOpponent==6)
replace TypeAccident=14 if (VehicleType2==2) & (TypeOpponent==8)
replace TypeAccident=. if (VehicleType2==2) & (TypeOpponent==9)

replace TypeAccident=15 if (VehicleType2==3) & (TypeOpponent==1)
replace TypeAccident=16 if (VehicleType2==3) & (TypeOpponent==2)
replace TypeAccident=17 if (VehicleType2==3) & (TypeOpponent==3)
replace TypeAccident=18 if (VehicleType2==3) & (TypeOpponent==4)
replace TypeAccident=19 if (VehicleType2==3) & (TypeOpponent==5)
replace TypeAccident=20 if (VehicleType2==3) & (TypeOpponent==6)
replace TypeAccident=21 if (VehicleType2==3) & (TypeOpponent==8)
replace TypeAccident=. if (VehicleType2==3) & (TypeOpponent==9)

replace TypeAccident=22 if (VehicleType2==4) & (TypeOpponent==1)
replace TypeAccident=23 if (VehicleType2==4) & (TypeOpponent==2)
replace TypeAccident=24 if (VehicleType2==4) & (TypeOpponent==3)
replace TypeAccident=25 if (VehicleType2==4) & (TypeOpponent==4)
replace TypeAccident=26 if (VehicleType2==4) & (TypeOpponent==5)
replace TypeAccident=27 if (VehicleType2==4) & (TypeOpponent==6)
replace TypeAccident=28 if (VehicleType2==4) & (TypeOpponent==8)
replace TypeAccident=. if (VehicleType2==4) & (TypeOpponent==9)

replace TypeAccident=29 if (VehicleType2==5) & (TypeOpponent==1)
replace TypeAccident=30 if (VehicleType2==5) & (TypeOpponent==2)
replace TypeAccident=31 if (VehicleType2==5) & (TypeOpponent==3)
replace TypeAccident=32 if (VehicleType2==5) & (TypeOpponent==4)
replace TypeAccident=33 if (VehicleType2==5) & (TypeOpponent==5)
replace TypeAccident=34 if (VehicleType2==5) & (TypeOpponent==6)
```

Analysing European in-depth data: Methodological framework and results

```
replace TypeAccident=35 if (VehicleType2==5) & (TypeOpponent==8)
replace TypeAccident=. if (VehicleType2==5) & (TypeOpponent==9)

replace TypeAccident=36 if (VehicleType2==6) & (TypeOpponent==1)
replace TypeAccident=37 if (VehicleType2==6) & (TypeOpponent==2)
replace TypeAccident=38 if (VehicleType2==6) & (TypeOpponent==3)
replace TypeAccident=39 if (VehicleType2==6) & (TypeOpponent==4)
replace TypeAccident=40 if (VehicleType2==6) & (TypeOpponent==5)
replace TypeAccident=41 if (VehicleType2==6) & (TypeOpponent==6)
replace TypeAccident=42 if (VehicleType2==6) & (TypeOpponent==8)
replace TypeAccident=. if (VehicleType2==6) & (TypeOpponent==9)
replace TypeAccident=43 if (VehicleType2==7) | (TypeOpponent==7)

label define TypeAccident 1 "Car-Car" 2 "Car-Van" 3 "Car-Heavy" 4
"Car-2Wheels" 5 "Car-Bicycle" 6 "Car-Pedestrian" 7 "Single Car" 8
"Van-Car" 9 "Van-Van" 10 "van-Heavy" 11 "Van-2-Wheels" 12 "Van-
Bicycle" 13 "Van-Pedestrian" 14 "Single Van" 15 "Heavy-Car" 16 "Heavy-
Van" 17 "Heavy-Heavy" 18 "Heavy-2Wheels" 19 "Heavy-Bicycle" 20 "Heavy-
Pedestrian" 21 "Single Heavy" 22 "2Wheels-Car" 23 "2Wheels-Van" 24
"2Wheels-Heavy" 25 "2Wheels-2Wheels" 26 "2Wheels-Bicycle" 27 "2Wheels-
Pedestrian" 28 "Single 2Wheels" 29 "Bicycle-Car" 30 "Bicycle-Van" 31
"Bicycle-Heavy" 32 "Bicycle-2Wheels" 33 "Bicycle-Bicycle" 34 "Bicycle-
Pedestrian" 35 "Single Bicycle" 36 "Pedestrian-Car" 37 "Pedestrian-
Van" 38 "Pedestrian-Heavy" 39 "Pedestrian-2Wheels" 40 "Pedestrian-
Bicycle" 41 "Pedestrian-Pedestrian" 42 "Single Pedestrian" 43 "Other"
label values TypeAccident TypeAccident

***Errors in database:

*Accident 4062, Vehicle 40624, User 406241: the number of occupants
for this observation is 99 in the database, based on accident
description, this value is replaced by "1":
replace numofoccupants=1 if ((AccidentID==4062)&(VehicleID==40624))

*All Single Pedestrian accidents are in fact accidents involving a
pedestrian and a vehicle classified as "other" (agricultural vehicle,
tram...), these are all recoded as "Pedestrian-Other"
replace TypeAccident=43 if
(AccidentID==6026)|(AccidentID==2396)|(AccidentID==2239)

*One Single Pedestrian accident is an uncorrectly coded Pedestrian-Car
accident:
replace TypeAccident=36 if (AccidentID==2195)

***Ambiguous accidents, accidents involving parked vehicles (cf
\\bivv03\Users Files$\Emmanuelle\My Documents\Projets\Safety Net -
WP7\D7.9\Ambiguous accidents for opponent analysis.doc)

replace TypeAccident=21 if (AccidentID==2066)&(vehiclenu==1)
drop if (AccidentID==2066)&(vehiclenu==2)

replace TypeAccident=13 if (AccidentID==2265)&(vehiclenu==3)
replace TypeAccident=37 if (AccidentID==2265)&(vehiclenu==2)

replace TypeAccident=3 if (AccidentID==2323)

replace TypeAccident=7 if (AccidentID==2387)&(vehiclenu==1)
replace TypeAccident=36 if
(AccidentID==2387)&(vehiclenu==2)|(vehiclenu==3)
drop if (AccidentID==2387)&(vehiclenu==4)
```



Analysing European in-depth data: Methodological framework and results

```
drop if (AccidentID==2444)

replace TypeAccident=7 if (AccidentID==2433)&(vehiclenumber==1)
drop if (AccidentID==2433)&(vehiclenumber==2)

replace TypeAccident=38 if (AccidentID==2473)&(vehiclenumber==2)
replace TypeAccident=17 if (AccidentID==2473)&(vehiclenumber==3)
drop if (AccidentID==2473)&(vehiclenumber==1)

replace TypeAccident=37 if
(AccidentID==4008)&(vehiclenumber==1)&(UserID==400811)

drop if ((AccidentID==4043)&(vehiclenumber==1))
replace TypeAccident=35 if (AccidentID==4043)&(vehiclenumber==2)

drop if (AccidentID==4062)

drop if (AccidentID==7048)

tabulate TypeAccident
tabulate TypeAccident Killed, row

edit eventdetail1 interactedwith1 IW1 eventdetail2 interactedwith2 IW2
eventdetail3 interactedwith3 IW3 eventdetail4 interactedwith4 IW4
eventdetail5 interactedwith5 IW5 eventdetail6 interactedwith6 IW6 if
(eventdetail1=="Parked vehicle (not travelling)")|
(eventdetail2=="Parked vehicle (not travelling)")|
(eventdetail3=="Parked vehicle (not travelling)")|
(eventdetail4=="Parked vehicle (not travelling)")|
(eventdetail5=="Parked vehicle (not travelling)")|
(eventdetail6=="Parked vehicle (not travelling)")

save "C:\DATA\D7.6\WP5 2d Data\WP5 Road-User File\Opponent\WP5 Road-
User Filemerge3.dta", replace
```