Deliverable D7.6

Analysis of the Fatal Accident Investigation Database

Martensen, H. (Ed.) (2007) Analysis of Fatal Accident Investigation Database. Deliverable D7.6 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
Loughborough, LE11 3UZ

Organisation name of lead contractor for this deliverable:
Belgian Road Safety Institute (IBSR)

Due Date of Deliverable: 31/07/2007

Submission Date:

Editor: H. Martensen (IBSR)

Contributing Authors: E. Dupont (IBSR), D. Katsochis (NTUA), H. Martensen (IBSR), E. Papadimitriou (NTUA), G. Yannis (NTUA)

Project Start Date: 1st May 2004 **Duration:** 4 years

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

Transport

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

Table of Contents

EXECUTIVE SUMMARY.....5

Table of Contents 2

CHAPTER 1 - INTRODUCTION 7

1.1 Pitfalls and solutions 8
1.1.1 The absence of exposure data 8
1.1.2 Accident Size Bias..... 9
1.1.3 Comparability of risks 10
1.1.4 Confounded Variables 11

1.2 Overview 14

CHAPTER 2 - DESCRIPTION OF THE IN-DEPTH ACCIDENT INVESTIGATION DATABASE..... 17

2.1 Structure of the database 17

2.2 Description of WP5 data with respect to accident severity 21
2.2.1 User-related variables: gender, age, seat belt use 22
2.2.2 Vehicle-related variables: vehicle type, number of vehicles 23
2.2.3 Road-related variables: speed limit, carriageway type, junction type, lighting. 24

2.3 Description of WP5 data with respect to fatal accident occurence 26

2.4 Summary and motivations for analysis..... 28

CHAPTER 3 - MODELLING ACCIDENT SIZE 41

3.1 The research question 41

3.2 The analytical problems..... 41
3.2.1 Dependent Variable 1: Number of crash participants 42
3.2.2 Dependent variable 2: number of victims..... 43
3.2.3 Selection criteria..... 43
3.2.4 The model 44

3.3 Results 45
3.3.1 Variables in the model..... 48
3.3.2 Variables not included in the model..... 50

3.4 Conclusion..... 53

Appendix 3A: Model implementation..... 55

Appendix 3B: Variables not in the final model 55

CHAPTER 4 - MODELLING THE FATALITY RISK FOR ALL ACCIDENTS .. 57

4.1 The research questions 57

4.2 The analytical problem..... 57

4.3 Towards an analytical solution 63

4.4 Discussion 67

CHAPTER 5 - THE FATALITY RISK IN CAR-CAR ACCIDENTS..... 69

5.1 The research question 69

5.2 The analytical problem..... 69

5.3 Results 74
5.3.1 The empty model..... 74
5.3.2 The final model 75

5.4 Conclusion..... 86

Appendix 5A: Variables not in the final model 88

CHAPTER 6 - MODELLING SEVERITY REPORTING RELIABILITY THROUGH IN-DEPTH DATA 91

6.1 The research question 91

6.2 The analytical problem..... 92

6.3 The analytical solution..... 94
6.3.1 Binomial modelling 96
6.3.2 Multinomial modelling 100

6.4 Conclusions – Future steps 103

APPENDIX 6A: Comparative injury severity distributions for other countries except Italy..... 106

CHAPTER 7 - CONCLUSION..... 109

7.1 Summary 109

7.2 A word of caution 111

7.3 Further questions and recommendations..... 112

7.4 In a nutshell 113

References 114

Executive Summary

The present deliverable is dedicated to the analysis of a preliminary version (December 2006) of the Fatal Accident Investigation Database (WP5, SafetyNet). As the database contains fatal accidents only, the analyses are focused on accident severity. In particular, the accident-size, the fatality risk and the reliability of the injury reporting were modelled.

The *accident size* was analysed in terms of two variables: Single vehicle accidents were compared to multi-vehicle accidents and accidents with one fatality were compared to those with more than one occupant killed or seriously injured. The results indicated that single vehicle accidents involve young, male, impaired drivers more often than older, female, and unimpaired drivers. The proportion of women among the occupants was higher in single vehicle accidents as compared to multi-vehicle accidents and multi-vehicle accidents took place after the execution of a manoeuvre more often than single vehicle accidents. Only two factors characterised accidents with more than one killed or seriously injured occupant (KSI): a higher average number of passengers per vehicle, and the fact that the accident took place in weekends rather than during the week.

The *fatality risk* is the probability to die, *given that one is involved in a fatal accident*. This was analysed for all accidents in the database and, in a second model, for car-car accidents only. The analysis of the fatality risk for all accidents indicated that the fatality risk was higher for vulnerable road users, as compared to occupants of motor vehicles and for seniors (65 and above), as compared to all other age groups. It also showed that drivers of the vehicle that contained the fatality had tried to avoid the accident by braking less often than the drivers of the vehicle that did not contain the fatality. Finally, it was shown that the proportion of fatalities was lower in accidents taking place on roads with a physical divide of the carriageway as compared to accidents occurring on other road types.

The analysis of the fatality risk in car-car accidents ensured maximal comparability between fatalities and survivors. In this analysis, the results from the global analysis were confirmed. Again, it was shown that as compared to occupants of vehicles whose drivers did not brake, occupants of vehicles whose drivers did brake had a much higher chance of surviving. Accidents on motorways were shown to exhibit a lower proportion of fatalities, and for those vehicles that contained the fatality more events were described than for others. Additionally, it was shown that in severe accidents, newer cars offer the occupants more protection than older cars; that side impacts are much more dangerous than frontal accidents and that there was an interaction between these two variables: While the protection from front impacts increased dramatically for newer cars, there is no significant increase in the protection from side impacts.

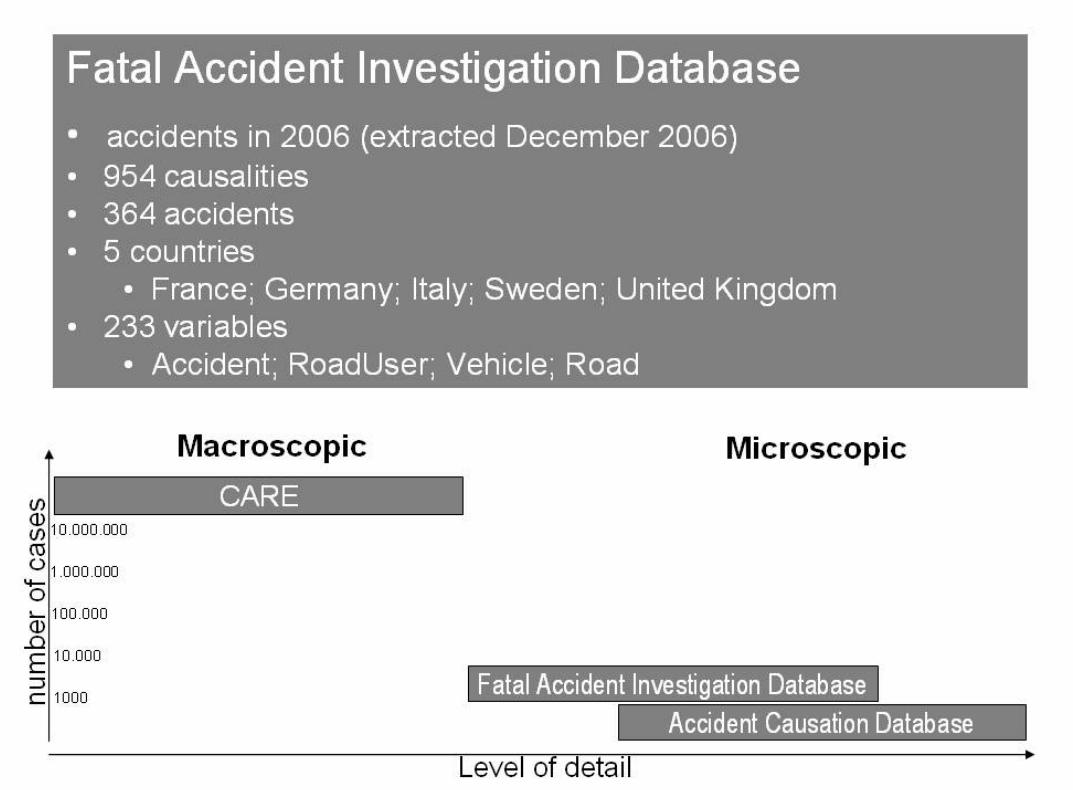
The Fatal Accident Investigation Database contains two different records of injury severity: The original police record and one revised by the SafetyNet team. These two records were not always in agreement, indicating a substantial number of reporting errors. These errors concerned predominantly victims who had initially been classified as “seriously injured”. An important finding was that differences between police and SafetyNet records were much more frequent in Italy than in all other countries. A systematic exploration of factors that predict reporting differences indicated that - for their largest part - the errors could not be related to characteristics of the accidents or of the victim. This suggests that they appear at random and are probably due to insufficient information for the recording officers. However, there were also a number of systematic biases identified. The exact factors differed for Italy and for the other countries (that were analysed jointly), but the two tendencies in biases that could be identified were: 1) For persons who could in some way be assumed to be weaker or less protected than others (children, seniors, women, vulnerable road users) the injury record changed during the revision more often than for others. 2) Complex accidents facilitate misreporting.

Chapter 1 - Introduction

(Heike Martensen and Emmanuelle Dupont, IBSR)

Accident data collection is often described in terms of macroscopic data versus microscopic 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 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 in disaggregate 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 to collect more detailed information has been served by a number of projects collecting microscopic accident data in which fewer accidents are described in greater amount of detail. In the present deliverable, a preliminary form of the *Fatal Accident Investigation Database*, collected in Workpackage 5 of SafetyNet is analysed. Figure 1.1, indicates some characteristic numbers for the data analysed and situates it at an intermediate level between macro- and microscopic data.

Figure 1.1 The data analysed in the context of macro- and microscopic data



Adapted from Morris (2007)

The database includes fatal accidents exclusively. In addition to the macroscopic accident data collected in those countries, this in-depth database includes a number of variables for which information is seldom available or reliable in national databases. Unlike macroscopic data, the in-depth data are available in a disaggregate format, as the information is recorded at the single-road-user level. More specifically, in the WP5 database, the chain of events of each accident is identified and described in detail, and important variables related to the road user are 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.), pedestrian facilities etc. In Chapter 2, the database and its structure is described in more detail.

1.1 Pitfalls and solutions

Because they allow for detailed and disaggregate analysis, the Fatal Accident Investigation Database can provide information on several key aspects and parameters of the accident process, which cannot be easily tackled by means of macroscopic data. In analyzing the data however, several methodological issues have to be considered. In the following, a number of problems that can occur in the analysis will be discussed and the solutions that have been applied in the present deliverable will be presented.

1.1.1 The absence of exposure data

In the fatal accident database,

- 59% of the cars was damaged most in the front
- 15% of the drivers was impaired by alcohol, drugs, or fatigue
- 58% of the drivers did not conduct an avoidance manoeuvre

These are examples for results that can be found in the Fatal Accident Investigation Database. It is, however, difficult to interpret these percentages without additional knowledge. The percentage of cars damaged in the front, for example, would become meaningful if it could be compared to the same percentage in non-fatal accidents. To say anything about the change of risk due to impairment of the drivers, one would need the percentage of impairments in the driving population. And to decide whether 58% of drivers not conducting an avoidance manoeuvre is a lot or a little, one would need to know how drivers reacted in successfully mastered crisis situations.

At this moment, it is impossible to identify differences between fatal accidents and more favourable situations (like non-fatal accidents, normal driving situations, or successfully mastered crisis situations). Consequently, it is not possible – on a statistical basis – to draw conclusions about the causation of accidents.

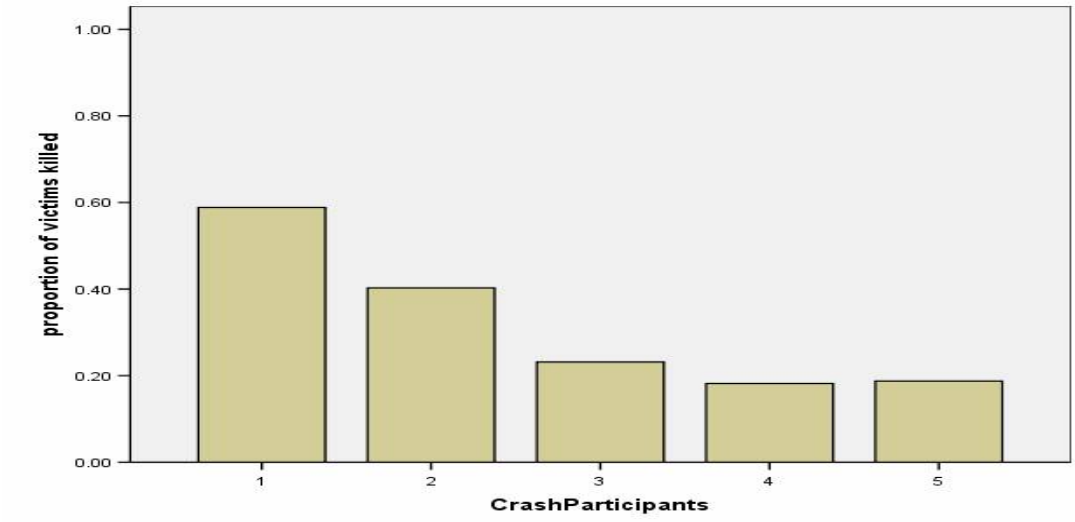
We can, however, identify factors differentiating between the occupants who were killed and those who survived. This is done by modelling the *fatality risk*, the number of people killed relative to the overall number of people involved (i.e. the proportion of people killed). This risk is determined for several variables (e.g. age; area of most damage; road type), or more specifically for their values (e.g. 65+; frontal damage; motorways). A high proportion of people killed indicate that the category in question bears a higher fatality risk than the categories it is compared to.

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 risk that road-users in general run to die in an accident.

1.1.2 Accident Size Bias

In Figure 1.2, the proportion of people killed (i.e. the fatality risk) is given for accidents with different numbers of crash participants. It can be noted that the fatality risk decreases with the number of participants. This does not mean, however, that larger accidents are less dangerous. What this figure reflects is the fact the vast majority of the fatal accidents contains exactly one fatality (for details see Chapters 2 and 3). As a consequence, the fatality risk is strongly related to the overall number of victims involved, and therefore to the size of the accident: *The larger the number of crash participants, the larger the number of persons involved and the smaller the fatality risk.* In the following, we will call this relation the *accident-size bias*.

Figure 1.2 Proportion of people killed per numbers of crash participants



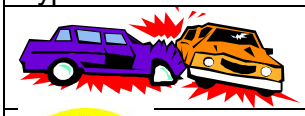
The possibility of an accident-size bias in the analysis of the fatality risk in fatal accidents has been recognized earlier (e.g., Evans, 2001; Mabe 2006) and is usually countered by focussing on one person category only (e.g., the driver) in one type of accident only (e.g., multi-car accidents). As this approach strongly reduces the number of cases that can be analysed, it is suitable for very large databases but problematic for smaller databases as the one analysed here. Another problem with this approach is that one looses all information about the cases that have been excluded (e.g., single vehicle accidents) which could be very interesting in itself. In the present deliverable, this problem was addressed in three ways: (a) by modelling the number of crash participants itself (Chapter 3), (b) by modelling the fatality risk in all accidents while statistically correcting for the accident size bias (Chapter 4), and (c) by modelling the fatality risk in accidents involving two participants (Chapter 5).

1.1.3 Comparability of risks

When comparing fatalities and survivors in fatal accidents, one wants to identify factors that can make the difference between surviving and dying in a severe accident. In order to draw this conclusion, the risk that the road users were running in the first place should be comparable. However, the simple comparison of different accident types in Table 1.1 shows that the risk a car occupant is running in a fatal accident crucially depends on the type of accident he or she is in and the type of opponent that has been encountered.

In the analysis of fatality risk, these baseline differences can be modelled directly, by taking up the type of accident into the model of analysis. This approach has been taken in the analysis of fatality risk presented in Chapter 4, where the type of accident (single vehicle vs. multi-vehicle) and the type of road-user (vulnerable vs. occupant of motor vehicle) have been included into the model of fatality risk. The advantage of this method is that all victims in the database can be included into the analysis.

Table 1.1 Distribution of risk in different accident types (fatal accidents)

Type of accident	Group of road-users	Distribution of risk
	Occupants Car 1 Occupants Car 2	50% 50%
	Occupants Car	100%
	Occupants Car Vulnerable road user	0% 100%

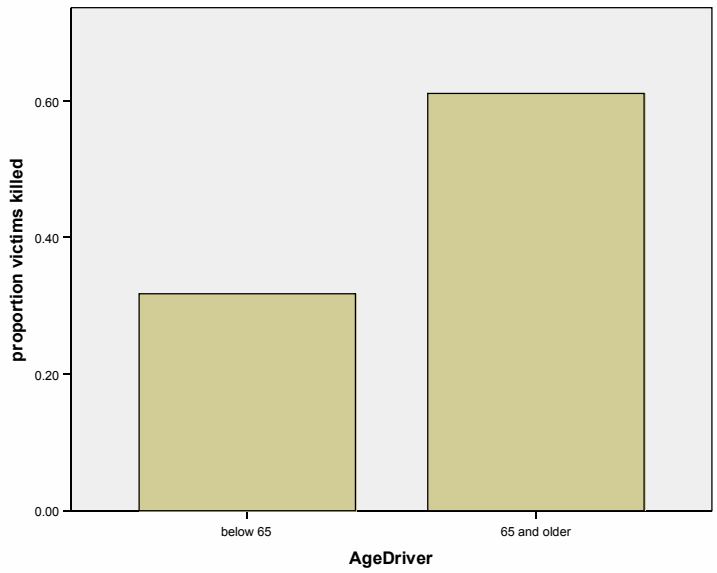
In Chapter 5, only victims of car-car accidents are analysed, assuring a maximal comparability of the victims. This analysis allowed a more detailed analysis of the factors that determine the fatality risk of car occupants. The disadvantage of this analysis is that it is restricted to a small number of victims. The advantage is the good control it allows on factors that affect the baseline risk and the possibility to include variables that are specific to cars and their occupants (e.g., seat-belt use or the area of main damage).

1.1.4 Confounded Variables

The problem of confounded variables is not specific to the analysis of accident data. Rather, it comes into play whenever observational data are analysed. It is discussed here, because it plays an important role in how the statistical models in Chapter 3, 4, 5, and 6 have been constructed. In the following an example from the analyses of car-car accidents (Chapter 5) is presented.

Figure 1.3 shows that the occupants of cars driven by senior drivers (65 and older) have a higher fatality risk when involved in fatal accidents than the occupants of cars with younger drivers. From such a graph one might conclude that senior drivers – when involved in a fatal accident – have slower reflexes, which make them less able to react in a way that might protect the occupants in their cars.

Figure 1.3 Fatality risk by driver's age



However from Figure 1.4, it can be seen that the age of the driver is highly related to that of the victims. Partly, this is because driver and victim are often identical (i.e. the driver is the victim). Even if this is not the case, though, there is a strong relation, because older passengers tend to travel with older drivers

and vice versa. Indeed, in Figure 1.5 we can see that differentiating the fatality risk between senior *victims* (65 and older) and younger ones, has more or less the same result as the differentiating it between senior *drivers* and younger ones.

In this example, the age of the driver and the age of the victim are *confounded* because they are highly related. The fact that they have the same effect on the fatality risk is therefore difficult to interpret: Do older people have a higher risk of dying when involved in a severe accident because of their own age (e.g. because their bones are more vulnerable to breaking and they heal more slowly) or because of the age of the driver they typically travel with?

Figure 1.4 Relation between age of driver and age of victim

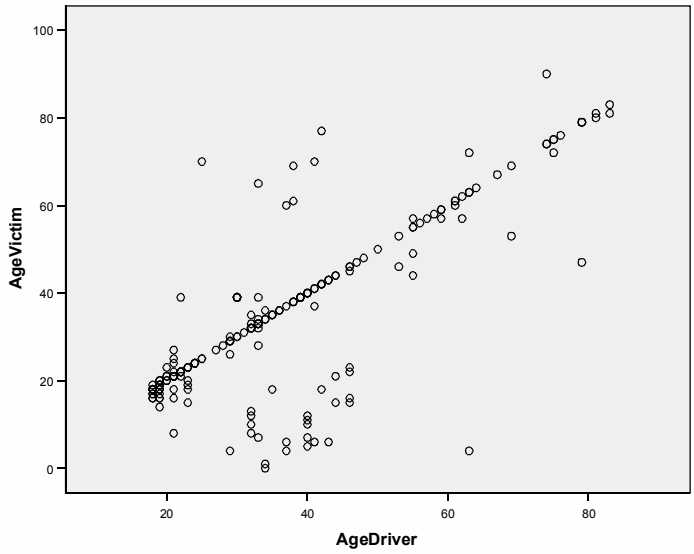
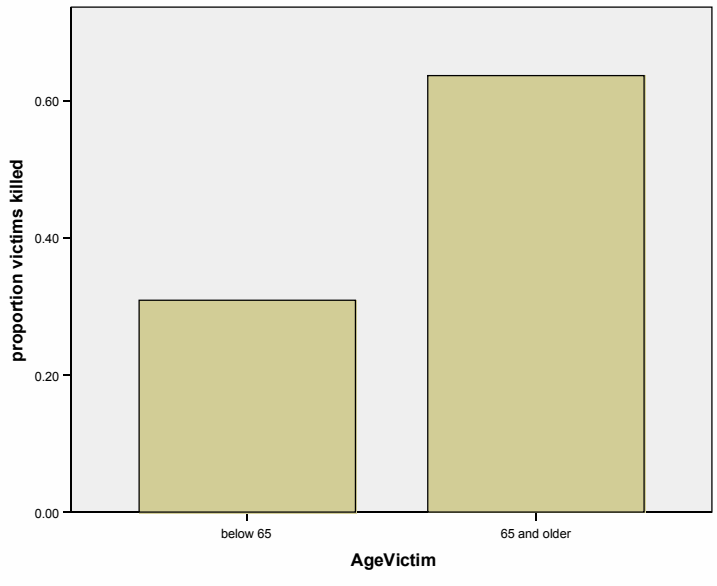


Figure 1.5 Fatality risk by age of victim



Simultaneously modelling the effect of both variables on the fatality risk causes extra problems, but can also give an indication of which of them is the better predictor of the fatality risk.

Table 1.2 shows three models (for the details of implementing such a model see Chapter 4 and 5) in which the fatality risk is predicted either by a variable indicating whether the driver was above 65 or not (SeniorDriver), or by a variable indicating whether the victim was above 65 or not (SeniorVictim), or by both.

Table 1.2 Predictors of fatality risk in car-car accidents

Predictors	Model 1		Model 2		Model 3	
	β (SE)	P	β (SE)	P	β (SE)	p
SeniorDriver	1.214 (.507)	.017			0.138 (.825)	.868
SeniorVictim			1.363 (.470)	.004	1.265 (.751)	.092

In the first two models, one can see that SeniorDriver (Model 1) and SeniorVictim (Model 2) have indeed very similar effects on the fatality risk. Both have a significant positive coefficient (β), indicating that the fatality risk is higher for senior drivers/victims. When entering both variables into the model together (Model 3), the estimated effects change dramatically. The size of the estimated coefficients (β) is lower and the size of the standard errors (SE) of the coefficients increases, indicating a lower reliability of the estimated coefficients. Accordingly, both predictors are not significant anymore. When two predictors in a multiple regression model are highly related, this is called *collinearity*, and the consequences shown in the present example are typical.

To understand why collinear predictors are not significant anymore, one must remember that in a multiple regression analysis it is tested for each predictor whether it explains a significant proportion of the variance in the dependent variable *over and above* what the other predictors in the model explain. The effect of each predictor is therefore corrected for that of all other predictors. When two predictors are too similar, they do not contribute anything that the other does not contribute as well and consequently none of them is significant anymore.

When two predictors show complete overlap (i.e. they show perfect collinearity), there are no cases where one shows a different value than the other. In such cases, it is impossible to say which of the two variables is actually responsible for the effect they both show. When entering them into the model jointly, they will both become (absolutely) non-significant.

In the present example, this is not quite the case. Although both predictors (SeniorDriver and SeniorVictim) become non-significant, the effect of entering them into the model jointly is different for each of the variables. While the coefficient for SeniorDriver is shrunk to almost one tenth of its original size, the size of the coefficient for SeniorVictim remains more or less the same. Accordingly, SeniorDriver becomes completely non-significant, while

SeniorVictim remains marginally significant (p<.10). The model containing both variables reflects only those cases where the two variables do not agree (i.e. a senior driver with a non-senior victim or vice versa). It shows that the original effect of SeniorDriver was entirely due to the correlation with SeniorVictim. When considering only those cases where the age of the driver and the age of the victim are dissociated, nothing of the SeniorDriver effect is left over. In contrast, SeniorVictim still shows the same effect on the fatality risk, which has become non-significant only because the standard error is doubled. SeniorVictim turned out to be the true predictor of the fatality risk, while SeniorDriver has been piggy-back riding on SeniorVictim due to their strong relationship. [1] To conclude, the probability to die when involved in a fatal accident is determined by the age of the victim rather than by the age of the driver.

To summarize on the problem of confounded variables, it is important to model variables simultaneously to avoid attributing the effect of one predictor to another one that is related to it. Confounded predictors require a careful investigation of the results from different combinations of these predictors.

1.2 Overview

The methodological considerations discussed in the previous section have guided the model building described in the following chapters. In Chapter 2, the database is described in its original form and a few descriptive statistics are given that formed the basis for the questions addressed in the modelling chapters.

In Chapter 3, accident size is modelled, identifying the factors that differentiate between multi-vehicle accidents and single vehicle accidents. Additionally the absolute number of victims that are killed or seriously injured in an accident is analysed.

In Chapter 4, a global analysis of the fatality risk for persons involved in fatal accidents is presented. Variables that describe the accident type (e.g. the number of crash participants, the type of crash participants) are included into the analysis.

In Chapter 5, the fatality risk in car-car accidents is analysed. While factors that affect the fatality risk in a general way (type road users involved, number of crash participants) are controlled for, this analysis zooms in on the factors that differentiate between those car occupants who have died and those who survived.

[1] See Chapter 5, for another variable confound involving SeniorVictim, the area of most damage in the vehicle.

Finally, in Chapter 6, the reliability of the injury reporting is analysed. The in-depth database contains two variables that indicate the consequence of an accident, the original police report and a revised report given by the SafetyNet investigator. The reliability analysis identifies factors that differentiate those cases where the investigators changed the original report and those where the original report was considered accurate.

==========
Page left blank intentionally
==========

Chapter 2 - Description of the in-depth accident investigation database

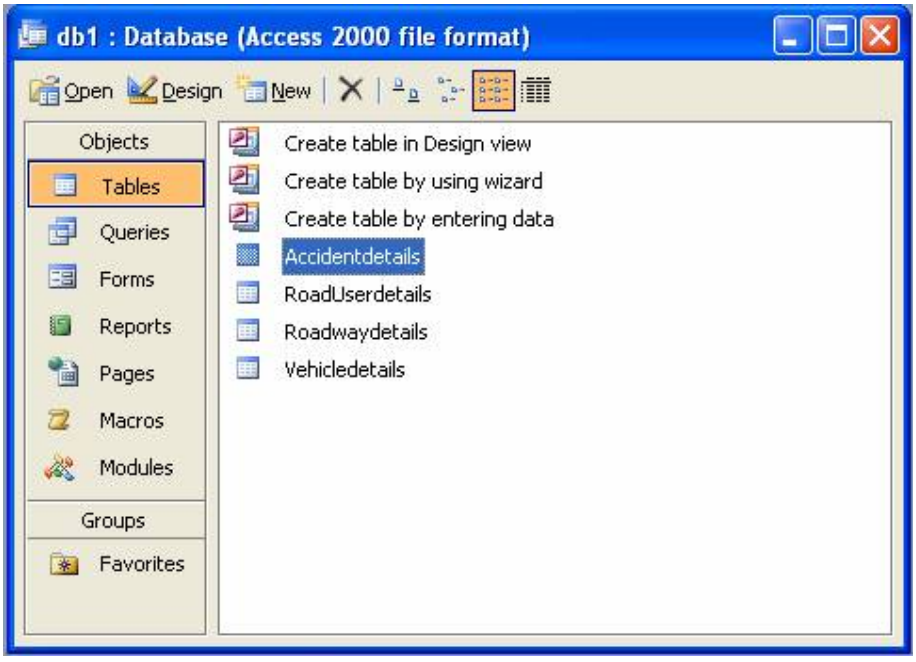
George Yannis, Dimitris Katsochis and Eleonora Papadimitriou, NTUA

Although limited to fatal accidents, the in-depth 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 proper variables allowing for the formation of models. Keeping in mind the major questions posed within the preliminary analysis of the data (absence of exposure data, accident size bias, interactions between variables), the objective of this section is to create an overall picture of the structure and main contents of the WP5 database and track those parameters that appear meaningful for the purposes of the analysis.

2.1 Structure of the database

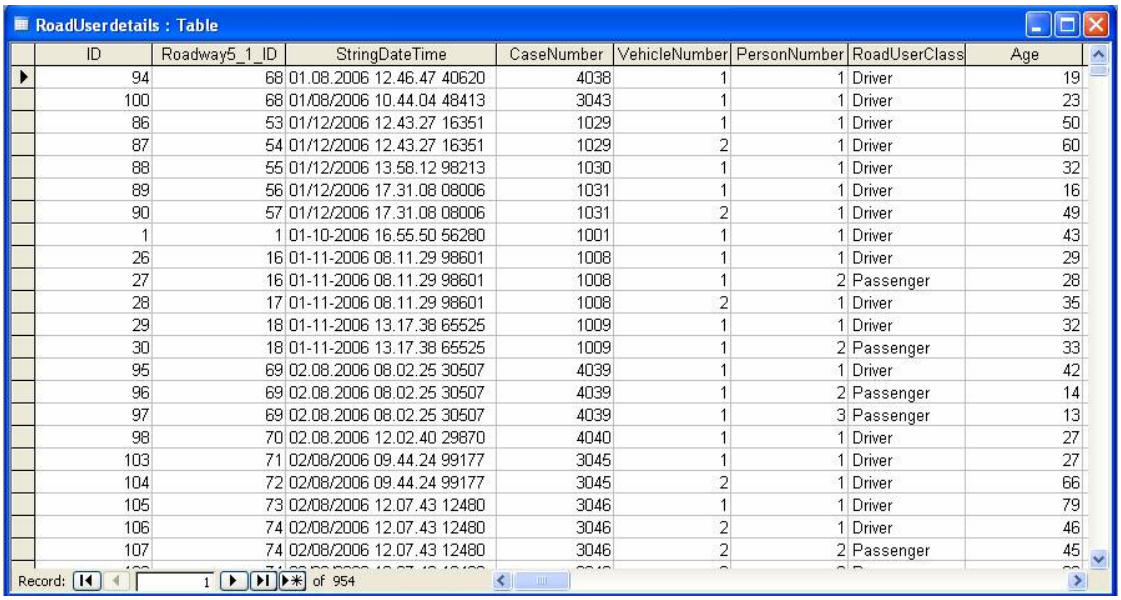
As most accident databases, the WP5 in-depth database consists of four separate yet linked Tables: Accident details, Road details, Vehicle details and Road user details (Figure 2.1), in MS Access format. These Tables are linked by means of the following "keys" (common variables): Date/time, Vehicle Number, Person Number.

Figure 2.1 Structure of the database



For instance, the Road User details Table includes the date/time and vehicle number keys, allowing to link the information with the accident, road and vehicle Tables (see Figure 2.2).

Figure 2.2 View of the Road User details Table



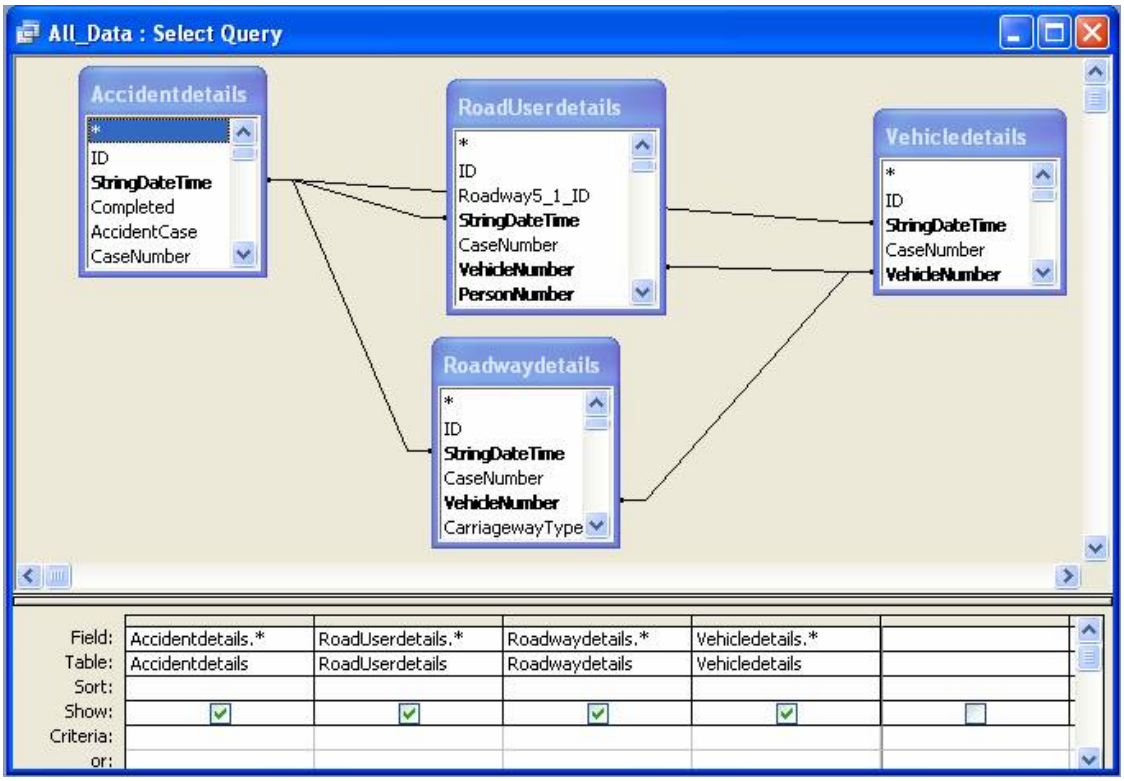
RoadUserdetails : Table

ID	Roadway5_1_ID	StringDateTime	CaseNumber	VehicleNumber	PersonNumber	RoadUserClass	Age
94	68	01.08.2006 12.46.47 40620	4038	1	1	Driver	19
100	68	01/08/2006 10.44.04 48413	3043	1	1	Driver	23
86	53	01/12/2006 12.43.27 16351	1029	1	1	Driver	50
87	54	01/12/2006 12.43.27 16351	1029	2	1	Driver	60
88	55	01/12/2006 13.58.12 98213	1030	1	1	Driver	32
89	56	01/12/2006 17.31.08 08006	1031	1	1	Driver	16
90	57	01/12/2006 17.31.08 08006	1031	2	1	Driver	49
1	1	01-10-2006 16.55.50 56280	1001	1	1	Driver	43
26	16	01-11-2006 08.11.29 98601	1008	1	1	Driver	29
27	16	01-11-2006 08.11.29 98601	1008	1	2	Passenger	28
28	17	01-11-2006 08.11.29 98601	1008	2	1	Driver	35
29	18	01-11-2006 13.17.38 65525	1009	1	1	Driver	32
30	18	01-11-2006 13.17.38 65525	1009	1	2	Passenger	33
95	69	02.08.2006 08.02.25 30507	4039	1	1	Driver	42
96	69	02.08.2006 08.02.25 30507	4039	1	2	Passenger	14
97	69	02.08.2006 08.02.25 30507	4039	1	3	Passenger	13
98	70	02.08.2006 12.02.40 29870	4040	1	1	Driver	27
103	71	02/08/2006 09.44.24 99177	3045	1	1	Driver	27
104	72	02/08/2006 09.44.24 99177	3045	2	1	Driver	66
105	73	02/08/2006 12.07.43 12480	3046	1	1	Driver	79
106	74	02/08/2006 12.07.43 12480	3046	2	1	Driver	46
107	74	02/08/2006 12.07.43 12480	3046	2	2	Passenger	45

Record: 1 of 954

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 2.3).

Figure 2.3 Linking and matching accident, road, vehicle and user information in the 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 2.1).

Table 2.1 Overview of accident, road, vehicle and user variables

Accidentdetails_ID	**RoadUserdetails_ID (cont.)**	**Roadwaydetails_ID (cont.)**	**Vehicledetails_ID (cont.)**
Accidentdetails_StringDateTime	CommentsAirbag	Junction	WasVehicleTowing
Completed	PoliceInjuryseverity	LocalArea	EnginePower
AccidentCase	SafetyNetMedicalOutcome	VerticalAlignment	YearOfManufacture
Accidentdetails_CaseNumber	BodyRegionMostInjured	HorizontalAlignment	KerbWeight
CentreName	Comments5	ConstrMaintZone	NumberOfAxles
AccidentDate	Ejection	RoadwaySurfaceType	SpecificSpeedLimit
AccidentDay	EntrapmentExtrication	PedestrianFacility	GeneralComments
TimeOfDay	TakenToHospital	CycleFacilities	AreDefects
HitAndRun	HospDuration	RoadConditions	Vehicledetails_Comments
AnimalInvolved	DiedAtScene	LightConditions	PassedInspection
AccidentTypeClass	NDaysUntilDeath	TrafficFlow	DriverManoeuvre
FirstEvent	Comments5b	WeatherConditions	TransientFactors
RelatedFactors	SuspicionAlcohol	StrongWinds	VehicleHeading
CrashParticipants	PoliceRepOtherDrug	Fog	HazardousCargo
CarMPV	FailureOfDriverRider	CommentCond	CargoDischarged
Van	WhatCausal	SurfaceContaminents	PreImpactSpeed
BusMinibus	Comments6	SignRelated	NumberOfEvents
Truck	ChildRestrFitted	TrafficCalming	MostHarmfulEvent
AgriculturalVehicle	ChildRestrUsed	WasTrafficCalm	AreaOfMostDamage
MotorcycleMoped	CRSType	CommentsFact	EventType1
Bicycle	Comments7	NumOfSigns	EventDetail1
TrainTram	MCycleHelmetWorn	Sign1	InteractedWith1
ShoeVehiclePedestrian	MHelmetType	ProblemWithSign1	CollisionType1
Other	Comments8	NotWorking1	EventType2
UnknownVehicle	PartialLeathersProtJack	Sign2	EventDetail2
Accidentdetails_Comments	PartialLeathersProtJackTrou	ProblemWithSign2	InteractedWith2
AccidentSummary	MGloves	NotWorking2	CollisionType2
Accidentdetails_FullFields	MBoots	Sign3	EventType3
Accidentdetails_EmptyFields	MReflItemWorn	ProblemWithSign3	EventDetail3
TotalFilledPercentage	Comments9	NotWorking3	InteractedWith3
StateString	BHelmetWorn	Sign4	CollisionType3
Accidentdetails_CreationDate	BHelmetType	ProblemWithSign4	EventType4
Accidentdetails_LastUpdateDate	Comments9b	NotWorking4	EventDetail4
Accidentdetails_DeleteDate	HighVisCloth	Sign5	InteractedWith4
CaseCheck	ThickCloth	ProblemWithSign5	CollisionType4
Accidentdetails_SessionID	Comments10	NotWorking5	EventType5
	PedVehInteraction	Roadwaydetails_OtherComments	EventDetail5
RoadUserdetails_ID	PedCompany	Roadwaydetails_FullFields	InteractedWith5
Roadway5_1_ID	PedDisabilities	Roadwaydetails_EmptyFields	CollisionType5
RoadUserdetails_StringDateTime	PReflectiveItemsWorn	Roadwaydetails_CreationDate	EventType6
RoadUserdetails_CaseNumber	Comments11	Roadwaydetails_LastUpdateDate	EventDetail6
RoadUserdetails_VehicleNumber	AnyOtherComment	Roadwaydetails_DeleteDate	InteractedWith6
PersonNumber	RoadUserdetails_FullFields	Roadwaydetails_SessionID	CollisionType6
RoadUserClass	RoadUserdetails_EmptyFields		ABS
Age	RoadUserdetails_CreationDate	**Vehicledetails_ID**	BAS
Gender	RoadUserdetails_LastUpdateDate	Vehicledetails_StringDateTime	ACS
Impairment	RoadUserdetails_DeleteDate	Vehicledetails_CaseNumber	ESP
Comments1	RoadUserdetails_SessionID	Vehicledetails_VehicleNumber	LDW
IsAResident		NumOfOccupants	CSS
IsFamiliar	**Roadwaydetails_ID**	VehicleType	TCS
Comments2	Roadwaydetails_StringDateTime	VehicleMake	ESafetyComments
CrashAvoidMan	Roadwaydetails_CaseNumber	VehicleModel	Vehicledetails_OtherComments
Comments3	Roadwaydetails_VehicleNumber	CarBodyStyle	Vehicledetails_FullFields
SeatPos	CarriagewayType	DrivenWheels	Vehicledetails_EmptyFields
SeatDir	NumberOfLines	DriveOfVehicle	Vehicledetails_CreationDate
SeatBelt	Motorway	VehicleColour	Vehicledetails_LastUpdateDate
Comments4	SpeedLimit	VehicleLength	Vehicledetails_DeleteDate
AirbagAvail	TypeOfSpeedLimit	VehicleWidth	Vehicledetails_SessionID
AirbagDeploy			

More details on variables, values and definitions can be found in the WP5 Glossary (SafetyNet WP5, 2006).

2.2 Description of WP5 data with respect to injury severity

This section is intended to present the profile of injury 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: road - vehicle - user:

- Road (e.g. speed limit, carriageway type, type of junction)
- Vehicle (e.g. number of vehicles in accident, type of vehicle)
- User (e.g. gender, age, seat belt use).

However, it is important to think about how injured persons end up in this database. Only accidents with at least one fatality were included, consequently the presence of injured persons means that these persons were *additional* victims. Therefore, it would be reasonable to consider that proportionally more injured people may not necessarily mean less severe accidents overall. In the analysis of the data, it will be shown how this "accident size bias", mentioned in the Introduction, can be corrected for. In this section, therefore, no conclusions on fatality risk are drawn and the data presented mainly serve the purpose of data description.

As demonstrated by the exploration of the data, two severity scores are available: a Police score, based on the Police accident records, and a SafetyNet score, updated or revised by the SafetyNet WP5 team. It was observed, however, that the two scores lead to considerably different distributions of the available sample of cases across examined parameters. The data exploration presented in this section also aimed to the identification of these differences, which are further analyzed in Chapter 6. In particular, it is interesting to note that these seem mainly to concern persons that had initially been classified as severely injured. Table 2.2 lists the difference between the numbers of persons between Police and SafetyNet categorization.

Table 2.2 Differences in severity scores between Police and SafetyNet

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

In the remaining of this section, the WP5 data are explored in terms of severity with relation to a number of basic road safety parameters, as mentioned above. All the figures presented below are based on the SafetyNet medical outcome severity classification.

2.2.1 User-related variables: gender, age, seat belt use

Overall, 44% of the individuals in the examined sample were killed, whereas non casualties involve around 27% of the individuals.

- Men account for the majority of casualties in the examined sample of road accidents. Moreover, men are slightly over-represented in fatalities and under-represented in serious and slight injuries, as compared to women (Table 2.3).

Table 2.3 Injury severity per gender (SafetyNet medical outcome)

Gender	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Male	46,4%	9,2%	14,7%	29,6%	**100,0%**
Female	39,4%	14,7%	26,6%	19,3%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

- Older individuals, especially those aged of 65 years and more, are killed more often, when involved in fatal accidents (Table 2.4). On the other hand, children and adolescents, aged of 15 years or less, are proportionally less-often killed. No clear patterns can be identified for younger individuals (15-34 years old).

- The 35-44 years age group appears to have the highest percentage of non casualties.

Table 2.4 Injury severity per age group (SafetyNet medical outcome)

Age	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
0-14	20,5%	27,3%	25,0%	27,3%	**100,0%**
15-24	43,6%	14,9%	17,0%	24,5%	**100,0%**
25-34	36,9%	11,4%	24,2%	27,5%	**100,0%**
35-44	37,2%	9,0%	18,6%	35,2%	**100,0%**
45-54	46,7%	8,9%	14,4%	30,0%	**100,0%**
55-64	49,4%	8,9%	20,3%	21,5%	**100,0%**
65+	61,7%	3,1%	13,0%	22,2%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

- The trends described above become even more pronounced when examining drivers alone (Table 2.5). In this case, however, young drivers (i.e. those aged less than 25 years) are also slightly over-represented in fatalities in relation to older drivers.

Table 2.5 Injury severity per age group - Drivers only (SafetyNet medical outcome)

Driver age	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
0-14	n.a	n.a	n.a	n.a	n.a
15-24	47,0%	10,6%	14,4%	28,0%	**100,0%**
25-34	41,5%	8,5%	22,0%	28,0%	**100,0%**
35-44	34,9%	7,3%	16,5%	41,3%	**100,0%**
45-54	44,3%	6,6%	14,8%	34,4%	**100,0%**
55-64	48,3%	10,3%	12,1%	29,3%	**100,0%**
65+	61,4%	2,9%	7,1%	28,6%	**100,0%**
Total	**45,1%**	**8,1%**	**15,3%**	**31,4%**	**100,0%**

n.a: not applicable

- Individuals who did not use seat belt are clearly over-represented in fatalities, relative to individuals who used a seat belt. However, the use and non use of seat-belt is equally over-represented in serious and slight injuries. It is noted that these results should be considered with caution, given the large number of cases without information on the use of seat belt (the variable was completed for 690 cases only, out of which 279 the value was "unknown").

Table 2.6 Injury severity per seatbelt use (SafetyNet medical outcome)

Seat belt	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Used	34,1%	16,5%	22,7%	26,7%	**100,0%**
Not used	48,3%	16,7%	23,3%	11,7%	**100,0%**
Unknown	30,0%	7,1%	16,5%	46,5%	**100,0%**
Total	**34,8%**	**12,5%**	**20,1%**	**32,6%**	**100,0%**

2.2.2 Vehicle-related variables: vehicle type, number of vehicles

- All pedestrians in the dataset were killed, whereas riders of mopeds, motorcyclists and bicyclists are highly over-represented in fatalities, their proportion being twice the average of all users. Accordingly, car or heavy goods vehicles occupants are clearly under-represented in fatalities in relation to other vehicle types (Table 2.7).

- Heavy vehicles occupants have the highest percentage of non casualties.

Table 2.7 Injury severity per vehicle type (SafetyNet medical outcome)

Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Bus / Minibus / Truck / Van	14,0%	8,0%	28,0%	50,0%	**100,0%**

Car / MPV	36,7%	13,5%	20,5%	29,3%	**100,0%**
Motorcycle / Moped	81,2%	5,9%	5,9%	7,1%	**100,0%**
Bicycle	87,2%	2,6%	7,7%	2,6%	**100,0%**
Shoe vehicle (pedestrian)	100,0%	0,0%	0,0%	0,0%	**100,0%**
Other	20,0%	0,0%	20,0%	60,0%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

- Single-vehicle accidents present the highest proportion of fatalities (Table 2.8). Obviously, in a fatal accident database, single-vehicle accidents with only one occupant (driver) would have a 100% proportion of fatalities. The presence of non-fatalities in single-vehicle accidents must consequently be attributed to the presence of more than one occupant in the vehicle, which is a clear example of the "accident size bias" mentioned above.

- When considering multi-vehicle accidents, the fatalities proportion is reduced with the number of vehicles involved.

Table 2.8 Injury severity per vehicle type (SafetyNet medical outcome)

Number of vehicles	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
1-veh	63,3%	13,3%	17,6%	5,7%	**100,0%**
2-veh	41,7%	9,6%	16,2%	32,5%	**100,0%**
(3+)-veh	23,5%	12,2%	28,7%	35,7%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

2.2.3 Road-related variables: speed limit, carriageway type, junction type, lighting

- There are two ranges of speed limits where fatalities are over-represented in relation to other speed limits; one around 65 and one around 100 km/h (Table 2.9). The first one may be explained by a possibly higher proportion of drivers exceeding the speed limit, although it is also likely that this speed limit reflects urban area conditions with more junctions. The second one may be attributed to the obviously more severe consequences of accidents at higher speeds. The relatively low proportion of fatalities at speed limits higher than 100 Km/h is somewhat surprising and should be considered with caution, given the relatively small number of cases in these speed limit categories.

- The highest percentages of non casualties are observed at speed limits lower than 50 Km/h.

Table 2.9 Injury severity per speed limit (SafetyNet medical outcome)

Speed limit	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**

20-30	42,9%	0,0%	14,3%	42,9%	**100,0%**
45-50	45,8%	5,2%	10,4%	38,6%	**100,0%**
60-70	47,9%	14,6%	21,9%	15,6%	**100,0%**
90	41,5%	13,7%	22,2%	22,6%	**100,0%**
97-100	55,8%	22,1%	15,6%	6,5%	**100,0%**
110-113	37,5%	8,3%	29,2%	25,0%	**100,0%**
120-130	31,1%	13,5%	33,8%	21,6%	**100,0%**
Total	**44,1%**	**10,9%**	**18,1%**	**26,9%**	**100,0%**

- Fatalities are over-represented in one-way roads and undivided two-way roads (Table 2.10) in relation to divided roads. A trend can be identified according to which, the less clear the separation of the opposed traffic streams, the more severe (on average) the casualties resulting from accidents.

- Two-way roads with physical separation have a disproportional high number of slightly injured people involved.

- The highest percentage of non casualties is observed at junctions (where most multi-vehicle accidents occur).

Table 2.10 Injury severity per carriageway type (SafetyNet medical outcome)

Carriageway type	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Two way physically divided roadway	37,5%	9,0%	29,2%	24,3%	**100,0%**
Two way traffic divided by painted line	45,5%	12,4%	17,0%	25,1%	**100,0%**
Two way traffic with no division markings	47,2%	13,2%	9,4%	30,2%	**100,0%**
One way traffic	52,9%	0,0%	5,9%	41,2%	**100,0%**
Junction	42,9%	3,6%	16,7%	36,9%	**100,0%**
Other	71,4%	0,0%	0,0%	28,6%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

- Crossroads (+ junctions) seem to perform slightly better compared to other junction types (Table 2.11).

Table 2.11 Injury severity per presence and type of junction (SafetyNet medical outcome)

Presence & type of junction	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
No junction	45.2%	12.9%	17.2%	24.7%	**100.0%**

Crossroads (+ junction)	39.0%	3.0%	29.0%	29.0%	**100.0%**
T junction	44.3%	9.2%	13.0%	33.6%	**100.0%**
Other	44,9%	4,1%	22,4%	28,6%	**100,0%**
Total	**44.4%**	**10.8%**	**18.2%**	**26.7%**	**100.0%**

- Lighting conditions do not appear to affect the proportion of casualties (Table 2.12)

- Seriously injured individuals are clearly over-represented in accidents at darkness, in relation to accidents at daylight, whereas the highest proportion of non casualties is observed in accidents at darkness with artificial light.

Table 2.12 Table 2.12. Injury severity per lighting conditions (SafetyNet medical outcome)

Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Darkness	45,4%	14,5%	17,4%	22,7%	**100,0%**
Darkness with artificial light	43,7%	4,2%	17,6%	34,5%	**100,0%**
Daylight	44,8%	10,3%	18,1%	26,7%	**100,0%**
Other	38,3%	15,0%	21,7%	25,0%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

The frequencies corresponding to these results are included in Appendix I.

2.3 Description of WP5 data with respect to fatal accident occurence

In this section, some combined explorations of the WP5 database are presented in the form of collision matrices, describing the occurrence of fatal accidents.

The first part of the following Table 2.13 includes a breakdown of vehicles participating in fatal accidents depending on the total number of involved vehicles per incident. The second part depicts how often each vehicle class appears in such accidents, alone or with other vehicle types.

Table 2.13. Collision matrix-like analysis: number of vehicles in accidents and vehicle type

	Car / MPV	Truck / Van	Bus / Minibus	Motorcycle / Moped	Bicycle	Shoe Vehicle Pedestrian	Other	Total
***Vehicles** involved in:*								
Fatal accidents	60,1%	10,9%	0,6%	12,6%	6,1%	8,9%	0,8%	100%
1-veh. fatal accidents	77,3%	5,9%	0,0%	16,0%	0,8%	0,0%	0,0%	100%
(2+)-veh. fatal accidents	56,1%	12,1%	0,8%	11,9%	7,3%	10,9%	1,0%	100%
***Accidents** in which:*								
A vehicle class is represented	81,6%	16,5%	1,1%	20,6%	9,6%	15,7%	1,4%	100%
A single vehicle class is represented	82,8%	5,6%	0,0%	11,1%	0,5%	0,0%	0,0%	100%

- Passenger cars and two-wheelers are relatively more often represented in single-vehicle accidents than bicycles and trucks. The opposite is observed in accidents with two or more collision partners.

- As far as the time of occurrence is concerned (Table 2.14), the hours from 13 to 21 appear the most common for fatal accidents to occur. It could be further investigated whether this is related to the traffic volume distribution.

Table 2.14. Distribution of fatal accidents per time of day

Time of day	1-5	5-9	9-13	13-17	17-21	21-1	Total
Fatal Accidents (%)	8,7%	13,3%	16,1%	23,8%	25,4%	12,7%	100,0%

- Finally, crossing the median / centre-line constitutes the most common 1st event in fatal accidents (Table 2.15).

Table 2.15. Distribution of fatal accidents per first event

First Event	Cross median/ centreline	Motor vehicle travelling on same roadway	Motor vehicle travelling on other roadway	Ran off road - near side	Ran off road - off side	Pedestrian	Pedal cycle	Other	Total
Fatal Accidents (%)	22,0	16,0	6,9	16,5	5,2	15,1	6,9	11,0	100,0

The frequencies corresponding to these results are included in Appendix I.

2.4 Summary and motivations for analysis

In this section, the WP5 in-depth fatal accidents database was briefly described with respect to its structure and contents. Some basic descriptive statistics were elaborated, in terms of accident severity and accident occurrence factors, in relation to user, road and accident characteristics.

The analysis of proportions presented in this section can by no means lead to conclusions as regards fatality risk, causation or other road safety conclusions. However, a number of issues warranting further investigation have been identified:

- In this dataset, which includes only fatal accidents, the overall severity of the accident decreases when the number of participants increases. Given that all accidents have at least one fatality, the baseline probability of fatality for a single-vehicle, single-occupant accident is 100%, whereas for a single-vehicle, two-occupant accident it is 50% for each occupant, and so on. Unless this particularity of a fatal accident database (accident size bias) is accounted for when modeling accident fatality risk, the results can be extremely misleading. For example, the data exploration showed that an increased proportion of fatalities is related to single-vehicle accidents; however, keeping in mind that in this dataset there is at least one fatality in each accident, it is obvious that the proportion of fatalities decreases with the number of participants.

- Several basic parameters, from those examined in this section, present some variation with respect to injury severity. The database includes many more parameters that deserve to be examined, especially those related to behavioural indicators (e.g. seatbelt, impairment) or to in-depth information

(e.g. first event of the accident), which are seldom available, or reliable, in macroscopic accident data.

- Important differences in reporting accident severity were observed between the two available severity scores, the Police outcome and the SafetyNet outcome. Intuitively, it would be more reasonable to use the (checked and confirmed) SafetyNet outcome. However, the investigation of these differences, and of the factors that might underlay them, would be very useful (see Chapter 6).

References

SafetyNet WP5, "Work Package 5 Database Glossary", version of 29 November 2006.

APPENDIX I - Detailed descriptive statistics

Note: Slight differences in the grand totals between Tables are due to different number of missing values in each case

Table 2.3. Injury severity per gender (SafetyNet medical outcome)

Gender	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Male	302	60	96	193	**651**
Female	102	38	69	50	**259**
Total	**404**	**98**	**165**	**243**	**910**
Gender	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Male	74,8%	61,2%	58,2%	79,4%	**71,5%**
Female	25,2%	38,8%	41,8%	20,6%	**28,5%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Gender	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Male	46,4%	9,2%	14,7%	29,6%	**100,0%**
Female	39,4%	14,7%	26,6%	19,3%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

Table 2.4. Injury severity per age group (SafetyNet medical outcome)

Age	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
0-14	9	12	11	12	**44**
15-24	105	36	41	59	**241**
25-34	55	17	36	41	**149**
35-44	54	13	27	51	**145**
45-54	42	8	13	27	**90**
55-64	39	7	16	17	**79**
65+	100	5	21	36	**162**
Total	**404**	**98**	**165**	**243**	**910**
Age	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
0-14	2,2%	12,2%	6,7%	4,9%	**4,8%**
15-24	26,0%	36,7%	24,8%	24,3%	**26,5%**
25-34	13,6%	17,3%	21,8%	16,9%	**16,4%**
35-44	13,4%	13,3%	16,4%	21,0%	**15,9%**
45-54	10,4%	8,2%	7,9%	11,1%	**9,9%**
55-64	9,7%	7,1%	9,7%	7,0%	**8,7%**
65+	24,8%	5,1%	12,7%	14,8%	**17,8%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Age	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
0-14	20,5%	27,3%	25,0%	27,3%	**100,0%**
15-24	43,6%	14,9%	17,0%	24,5%	**100,0%**
25-34	36,9%	11,4%	24,2%	27,5%	**100,0%**
35-44	37,2%	9,0%	18,6%	35,2%	**100,0%**
45-54	46,7%	8,9%	14,4%	30,0%	**100,0%**
55-64	49,4%	8,9%	20,3%	21,5%	**100,0%**
65+	61,7%	3,1%	13,0%	22,2%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

Table 2.5. Injury severity per age group - Drivers only (SafetyNet medical outcome)

Driver age	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
0-14	3	1	1	1	**6**
15-24	62	14	19	37	**132**
25-34	49	10	26	33	**118**
35-44	38	8	18	45	**109**
45-54	27	4	9	21	**61**
55-64	28	6	7	17	**58**
65+	43	2	5	20	**70**
Total	**250**	**45**	**85**	**174**	**554**
Driver age	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
0-14	1,2%	2,2%	1,2%	0,6%	**1,1%**
15-24	24,8%	31,1%	22,4%	21,3%	**23,8%**
25-34	19,6%	22,2%	30,6%	19,0%	**21,3%**
35-44	15,2%	17,8%	21,2%	25,9%	**19,7%**
45-54	10,8%	8,9%	10,6%	12,1%	**11,0%**
55-64	11,2%	13,3%	8,2%	9,8%	**10,5%**
65+	17,2%	4,4%	5,9%	11,5%	**12,6%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Driver age	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
0-14	50,0%	16,7%	16,7%	16,7%	**100,0%**
15-24	47,0%	10,6%	14,4%	28,0%	**100,0%**
25-34	41,5%	8,5%	22,0%	28,0%	**100,0%**
35-44	34,9%	7,3%	16,5%	41,3%	**100,0%**
45-54	44,3%	6,6%	14,8%	34,4%	**100,0%**
55-64	48,3%	10,3%	12,1%	29,3%	**100,0%**
65+	61,4%	2,9%	7,1%	28,6%	**100,0%**
Total	**45,1%**	**8,1%**	**15,3%**	**31,4%**	**100,0%**

Table 2.6. Injury severity per seatbelt use (SafetyNet medical outcome)

Seat belt	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Used	93	45	62	73	**273**
Not used	58	20	28	14	**120**
Unknown	89	21	49	138	**297**
Total	**240**	**86**	**139**	**225**	**690**
Seat belt	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Used	38,8%	52,3%	44,6%	32,4%	**39,6%**
Not used	24,2%	23,3%	20,1%	6,2%	**17,4%**
Unknown	37,1%	24,4%	35,3%	61,3%	**43,0%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Seat belt	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Used	34,1%	16,5%	22,7%	26,7%	**100,0%**
Not used	48,3%	16,7%	23,3%	11,7%	**100,0%**
Unknown	30,0%	7,1%	16,5%	46,5%	**100,0%**
Total	**34,8%**	**12,5%**	**20,1%**	**32,6%**	**100,0%**

Table 2.7. Injury severity per vehicle type (SafetyNet medical outcome)

Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Bus / Minibus / Truck / Van	14	8	28	50	**100**
Car / MPV	229	84	128	183	**624**
Motorcycle / Moped	69	5	5	6	**85**
Bicycle	34	1	3	1	**39**
Shoe vehicle (pedestrian)	57	0	0	0	**57**
Other	1	0	1	3	**5**
Total	**404**	**98**	**165**	**243**	**910**
Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Bus / Minibus / Truck / Van	3,5%	8,2%	17,0%	20,6%	**11,0%**
Car / MPV	56,7%	85,7%	77,6%	75,3%	**68,6%**
Motorcycle / Moped	17,1%	5,1%	3,0%	2,5%	**9,3%**
Bicycle	8,4%	1,0%	1,8%	0,4%	**4,3%**
Shoe vehicle (pedestrian)	14,1%	0,0%	0,0%	0,0%	**6,3%**
Other	0,2%	0,0%	0,6%	1,2%	**0,5%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Vehicle type	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Bus / Minibus / Truck / Van	14,0%	8,0%	28,0%	50,0%	**100,0%**
Car / MPV	36,7%	13,5%	20,5%	29,3%	**100,0%**
Motorcycle / Moped	81,2%	5,9%	5,9%	7,1%	**100,0%**
Bicycle	87,2%	2,6%	7,7%	2,6%	**100,0%**
Shoe vehicle (pedestrian)	100,0%	0,0%	0,0%	0,0%	**100,0%**
Other	20,0%	0,0%	20,0%	60,0%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

Table 2.8. Injury severity per vehicle type (SafetyNet medical outcome)

Number of vehicles	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
1-veh	133	28	37	12	**210**
2-veh	244	56	95	190	**585**
(3+)-veh	27	14	33	41	**115**
Total	**404**	**98**	**165**	**243**	**910**
Number of vehicles	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
1-veh	32,9%	28,6%	22,4%	4,9%	**23,1%**
2-veh	60,4%	57,1%	57,6%	78,2%	**64,3%**
(3+)-veh	6,7%	14,3%	20,0%	16,9%	**12,6%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Number of vehicles	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
1-veh	63,3%	13,3%	17,6%	5,7%	**100,0%**
2-veh	41,7%	9,6%	16,2%	32,5%	**100,0%**
(3+)-veh	23,5%	12,2%	28,7%	35,7%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

Table 2.9. Injury severity per speed limit (SafetyNet medical outcome)

Speed limit	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
20-30	6	0	2	6	**14**
45-50	158	18	36	133	**345**
60-70	46	14	21	15	**96**
90	112	37	60	61	**270**
97-100	43	17	12	5	**77**
110-113	9	2	7	6	**24**
120-130	23	10	25	16	**74**
Total	**397**	**98**	**163**	**242**	**900**
Speed limit	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
20-30	1,5%	0,0%	1,2%	2,5%	**1,6%**
45-50	39,8%	18,4%	22,1%	55,0%	**38,3%**
60-70	11,6%	14,3%	12,9%	6,2%	**10,7%**
90	28,2%	37,8%	36,8%	25,2%	**30,0%**
97-100	10,8%	17,3%	7,4%	2,1%	**8,6%**
110-113	2,3%	2,0%	4,3%	2,5%	**2,7%**
120-130	5,8%	10,2%	15,3%	6,6%	**8,2%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Speed limit	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
20-30	42,9%	0,0%	14,3%	42,9%	**100,0%**
45-50	45,8%	5,2%	10,4%	38,6%	**100,0%**
60-70	47,9%	14,6%	21,9%	15,6%	**100,0%**
90	41,5%	13,7%	22,2%	22,6%	**100,0%**
97-100	55,8%	22,1%	15,6%	6,5%	**100,0%**
110-113	37,5%	8,3%	29,2%	25,0%	**100,0%**
120-130	31,1%	13,5%	33,8%	21,6%	**100,0%**
Total	**44,1%**	**10,9%**	**18,1%**	**26,9%**	**100,0%**

Table 2.10 Injury severity per carriageway type (SafetyNet medical outcome)

Carriageway type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Two way physically divided roadway	54	13	42	35	**144**
Two way traffic divided by painted line	275	75	103	152	**605**
Two way traffic with no division markings	25	7	5	16	**53**
One way traffic	9	0	1	7	**17**
Junction	36	3	14	31	**84**
Other	5	0	0	2	**7**
Total	**404**	**98**	**165**	**243**	**910**

Carriageway type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Two way physically divided roadway	13,4%	13,3%	25,5%	14,4%	**15,8%**
Two way traffic divided by painted line	68,1%	76,5%	62,4%	62,6%	**66,5%**
Two way traffic with no division markings	6,2%	7,1%	3,0%	6,6%	**5,8%**
One way traffic	2,2%	0,0%	0,6%	2,9%	**1,9%**
Junction	8,9%	3,1%	8,5%	12,8%	**9,2%**
Other	1,2%	0,0%	0,0%	0,8%	**0,8%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**

Carriageway type	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Two way physically divided roadway	37,5%	9,0%	29,2%	24,3%	**100,0%**
Two way traffic divided by painted line	45,5%	12,4%	17,0%	25,1%	**100,0%**
Two way traffic with no division markings	47,2%	13,2%	9,4%	30,2%	**100,0%**
One way traffic	52,9%	0,0%	5,9%	41,2%	**100,0%**
Junction	42,9%	3,6%	16,7%	36,9%	**100,0%**
Other	71,4%	0,0%	0,0%	28,6%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

Table 2.11. Injury severity per presence and type of junction (SafetyNet medical outcome)

Presence & type of junction	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
No junction	284	81	108	155	**628**
Crossroads (+ junction)	39	3	29	29	**100**
T junction	58	12	17	44	**131**
Other	22	2	11	14	**49**
Total	**403**	**98**	**165**	**242**	**908**

Presence & type of junction	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
No junction	70,5%	82,7%	65,5%	64,0%	**69,2%**
Crossroads (+ junction)	9,7%	3,1%	17,6%	12,0%	**11,0%**
T junction	14,4%	12,2%	10,3%	18,2%	**14,4%**
Other	5,5%	2,0%	6,7%	5,8%	**5,4%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**

Presence & type of junction	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
No junction	45.2%	12.9%	17.2%	24.7%	**100.0%**
Crossroads (+ junction)	39.0%	3.0%	29.0%	29.0%	**100.0%**
T junction	44.3%	9.2%	13.0%	33.6%	**100.0%**
Other	44,9%	4,1%	22,4%	28,6%	**100,0%**
Total	**44.4%**	**10.8%**	**18.2%**	**26.7%**	**100.0%**

Table 2.12. Injury severity per lighting conditions (SafetyNet medical outcome)

Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	Total
Darkness	94	30	36	47	**207**
Darkness with artificial light	52	5	21	41	**119**
Daylight	235	54	95	140	**524**
Other	23	9	13	15	**60**
Total	**404**	**98**	**165**	**243**	**910**
Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Darkness	23,3%	30,6%	21,8%	19,3%	**22,7%**
Darkness with artificial light	12,9%	5,1%	12,7%	16,9%	**13,1%**
Daylight	58,2%	55,1%	57,6%	57,6%	**57,6%**
Other	5,7%	9,2%	7,9%	6,2%	**6,6%**
Total	**100,0%**	**100,0%**	**100,0%**	**100,0%**	**100,0%**
Lighting	Killed	Seriously Injured	Slightly Injured	Not Injured	**Total**
Darkness	45,4%	14,5%	17,4%	22,7%	**100,0%**
Darkness with artificial light	43,7%	4,2%	17,6%	34,5%	**100,0%**
Daylight	44,8%	10,3%	18,1%	26,7%	**100,0%**
Other	38,3%	15,0%	21,7%	25,0%	**100,0%**
Total	**44,4%**	**10,8%**	**18,1%**	**26,7%**	**100,0%**

Table 2.13 Collision matrix-like analysis: number of vehicles in accidents and vehicle type

	Crash Participants	Car / MPV	Truck / Van	Bus / Minibus	Motorcycle / Moped	Bicycle	Shoe Vehicle Pedestrian	Other
***Vehicles** involved in:*								
Fatal accidents	641	385	70	4	81	39	57	5
1-veh. fatal accidents	119	92	7	0	19	1	0	0
(2+)-veh. fatal accidents	522	293	63	4	62	38	57	5
***Accidents** in which:*								
A vehicle class is represented	364	297	60	4	75	35	57	5
A single vehicle class is represented	198	164	11	0	22	1	0	0

Table 2.14. Distribution of fatal accidents per time of day

Time of day	1-5	5-9	9-13	13-17	17-21	21-1	Total
Fatal Accidents	28	43	52	77	82	41	323
Fatal Accidents (%)	8,7%	13,3%	16,1%	23,8%	25,4%	12,7%	100,0%

Table 2.15. Distribution of fatal accidents per first event

First Event	Cross median/ centreline	Motor vehicle travelling on same roadway	Motor vehicle travelling on other roadway	Ran off road - near side	Ran off road - off side	Pedestrian	Pedal cycle	Other	Total
Fatal Accidents	80	60	25	60	19	55	25	40	364
Fatal Accidents (%)	22,0	16,0	6,9	16,5	5,2	15,1	6,9	11,0	100,0

==========
Page left blank intentionally
==========

Chapter 3 - Modelling Accident Size

Heike Martensen and Emmanuelle Dupont (IBSR)

3.1 The research question

In the introduction, it has been described how the size of an accident affects the fatality risk, i.e. the proportion of people killed among those involved in the fatal accidents. The following chapters will be dedicated to identifying the factors that affect the fatality risk while controlling for the size of the accident. In Chapter 4, accident size will be controlled for statistically and in Chapter 5 by considering only accidents with two participants. 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, accidents involving only a single vehicle differ in many variables from accidents involving two or more participants. One reason why it is interesting to differentiate between single and multi-vehicle accidents is 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 accident 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 distributed between the various drivers involved). On average, the drivers in single-vehicle accidents have made more (severe) mistakes than the drivers in multi-vehicle accidents. The variables that differentiate the two types of accidents could, therefore, be seen to characterize drivers that are more prone to make errors than others. One should, however, keep in mind that this is a very indirect way of reasoning, as the drivers in multi-vehicle accidents that were taken up in the fatal accident investigation database have been involved in fatal accidents as well and should certainly not be seen as “exemplary”.

The goal of the present chapter is threefold: In the first place, factors will be identified that differentiate between single- and multi-vehicle accidents. Second, it will be simultaneously analysed how accidents in which only one person is killed differ from those with more than one victim who is either killed or severly injured. These two aspects are related to some extent. For instance, if a lonely driver hits a tree with his/her car, there is no possibility to observe more than one person who is killed or seriously injured. The third goal of this chapter is therefore to investigate the relation between the number of crash participants and the number of victims who are killed or severly injured.

3.2 The analytical problems

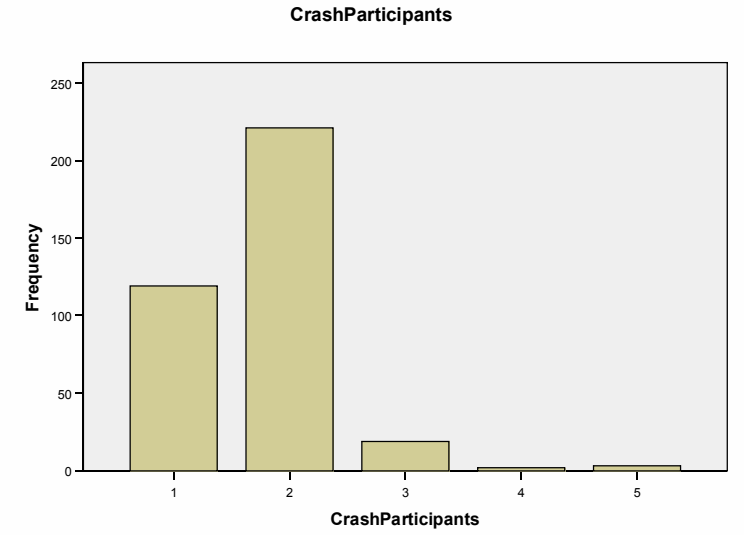
The three goals described above are pursued in a bivariate analysis where the number of participants and the number of victims are simultaneously entered as

dependent variables. In the following, we will at first describe the selection of the dependent variables in more detail, then the selection criteria for the accidents included and, finally, the structure of the bivariate model.

3.2.1 Dependent Variable 1: Number of crash participants

In the present database, 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. The number of crash participants is certainly not normally distributed. As a consequence, it is not advisable to use linear regression, t-tests, or other methods that are based on the normality assumption.

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



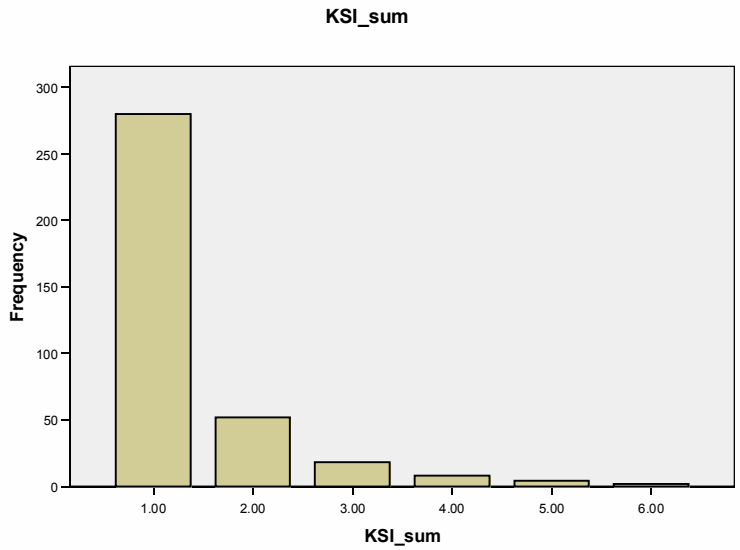
As can be seen in the chart, the majority of the accidents involved either one or two participants. There are only a few accidents with 3 or more participants. One option would be to model the probability to be involved in those three types of accidents (1, 2, 3+ participants) in an ordered category model. Given the small number of accidents in the 3+ category, it was however decided to model only the difference between accidents with one participant and those with two or more participants. A new variable, <Multi-vehicle Accident> (in the following MultiVehicle) was generated, that is 0 for single-participant accidents and 1 for multiple-participants accidents[2].

[2] As described below, accidents involving non-motor vehicles were excluded from the analysis. Therefore all crash-participants were in fact 'vehicles'.

3.2.2 Dependent variable 2: number of victims

In a fatal accident database, every accident has at least one person killed. There were only 31 accidents with more than one person killed. As this number is too small to statistically analyse them, the analysis was based on the number of victims who were *killed or seriously injured* (KSI). The distribution of the number KSI is presented in Figure 3.2. Although the majority of the accidents have only one killed or seriously injured victim, there are also a few cases (84) with more than one KSI.

Figure 3.2 The number of accidents with 1, 2, 3… victims killed or seriously injured



A new variable, <Multi-KSI Accident> (in the following Multi-KSI), was generated that is 0 for accidents with only one victim killed or seriously injured, and 1 for accidents with more than one person killed or seriously injured.

3.2.3 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 and no seriously injured victims. To separate factors that affect the accident size from those that differentiate between accidents involving vulnerable road users and those that do not, all accidents involving non-motor vehicles have been excluded from the following analyses, leaving 272 of the 364 accidents. Moreover, all accidents for which there was no information on whether the drivers had been drinking, were otherwise impaired, or were unfamiliar with the area were excluded, leaving 233 cases with 65 accidents with more than one KSI for the subsequent analysis.

3.2.4 The model

The variables <Multi-vehicle Accident> and <Multi-KSI Accident> were the dependent variables in a multivariate logistic regression analysis (c.f. Yannis, Papadimitiriou, & Antoniou, 2007a; see Appendix 3A for the implementation details). Modelling the number of crash participants and the number of KSI simultaneously offers the advantage of taking account of the interrelation that possibly exists between these two variables.

Variables that might affect either the number of vehicles or the number of KSI, were entered as independent variables. All independent variables that have become significant in the model can explain a significant part of the variance in the respective dependent variable (Multi-vehicles or Multi-KSI) *over and above* what the other variables in the model can explain. That means that the effect of each variable is corrected for the effects of all the others. Independent variables can be either continuous or categorical. For continuous variables, a positive coefficient indicates a higher score of multi-vehicle/KSI accidents as compared to single vehicle accidents. A negative coefficient indicates a lower score. For categorical independent variables, a positive coefficient means that multi-vehicle/KSI accidents have a higher probability of the feature being present as compared to single vehicle/KSI accidents. For this first global analysis, we tried to simplify the variables before entering them into the logistic model so as to avoid dealing with too many contrasts that become difficult to interpret. In Table 3.1, all variables that have been considered are presented.

Table 3.1 Variables considered in analysis

Variable	Values
MultipleVehicles	1: Multi-vehicle accident, 0: Single-vehicle accident
MultipleKSI	1: Multi-KSI, 0: One fatality
FirstEventType	1: Noncollision 2: Collision with other vehicle 3: Collision with object non/fixed 4: Collision with fixed object
DriverAge	Mean age of drivers involved in accident
pWomen	Percentage women involved in accident
pWomenDriver	Percentage women drivers involved in accident
Alcohol	1: Suspicion alcohol for any driver 0: No suspicion
Impairment	1: Suspicion impairment for any driver 0: No suspicion
Unfamiliar	1: Any driver unfamiliar, 0: All drivers familiar
AvoidManoeuvre	1: Avoidance manoeuvre conducted by any driver, 0: None
ExecutedManoevre	1: Manoeuvre executed by any driver, 0: All driving straight
LostControl	1: At least one of the driver lost control over vehicle, 0: no loss of control
[pSeatbelt	Percentage victims that used a seatbelt]
Carriageway	1: Carriageways physically divided, 0: Not divided
Motorway	1: Yes, 0: No
Speedlimit	1: <50, 2: 51-80, 3: 81-100, 4: >=100
Junction	1: Yes, 0: No
RuralArea	1: Rural, 0: Urban or Mixed
Light	1: Daylight, 0: Darkness or artificial light
Weekend	1: Weekend, 0: Week
NumberOfEvents	Number of events in the accident
AveOccVeh	The average number of occupants in the vehicles involved in the accident
VehicleAge	Mean age of vehicles involved in accident
ABS	1: Any of the vehicles had ABS, 0: no ABS

3.3 Results

The bivariate structure allows to test whether the two dependent variables, in this case MultipleVehicles and MultipleKSI are correlated and, if so, to correct for this correlation in the joint analysis. In the present case it is interesting to consider the empty model, the model in which the bivariate structure is implemented without any explanatory variables. The results of this empty model are presented in Table 3.2.

Table 3.2 Results empty model

	Multi-vehicles			Multi-KSI		
	Σ	SE	P	Σ	SE	P
Random parameters						
Variance	1.004	0.093	.650	1.004	0.093	.650
Covariance	0.105	0.066	.113			

Note – The p values of the variance parameters result from the test of whether they differ from 1.

As the dependent variables are assumed to follow the Binomial distribution, they are expected to have a variance of 1. To test whether this assumption holds, we have estimated the model under the assumption of an extra-binomial distribution, variances are allowed to differ from 1. As can be seen in Table 3.2, the estimated variance for MultiVehicles as well as for MultiKSI is indeed close to one.

Note that the covariance between MultiVehicles and MultiKSI is not significant. This is surprising. Generally speaking, there should be a relation between these two variables, because as mentioned above, accidents involving several vehicles also involve more occupants (3.38 on average as opposed to 1.89 on average for single vehicle accidents) and therefore offer a larger chance to observe several occupants who are killed or seriously injured. The fact that the covariance between both variables is non-significant suggests that single-vehicle accidents have on average as many KSI as multi-vehicle accidents. This suggests that they are relatively speaking more harmful to those involved.

InTable 3.3, the results of the model with those variables that were significant are presented. For most of the predictor variables the coefficients take opposite values. This can be seen more easily in Figure 3.3, where the reliability and the direction of the coefficients are plotted for multiple-vehicles and multiple-KSIs accidents. Bars that exceed the dotted lines are significant. An upwards directed bar means the category that gave the variable its name (e.g., “weekend” or “impairment”) is more likely for multi-vehicle/KSI accidents than for single-vehicle/KSI accidents, while for the opposite category (e.g. “week” or “No driver impaired”) it is the other way round. For downwards directed bars the relations are vice versa. Before looking at the details, it is important to note that the patterns for multi-vehicles and multi-KSI are very different. The bars usually point in opposite directions and those coefficients that are significant for one are often not significant for the other. Moreover, it can be noted that the bars for multi-vehicles accidents (blue bars) are generally longer than those for multi-KSI accidents, indicating that we can identify a number of factors that differentiate between multi- and single-vehicle accidents, but only a few that differentiate between multi- and single-KSI accidents.

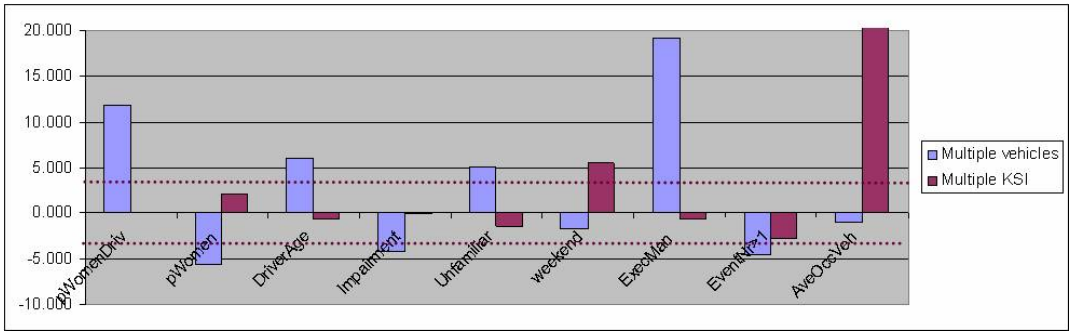
Table 3.3 Model with predictors

	Multi-vehicles				Multi-KSI			
	B	SE	P	exp	β	SE	P	exp
Predictor Variables								
pWomanDriver	2.567	0.746	0.001	13.03	0.148	0.592	0.803	1.16
pWoman	-1.514	0.639	0.018	0.22	0.802	0.55	0.145	2.23
DriverAge	0.485	0.197	0.000	1.62	-0.154	0.196	0.432	0.86
Impairment	-7.410	0.365	0.024	0.00	-0.098	0.405	0.809	0.91
Unfamiliar	1.020	0.453	0.024	2.77	-0.523	0.442	0.237	0.59
Weekend	-0.460	0.356	0.196	0.63	0.841	0.358	0.019	2.32
ExecutedManoeuvre	1.609	0.367	0.000	5.00	-0.311	0.385	0.419	0.73
EventNr>1	-1.906	0.886	0.031	0.15	-0.909	0.538	0.091	0.40
AveOccVeh	-0.209	0.199	0.294	0.81	0.983	0.215	0.000	2.67
Random parameters								
Variance	1.004	0.113	0.635		1.046	0.097	0.635	
Covariance	0.286	0.076	0.000					

Note that in the final model, presented in Table 3.3, the covariance between MultipleVehicles and MultipleKSI is significant, which was not the case in the empty model. This means that if – and only if – factors are taken into account that differentiate multi-vehicle/KSI accidents and single-vehicle/KSI accidents, a positive relation between the two dependent variables can be evidenced: Accidents with more vehicles involved are more likely to produce more KSIs.

The comparison of the coefficients estimated in the model for KSIs and for number of vehicles gives an indication why initially (i.e. in the empty model) these two variables appeared to be unrelated. One could say that we have two conflicting tendencies for the relation between the two dependent variables. Intuitively, one would expect to observe an overall relationship between the number of vehicles and the number KSI, simply because more people are involved in multi-vehicle accidents. However, as becomes clear in Figure 3.3, multi-vehicle accidents and multi-KSI accidents are oppositely related to most of the predictor variables in the model (e.g., multi-vehicle accidents tend to take place during the week, while multi-KSI accidents tend to take place in the weekend). In the empty model, these two tendencies compensate each other, leaving no significant covariance. In the final model, one of those tendencies, the opposite relation to the predictor variables, is explained by the model, leaving the other tendency (more people in multi-vehicle accidents offering a larger chance to observe more KSIs) in the unexplained part and making the covariance between the two dependent variables significant.

Figure 3.3 Reliability (chi-square statistic) and direction of coefficients



Note – Bars exceeding the dotted line indicate a significant predictor.

In the following, the results will be discussed with respect to each variable in the model. Subsequently we will also discuss the variables that have not been included in the models, because there can be different reasons for not doing so: Either the variable does not have a significant relation with either of the dependent variables in the first place, or it shows a high degree of overlap with another variable that is already taken up.

3.3.1 Variables in the model

3.3.1.1. Person Variables

As this analysis is conducted at the accident level, all person variables are aggregated in some way across all persons or all drivers involved in the accident.

Gender

The variable <pWomen> indicates the proportion of women among all persons in the accident; the variable <pWomenDriver> indicates the proportion of women among the drivers. Interestingly, we see opposite effects for these two variables. A fatal accident where a woman has been steering is more likely to be a multi-vehicle accident but a fatal accident in which women were present as passengers is more likely to be a single vehicle accident.

For the chance of observing an accident with more than one KSI, the proportion of women does not play a role – neither the proportion of women among the drivers, nor the proportion of women among all occupants.

Age of the driver

On average, drivers involved in single-vehicle accidents are younger than those involved in multi-vehicle accidents. The finding that drivers in single-vehicle accidents are younger and therefore less experienced is consistent with the idea that these have made more (or more severe) errors than drivers in multi-vehicle accidents.

There is no relation between the age of the drivers and the number of victims in an accident.

Impairment

Single-vehicle accidents involve more often than multi-vehicle accidents a driver who is impaired by alcohol, drugs, fatigue or other factors. Again, this result is in accordance with the idea that single-vehicle accidents are a more direct result of the driver's errors. These occur more often with impaired drivers.

Impaired drivers, who have a fatal accident, do not seem to cause more victims than unimpaired ones who have a fatal accident.

Driver Manoeuvre

The variable <DriverManoeuvre> indicates the type of manoeuvre each driver had been executing. The results for all drivers in the accident were summarised into the variable <ExecutedManoeuvre> which indicates whether any of the drivers involved executed any kind of manoeuvre (like turning, overtaking, etc.) or whether all of them were simply driving along the road.

Drivers in single-vehicle accidents had been all driving along the road much more often than drivers in multi-vehicle accidents..Put differently: In multi-vehicle accidents usually at least one of the drivers had been executing a manoeuvre before the accident.

Single- and multi-KSI accident do not differ with respect to the execution of manoeuvres right before.

Familiarity

Multi-vehicle accidents involve more often than single-vehicle accidents a driver who is unfamiliar with the area. On the one hand, it may be the case that drivers who are not familiar with the area might have problems to master junctions, because they are distracted from the traffic by finding out which route to take. This is one way to explain the tendency for unfamiliar drivers to be involved in multi-vehicle accidents more often. On the other hand, drivers who are more familiar with the area might have a stronger tendency to speed than those who are on unfamiliar grounds, and therefore have a higher risk of losing control over the vehicle which might lead to single vehicle accidents more often than to multi-vehicle accidents.

3.3.1.2. Situation and Accident variables

Day of the week

In the weekend, accidents involve only a single vehicle more often than during the week. This trend is, however, not significant.

In contrast, fatal accidents in the weekend involve multiple KSIs more often than those during the week. Interestingly, this effect is significant, while the number of occupants per vehicle as well as the impairment and age of the drivers was controlled for. This means that there is something that makes weekend accidents more harmful that cannot be accounted for by these factors.

Number of events

Each accident in the Fatal Accident Investigation Database is described in a chain of events, such as “cross median/centre line” or “collision with vehicle travelling on the same road-way” There are a number of accidents (26) that are described by only one event. These are almost never single vehicle accidents and also less often than expected single-KSI accidents. It is an interesting question whether these accidents consisted indeed of so few events or whether there might be a bias in the reporting (some accidents might capture the attention of the reporting persons more than other and therefore end up to be described in more detail). It is not clear, however, how the present result (more details if the accidents involved more vehicles or more killed or seriously injured victims) could result from a reporting bias.

Number of occupants per vehicle

The average number of occupants per vehicle (<AveOccVeh>) does not differ for single and multi-vehicle accidents. In contrast, accidents with more than one victim have – on average – more car occupants than accidents with only one victim. This is not very surprising as the presence of more persons increases the chance of one of them to get seriously injured, or to die. This variable is taken up into the model mainly in order to control for this effect. As an example, it helps us to rule out the “more occupants per vehicle” hypothesis as a possible reason for the higher number of multi-KSI accidents during the weekend.

3.3.2 Variables not included in the model

Variables can be excluded from the model for three reasons: The first, simplest reason is that the variable is not significant. Apart from that, it may also be the case that the predictor is significant when entered by itself, but shows a high degree of overlap with (an) other predictor(s) also included in the model. Finally, it can also be the case that the inclusion of a given predictor appeared impossible because of technical problems.

Normally, one should not talk about non-significant results. However, the present analysis concerns a data set that has yet to be completed. Due to the higher number of cases in the final data set, some of the variables that are not significant now, may become so in subsequent analysis. Therefore, marginal results (.050<p<.300) are also described. All variables are listed in Appendix 3B, along with the results associated with their individual inclusion in the empty model.

3.3.2.1. Variables excluded for technical reasons

Seatbelt

The percentage of persons who used a seatbelt is higher in multi-vehicle accidents as compared to single-vehicle accidents. This could mean that people in single-vehicle accidents are less concerned about safety than people in multi-vehicle accidents and therefore do not use protection as often.

The percentage of persons using seatbelts also tends to be higher in multi-KSI accidents as compared to single-KSI accidents. This contra intuitive result is, however, non-significant.

The variable Seatbelt has a lot of missing values (i.e. “unknown”). When including this variable into the analysis another, 75 accidents have to be omitted, which makes the results for a couple of other variables non-significant. Seatbelt was therefore not included in the final model (for the limited dataset for which we have seatbelt data, the result for this variable seems to be stable though and remains relatively unchanged when combined with the other model variables).

First Event Type

Multi-vehicle accidents often start with collisions between vehicles, which single-vehicle accidents never do. In contrast, single-vehicle accidents sometimes start with collisions with fixed objects, which multi-vehicle accidents never do. As this result is without exceptions[3], it is not necessary (and not possible) to test its significance in a binomial model.

3.3.2.2. Variables excluded because of overlap with other variables

Alcohol

Single-vehicle accidents involve drivers impaired by alcohol more often than multi-vehicle accidents do. This result is consistent with the outcome for the more general variable <Impairment>, which includes other types of impairments (e.g. drugs and fatigue). When including both variables into the model simultaneously, <Alcohol> becomes non-significant, while the coefficient for

[3] In fact there is one exception where a single vehicle accident started with a collision with a parked vehicle.

<Impairment> is reduced but remains significant. This suggests that all types of impairments (not just alcohol) can be responsible for involving the driver in a single-vehicle accident.

Drivers impaired by alcohol, when having a fatal accident, also tend to cause more victims. This result is, however, not significant.

Junction

Unsurprisingly, multi-vehicle accidents take place at junctions more often than single-vehicle accidents do. Accidents that take place at junctions involve usually at least one driver who was executing a manoeuvre. Consequently, junction is not significant anymore in combination with the variable <ExecutedManoeuvre>.

With respect to the number of victims, there is no difference between junction and non-junction accidents.

Speed Limit

The probability of observing single-vehicle accidents is higher at speed limits between 51 and 80 as well as above 100, as compared to speed limits of 50 and below and between 81 and 100. The effect of speed limit on the probability of single versus multiple-vehicles accidents can partly be explained by other variables as well, in particular whether drivers had been executing an avoidance manoeuvre or not, and whether a manoeuvre had been executed by one of the accident participants (as opposed to all of them driving straight). Consequently, the variable speed limit is not significant as a predictor any more when taken up together with the other variables and was deleted.

The speed limit does not differentiate between single- and multi-KSI accidents.

Area

Fatal accidents in rural area are more likely to be single-vehicle accidents than those in urban or mixed areas. Similar to speed limit, this variable has a strong overlap with the variable execution of manoeuvre and becomes non-significant together with the other variables in the model.

There is a tendency for multi-vehicle accidents to take place in rural areas rather than in urban or in mixed areas. However, this result is not significant.

Light

Fatal accidents taking place in daylight are more often multi-vehicle accidents than those occurring at night (darkness, dusk, or artificial light). There is also a non-significant tendency for multi-victim accidents to occur in the darkness rather than at daylight. This variable shows a strong overlap with the mean driver age and the impairment of any of the drivers and is not significant in combination with these.

3.3.2.3. Nonsignificant Variables

Lost Control

One of the categories in the variable "DriverManoeuvre" is "Lost Control". This category actually describes a first event in the accident rather than the last executed manoeuvre. Therefore, a dichotomous variable was coded which is 1 if DriverManoeuvre is "Lost Control" and 0 for all other categories. Although this variable was not significant, there is a tendency for drivers who lost control to be in a single-vehicle rather than multi-vehicle accident and in multi-KSI rather than single-KSI accidents.

Motorway and Carriageway

The results for motorway and carriageway (physically divided or not) are very similar. There is a non-significant tendency for fatal accidents on motorways or roads with physically divided carriageways to be single rather than multi-vehicle accidents. This result is marginally significant for carriageways.

In contrast, accidents on motorways or on roads with divided carriageways have multiple victims who are killed or seriously injured more often than other accidents. This tendency is marginally significant for motorways.

Although a physical division between carriageways should increase safety, they tend to produce a disproportionately high number of multi-KSI accidents. This can probably be explained by the higher speed regimes on these roads.

ABS

The variable ABS indicates whether an accident involved at least one car with ABS. Accidents with more than one KSI involved cars with ABS relatively more often than accidents with only one fatality. This result is marginally significant.

Vehicle Age

The vehicles in single-vehicle accidents tend to be older than those in multi-vehicle accidents. Vehicle age does not affect the number of victims.

Avoidance Manoeuvre

Fatal accidents in which one of the participants executed an avoidance manoeuvre tend to be multi- rather than single-vehicle accidents. They also have the tendency to have one rather than many victims who are killed or seriously injured.

3.4 Conclusion

The size of fatal accidents involving exclusively motor vehicles was modelled on the basis of two - binary – dependent variables. MultiVehicles indicated whether an accident involves several vehicles as opposed to only one, and MultiKSI

indicated whether an accident involved more than one victim who was killed or seriously injured, as opposed to only one. These variables were modelled simultaneously in a multivariate binomial regression model.

We can conclude that fatal single vehicle accidents *tend* to be caused by young male drivers impaired by alcohol, drugs, or fatigue who are familiar with the area. The drivers are often accompanied by female passengers and often the car occupants do not wear seatbelts. The accidents take place outside the area of junctions in the darkness and are usually recorded the accident database in more than one event.

Please note that the above statement on single-vehicle accidents is not to be understood as an absolute description but is *valid only in comparison to multi-vehicle accidents*. These involve older people and more female drivers, but fewer female passengers than single-vehicle accidents. The occupants tend to wear seatbelts more often and the driver is more often unfamiliar with the area. Multi-vehicle accidents take place in daylight at junctions more often than single vehicle accidents and they are sometimes reported in the accident database in only one event.

Single-vehicle accidents produce as many victims who are killed or seriously injured as multi-vehicle accidents. Given that they involve fewer people on average, this suggests that they are relatively speaking more dangerous.

The large majority of the accidents in the database involve only one fatality and mostly also only one victim that is killed *or* seriously injured. There are only two factors that reliably characterize the accidents with more than one victim who is killed or severly injured: They take place during the weekends more often and they involve vehicles with more participants. Possibly, the small number of accidents with more than one victim resulted in a lack of power to identify other factors that might have an effect.

Appendix 3A: Model implementation

The bivariate binomial model was implemented with the MLwin software (Rasbash, Steele, Brown, & Prosser, 2004). In structure this is a two level model, where the first level is defined by a response indicator that defines whether a response belongs to the first or the second dependent variables. The second level was defined by the accident identifier. For a more complete description, see Yannis, Papadimitriou, & Antoniou (2007a) who describe a bivariate poisson model of a similar structure.

The responses were assumed to follow the extra-binomial distribution and the logit link function was used to implement a generalized linear model. The restricted iterative generalized least squares (RIGLS) estimation was used with first order linearization and marginal quasi likelihood (MQL). The significance of the predictors was estimated using the Wald test.

Appendix 3B: Variables not in the final model

InTable 3.4, the results for all variables excluded from the final model are given. Each variable has been entered into the empty model by itself. In contrast to the results in Table 3.3 the coefficients below are therefore not corrected for the effect of the other variables.

Table 3.4 Results excluded variables in empty model

		Multi-vehicles			Multi-KSI		
		B	SE	P	β	SE	p
First event (compared to non-collision)	Collision with vehicle	4.509	1.027	0.000	-0.260	0.323	0.421
	Collision obj. nonfixed	0.165	1.434	0.908	-0.874	1.438	0.543
	Collision fixed object	*0.000	*0.000	*	-0.736	1.120	0.511
	pSeatbelt**	1.643	0.369	0.000	0.653	0.399	0.102
	pAlcohol	-1.174	0.350	0.001	0.405	0.361	0.262
	Junction	2.048	0.43	0.000	-0.612	0.353	0.083
Speed limit (compared to 50 kmh)	80kmh	-1.163	0.443	0.009	-0.034	0.469	0.942
	100kmh	-0.681	0.358	0.057	-0.032	0.360	0.929
	130kmh	-1.163	0.523	0.026	0.628	0.523	0.230
	RuralArea	-0.649	0.332	0.051	0.317	0.349	0.364
	Light	0.768	0.279	0.006	-0.408	0.297	0.170
	Lost control	-1.244	0.881	0.158	0.979	0.834	0.240
	Motorway	-0.647	0.428	0.131	0.807	0.435	0.064
	Div. carriageway	-0.680	0.380	0.074	0.461	0.397	0.246
	pABS	-0.208	0.309	0.501	0.599	0.318	0.060
	Vehicle age	-0.046	0.030	0.125	-0.006	0.033	0.856
	Avoid. Manoeuvre	0.312	0.283	0.270	-0.302	0.298	0.311

Note -- * estimation not possible ** based on reduced data se

==========
Page left blank intentionally
==========

Chapter 4 - Modelling the fatality risk for all accidents

George Yannis and Eleonora Papadimitriou, NTUA

4.1 The research questions

In this section, the WP5 in-depth accident data are exploited for modelling accident fatality risk, in terms of the probability of being killed in a road accident, in relation to person, vehicle and accident characteristics. The present database contains disaggregate casualty data, which are seldom available in the national or international databases, on the basis of sample datasets of seven EU countries. Apart from the obvious interest of modelling fatality risk in itself, this analysis can be useful in the following ways: first, the database includes new information which is not usually collected at national level, and second, the standard information that can also be found in national databases is considered to be more reliable in the in-depth database.

Despite these advantages of analyzing an in-depth accident database, several issues need to be addressed for obtaining unbiased and meaningful results. For instance, the fact that only fatal accidents are included in the database induces the "fatal accident size" effect, which was extensively analyzed in the previous sections, and needs to be controlled for in fatality risk modelling.

Moreover, several dependencies may be hidden in the data and need to be examined and accounted for. The accident process, for example, is typically hierarchically structured, as victims are nested into vehicles, which are in turn nested into accidents. Additionally, the data were sampled from different countries, and this may induce some additional variation in the examined effects. Therefore, all possible dependencies in the dataset need to be examined.

4.2 The analytical problem

Within the preliminary analysis of the WP5 data in terms of modelling fatality risk, a number of questions need to be addressed, in order to define the overall framework of this first set of results.

Police severity or SafetyNet severity?
In the WP5 database, accident severity is considered according to four categories: killed, seriously injured, slightly injured and not injured. For these categories, two scores are available: a "Police" score and a "SafetyNet" score. The definitions of - and differences between - these two scores are extensively presented in a separate section of the present document (see chapter 5). The "SafetyNet" score is considered to be more reliable, given that it is checked and

confirmed by the SafetyNet team. It thus is selected for use in the present analysis.

Binomial or multinomial analysis?
In the framework of preliminary analysis of the WP5 in-depth data, a less detailed classification of injury severity is opted for. In particular, the dependent variable in the analysis will be a binary one, indicating whether the person was killed or not. A more detailed (multinomial) analysis may be pursued in the next stages of the analysis, once the main data and methodological issues have been successfully handled.

Geographical hierarchy or accident process hierarchy?
The accident process follows a typically hierarchical structure, as persons are nested in vehicles and vehicles are nested in roads or in accidents, and there may be dependencies among these levels (Jones and Jørgensen, 2003). Moreover, given that the database cases were sampled from several different countries, there may be dependencies due to this geographical structure of the sampling (Yannis et al. 2007b). Given that it is not possible to conclude beforehand on which hierarchy (if any) is most likely to affect the data, both hierarchical structures will be tested in this analysis, and all significant effects will be accounted for.

Given the above overall framework, a number of additional issues need to be addressed, with respect to data handling.

Accident size bias
In the previous section, the particularities of analyzing fatal accidents databases were extensively discussed and it was concluded that ignoring or failing to account for the accident size effect in fatal accident databases can bring important bias to the results. In the framework of fatality risk modelling, this effect can be accounted for by incorporating a variable related to the size of the accident, such as the total number of persons or of participants in the accident, together with the number of occupants in the vehicle.

Treatment of missing values
As shown in Chapter 2, the WP5 database includes numerous variables describing the conditions of the accident, some of which can seldom be found in accident files. Unfortunately, a closer look at the database indicated that several of these interesting variables include a non negligible number of missing values. Part of this is due to the definition of the variable (e.g. impairment is only recorded for drivers, and so is the familiarity with the road network) and part of it is probably due to unavailability of the necessary information (e.g. vehicle age, seat belt use). Including these variables in the analysis results, during the statistical processing, in the exclusion of all the cases where the values are missing. In order to deal with this, two options were followed:

- for the variables that were only available for drivers (e.g. impairment, familiarity) a recoding was implemented, so that for each accident, the

condition of the driver is associated to all accident participants (e.g. driver impairment, driver familiarity)
- other variables with many missing values were not used in the analysis

Variables coding
An extensive effort was devoted to the appropriate coding of the explanatory variables, not only for the treatment of missing values, but also for an efficient statistical processing. Some numerical variables, such as person age, were converted to categorical ones. Moreover, more than one coding was applied for some variables, in order to have more options (i.e. in some cases, a detailed classification was inefficient, whereas a general one proved more useful). Finally, variables related to only one type of road user, such as impairment, were recoded so that, for the analysis, the condition of this road user is assigned to all other users in the same vehicle.

The final dataset resulting from the variables coding is presented in Table 4.1. This final dataset includes 817 cases appropriate for analysis.

Selection of variables for the analysis
From Table 4.1 it can be seen that an important number of variables is available for analysis, reflecting a lot of detailed information for each road user, vehicle, and accident. However, including all the variables in the model and rejecting those that are non-significant could bring extremely misleading results. In particular, it can be seen that several variables in the dataset are directly related to each other (e.g. "motorway" and "carriageway physically divided", "speed limit" and "area type"), whereas others are indirectly related (e.g. "traffic volume" and "weekend", "junction" and "number of participants"). Such relations could bring multicollinearity in the model, affecting the parameter estimates, their statistical significance and the overall performance of the model. In order to avoid this, the correlations between variables were tested beforehand. In addition to the statistical tests, predictors that are stringly related to each other were not included together in the model.

Table 4.1. Variables and values of the final dataset

Variable	Description	Values*	Missing
Cons	Constant term	1	0
Country	The country code	From 1 to 7	0
AccidentID	The accident ID number	From 1001 to 7021	0
VehicleID	The vehicle ID number	From 10011 to 70212	0
UserID	The road user ID number	From 100111 to 702121	0
Consequence	The severity of the accident	1: uninjured, 2: slightly injured, 3: seriously injured, 4: killed	0
Killed	The person is killed	1: yes, 0: no	0
Occinveh	The number of occupants in the vehicle	Form 1 to 9	0
CrashpartCat	The number of vehicles in the accident	1: one, 2: two, 3: more than two	0
Number of participants	The number of participant vehicles	1: one vehic;e, 0: two or more	0
Vehicecat	The type of vehicle	1: heavy, 2: car, 3: motorcycle, 4: pedal cycle, 5: pedestrian	0
Moto	The vehicle is a motorcycle	1: yes, 0: no	0
Vulnerable	The person is vulnerable	1: yes, 0: no	0
Heavy	The vehicle is heavy	1: yes, 0: no	0
UserClass	The user class of the person	1: driver, 2: passenger, 3: pedestrian	0
AgeCat	The age category of the person	1: 0-17, 2: 18-25, 3: 26-34, 4: 35-54, 5: 55-64, 6: 65+	0
senior	The person is over 65 years old	1: yes, 0: no	0
Female	The person is female	1: yes, 0: no	0
Unfamiliar	The person is not familiar with the road	1: yes, 0: no	340
Alcohol	There is suspicion of alcohol influence	1: yes, 0: no	270
Impairment	There is suspicion of impairment	1: yes, 0: no	270
Backseat	The person was at the back seat	1: yes, 0: no	174
SeatBeltU	The person used seat belt	1: yes, 0: no	180
AirbagU	The air bag was used	1: yes, 0: no	0
FirstEventCatControl	The category of the first event of the accident with respect to control of the vehicle	1: Loose control, 2: interaction with other vehicle, 3: interaction with vulnerable user, 4: other	0
FirstEventCatCollision	The category of the first	1: non collision, 2: collision	0

	event of the accident with respect to the type of collision	other vehicle, 3; collision with object non fixed, 4: collision with fixed object	
Traffic flow3	The traffic flow at the accident site	1: Light, 2: Normal, 3: Heavy	0
SpeedLimitCat	The speed limit at the location of the vehicle	1: Lower then 50, 2: 50-80, 3: 80-100, 4: 100-130	0
SLfaster50	The speed limit is higher than 50	1: yes, 0: no	0
SLfaster99	The speed limit is 100 or higher	1: yes, 0: no	0
Area	The area type	1: urban, 2: rural, 3: mixed	0
RuralArea	The area type is rural	1: yes, 0: no	0
Junction01	The accident is at junction	1: yes, 0: no	0
CarPhysDivided	The carriageway is physically divided	1: yes, 0: no	0
Motorway01	The accident is on motorway	1: yes, 0: no	0
Weekend	The accident is on weekend	1: yes, 0: no	0
Light	The accident takes place during daylight	1: yes, 0: no	0
Day	The accident takes place during day-time	1: yes, 0: no	85
AreaDamaged	The area most damaged of the vehicle	1: front, 2: left, 3: right, 4: back, 5: roof	0
MoreThanOneEvent	There was more than one event in the accident	1: yes, 0: no	0
Heaviest vehicle	The heaviest vehicle in the accident	1: heavy, 2: car, 3: motorcycle, 4: pedal cycle,	0
Lightest vehicle	The lightest vehicle in the accident	1: heavy, 2: car, 3: motorcycle, 4: pedal cycle, 5: pedestrian	0
Victimsum	The total number of vehicles in the accident	From 1 to 12	0
sumAlcohol	The total number of persons with alcohol influence in the accident	From 0 to 1	2
sumImpairment	The total number of persons with impairment in the accident	From 0 to 1	2
sumUnfamiliar	The total number of persons unfamiliar with the road in the accident	From 0 to 4	69
Multkilled	More than one persons in the accident was killed	1: yes, 0: no	0
KSIsum	The total number of	From 1 to 6	0

	persons killed or seriously injured in the accident		
MultKSI	More than one persons were killed or seriously injured in the accident	1: yes, 0: no	0
AgeDriver	The age of the driver of the vehicle	From 7 to 92	0
SeniorDriver	The driver of the vehicle is senior	1: yes, 0: no	0
AgeCatDriver	The age category of the driver of the vehicle	1: 0-17, 2: 18-25, 3: 26-34, 4: 35-54, 5: 55-64, 6: 65+	0
GenderDriver	The gender of the driver of the vehicle	1: female, 0: male	0
AlcoholDriver	There is suspicion of alcohol influence of the driver of the vehicle	1: yes, 0: no	0
ImpairmentDriver	There is suspicion of impairment of the driver of the vehicle	1: yes, 0: no	0
UnfamiliarDriver	The driver of the vehicle is unfamiliar with the road	1: yes, 0: no	119
AvoidMan	The driver made a crash avoidance maneuver	1: yes, 0: no	0
Braking	The driver was braking before the accident	1: yes, 0: no	0
Steering	The driver was steering	1: yes, 0: no	0
CrashParticipants	The number of vehicles in the accident	From 1 to 5	0
Van	The vehicle is a van	1: yes, 0: no	0
BusMinibus	The vehicle is a bus or a minibus	1: yes, 0: no	0
Truck	The vehicle is a truck	1: yes, 0: no	0
AgriculturalVehicle	The vehicle is an agricultural vehicle	1: yes, 0: no	0
MopedMotorcycle	The vehicle is a moped or a motorcycle	0: unknown, 1: moped, 2: motorcycle	0
Bicycle	The vehicle is a bicycle	1: yes, 0: no	0
TrainTram	The vehicle is a train or a tram	1: yes, 0: no	0
ShoeVehiclePedestrian	The person is a pedestrian	1: yes, 0: no	0

*All unknown values are coded as -999

4.3 Towards an analytical solution

The dependent variable is a binary variable (1: killed, 0: not killed) of the injury severity of each road user. The explanatory variables mainly include categorical variables, as well as a couple of continuous ones. A binary logistic regression model (i.e. a generalized linear model with a logit link function) was fitted to the data using the MLwiN statistical package. Initially, an "empty" single-level model (i.e. including a constant term only) is created (Model 1). This model will be considered as a baseline for comparing more analytical models in terms of fit.

However, standard estimation methods for discrete choice models, as the RIGLS (restricted iterative generalized least squares) method, can not provide reliable estimates of the likelihood statistic (Papadimitriou, Antoniou, & Yannis, 2007). For this reason, Bayesian modelling is exploited, which is based on simulation techniques and produces estimates on the basis of randomly generated numbers (Rasbash et al. 2000): After the standard estimation procedure, an MCMC (Markov Chain Monte Carlo) estimation method is applied by means of Metropolis Hastings sampling, in order to obtain more accurate (interval) parameter estimates and likelihood statistics (for more details see Papadimitriou et al., 2007). The starting values of the process are those obtained from the RIGLS estimation, whereas a diffuse (weakly-informative) Gamma prior is used (Browne, 2003). The results for Model 1, obtained after 15,000 iterations, are presented in Table 4.2.

In this framework, the possible hierarchies in the dataset are tested, in terms of geographical dependencies. A two-level model is considered, in which road users are nested into countries (Model 2).

$\text{logit}(\pi_{ij}) = \beta_{0j}\ \text{cons}$
$\beta_{0jk} = \beta_0 + u_{0j}$

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

Table 4.2 shows that the random variation at the country level is not significant after 400,000 MCMC iterations, whereas only a marginal improvement in the likelihood is obtained.

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 structure is considered, according to which persons are nested into vehicles and vehicles are nested into accidents (Model 3).

$\text{logit}(\pi_{ijk}) = \beta_{0jk}\ \text{cons}$
$\beta_{0jk} = \beta_0 + u_{0jk} + v_{0k}$

$u_{0jk} \sim N(0, \sigma^2_{u0})$
$v_{0k} \sim N(0, \sigma^2_{v0})$

In this case, because of the multilevel structure and also because of the fact that discrete response models tend to produce highly autocorrelated chains, results are stabilized after 400,000 iterations[4]. It can be seen in Model 3 of Table 4.2 that the random variation at the accident level (level 3) is not significant and the variation at the vehicle level (level 2) is marginally significant. The improvement of the likelihood statistic (residual deviance) compared to Model 1 is equal to 164, which is non significant for 145 residual degrees of freedom[5]. It is thereby indicated that the probability of a person being killed in a fatal accident does not vary systematically across different vehicles and / or different accidents (only a marginal user-within-vehicle dependency can be identified), and that the consideration of such a hierarchical structure does not improve the fit of the model.

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

Fixed effects	**Model 1**	**Model 2**	**Model 3**
Constant	-0.323 (0.072)	-0.293 (0.117)	-0.276 (0.095)
Random effects			
σ^2_{u0} (country level)		0.037 (0.100)	
σ^2_{u0} (vehicle level)			1.212 (0.577)
σ^2_{v0} (accident level)			0.011 (0.015)
-2*loglikelihood	1112.53	1110.97	948.16

Consequently, the testing and selection of the explanatory variables will be made within a single-level model. Random effects can then be tested again on the final model for confirmation.

The first step for building an analytical model involves accounting for the "accident size" effect. This can be achieved by including a variable representing the number of participants in the accident, and the categorical variable "number of participants" (one vehicle / two or more) is selected on that purpose. Model 4 is a single-level model that includes a constant term and this control variable; 115,000 MCMC iterations were required for estimating the posterior distributions of the parameters. The results presented in Table 4.3 indicate a significant negative effect of the "number of participants" control variable, suggesting that when there are two or more vehicles in the accident, each person involved in the accident has a lower probability of being killed, which is intuitive.

[4] Two diagnostics are used to ensure that the number of iterations is sufficient to obtain accurate estimates: the Raftery-Lewis (Nhat), indicating the necessary number of iterations for accurate quantile estimates of the parameter posterior distribution, and the Brooks-Draper (Nhat), indicating the necessary number of iterations for accurate mean estimates of the parameter posterior distribution. For details see Browne (2003).

[5] This indicates that 146 separate intercepts are calculated, which is a difference in effective parameters of 145 in relation to Model 1.

The next step concerns the selection and incorporation of explanatory variables in the model. All the variables of Table 4.1 were tested on that purpose. First, the correlations between variables were examined; an initial idea of the interrelations between variables was obtained by their joint incorporation in the model, and the correlations were in most cases further validated by statistical testing. A detailed presentation of the associated results for all variables is beyond the scope of this report. However, the following conclusions that were drawn from this process:

- "Traffic flow" is correlated with several variables, such as the number of vehicles in the accident (e.g. multi-vehicle accidents are more likely to occur in denser traffic), the area type (e.g. urban areas have higher traffic volumes), the motorway (e.g. motorways have higher traffic volumes) and the type of vehicle (e.g. it is possible that more traffic means more passenger cars).
- "Speed limit" is correlated with "area type" and "rural area" (e.g. urban areas have lower speed limits). It also appears to be associated with "divided carriageway" (when including both variables in the model, the otherwise significant effect of "divided carriageway" was eliminated).
- "Junction" is correlated with accident size variables, such as the number of vehicles in the accident (e.g. the majority of accidents at junctions are multi-vehicle accidents).
- "First event category control" is correlated with "vulnerable". In fact, one of the values of this variable is "interaction with vulnerable", and therefore when entering both variables in the model, the (otherwise significant) effect of "vulnerable" is completely eliminated. It is also noted that only one of the four categories of this variable is significant.
- Accordingly, "first event category collision" is correlated with "vulnerable", given that one of its values is "collision with object not fixed" (which always corresponds to a pedestrian or a cyclist). Including both variables in the model eliminates the effect of "vulnerable". It is also noted that only one of the four categories of this variable is significant.
- "Motorway" is correlated with "carriageway physically divided"
- "User class" is strongly correlated with "vulnerable", as it includes the case of pedestrians, and so is the case for the variable "pedestrian". Moreover, "user class" is correlated with "back seat" (drivers are always seated at the front).
- Apart from the correlations already mentioned, "back seat" appears to be associated with "vulnerable", because all the values corresponding to pedestrians are missing (the variable is not applicable for pedestrians). Including both variables in the model resulted in elimination of the effect of "vulnerable" and in the weakening of the (otherwise significant) "back seat" effect.
- "Age categories" is obviously correlated with "senior".
- In the same way, "Vehicle category" is correlated with all binary variables describing vehicle type ("heavy", "mopedmotorcycle", "busminibus" etc.).

These findings suggest that the respective sets of correlated variables should not be used jointly for modelling fatality risk. When choosing the predictors to be included in the final model, the two following criteria were taken into account:

their theoretical interest and the stability of the associated coefficients in the different model. It was also pursued to have as many simple (e.g. binary) variables in the model as possible.

Apart from the correlations tests, several other variables were removed from the final model, either because they were non significant or because they were not available for all road users. These include:

- "Vehicle category", "heaviest and lightest vehicle" were non significant
- "Female" was non significant
- "Area type", "weekend", "light" were non significant.
- "Alcohol", "impairment", "seat belt", "unfamiliarity with the road": This group of variables has several particularities; they are available only for drivers and pedestrians, and some of them have a high number of missing or unknown values.
- All the variables related to the driver of the vehicle associated with the person (age category driver, gender driver, senior driver, alcohol driver, impairment driver, unfamiliar driver) were non significant: It is thereby indicated that, given that one ends up in a fatal accident, the characteristics of the driver do not affect the probability of another road user (passenger) in the same vehicle being killed..

The final model (Model 5), including 6 significant variables, is presented in Table 4.3. It took 140,000 MCMC iterations to obtain stabilized posterior parameter estimates. This model is significantly improved compared to the "empty" Model 1, with residual deviance equal to 283.91 and 6 residual degrees of freedom.

Table 4.3. Accident fatality risk model building (single level model)

Fixed effects		**Model 4**	**Model 5**
Constant	1.000	0.357 (0.152)	-1.125 (0.286)
Number of participants	One Vehicle	.	.
	Two or More	-0.876 (0.172)	-0.791 (0.207)
Senior	> 65 years old		1.157 (0.284)
	Younger		.
Number of Events	More than one		1.827 (0.219)
	One		.
Divided carriageway	Yes		-0.705 (0.241)
	No		.
Vulnerable	Yes		4.103 (0.521)
	No		.
Braked	Yes		-0.532 (0.193)
	No		.
-2*loglikelihood		1087.69	828.62

As regards the parameter estimates, these suggest the following findings:

- As mentioned previously, accidents with more participants have a lower overall severity.
- Persons with more than 65 years of age have a higher probability of being killed in an accident, obviously due to their physical vulnerability.
- The probability of being killed increases when there is more than one event for the vehicle in the accident.
- Accidents on divided carriageways are less severe for the persons involved
- Vulnerable road users (i.e. pedestrians) have significantly higher probability of being killed in road accidents.
- The probability of being killed is reduced when the vehicle braked before the collision.

The proposed final model includes a limited yet sufficient number of variables, with stable and significant parameter estimates. Although it was initially indicated that no significant random effect was to be expected, random intercept models were attempted at for confirmation. First, a two-level model with users nested into vehicles was considered, including the seven fixed effects of Model 5, plus a random intercept. After 246,000 MCMC iterations, a non significant estimate of the random intercept was obtained (0.072 with a standard error of 0.202); it is noted that the effect is also non significant when applying the RIGLS estimation method. Moreover, a two-level model with persons nested into countries was considered; the random intercept was non significant (0.075 with standard error 0.240) after 200,000 MCMC iterations.

It is noted that, although non significant, the random intercept estimates of the two models are very similar. It appears that the minor higher level variation that is present in the data can be equally captured by a vehicle- or a country- higher level. Nevertheless, the above results show that single level models are efficient for modelling fatality risk in the particular dataset. The analysis of a larger dataset might produce more significant results with respect to hierarchical dependencies, allowing for conclusions to be drawn.

4.4 Discussion

In this section, preliminary results of modelling accident fatality risk, as the probability of being killed in a road accident, were presented. Particular emphasis was given on the "accident size" effect, which needs to be dealt with when working with fatal accidents only. Moreover, a lot of effort was devoted on revealing the relationships and correlations between the numerous available variables and selecting those that were appropriate for the models.

The final model includes six explanatory variables (plus the intercept) and is significantly improved compared to the "empty" model. It was shown, however, that this limited number of variables bear various additional effects, due to their correlations with other variables. For instance, the variable "carriageway

physically divided" is strongly related to motorways, which are in turn related to traffic volume, in turn related to speed limit, vehicle type and area type, and so on. From this whole set of associated variables, "carriageway divided" was selected as the most powerful and well behaving one; this variable reflects to some degree the effect of all the variables in the set.

It is also interesting to note that most of the variables of the final model concern characteristics of the accident and the related events. In particular, accident and event characteristics seem to be quite more powerful than user and vehicle characteristics, especially those being typically considered in road accident analyses, such as age and vehicle type. It is possible that the detailed additional information available in an in-depth database is a stronger determinant of fatality risk than the standard variables available in national databases. However, it should be underlined that "pedestrian" and "senior" have very strong effects on fatality risk and are the only person characteristics that seem to outperform the accident characteristics in this analysis.

Another interesting finding of the present preliminary analysis concerns the fact that driver characteristics do not appear to affect the consequences of the accident for the other road users involved in the accident. A more focused analysis, examining drivers and passengers only would be interesting. Moreover, behavioural variables such as alcohol, impairment and seatbelt use were not found to be significant in this fatality risk analysis; it is likely that such variables would be more meaningful for accident risk analysis, however, their non significance may be also partly due to the important number of missing values for these variables.

Finally, although no significant dependences, calling for multilevel modeling, were identified in this particular dataset, it is very important to always examine the possible dependences that may be hidden in the data.

Chapter 5 - The fatality risk in car-car accidents

Heike Martensen and Emmanuelle Dupont (IBSR)

5.1 The research question

In the previous section, a large number of variables that could affect the risk of dying given that one is involved in a fatal accident were analysed. A global picture of those factors that can effectively predict the fatality risk factors is given. One of the factors identified in the previous chapter clearly stands out in the size of its effect: <vulnerable>, i.e. whether a crash participant is a pedestrian or a pedal cyclist, or whether he/she is the occupant of a motor vehicle. It was shown that vulnerable road users have an increased risk as compared to the occupants of motor vehicles, so strongly that whenever motor-vehicle occupants and vulnerable road users are involved in fatal accidents together it is the vulnerable road user who dies.

In this section we will zoom in on the occupants of motor-vehicles, more specifically cars occupants and evaluate the factors that contribute to their fatality risk. The question asked is basically: When two cars crashed; which person, vehicle or event characteristics are responsible that this accident became fatal?

5.2 The analytical problem

The goal in collecting accident data is to learn from the past and gain information that can help to prevent future accidents. The main 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, the survivors (and their vehicles) in fatal accidents serve as a control group for the fatalities in the present analysis. However, extreme caution is necessary when interpreting the results. In the first place, it is important to ensure that the risk that survivors and fatalities were running is indeed comparable. As presented in Chapters 2 and 4, this is not the case for different types of road users (i.e. vulnerable road users vs. vehicle occupants). Moreover, this is also not the case for car occupants who were involved in accidents with different vehicle types.

As mentioned in Chapter 3, accidents in the fatal accident database that involved a vulnerable road user followed the same pattern: The vulnerable road user was killed and there were no victims among the car occupants. This means that the risk a car occupant runs when involved in an accident with a vulnerable road-user is not comparable to that of car occupants involved in an accident with another car. In a similar way the risks of car occupants involved in accidents with either motor cycles or heavy good vehicles are difficult to

compare to those of occupants in car-car accidents. The first show a much decreased and the latter a much elevated risk.

An accident between two cars offers maximal comparability between the fatality and the survivors. All other accidents were consequently excluded from the analysis. All accidents with two participants, both of which had to be cars, were selected, leaving 216 victims from 67 accidents.

The fatality risk was modelled in a binomial model. The dependent variable was a binary one, indicating for each victim whether he or she survived the accident (0) or was killed in it (1). The explanatory variables reflected the hierarchical structure of accident data: they were road-user, vehicle, or accident variables. For hierarchically structured data, it is possible that there is significant variation at the higher levels (Martensen and Dupont, 2007; Jones and Jørgensen, 2003). For the accident data analysed here, this could be variation at the vehicle level, for instance, indicating that victims within the same vehicle have a more similar fatality risk than those in other vehicles. The Fatal Accident Investigation Database is also collected within different countries which might also account for some variation in the data. It is important to check for such higher-level variation and if it is present, to take it into account in the statistical model applied. The data were therefore tested for geographical and accident process hierarchies. Eventually, a two level binomial model was implemented as described below. The model was estimated with the MCMC algorithm as described in the previous section (Browne et al. 2001, see also Papadimitriou et al., 2007).

The price for the good comparability of the fatalities and the survivors is a very restricted sample. This is especially problematic in this preliminary analysis where the data set is not yet complete. Consequently, the power of this analysis is quite low and non-significant effects should be interpreted with caution. It cannot be assumed that a variable that does not become significant does not play a role in the accident process. In such circumstances, it is advisable not to be too conservative in selecting the variables to be included in the model or not.

The variables that were considered are presented in Table 5.1. Some of the variables listed below are based on the same variable from the original database and differ only in the way the original categories were recoded (e.g., AgeVictim, AgeCatVictim, & SeniorVictim). These variables were considered alternatively and eventually the most parsimonious way of coding was chosen (i.e. the fewest number of categories that could still explain the data well).

Table 5.1 Variables in the car-car analysis

Level/Type	Variable	Description	Values
Dependent Variable	Killed	The person is killed	1: yes, 0: no
Identifiers	Constant	Constant term	1
	CountryID	The country code	1 to 7
	AccidentID	The ID number	1001 to 7021
	VehicleID	The vehicle ID number	10011 to 70212
	UserID	The road user ID number	100111 to 702121
Road-User Variables	UserClass	The user class of the person	1: driver, 2: passenger, 3: pedestrian
	AgeVictim	The (z-transformed) age of the victim	-1.74 to 2.76
	AgeCatVictim	The age category of the person	1: 0-17, 2: 18-25, 3: 26-34, 4: 35-54, 5: 55-64, 6: 65+
	SeniorVictim	The person is 65 or older	1: yes, 0: no
	$AgeVictim^2$	The square of the victims age	0 to 7.5
	Female	The person is female	1: yes, 0: no
	Backseat	The person was at the back seat	1: yes, 0: no, -999: unknown
	SeatBelt	The person used seat belt	1: yes, 0: no, -999: unknown
	SeatBelt01	The person used seat belt	1: yes, 0: no or unknown
	AirbagUsed	The air bag was used	1: yes, 0: no, -999: unknown
Vehicle Variables	OccupantsInVehicle	The number of occupants in the vehicle	1 to 9
	VehicleAge	Age of vehicle in years (z-transformed)	-1.83 to 5.27
	ABS	Does the car have ABS?	1: yes, 0: no or unknown
	AgeDriver	The age of the driver of the vehicle in years (z-transformed)	-1.14 to 2.66
	AgeCatDriver	The age category of the driver of the vehicle	1: 0-17, 2: 18-25, 3: 26-34, 4: 35-54, 5: 55-64, 6: 65+

	SeniorDriver	The driver is 65 or older	1: yes, 0: no
	AgeDriver^2	AgeDriver squared	0 to 7.08
	GenderDriver	The gender of the driver of the vehicle	1: female, 0: male
	AlcoholDriver	There is suspicion of alcohol influence of the driver of the vehicle	1: yes, 0: no
	ImpairmentDriver	There is suspicion of impairment of the driver of the vehicle	1: yes, 0: no
	UnfamiliarDriver	The driver of the vehicle is unfamiliar with the road	1: yes, 0: no
	Manoeuvre	The driver-manoeuvre at the beginning of the accident	1: straight, 2: bend, 3: turning, 4: changing lanes, 5: overtaking, 6: loss control, 7:illegal, 8: reversing
	Manoeuvre5	The driver-manoeuvre at the beginning of the accident	1: driving straight or in bends, 2: turning, 3: pulling out, 4: loss control, 5:other
	ExecutedManoeuvre	Any manoeuvre has been executed.	1: yes, 0: driving straight or "lost control"
	LostControl	Driver-manoeuvre was "lost control"	1: yes, 0: no
	AvoidanceManoeuvre	The driver made a crash avoidance maneuver	1: yes, 0: no
	Braking	The avoidance manoeuvre was braking	1: yes, 0: no
	Steering	The avoidance manoeuvre was steering	1: yes, 0: no
	AreaDamaged	The area most damaged of the vehicle	1: front, 2: left, 3: right, 4: back, 5: roof
	AreaDamaged3	The area most damaged of the vehicle	1: front, 2: side, 3: other
	FrontDamaged	The area most damaged of the vehicle was the front	1: yes, 0: no
Accident Variables	MoreThanOneEvent	There was more than one event in the accident	1: yes, 0: no
	ExecManoeuvreAcc	Any manoeuvre has been executed by any driver in accident	1: yes, 0: no

	LostControlAcc	Control has been lost by any driver in accident	1: yes, 0: no
	AvoidManoeuvreAcc	Any driver made an avoidance manoeuvre	1: yes, 0: no
	UnfamiliarAcc	Any driver was unfamiliar	1: yes, 0: no
	SpeedLimitCat	The speed limit at the location of the vehicle	1: Lower then 50, 2: 50-80, 3: 80-100, 4: 100-130
	SLfaster50	The speed limit was higher than 50	1: yes, 0: no
	Area	The area type	1: urban, 2: rural, 3: mixed
	RuralArea	The area type was rural	1: yes, 0: no
	Junction	The accident was at junction	1: yes, 0: no
	Carriageway	The carriageway was physically divided	1: yes, 0: no
	Motorway	The accident was on motorway	1: yes, 0: no
	Weekend	The accident occurred in the weekend	1: yes, 0: no
	Light	The accident took place during daylight	1: yes, 0: no

Note-- N=216. There were no missing values

As described in the previous section, there is no straightforward way to select the variables to be included in the model. In the first place variables were grouped on the basis of theoretical ideas according to their overlap. (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 outside rural areas than inside). Within each of these groups, it was then tested which variable(s) could best predict the fatality risk. Subsequently, the “winners” from each group were put into one model together. This revealed many unexpected collinearity problems. It was thoroughly explored which variables influenced the significance of which other variables (among all variables, not only the ones that had been identified as “winners”) when taken up into the model together. These exploratory analyses were done with the RIGLS estimation procedure, which provides parameter estimates but no reliable goodness-of-fit measure for binomial models. Finally the different candidate models were systematically compared to each other on the basis of their DIC-score (Deviance Information Criterion, Spiegelhalter et al. 2002). The DIC criterion is a value that can be calculated for a whole model (i.e., for all the predictors it includes) on the basis

of the results of MCMC estimation. It favours models with a good fit but it penalizes complexity. Of two models, the model with the lowest DIC is to be preferred because it explains the data better with fewer variables.

A Wald test was also performed for each predictor within a model. This 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 is at least larger than 0. It turned out that the results of the DIC sometimes were in conflict with those of the Wald test. In other words, the DIC criterion sometimes indicated that a model with a variable that was non-significant according to the Wald test was to be preferred above the same model not containing that variable. Generally speaking, both criteria should be met when including a variable into a model. In the present case, however, given the few number of cases and the low power of this preliminary analysis, it was considered important not be too stringent for variable inclusion. Therefore, we based the decision to include a variable into the model on the DIC only. 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 were kept. This means that all variables presented in the final model increased its fit to the data (more than they increased its complexity).

5.3 Results

At first, the structure of the basic model ("the empty model") in which the variables have been entered will be described. Subsequently, the results for the best fitting model are given and finally the results for the variables that have not been included into the model will be summarized.

5.3.1 The empty model

The empty model contains only a constant which provides the intercept and the variable <OccInVeh>. As described earlier it is important to correct all estimates for the accident size bias. The more persons involved, the higher the chance for each of them not to be the fatality. In the subset that is analysed here, the number of participants was always 2, so the only size factor that needed to be corrected for was the number of occupants in each vehicle. As this factor was selected a-priori, it was included into the empty model. This empty model was at first estimated with a 4-level hierarchical structure. The first level corresponds to the road-user, the second to the vehicle, the third to the accident, and the fourth to the country in which the accident was recorded. The variance of the first level is given in a binomial model, the variance of the other three levels (for detailed description see Vanlaar, 2007) was estimated with the MCMC estimation technique (see Papadimitriou et al., 2007). The results after 200,000 iterations are given in Table 5.2.

Table 5.2 Empty four-level model

Variables		β/σ	SE	Chi2	p	exp
Predictors:	Constant	0.437	0.472	0.857	0.355	0.437
	OccpantsInVehicle	-0.468	0.159	8.664	0.003	-0.468
Levels:	Road-user	Fixed				
	Vehicle	0.267	0.989	0.073	0.787	
	Accident	0.032	0.05	0.410	0.522	
	Country	0.847	0.978	0.750	0.386	
Deviance	243.86					
DIC	272.98					

There was some variation at the country, accident, and vehicle levels, which was however not significant. The safer method to decide whether a particular level is necessary in the model is however, to compare the DICs for the model with and without this level. Generally, the DIC scores of the different models were all very similar (271.71 to 274.00), indicating that the complexity added by the extra level and the improvement of the fit were just about in balance for the three possible extra levels. The model with the lowest DIC score was the model with a country level only.

5.3.2 The final model

The two level empty model (with country as only extra-level) was used to introduce possible predictors of the fatality risk in a fatal car-car accident. The model with the lowest DIC is presented in Table 5.3.

5.3.2.1. Differences between countries

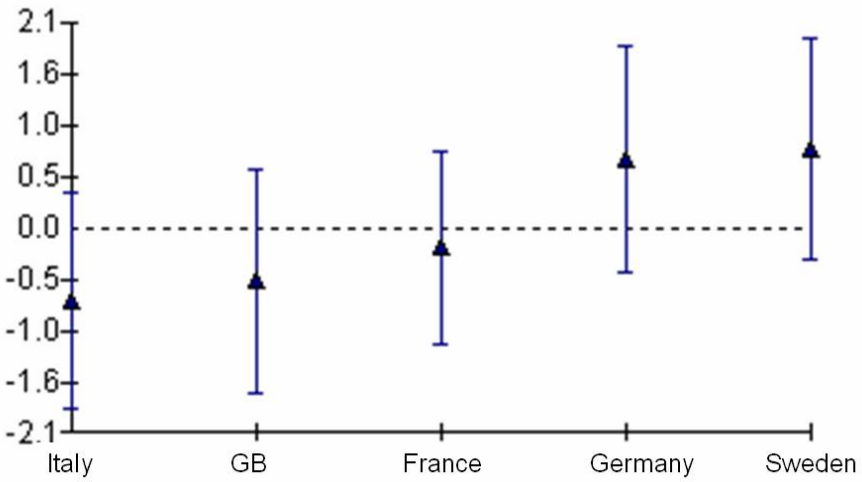
The five different countries in which the accident data have been collected were included as a potential source of variation in this model. The aim is to examine whether the fatality risk differs across countries, without making any further attempt to determine which factors would be responsible for such a variation. In other words, unexplained, or residual, variation of the fatality risk is examined as a function of the country in which the data have been recorded. Should the fatality risk not be affected by the country, then the results should indicate that the residuals do not significantly differ across countries. Figure 5.1 displays the residuals for each of the countries. These residuals can be interpreted as differences between the proportion of victims killed in the fatal accidents after all variables in the model have been accounted for. Although there is some variation, no country differs significantly from any other.

Table 5.3 Two-level model of fatality risk for car-car accidents

		β/σ	SE	chi2	P	Exp
Predictors:	Constant	2.845	1.265	5.058	0.025	17.202
	OccupantsInVehicle	0.788	0.201	15.370	0.000	2.199
	Motorway	-3.434	1.288	7.108	0.008	0.032
	ExecutedManoeuvreAcc	-1.221	0.494	6.109	0.013	0.295
	LostControl*	1.387	1.585	0.766	0.382	4.003
	EventNr>1	1.498	0.537	7.782	0.005	4.473
	Braking	-2.294	0.586	15.325	0.000	0.101
	FrontDamaged	-2.072	0.573	13.076	0.000	0.126
	AgeVehicle	1.002	0.270	13.772	0.000	2.724
	AgeVictim^2	0.400	0.183	4.778	0.029	1.492
	SeatBelt01*	-0.724	0.626	1.338	0.247	0.485
Level:	Country*	2.421	6.398	0.143	0.705	
Deviance	172.853					
DIC	187.96					

Note -- * The effect of these predictors was not significant according to the Wald test, yet, they were included into the model because this led to a decrease of the DIC score.

Figure 5.1 Differences between countries in proportion victims killed in fatal accidents



Note – The bars present 95% confidence intervals

5.3.2.2. Variables in the final model

Number of Occupants

The variable <OccupantsInVehicle> indicates the number of occupants in the vehicle and was entered a priori into the model to correct for the accident size bias. As discussed earlier, fewer occupants increase the chance of each

occupant to be the fatality and vice versa. This is reflected in a significant positive coefficient for this variable.

Number of Events

The chain of events for vehicles without a fatality are most often recorded in only one event. Participants with at least one fatality are most often recorded in two or more events. This result is reflected in the positive coefficient for this variable and replicates the result from the model for all road-users (Section 4.1). It is unclear, however, what the reason for this result is. On the one hand it is possible that more complex chains of events for a vehicle make it more likely that this is the vehicle that contains the fatality. On the other hand, the difference could also be the result of a reporting bias. It would be natural for the accident investigators to devote their attention most strongly to the vehicle that contains the fatality and therefore describe their chain of events in more detail.

Motorway

Generally speaking there are very few accidents on motorways included in the fatal-accident database (which is probably a much better indicator for the safety of motorways than the result explained here, but needs to be compared to the appropriate exposure data). The negative coefficient for the variable in this analysis indicates that the fatality risk for occupants in fatal car-car accidents is lower on motorways than elsewhere. This result is in agreement with the results from the accidentsize analysis in Chapter 3 and just like there the variable <Motorway> is not significant by itself. It becomes significant only when entered together with the variable <FrontDamaged>. The reason for this is that the motorway-effect can best be seen when separately considering those cases where the area of most damage was the front of the car and those with another area of most damage (see Appendix 5B for a numerical explanation of the interaction between <Motorway> and <FrontDamaged>).

Driver Manoeuvre

This variable has been examined at two levels, at the accident level and at the vehicle level. At the accident level, it has been aggregated to the variable <ExecutedManoeuvreAcc> which indicates whether any driver in the accident had just been executing a manoeuvre at the beginning of the accident. This variable has a significant negative effect, indicating that accidents where none of the participants had been executing a manoeuvre (i.e. they had either been driving along the road or had been losing control) show a higher proportion of people killed than accidents in which at least one of the participants had been executing a manoeuvre. It is important to note, however, that this variable is only significant when entered together with the variable <FrontDamaged>. The two variables are related because accidents in which no manoeuvre had been executed lead more often to frontal damage than other accidents. Generally speaking accidents with frontal impact are less dangerous than those with side impact. However, this effect is modified by the variable <ExecutedManoeuvreAcc> because two cars driving straight into each other

(i.e. none of them executed a manoeuvre) is an exception to the rule that frontal impacts are not very dangerous to the occupants.

In the database, the variable <DriverManoeuvre> is coded as a vehicle variable and it has been analysed at the vehicle level in several ways (see section 4.3.3 for other results for this variable). One of its categories <Lost Control> was coded into a dichotomous variable (1 if <DriverManoeuvre> is "Lost Control" and 0 for all other categories). This variable was included into the model because it increased the fit of the model more than it increased its complexity. The variable is however, not significant in terms of the Wald test. The positive coefficient suggests that occupants in vehicles of which the driver lost control have a larger chance to be the fatality than the occupants in those vehicles of which the driver did not loose control. But this result would have to be confirmed with a larger sample.

Avoidance manoeuvre

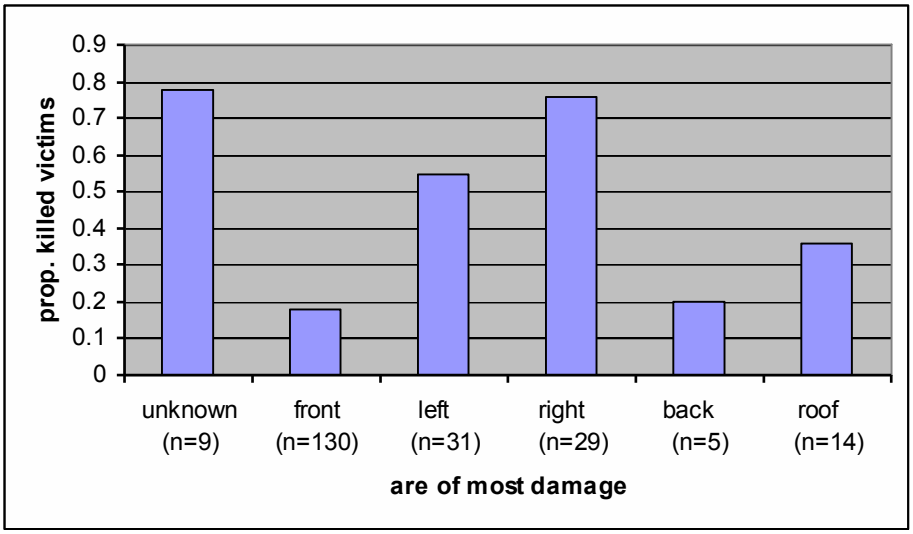
The variable <AvoidanceManoeuvre> indicates whether the driver of the vehicle executed an avoidance manoeuvre and whether this was braking, steering, braking & steering or another. This variable had been recoded into a number of different variables. At the vehicle level: <AvoidManoeuvre> (indicating whether the driver had executed any avoidance manoeuvre), <braking> (indicating whether or not he braked) and <steering>. At the accident level it was coded <AvoidManoeuvreAcc> indicating whether any driver had executed any avoidance manoeuvre. All these variables, except <steering> have a negative effect, suggesting that avoidance manoeuvres reduce the fatality risk. The variable <braking> increases the fit most, indicating that it is the most effective avoidance manoeuvre.

Area of most damage

The area of most damage is an importing factor in determining the fatality risk for the occupants in a particular vehicle. In the preliminary dataset analysed here, no significant difference was between right or left impacts – neither between the absolute number of occurrences nor between the fatality risk associated with each side. For this reason side-impacts were analysed as one category[6]. In Figure 5.2, it can be seen that side impacts are more fatal than front and back impacts. The category "unknown" also shows a very high proportion of killed occupants. These are mostly cases where the car was damaged in several ways making it impossible to tell which area was the one most damaged.

[6] This is also true when considering cases from the UK (left-hand driving) and from other countries (right-hand driving) separately.

Figure 5.2 Proportion of victims killed according to area of most damage

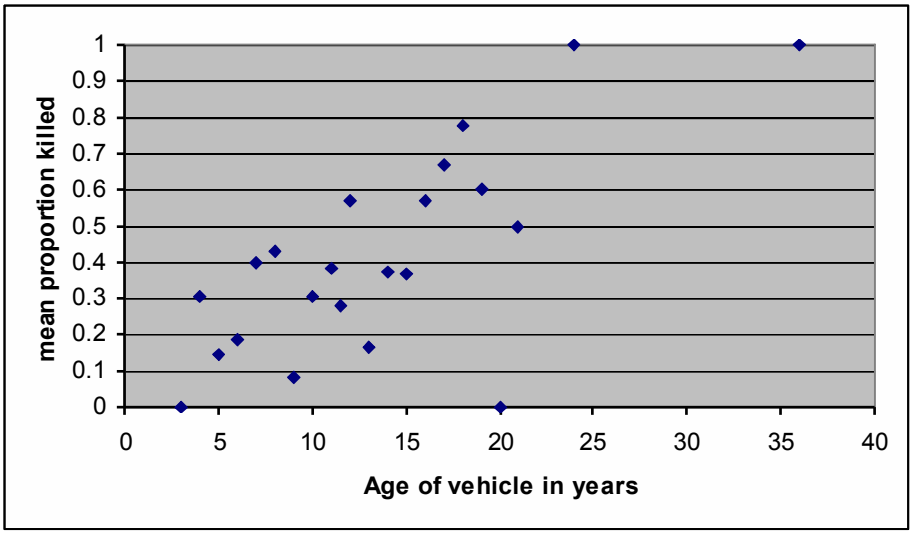


In the model a variable differentiating only between vehicles with frontal damage and all other categories was the most efficient predictor.

Vehicle Age

The negative coefficient for the variable <VehicleAge> indicates that persons involved in a fatal accident have a higher chance of being the fatality when the car they are seated in is older. In Figure 5.3, the mean proportion of occupants killed is plotted each car age in years.

Figure 5.3 Mean proportion of occupants killed by vehicle age

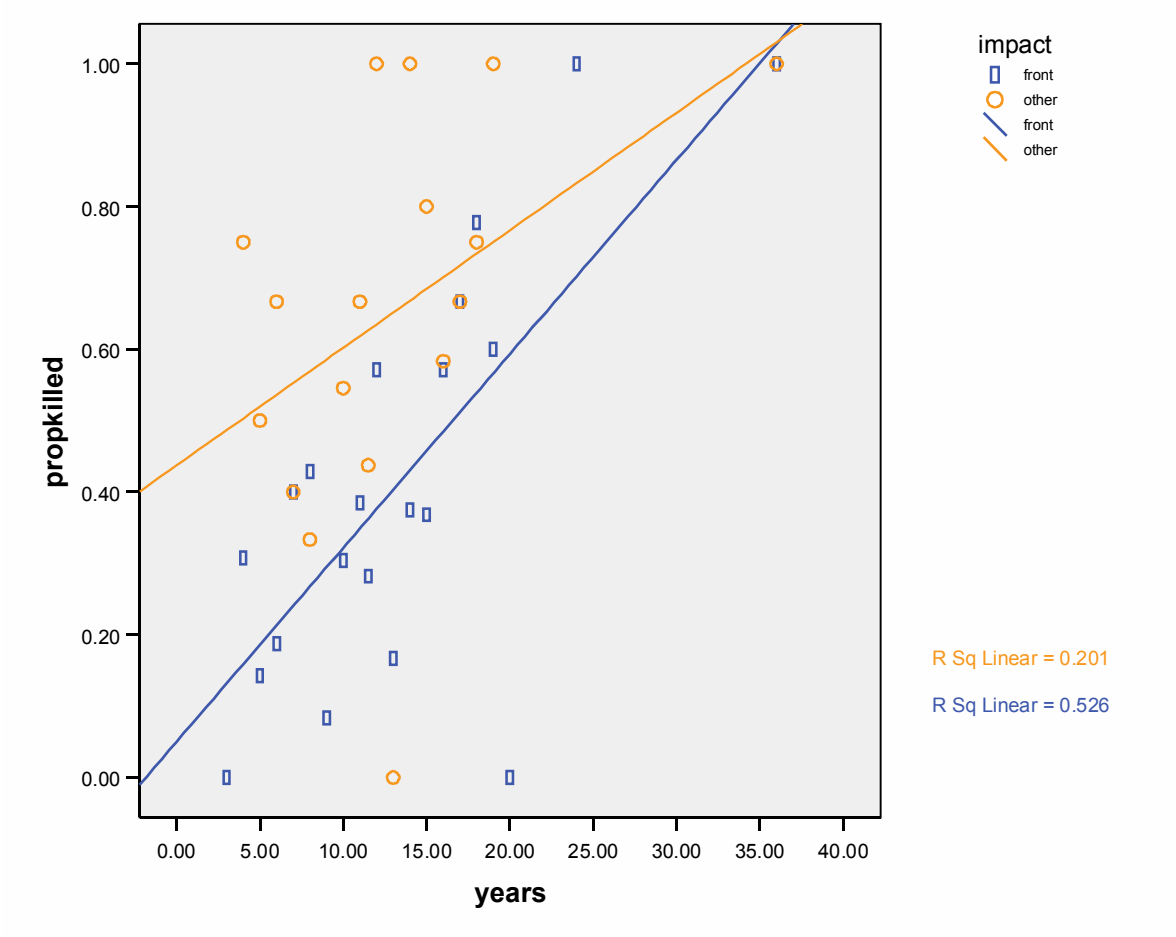


It can be seen clearly that the relation between age of vehicle and fatality risk, although not perfect, is quite strong. Note however, that this relation only addresses the fatality risk associated with older or newer cars *once one is involved in an accident*. It may very well be that drivers of older cars do not run a higher risk altogether, e.g. because they drive more carefully and do not get involved in severe accidents as often in the first place. This question cannot be addressed on the basis of the present data. In this analysis we have to restrict ourselves to the risk assessment for those car-occupants who are involved in a fatal accident and for those cases, the data show clearly that the protection a car offers its occupants strongly decreases with its age.

Interaction between vehicle age and area of most damage.

In Figure 5.4 we have split the relation between vehicle age and fatality risk (as presented in Figure 5.3) according to the area of most damage. The blue blocks indicate the mean proportion of killed occupants of cars with mostly frontal damage. For this group we see a strong decrease of risk for newer cars. The orange circles indicate the relation between vehicle age and fatality risk for cars where the most damaged area was not the front. The difference is striking in several ways: a) The orange circles lie generally higher than the blue blocks, indicating the increased fatality risk for non-frontal impacts. b) The orange circles are not so close to their regression line, indicating a much weaker relation between vehicle age and fatality risk, and c) the orange line is less steep, suggesting that the fatality risk due to side impacts has not seen as strong a decrease for newer cars as the fatality risk due to front impacts.

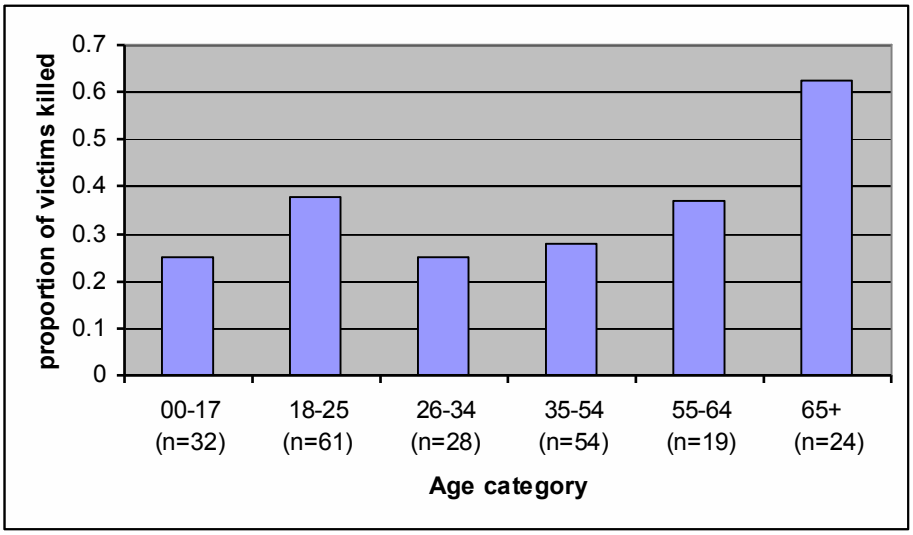
Figure 5.4 Improvement of vehicle safety with respect to frontal and other damage



Age of the victim

In Figure 5.5, the proportion of occupants killed for six age groups is plotted. Young car occupants (18-25) who form the largest group in the fatal-accident database have a higher chance to be the fatality than occupants of medium age. Old car occupants (especially those above 65) show an even more elevated risk to be the fatality.

Figure 5.5 Proportion of victims killed by age category[7]



Seniors above 65 show a strongly increased fatality risk when involved in a severe accident. However, this effect is confounded with the area of most damage and becomes nonsignificant, once <FrontDamaged> is included into the model (see 5.2.3.2 for a detailed explanation). When ignoring the 65+ category, there is still an approximately U-shaped relation between age and fatality risk. This relation is captured by the variable <VictimAge^2>, which stays significant in combination with <FrontDamaged>. The variable age was not significant, but the squared age was.

It can be assumed that there are different reasons for the elevated risks for the two extreme age groups (young and old people). As mentioned in the previous section the higher fatality risk for older people might be attributed to them being more frail (see however the discussion of the variable “senior” in section 1.3.3), while the elevated risk for young people is probably rather due to behavioural aspects. It is questionable however, whether these aspects concern the driving behaviour. In that case one would expect the age of the driver to be the more relevant variable. The age of the driver is highly correlated to that of the victim (r=.79**) and consequently shows similar effects. <DriverAge> does, however, not have a significant effect, while <VictimAge> does.

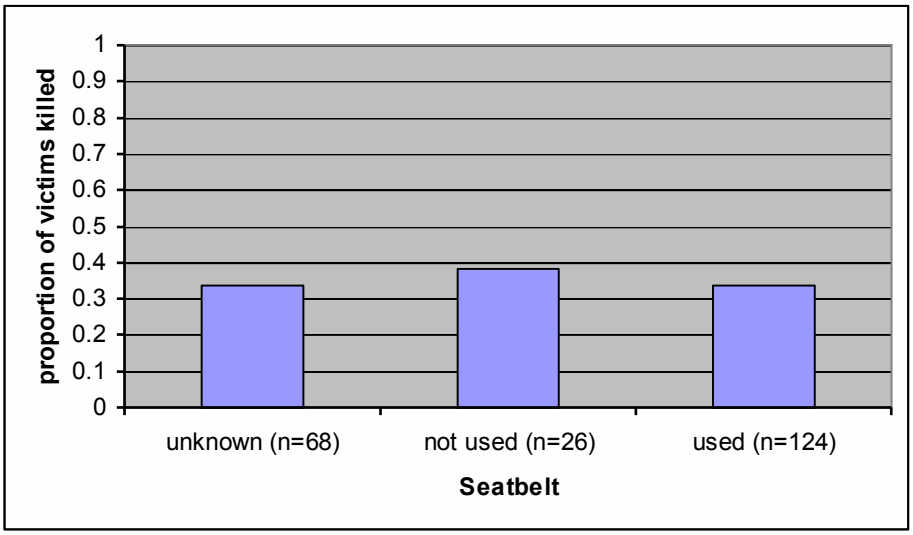
Seatbelt use

Car occupants who wore a seatbelt had a slightly lower probability to be the fatality than those car occupants who did not. The variable seatbelt improved the fit of the model more than it increased its complexity. Nevertheless, the

[7] Note that the proportions presented here, do not necessarily agree with those presented in Chapter 2, as the data analysed here form only a subset (the accidents involving two cars) of the complete dataset described in Chapter 2.

positive coefficient is not statistically reliable. In Figure 5.6 it can indeed be seen that the effect of seatbelt use is quite small.

Figure 5.6 Proportion of victims killed by seatbelt use



The reason for this might be that there is only a very small group of whom it is known that they were not wearing a seatbelt. This small number (n=26) does not allow a very reliable estimation of the fatality rate in this group.

5.3.2.3. Alternative variables

When possible predictors of the fatality risk are related among each other, they cannot be taken up into the model simultaneously. In the final model, from a group of related variables those with the greatest power to predict whether a person died or not was chosen. However, having the greatest predictive power does not necessarily mean that the variable represents the true reason for an increased or reduced fatality risk. Although the variable that actually lies at the basis of the effect will usually be related to the dependent variable most strongly, it is also possible that a variable that actually combines a number of conceptually different reasons ends up to be the better predictor.

Because of this uncertainty, we will also describe the variables that show a significant relation with the fatality risk when considered by themselves but not in combination with other variables. All variables are listed in Appendix 5A with their results when entered into the empty model by themselves.

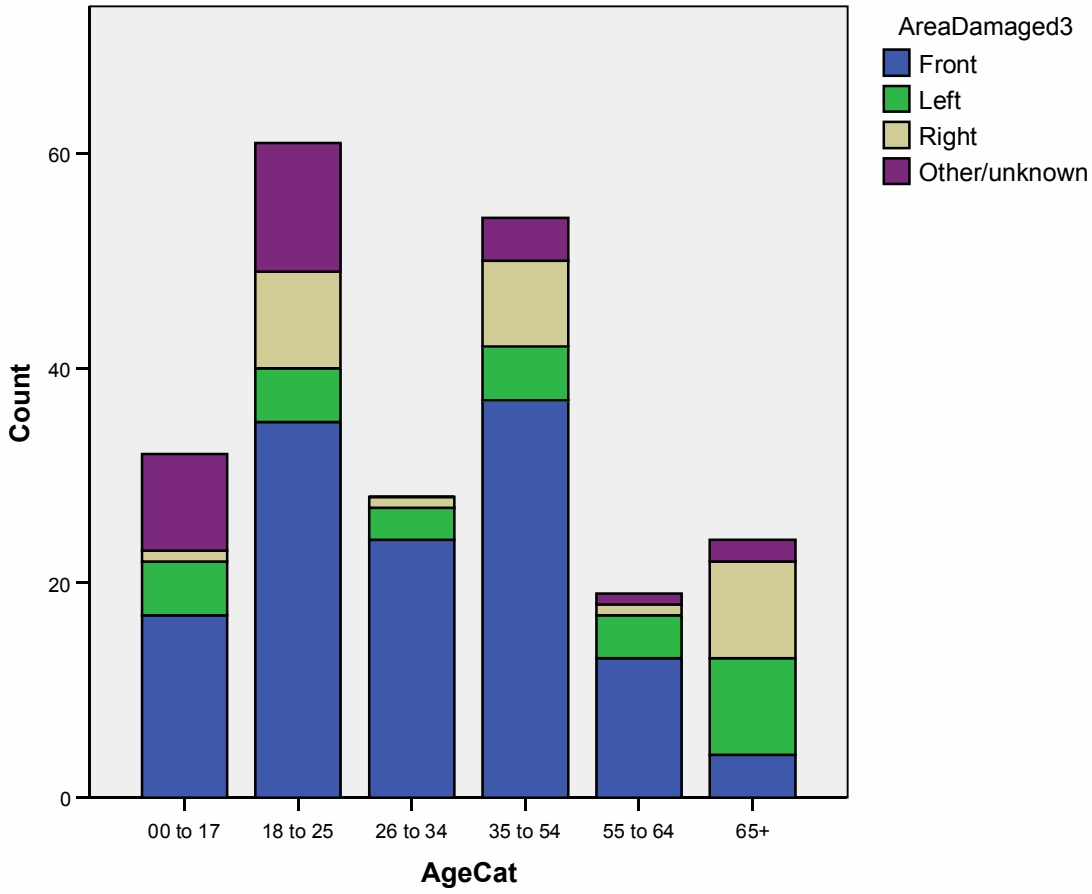
Age

As we have seen already, if you are 65 years or older your chances of dying when involved in a fatal accident are higher than if you are younger than that. By itself, the variable <senior> shows a highly significant relation with the

probability of being killed. It becomes nonsignificant, however, once <FrontDamaged> is taken up into the model.

In Figure 5.7, the number of victims (all persons involved in the accidents) is split up by age category and the area of most damage. It can be seen that the category of seniors (65+) has a particularly large proportion of victims who were seated in vehicles where the side (left as well as right) was the area most damaged (beige and green parts of the bar)[8]. We have seen that damage to the side of the vehicle is much more harmful for the occupants than damage to the front. Consequently it is impossible to say whether the seniors are so much at risk because of their age or because they tend to be seated in vehicles with side impacts or whether a third variable that is not considered here is actually responsible for both effects.

Figure 5.7 Number of victims by age and vehicle area most damaged



[8] The same pattern is found when investigating the age of the driver and the area of most damage.

Airbag use

Victims for whom there was an airbag present and actually deployed had a lower risk of dying in the fatal accident they had been involved in than victims for whom there either was no airbag or where it did not deploy. This initially significant relation becomes however nonsignificant once the variable <FrontDamaged> was taken up. This can be explained by the fact that airbags deployed more often in frontal impacts. The effect of airbag-deployment can therefore be explained as an effect of the generally less dangerous accidents where the area of most damage was the front of the car.

ABS

The presence of ABS in the vehicle decreases the probability that the vehicle contains the fatality in the accident. This variable is, however, strongly related to the age of the vehicle as it is usually the newer vehicles that have ABS. <ABS> becomes non significant once <VehicleAge> is included into the model, while <VehicleAge> is still significant.

Driver manoeuvre

Different types of manoeuvres (turning, overtaking, and loss of control) are more dangerous than driving along the road (either straight or in bends). Note that the driver manoeuvre variable at the vehicle level has the opposite sign as that at the accident level (see section 4.3.2). If you are in an accident in which anyone has been executing a manoeuvre, it increases you chance of survival. However, given that somebody has executed a manoeuvre, it is better for you if it has been the other one. Especially “pulling out” and “loss of control” increases your fatality risk as compared to driving straight. The DriverManoeuvre variables at the vehicle level become non-significant together with the other variables (mainly <FrontDamaged> and <Braking>).

5.3.2.4. Nonsignificant variables

As mentioned before, the analysis on the car-car accident subset does not have as much power as one may wish for. In such a situation, as a rule of the thumb, p values smaller than .400 should be considered ambiguous. They do not give a clear indication that there is an effect. However, neither should they be seen as an indication that there is no effect. We will therefore first report those variables that have an ambiguous effect and subsequently those for which we can state relatively clearly that there is no effect.

Ambiguous results

With respect to road user characteristics there is a tendency for older male passengers who are seated in the front to be the fatality more often than younger female drivers or passengers who are seated at the backseat. Vehicles with impaired, very young, or very old drivers show the tendency to contain the

fatality more often than vehicles in which the driver is middle-aged and not impaired. All these tendencies are not significant and would have to be confirmed in an extended data set.

Null effects

There does not seem to be a difference between male and female drivers, nor between drunk and sober drivers, or between drivers who are familiar with the area and those who are unfamiliar. Moreover, whether drivers tried to avoid the accident by steering or not does not seem to have an effect. This does not mean that these factors do not contribute to the accident risk (which they probably do). It means that once in a fatal accident these factors do not determine whether the fatality was in the car of the driver or in the other car.

Most variables at the accident level (Light, Carriageway, Junction, Area, SpeedLimit, and whether a driver unfamiliar with the area was involved) do not have an effect in this analysis. As noted earlier, most accidents had exactly one fatality. Consequently there is very little variation between accidents with respect to the fatality risk. Again, this does not mean that these factors do not contribute to the accident risk, but as they always have the same value for each victim in a particular accident, they cannot help identifying the characteristics of the fatality.

5.4 Conclusion

A subset of cases from the fatal accident database has been selected to investigate the factors that affect the fatality risk for the occupants in fatal accidents between two cars. This restricted data set offered the advantage that the variables that otherwise have many missing values were all complete so that the cases included did not vary according to the variables that were taken up into the model. Moreover the values of a large set of variables could be meaningfully compared for those people who died in the accident and those who did not.

This analysis does not allow identifying risk factors that cause fatal accident to happen in the first place. Variables that describe the accident as whole could only become significant in this analysis if they contribute to the number of fatalities in the accident. As the large majority of the accidents contained only one fatality, there was not much variation that could be predicted this way. What this analysis is really about is the identification of protective factors that – once in the accident – made the difference between staying alive or entering the statistics as a fatality.

Most of the factors that differentiated fatalities from survivors are situated at the vehicle level. The most important factor is whether the driver had been braking to avoid the accident. In the first place it is striking that more than half of the drivers did not conduct any avoidance manoeuvre at all. However, if they

braked, it reduced the fatality risk in their car by two thirds. This result shows the potential of early warning systems that would alert the driver if he/she is on collision course.

The next important factor is the age of the vehicle. Figure 5.3 gives a clear indication how the safety of the car occupants in severe accidents decreases with the age of the car. This result demonstrates that the attempts of the car industry to make cars safer for their occupants have been quite fruitful.

The third big factor is the area of most damage in the vehicle. Figure 5.2 shows that front impacts are far less dangerous to the occupants than side impacts. Moreover, in Figure 5.4 it is demonstrated that the safety of cars with side impacts has been much less improved than the safety of cars with frontal impacts. This result somehow modifies the conclusion above and identifies the protection to side impacts as an area where more improvement is still necessary.

A number of person and vehicle characteristics have been identified that affect the probability to be/contain the fatality when involved in a fatal accident. The most important factors (in terms of their predictive power) are whether the driver has been breaking to avoid the accident, the age of the vehicle, and the area where the vehicle is damaged most. Moreover young and old people are at greater risk to be the fatality in the accident than middle aged people.

At the accident level only two factors became significant. Fatal accidents on motorways on the one hand and fatal accidents that have been preceded by a manoeuvre from at least one of the participants on the other hand are both less dangerous to the occupants than other accidents.

At the victim level it has been established that young and old persons run a higher risk to be the fatality than middle aged persons. Protection by seatbelts showed only an unstable effect and would have to be confirmed in a larger dataset.

We can conclude that the Fatal Accident Investigation Database offers a wealth of information. The analysis of the fatality risk in car-car accidents exploits some of this information to give a global picture of the factors that contribute to the safety of car occupants who are involved in severe crashes. By comparing the person and vehicle characteristics of the fatalities and the survivors it was possible to identify possible areas for improvement in car safety.

Appendix 5A: Variables not in the final model

All variables that were not included into the final model were entered into the empty model by themselves. The results are given in Table 5.4.

Table 5.4 Variables not in the final model entered into empty model

		β/σ	SE	chi2	p	exp
UserClass		0.376	0.354	1.128	0.288	1.456
AgeVictim		0.195	0.152	1.646	0.200	1.215
senior		1.164	0.461	6.375	0.012	3.203
Female		-0.37	0.314	1.388	0.239	0.691
Backseat		0.001	0.001	1.000	0.317	1.001
AirbagU	yes	-0.947	0.407	5.414	0.020	0.388
	unknown	-0.568	0.459	1.531	0.216	0.567
ABS	yes	-1.127	0.38	8.796	0.003	0.324
	unknown	-0.877	0.373	5.528	0.019	0.416
AgeDriver		0.111	0.147	0.570	0.450	1.117
$AgeDriver^2$		0.174	0.098	3.152	0.076	1.190
AgeCatDriver				7.039	0.134	
	18-25	0.737	0.438	2.831	0.092	2.090
	35-54	0.259	0.46	0.317	0.573	1.296
	55-64	0.228	0.589	0.150	0.699	1.256
	65+	1.343	1.343	1.000	0.317	3.831
GenderDriver		-0.016	0.341	0.002	0.963	0.984
AlcoholDriver		0.34	0.484	0.493	0.482	1.405
ImpairmentDriver		0.635	0.424	2.243	0.134	1.887
UnfamiliarDriver		0.126	0.392	0.103	0.748	1.134
Manoeuvre				14.310	0.046	
	bend	0.534	0.392	1.856	0.173	1.706
	turning	0.931	0.419	4.937	0.026	2.537
	overtaking	1.5	0.627	5.723	0.017	4.482
	lossControl	2.745	1.117	6.039	0.014	15.565
	illegal	1.52	1.239	1.505	0.220	4.572
	reversing	1.354	0.755	3.216	0.073	3.873
ExecutedManoeuvre		0.847	0.312	7.370	0.007	2.333
AvoidMan		-1.188	0.325	13.362	0.000	0.305
Steering		0.001	0.464	0.000	0.998	1.001
LossControlAcc		1.393	0.907	2.359	0.125	4.027
AvoidManAcc		-0.486	0.3	2.624	0.105	0.615
UnfamiliarAcc		-0.023	0.361	0.004	0.949	0.977
SpeedLimitCat	80	-0.161	0.531	0.092	0.762	0.851
	100	0.039	0.383	0.010	0.919	1.040
	130	-1.204	1.182	1.038	0.308	0.300
RuralArea		-0.186	0.374	0.247	0.619	0.830
Junction01		-0.157	0.287	0.299	0.584	0.855
CarPhysDivided		-0.197	0.631	0.097	0.755	0.821
Weekend		0.679	0.335	4.108	0.043	1.972
Light		0.057	0.308	0.034	0.853	1.059

Appendix 5B: Motorway and area of most damage

In Table 5.5, the proportion of victims killed (i.e. the fatality risk) is given for different road types (motorway and other roads). The motorway-effect (i.e. the difference in proportion killed between motorways and other road types) is given for cars that had been damaged mostly in the front (first column), mostly at other areas (second column), and for all cars (third column). When comparing the motorway effect in the three columns, one can see that in terms of percent the effect is larger when considering cars damaged mostly in the front and cars damaged elsewhere separately.

Table 5.5 Proportion of victims killed by Road type and Area of most damage

	FrontDamaged	OtherAreaDamaged	Total
No Motorway	.18 (128)	.62 (79)	.35 (207)
Other roads	.00 (2)	.29 (7)	.22 (9)
Size effect	.18 (130)	.33 (86)	.13

Note – The number of victims the proportion is based on is given in parenthesis.

The effect of Motorway is not significant when entered into the model by itself, but it becomes so, when entering FrontDamaged into the model simultaneously. The latter has an effect comparable to considering the motorway effects in the first two columns rather than considering the motorway in the third column.

==========
Blank page left intentionally
==========

Chapter 6 - Modelling severity reporting reliability through in-depth data

George Yannis, Dimitris Katsochis, Eleonora Papadimitriou, NTUA

6.1 The research question

In most European countries, road accident data are collected by the Police, who is responsible for determining injury severity. As regards road fatalities, the common EU definition of fatalities within 30 days from the accident has contributed significantly in the reduction of fatality under-reporting. However, no such definitions are available with respect to injuries (serious or slight). Therefore, injury under-reporting or inappropriate reporting is a critical issue towards the full harmonization of EU road accident data.

As regards injury under-reporting, the issue 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 a reporting inaccuracy problem exists, accounting in many countries for over 50% of all injuries (especially slight ones) (ETSC, 2007). This serves as a basis to proceed to further investigation in the framework of SafetyNet WP7.

In the WP5 in-depth database, 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 WP5, 2006), there are:

- “Police injury severity”, i.e. injuries or complications directly due to the accident within 30 days of the crash
- “SafetyNet medical outcome”, i.e. overall outcome of the crash, as police only follows the situation of each individual’s health for a limited period of time

In both cases, there are four possible outcomes, namely "killed", "seriously injured", "slightly injured" and "not injured".

Therefore, the objective of this analysis is the identification of the degree of mismatch between "Police" and "SafetyNet". Moreover, it aims to investigate whether any prevailing factors emerge that are related to these differences, making the initial "Police" outcome to change by "SafetyNet". This research question is rather critical, especially since there have been in practice certain problems in the determination of the degree of injury severity across different countries, as a result of the application of police definitions. The results of the analysis may be a very first step towards the development of correction coefficients for inappropriate severity recording, like those that are under development in WP1 for injury under-reporting.

6.2 The analytical problem

SafetyNet WP5 has contributed truly to the collection of in-depth data that allow for detailed, disaggregate analysis of various injury-related aspects. It is therefore possible to examine the effect of several interesting parameters on the probability to record eventually 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.

It is important to note that a careful examination of subsets of cases is required, as there are several cases (categories) of initial outcomes eventually changing category (e.g. from slightly injured to not injured). It could be that the most promising field for further analysis entails cases recorded by the police as serious injuries; in fact, this is the initial outcome (police records) changing more often to some other injury severity score - fatality or slight injury - in the final SafetyNet outcomes (see Tables 6.1-6.2).

Table 6.1. Distribution of casualties recorded in different severity scores

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

Table 6.2. Corresponding injury severity for the two severity scores

Police Injury Severity	SafetyNet Medical Outcome					
	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
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**

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

It appears that there are quite a few cases in which the severity score changes from the police to the SafetyNet recording system. 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 category, with almost 60% of initial scores changing to some other severity category. It appears that part of the cases initially rated as "serious" by the police are afterward categorized either as "slightly" injured (20%), as "fatal" (30%) or as "unknown" (10%) by the Safetynet team.

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 more than 40% of all cases in the database, reveals some striking differences when compared to the other countries (see Tables 6.3 - 6.4).

Table 6.3. Corresponding injury severity for the two severity scores: all countries except from Italy (541 cases)

Police Injury Severity	SafetyNet Medical Outcome					
	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
Fatal	44,5%					44,5%
Not Injured		18,9%		0,7%		19,6%
Serious	0,9%		14,8%	0,9%	1,1%	17,7%
Slight		0,4%	0,6%	16,1%	0,7%	17,7%
Unknown	0,2%				0,2%	0,4%
Grand Total	45,7%	19,2%	15,3%	17,7%	2,0%	100,0%

The values of Table 6.3 include Sweden, Germany, France, Finland and United Kingdom. It is acceptable to treat those records in a uniform way, since the distributions of injury severity for each of these countries are most similar (see Appendix 6A). On the other hand, Table 6.4 confirms that Italy follows a different pattern with respect to the accuracy of injury severity recording.

Table 6.4. Corresponding injury severity for the two recording systems in the WP 5.1 database: Italy (413 cases)

Police Injury Severity	SafetyNet Medical Outcome					
	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
Fatal	21,1%					21,1%
Not Injured		24,0%				24,0%
Serious	16,9%	0,5%	3,6%	10,9%	4,6%	36,6%
Slight		9,2%		5,8%	1,2%	16,2%
Unknown					2,2%	2,2%
Grand Total	38,0%	33,7%	3,6%	16,7%	8,0%	100,0%

Three main conclusions may be drawn, based on these comparative tables:

- In all countries except Italy, the large majority of cases is in 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 mismatches are relatively more frequent for the entries that have been rated as “serious” by the police; approximately 11% of these cases ended up in a different severity category for the SafetyNet score.
- Italy constitutes a striking exception to the generally satisfactory picture, as a large proportion of 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 correctly appointed this score. As mentioned, changes in score involve both directions (i.e. to a status of either heavier or less severe type of casualty).

The dependent variable considered in the present analysis is at first a binary one, indicating whether the two severity scores (Police and SafetyNet) are the same or different. Dealing with a not so common dependent variable calls for a 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. 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 shown in previous chapters (severity modelling), quite a few variables in the 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 variables 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. In this report the presentation is limited to the most interesting subset of variables.

6.3 The analytical solution

In this section, logistic regression models are developed to compare alternative combinations of scores of injury severity between the police and SafetyNet team records. First, binomial regression models are presented, for the probability of the occurrence of mismatches between the police and SafetyNet severity scores. Then, multinomial models are developed, in order to examine whether the different severity scores are an overestimation or underestimation of the injury severity. This is a meaningful order to follow, as some factors may only work towards one direction (over- or under-estimation), while others may

work towards both over- and underestimation. In each case, separate models are developed for Italy and all the other examined countries[9].

In the case of the binomial models, the dependent variable is a binary variable (1: same outcome, 0: different outcome) of the record of injury severity for each road user. All explanatory variables have been defined as categorical (see Table 6.13). Categories were formed by means of independent contrasts to enhance 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.

In the case of multinomial models, the dependent variable is a multinomial 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)

[9] 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 were the reporting changed.

Table 6.13. Variables and values

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: 1st event, 1: 2nd-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

6.3.1 Binomial modelling

In this analysis, three different binomial models based solely on those individuals that were initially classified as “Seriously Injured” by the Traffic Police have been examined, namely:

- Two alternative models (in terms of explanatory variables) for this type of casualties for Italy
- A model including this type of casualties for all other examined countries

6.3.1.1. Binomial models for the case of Italy

As mentioned previously, the dependent variable is a binary variable (1: same outcome, 0: different outcome) of the record of injury severity for each road

user. There are 132 known cases (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 6.14.

Table 6.14. Parameter estimates of the best fitting binary logit model of the probability of matching Police and SafetyNet injury severity scores (Italy - 1)

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

There is notable improvement from the empty model as far as the likelihood ratio is concerned. 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) [10]. These results are quite satisfactory. The 'empty' model predicted correctly only non-matching cases (around 88% of total cases).

The selection of Light Conditions instead of Vehicle Type yields a slightly different, practically equivalent model in terms of performance.

[10] 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”.

Table 6.15. Parameter estimates of a second binary logit model of the probability of matching Police and SafetyNet injury severity scores (Italy - 2)

Variables	Parameter estimates			
	B	S.E.	Sign.	Exp(B)
Traffic(normal/heavy)	-1.525	0.621	0.014	0.218
Traffic(light)	.	.	.	.
Light Conditions(dusk/night)	1.145	0.632	0.070	3.143
Light Conditions(daylight)				.
Junction(yes)	-1.288	0.662	0.051	0.276
Junction(no)	.	.	.	.
Gender(female)	-1.582	0.842	0.060	0.206
Gender(male)	.	.	.	.
Constant	-0.919	0.646	0.155	0.399

Although the decrease in the likelihood statistic is slightly smaller (reduction from 93.470 to 75.295 compared to the empty model), the latter model includes parameters that are more significant. More than 92% of the outcomes are correctly predicted. These results are slightly better than the previous (first) version of the model. Nevertheless, the two models can be considered to be equivalent, given that the above differences are minor.

It is noted that the two variables (Light conditions and Vehicle type) may be associated, in the sense that there may be less vulnerable road users at night. On the other hand, accidents with vulnerable road users are more likely to occur at night. Therefore, the night-time effect may not be fully due to some association with vulnerable road users, but may also have a meaning of its own with respect to reporting.

The examination of these two versions of a model that share a common core group of variables (Traffic – Junction – Gender) in parallel produces some interesting conclusions:

- The heavier the traffic, the more likely it becomes to obtain different scores 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
- 2-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 presented in section 6.4.1.
- On the other hand, the absence of daylight appears to enhance matching scores between the two recording systems. This is not a fully intuitive finding (neither a very strong one), as there is a general idea that difficult natural

conditions make the recording process more difficult. However, it can be said that the police may be more careful in data recording during night.

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.

6.3.1.2. A binomial model for all other countries

The second dataset, containing all other countries except Italy, contained 90 known cases (6 more were excluded from the analysis as unknown). The results of the analysis are presented in Table 6.16

Table 6.16. Parameter estimates of the best fitting binary logit model of the probability of matching Police and SafetyNet injury severity recording (all countries except Italy)

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

Since the likelihood statistic 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).

In terms of percentage of correctly classified cases, 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, 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. The same interpretation than the one suggested in the case of Italy may hold (i.e. more careful recording 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.

6.3.2 Multinomial modelling

6.3.2.1. A multinomial model for Italy

A multinomial logit model was formed for the set of individuals classified as seriously injured by the police in Italy, in accordance to the respective binary models presented in section 6.3.1. This is an unordered multinomial model with four independent variables.

It is noted that a second model that involves Light Conditions instead of Vehicle Type was also tested and did not yield similarly encouraging results in terms of classification, while the proportions of correct predictions for the three possible outcomes were quite different Therefore, it is not presented here. At a first step of investigation, the explanatory variables that were included in the model of Table 6.17 were selected from those that appeared to have a significant effect in the respective binomial model.

As regards model fit, a reduction of 70.83 in the likelihood ratio is obtained, which is significant for 8 degrees of freedom. Finally, almost 73% of the outcomes are correctly predicted by the model. These results are quite satisfactory, as the proportions of correct predictions for the three possible outcomes are more balanced than with any other combination of variables.

Table 6.17. Parameter estimates of a multinomial logit model of the probabilitiy of over- and under-estimation of injury severity (Italy)

Variables		Parameters estimates			
		B	S.E.	Sign.	Exp(B)
Probability to switch from seriously injured (1) to killed (0)					
0	Intercept	5.152	1.354	0.000	
	Traffic(light)	-1.463	0.678	0.031	0.232
	Traffic(normal/heavy)	.	.	.	.
	Vehicle Type(occupants)	-2.346	0.835	0.005	0.096
	Vehicle Type(pedestrians-riders)	.	.	.	.
	Junction(no)	-0.923	0.715	0.197	0.397
	Junction(yes)	.	.	.	.
	Gender(male)	-1.161	0.909	0.201	0.313
	Gender(female)	.	.	.	.
Probability to switch from seriously injured (1) to slightly/not injured (2)					
2	Intercept	3.074	1.490	0.039	
	Traffic(light)	-2.176	0.711	0.002	0.114
	Traffic(normal/heavy)	.	.	.	.
	Vehicle Type(occupants)	1.439	1.106	0.193	4.215
	Vehicle Type(pedestrians-riders)	.	.	.	.
	Junction(no)	-1.318	0.726	0.070	0.268
	Junction(yes)	.	.	.	.
	Gender(male)	-2.061	0.892	0.021	0.127
	Gender(female)	.	.	.	.

The parameter estimates yielded the following findings and ideas:

- The part of the model that explains the probability to switch scores from seriously to slightly/not injured road users (bottom part of table 6.17) is somewhat more reliable than the part examining the probability to switch from seriously injured to killed road users (top part of Table 6.17), with two predictors that are significant at 95% (and three at 93%).
- It should be stressed, though, that Vehicle Type exhibits different impact in the two parts of the model (a negative and highly significant coefficient for the probability of serious injury scores to switch to fatalities and a positive but non-significant one for their probability to turn to slight injuries eventually). This seems to imply that the effect of vehicle type concerns the probability of injury deterioration only.
- The severity score in light traffic is much less likely to change either way (especially to lighter injuries) compared to normal/heavy traffic. This is in accordance with the respective binomial model; however, additional

information is given by this model, in the sense that it is proved that the effect works in both directions.

- The same is true in the case of Gender. This is only significant when serious injuries are compared to changes to lighter injuries (2nd part of the model) and also agrees with the observation of the binomial model (female road users where found to favour non-matching scores, although this was marginally significant).
- The presence of a Junction is associated to a higher probability for individuals to undergo a change of injury severity. This is only marginally significant and only for changes from severely injured to lighter injuries. The finding agrees with the observation of the binomial model (presence of junction was found to favour non-matching scores, although this was only significant at 90%).
- The Vehicle Type reveals significant impact when change of status from serious injury to fatality is considered (top part of the model). Car occupants are much less likely to have such a change compared to pedestrians and 2-wheelers riders.

6.3.2.2. A multinomial model for all other countries

Similarly to Italy (see previous section), a multinomial model is also built for the other examined countries. This is an unordered multinomial model with two independent variables, namely Age and Area type. The dependent variable is a multinomial one of the record of injury severity for each road user (same coding as in Italy's model is followed). Compared to the model presented in section 6.3.2, the only variable that does not seem to remain significant at all in terms of a multinomial model is Light Conditions.

With respect to the model's quality, the above best fitting model is relatively improved compared to the empty model in terms of likelihood (likelihood is reduced by 8.78, which is marginally significant with 4 degrees of freedom).
90% of the outcomes are correctly predicted by the model.

The main conclusions derived from this analysis may be summarized as follows:

- The two variables are marginally significant as far as the change from seriously injured to killed is concerned. This does not hold when the change from seriously to slightly/not injured is considered.
- Moreover, while all coefficients are clearly negative (i.e. the direction of the effects is the same for switching to "killed" and switching to "slightly/not injured"), in the former case they are quite larger as well.
- The Age group 15-54 years old is much less likely to change either way (especially towards fatalities) compared to the other group (0-15, 55+). This is in accordance with the respective binomial model, where group 1 was found to be associated to changing score of injury severity.

- The same is true in the case of rural area. This also is in accordance with the observation of the binomial model (urban/mixed areas where clearly found to favour non-matching scores).

Table 6.18. Parameter estimates of the best fitting multinomial logit model of the probability of the over- and under-estimation of injury severity (all countries except from Italy)

Variables		Parameter estimates			
		B	S.E.	Sign.	Exp(B)
Probability to switch from seriously injured (1) to killed (0)					
0	Intercept	0.010	1.111	0.993	
	Age(15-54)	-1.955	1.060	0.065	0.142
	Age(0-14 / 55+)	.	.	.	.
	Area(rural)	-2.189	1.107	0.048	0.112
	Area(urban/mixed)	.	.	.	.
Probability to switch from seriously injured (1) to slightly/not injured (2)					
2	Intercept	-0.859	1.284	0.503	
	Age(1-54)	-1.671	0.990	0.092	0.188
	Age(0-14 / 55+)	.	.	.	.
	Area(rural)	-1.120	1.245	0.368	0.326
	Area(urban/mixed)	.	.	.	.

Overall, the added value of the multinomial logistic regression models lies on the differentiation of the relation between the two possible alternative outcomes of the dependent variable (killed or slightly/not injured) and the respective reference value (seriously injured). For almost all independent variables that were significant in the binomial model, it is verified that the probabilities to change to either heavier or lighter injury severity show the same direction (i.e. either an increased or a decreased probability).

What does really differ is the strength of each impact, as expressed through the coefficients of the explanatory variables. For instance, women involved in road accidents in Italy tend to change their severity status from seriously injured to killed more often than men, but this relative difference between the two groups of the “Gender” variable becomes much larger when a switch to lighter injuries is considered. In other words, the finding of the corresponding binomial model is practically due to the later effect –rather than the former.

6.4 Conclusions – Future steps

As revealed from the preceding analysis, two problems mainly appear with respect to the investigation of accurate injury severity reporting:

- The selection of an appropriate subset of variables

- The interpretation of conclusions or implications provided by such an analysis

From the existing results, a general trend can 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 the differences in scores are mainly due to recording bias (e.g. the Police tends 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 Age variable was very significant for all countries but non-significant in the model for Italy. It may be that in the other countries 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 that there may be some recording bias present.

The multinomial models aimed at the further analysis of the major conclusions drawn through the respective binomial models. Apart from that, they are aiming to highlight any differences in variables behaviour by comparing the dependent variable's reference category (same scores) with two possible changes (improvement or deterioration of injury severity).
It is noted that all predictors that have been tested in the multinomial models are those that already appeared to be significant in the respective binomial models. 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. For example, it was shown that Vehicle Type in Italy has different effect with respect to improvement or deterioration of injury severity. The negative tendency of car occupants to switch from seriously injured to fatalities is quite large and most significant. This implies that the effect found in the binomial model practically comes from the increased probability of seriously injured vulnerable road users (i.e. pedestrians and 2-wheelers riders) to die after all.

However, there may be additional predictors that were not found significant in the binomial models, but could be significant in the multinomial models. For example, if "females" have an increased probability of "overestimation" of severity in relation to "males", and at the same time a decreased probability of an "underestimation" in relation to "males", then because the two types of mismatches are pooled in the binomial model, the effect of "gender" would not appear significant in the case of the binomial model. For this reason, in the next stages of this analysis, more variables will be tested in the multinomial models.

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, which will be taken into account once available, in the next stages of this analysis.

APPENDIX 6A: Comparative injury severity distributions for other countries except Italy

Table 6A.1. Corresponding injury severity for the two different scores: Sweden

Police Injury Severity	SafetyNet Medical Outcome					
	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
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 6A.2. Corresponding injury severity for the two different scores: Germany

Police Injury Severity	SafetyNet Medical Outcome					
	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
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 6A.3. Corresponding injury severity for the two different scores: France

Police Injury Severity	SafetyNet Medical Outcome					
	Fatal	Not Injured	Serious	Slight	Unknown	Grand Total
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 6A.4. Corresponding injury severity for the two different scores: Finland

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

Table 6A.5. Corresponding injury severity for the two different scores: United Kingdom

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

==========
Blank page left intentionally
==========

Chapter 7 - Conclusion

(Heike Martensen, IBSR)

7.1 Summary

In the present deliverable, a number of analyses on a preliminary version of the Fatal Accident Investigation Database (extracted in December 2006) have been presented. The accident outcomes that were analysed were the accident size, the fatality risk in the accidents, and the reliability of the injury reporting. The enormous amount of information contained in the database has been reduced by selecting those variables for the models that have the highest predicting power and by recoding them into a few broad categories (mostly two) that differentiated best between the different outcomes.

In Chapter 3, the accident size was analysed in terms of two dependent variables: Single vehicle accidents were compared to multi-vehicle accidents and accidents with one fatality were compared to those with more than one victim killed or seriously injured. The results indicated that single vehicle accidents involve young, male, impaired, drivers more often than older, female, and unimpaired drivers. While drivers in single vehicle accidents tended to be male, the proportion of women among the occupants was higher as compared to multi-vehicle accidents. Furthermore multi-vehicle accidents involved more often a driver who had just executed a manoeuvre, while the drivers in single vehicle accidents tended to have driven along the road.

While a number of factors differentiated between single and multi-vehicle accidents, only two factors differentiated between accidents with one or more victims killed or seriously injured (multi-KSI). The only characteristics that were reliably associated with multi-KSI accidents were a higher average number of passengers per vehicle, and the fact that the accident took place in weekends rather than during the week. One reason for this meagre result could be the small number of cases where there was more than one person killed or seriously injured in the accident.

In Chapter 4, the fatality risk was analysed for all accidents in the database. This global analysis indicated that the fatality risk was higher for vulnerable road users as compared to occupants of motor vehicles, and for seniors (65 and above) as compared to all other age groups. It also appeared that participants who had tried to avoid the accident by braking had a lower risk of being/containing the fatality in the accident, and that for participants who contained the fatality more events were described in the database. Finally it was shown that for accidents that took place on roads where the carriageways were physically divided the proportion of fatalities was lower than for other accidents.

In Chapter 5, the fatality risk was analysed for car-car accidents, ensuring maximal comparability between fatalities and survivors. In this analysis the

results from the global analysis were confirmed. Again, it was shown, that occupants in those vehicles where the driver had braked had a much higher chance of surviving (more than three times as large) than those in vehicles where the driver did not brake. Accidents on motorways[11] were shown to exhibit a lower proportion of fatalities and those participants who contained the fatality were described with more events than others.

Additionally, it was possible to analyse safety factors that were specific to cars and their occupants. It turned out that when involved in severe accidents newer cars are much more secure than older cars; that side impacts are much more dangerous than frontal impacts and interestingly that there was an interaction between these two variables: While the protection from front impacts increased dramatically for newer cars, there is no significant increase in the protection from side impacts.

The only difference in result between the global analysis fatality risk and the car-specific one concerned the senior citizens. While they showed a higher fatality risk in the complete data set as well as in car-car accidents, their increased fatality risk in car-car accidents could be completely explained by the fact that they suffered side impacts (which are more dangerous) much more often than other road users. Because old age of the victim is so often associated with cars where the side is the main impact area, it becomes impossible to statistically disentangle these two possible causes of a higher fatality risk.

The highly significant relation between age and impact area, could be due to the fact that old people have more problems to master junctions than younger ones (which lead to accidents with side impacts more often). Another reason may be related to senior citizens being less likely to show aggressive driving behaviour, which might lead to a higher proportion of front impacts (assuming that the ones who hit someone else will have a frontal impact more often while the ones who are hit by another car may suffer side impacts more often). This issue should be investigated more closely.

In Chapter 6, two different records of injury severity were compared: The original police record and one revised by the SafetyNet team. These two records were not always in agreement, indicating a substantial number of reporting errors. These errors concerned predominantly victims who had initially been classified as “seriously injured”. An important finding was that differences between police record and the SafetyNet record were much more frequent in Italy than in all other country. In a systematic exploration of factors that predict reporting differences, it turned out that - for their largest part - the errors could not be related to characteristics of the accidents or the victim, suggesting that they appear at random and are probably due to insufficient information for the recording officers. However, there were also a number of systematic biases

[11] Note that the variable “motorway” that was the best predictor in the car-car analysis is strongly related to the variable “DividedCarriageways” that was the best predictor in the model for all users.

identified. The exact factors differed for Italy and for the other countries (that were analysed jointly), but the two tendencies in biases that could be identified were: 1) For persons who could in some way be assumed to be weaker or less protected than others (children, seniors, women, vulnerable road users) the injury record changed during the revision more often than for others. 2) Complex accidents facilitate misreporting.

7.2 A word of caution …

is necessary, especially for those who are only scanning the results. When the term *fatality risk* is used in the present analyses, this concerns exclusively the risk of dying, *given that one is already involved in a fatal accident*. This risk is in no way comparable to the risk that road-users in general run to die in an accident (which is much lower of course, but also shows a very different pattern). With some care one can extrapolate the fatality risk results (especially those from the car-car analysis) 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.

It must be kept in mind that the variables that were found to be effective in differentiating between the survivors and the fatalities, do not say *anything* about the *causation* of the accident. As an example, the variable "DriverImpairment" did not differentiate between the survivors and fatalities, meaning that in an accident between an impaired and a non-impaired driver, the occupants in the car of former had the same fatality risk as those in the car of the latter. This does not exclude the possibility, however, that the impaired driver would usually have been the one who had *caused* the accident.

The fact that characteristics of the vehicle rather than those of the driver seem to differentiate best between fatalities and survivors should therefore not be seen as an indication that vehicle characteristics are the ones that determine the overall safety of the road user most. The best protection against dying in a fatal accident is, after all, not being involved in one.

The only analysis that might at least hint at the factors that affect accident causation are the comparison between single- and multi-vehicle accidents. The logic behind this is the assumption that in multi-vehicle accidents the responsibility is distributed between all participants (either one is responsible and the other is innocent or, more often, both are partly responsible). All drivers in single vehicle accidents, however, had been fully responsible for causing the accident. Consequently, characteristics which are more strongly present among drivers in single-vehicle accidents might be those that are associated with "drivers responsible for the accident". These characteristics were indeed the usual suspect for high-risk drivers: impaired young males in the presence of female passengers, who were familiar with the area and were driving either straight along the road or had lost control. One should, however, keep in mind that this is a very indirect way of reasoning, as the drivers in multi-vehicle

accidents had been in fatal accidents as well, and should certainly not be seen as “exemplary”.

7.3 Further questions and recommendations

Obviously, these results of a quantitative analysis of the Fatal Accident Investigation Database should be seen complementary to a qualitative analysis. The outcomes of the present analyses are meant to point to areas that are worthwhile being explored in more depth. The results of the analyses pose questions rather than they give answers. Examples are:

- What makes weekend accidents more harmful in terms of victims killed or seriously injured than accidents occurring during the week? The question is less trivial than it seems, as the usual “suspects” for answers, age of the driver, impairment, and the number of occupants have been controlled for in the analysis.
- Why are there so many crash participants who did not attempt an avoidance manoeuvre, although braking seems to reduce the fatality risk in an accident so strongly?
- Why did such a large proportion of senior citizens suffer from side impacts?
- Why are there so many misclassifications of injury severity in Italy?
- How come that the severity of the injuries of women is incorrectly reported more often than the one of men’s?

Additional to guiding the qualitative analyses, the exploration of the Fatal Accident Investigation Database indicated a few issues where there is room for improvement in the database.

Lost Control

The variable <DriverManoeuvre> contains the category <LostControl>. This is a potentially very interesting category, because it indicates accidents that are associated with inappropriate speeding. Apparently however, this category was not systematically indicated for vehicles that had been described in the summary as having lost control. This is probably due to the fact that <LostControl> is not a manoeuvre and that some investigators chose to indicate the manoeuvre the driver was executing when loosing control (e.g. “Driving round a right hand bend”). We would therefore suggest dedicating a new variable to the question of whether the driver had lost control or not.

Driver responsibility

It might be interesting to try identifying characteristics that differentiate between the drivers that were responsible for the accidents and those that were not. Obviously, this question does not have a clear answer in many cases. Moreover, it is understood that indicating “blame” is against the spirit of in-depth investigations. Nevertheless, the summaries often seem to imply that the accident was mainly initiated by one participant and it can be imagined that the investigators often do have a clear idea about the share of responsibility. A

variable <Responsible> with, for instance, the categories: <Yes>, <No>, and <Partly> would allow some very interesting analyses.

Non-fatal accidents
The fatal accident database shows very little variation with respect to severity. The large majority of the accidents contain exactly one fatality (not more and not less). Including non-fatal accidents would allow much more interesting analysis on the factors that differ between more or less severe accidents.

Risk-exposure data
The possibility to exploit the wealth of information in the Fatal Accident Investigation Database is limited as long as there are no exposure data at the same level of detail. For each variable and each category, knowing its frequency of occurrence in the general driving population would help to evaluate whether a particular frequency in the fatal-accident population is different from what should be expected. As an example: The majority of the vehicles involved in fatal accidents (34%) has been driving straight along the road. Should we launch campaigns warning road-users that seemingly “safe” situations (such as driving straight) are really the most dangerous ones? Or is 34% the percentage that should be expected, given that most of the road-users drive straight along the road most of the time? To answer this question, it is necessary to compare this accident percentage to the percentage of vehicles driving straight along the road in normal traffic. Exposure data that allow such comparisons would enable conclusions about factors that influence the safety of road-users in general, rather than the conclusions presented here that are restricted to road-users who are already involved in a fatal accident.

7.4 In a nutshell

A number of factors have been identified that differentiate between single and multi-vehicle accidents and between the survivors and the fatalities in fatal accidents. Moreover, it has been indicated that there are a substantial number of misclassifications of injury severity in the original police records and a few reporting biases have been identified. In the majority, the results raise further questions rather than providing definite answers. They indicate areas in the Fatal Accident Investigation Database that deserve further (and a more qualitative) exploration.

References

Browne W, Goldstein H, Rabash J. (2001). Multiple membership multiple classification (MMMC) models. Statistical Modelling 2001;1, pp.103–24.

Dupont, E. & Martensen, H. (2007). Introduction to discrete response models. In: 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.

Evans, L. (2001), Age and fatality risk from similar severity impacts. *Journal of Traffic Medicine, 29*, 10-19.

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.

Morris, A (2007). *WP5 – Independent Accident Databases*. Presentation at SafetyNet Members Plenary meeting, Arcachon, France.

Mabe, K. (2006), The impact of safety belts on driver mortality in multi vehicle fatal accidents. *The Economist*. Farmers Group.

Martensen, H. & Dupont, E. (2007). The added value of multilevel models in road safety analysis. In: 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.

Papadimitriou, E., & Antoniou, C., & Yannis, G. (2007). Bayesian estimation in multilevel modelling. In: 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.

Rasbash, J., Steele, F., Browne, W, Prosser, B. (2004). *A User's Guide to MLwiN. Version 2.1e*, Centre for Multilevel Modeling, Institute of Education, University of London, UK.

Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B, 64*, 583-639.

SafetyNet WP5, "Work Package 5 Database Glossary", version of 29 November 2006.

Vanlaar, W. (2007). Binary and General Binomial Responses. In: 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.

Yannis, G., Papadimitriou, & E., & Antoniou, C. (2007a). Multivariate Models. In: 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.

Yannis G., Papadimitriou E., Antoniou C., (2007b). Multilevel modeling for the regional effect of enforcement on road accidents. In press, Accident Analysis and Prevention.