EFFECT OF VEHICLE CHARACTERISTICS ON CRASH SEVERITY: PORTUGUESE EXPERIENCE

Guilhermina A. Torrão, MSc.
Graduate Student, Mechanical Engineering
Centre for Mechanical Technology and Automation and Department of Mechanical Engineering, University of Aveiro
Campus Universitário de Santiago
3810-193 Aveiro, Portugal Portugal
Phone: (+351) 234 378 181
E-mail: guilhermina.torrao@ua.pt

Margarida C Coelho, Ph.D.
Invited Assistant Professor, Mechanical Engineering
Centre for Mechanical Technology and Automation and Department of Mechanical Engineering, University of Aveiro
Campus Universitário de Santiago
3810-193 Aveiro, Portugal
Phone: (+351) 234 370 830
E-mail: margarida.coelho@ua.pt

Nagui M. Rouphail, Ph.D.
Director, Institute for Transportation Research and Education (ITRE)
Professor of Civil Engineering, North Carolina State University
Raleigh, NC, 27695-8601, USA
Phone: (919) 515-1154
E-mail: rouphail@eos.ncsu.edu

ABSTRACT

Road accidents and the resulting public health impacts is a critical issue in Portugal where mortality rates from vehicle crashes exceed the European Community average. Reductions in traffic fatalities and injuries, as well as transport- generated emissions are currently problems of global interest and represent two very important factors in setting national transportation policy. Reports from the Portuguese National Authority for Road Safety show that, in 2009, 737 individuals lost their lives in road crashes, and 2624 persons were seriously injured, from a total population of 10,627,250.

This research explores the conditional probability of crash severity levels for the population of crashes resulting in injuries and/or fatalities. Real world crash data were collected from the Portuguese Police Republican National Guard crash records for the Porto metropolitan area, for the period 2006-2008. From a total of 1925 gathered report crashes, vehicle technical data was available for 314 crash observations. This study has the main purpose of developing a comprehensive database and analysis methodology taking into account the vehicle characteristics effects on crash severity. Ultimately, the goal is the development of a crash-severity prediction model with application to crash analysis and prevention. In this paper, the effect of vehicle characteristics, such as weight, engine size, wheelbase and registration year (age of vehicle) were analysed with data mining methodology to extract patterns from the predictors and relate them to the dependent variables, including the number of injuries and fatalities in a crash, or a crash severity index. The research presented in this paper targeted at studying the effect of the vehicle characteristics on its occupants injuries and/or fatalities.

CART (Classification and Regression Trees) methodology was selected to analyse the independent variables that are more significant predictors of the dependent variables. The dependent variables used as targets were derived from the crash original variables. From this preliminary study, the following independent variables were selected: engine size of the vehicle, wheelbase of vehicle, Age of Vehicle and weight differential amongst two vehicles involved in a crash. The decision trees for the dependent variables indicated that selected differential vehicle characteristics provided some explanatory power of the injury severity levels.

This research is intended to support decision-making for safe and sustainable transportation policy and mobility in Portugal. The findings will provide meaningful interpretations that can be used to identify potential correlations amongst vehicle characteristics and injury risk. It will also provide important information to the automotive industry to produce low emission vehicles without compromising many of the basic vehicle functions of performance and safety.

Keywords: CART, Crash, Engine Size, Injured, Killed, Vehicle, Weight, Wheelbase

1. INTRODUCTION

During the past decennia there has been a steady increase in traffic volume, which has resulted in continuously increasing traffic congestion-related problems, such as pollutants emissions, delays, crashes, injuries and casualties. Recent status on road safety from Worlds Health Organization (WHO) showed that more than 1.2 million people die on the world´s roads every year (WHO, 2009a). Road traffic injuries are a major public health problem in Europe and cause the premature deaths of 120 000 people per year. In addition,

almost 2.4 million people are estimated to be seriously injured as to require hospital admission each year (WHO, 2009b).
The European Union had set ambitious target of reducing the yearly number of road deaths by 50% in 2010 compared to 2001. A recent study of 21 European Countries showed that Portugal had the lowest safety performance score (Hermans et al., 2009). The study suggests that Portugal should try to invest more in vehicle technology and in promoting new(er) cars (Hermans et al., 2009). The latest road safety indicators from the Portuguese National Authority for Road Safety (ANSR) show that during the year 2009, there have been a total of 35,484 crashes with injuries and fatalities on the Portuguese mainland roads. From those crashes, there were 737 fatalities and 2624 serious injuries (ANSR, 2010). When compared with 2008 data, these results show the following trend: an increase of 5.6% for crash with victims, a decrease of 5% in fatalities, and an increase of 0.7% in seriously injured. However it should be noticed that, until 2010, ANSR statistics did not include the number of fatalities which resulted from serious injuries during the 30 days time period after the crash. Therefore, the number of fatalities would be greater than would indicate the statistics.

There has been an increase in the amount of consumer interest in the safety performance. However, consumers tend to equate vehicle safety with the presence of specific features or technologies rather than with vehicle crash safety/test results or crashworthiness (Koppel et al., 2008). Crash testing is a resource for consumer regarding vehicle crash safety and credits a car manufacturer for focusing on safety. Two major providers of crash testing are the New Car Assessment Program (EuroNCAP) in Europe and the Insurance Institute for Highway Safety (IIHS) in the US. EuroNCAP discourage consumers from comparing ratings of cars from different segments, and in real crashes, there is obviously no control on the vehicle categories involved (EuroNCAP, 2009a, 2009b). Despite the rigorous scientific conditions under which crash tests are conducted, they have limitations: first, they do not account for mass differential between the vehicles involved within the collision; second, on real roads the speed of the crash impact frequently is higher than the 64 km/h, which is the speed at the frontal impact takes place in crash testing conducted by EuroNCAP and IIHS (EuroNCAP, 2009a; 2009b; IIHS-IHLDI 2008). Therefore, crash testing results must be viewed with some caution when it comes to predict car crashworthiness in crashes involving vehicles of different weights and sizes.

The research tasks addressed in this paper are indented to:
- Analyze the effect of vehicles characteristics in injuries and fatalities outcomes at occupants following a two-vehicle collision;
- Identify critical explanatory variables to include in a crash severity prediction model.

The paper is organized as follows: section 2 provides a literature review, followed by the methodology in section 3, results and discussion in section 4, and finally conclusions in section 5.

2. REVIEW OF TECHNICAL LITERATURE

A number of studies have attempted to correlate safety and vehicle design features. Evans (2004) explored vehicle mass and size, and concluded that those variables are strongly correlated, which makes it difficult to determine the separate contribution of mass and size on crash risk. Wood and Simms (1997) showed that in collisions between cars of similar size and in single vehicle crashes the fundamental parameters which determine the injury risk are associated to the size, i.e. the length of the vehicle. However, in collision between dissimilar sized cars the fundamental parameters are the weight and the structural energy absorption of the vehicle. Wenzel and Ross (2005) research suggested the quality of cars may be more correlated to the risk than weight, but this correlation is not strong. Most of the range in risk in cars must be attributed to vehicle design and to the difficulty to quantify driver characteristics and/or behaviour (Wenzel and Ross, 2005). Broughton (2008) showed that the driver casualty rate decreases with the size of his car, however the driver casualty increases with the size of the other car involved in the collision. When cars were grouped by year of first registration rather than type, the driver of the older car tends to be at greater risk than the driver of the newer car (Broughton, 2008). More recently, Tolouei et al. (2009) showed that increasing vehicle mass generally decreases the risk of injury to the driver.

Previous studies related to crash analyses have used a broad spectrum of statistical models to reach conclusions. For example, statistical regression models are very popular for analyzing contributing factors to injury severity (Li and Bai, 2008; Boufous et al., 2008; Bedard et al., 2002; Al-Ghamdi, 2002). However, regression models have many assumptions and pre-defined underlying relationships between the dependent and independent variables (Chang and Wang, 2006; Das et al., 2009). A more advanced data mining technique is the Classification and Regression Trees Analysis (CART). CART methods do not require predefined causal relationship between target (dependent variable) and predictors (independent variables). Chan and Wang (2006) have classified CART as a flexible non-parametric technique which can provide more informative and smart set of models, and its application is a valuable precursor to a more detailed logistic regression analysis in crash injury data.

In summary, much has been said about the high risk of low-mass cars in certain kinds of collisions. Various analyses have attempted to address this fundamental difficulty but the explanations are not completely satisfactory. There is no doubt that vehicle design can influence not only avoidance and crashworthiness, but also whether it endangers the occupants of the other vehicle(s) involved in the collision. In the preceding studies, the effect of differential vehicle weight and size and its impact on overall crash severity has not been studied. Hence, further research is needed to address the issue of the vehicle mass effect on road crash risk, leading to better strategies for traffic safety policies.

3. METHODOLOGY

One motivation for this research is to develop a prediction model for crash severity (both in terms of drivers and passengers) which focuses on the vehicle fleet characteristics and uses real-world crash data. Such model will have the potential to estimate the safety effects of many vehicle related independent variables: make and model, weight (mass), engine size, wheelbase, size (length), year of registration (age), mileage and fuel type. Dependent variables are related to the occurrence or number of severe injuries and/or fatalities of the vehicles occupants. The research focused exclusively on post-crash consequences rather than on pre-crash contributing factors to the event, such as driver behaviour or driver´s age.

3.1 Crash Data Collection

Data for the crash severity model development were collected from the road traffic Departments of the Portuguese Road Safety Police National Republican Guard (GNR) and Portuguese Public Safety Police (PSP). Crash reports that involved injuries and/or fatalities outcomes were exclusively selected and linked to passengers' cars, sport utility vehicles (SUVs), vans and pick-up trucks. From an extensive database of the Police Road Accident Records Reports, a total of 1925 reports were extracted, as indicated in Table I. The reports were gathered for a time period of three years, 2006 to 2008, for the urban areas of Aveiro and Porto, in Portugal.

Table I – Crash Frequencies

Data Source	**Crash reports recorded by Police Force, by Year**			**Sum of Recorded Crashes by Source**
	2008	2007	2006	
GNR Porto, PT	508	548	298	1354
PSP Aveiro, PT	65	65	-	130
PSP Porto, PT	275	166	-	441
Total	848	779	298	1925

From a total of 1925 gathered report crashes, vehicle technical data were available for 314 crash observations with exploratory variables to analyse the causative effect into crash severity outcomes, predictive variables, such as number of serious injuries and fatalities. The crash dataset for those completed crash observations is described next.

3.2 Crash Dataset Development

GNR reports produce a higher incidence of crash severity than PSP reports. GNR is responsible for the traffic safety enforcement on roads where the speed limit is higher, such as freeways and motorways; these reports were selected and analyzed in this study. A sample of 23% of crashes (314/1354) collected from the GNR records was used to develop the crash dataset. Those records provide the following information about the crash event: date, location, driver's identification ID, crash type and crash outcomes. On the other hand, the information related to the vehicles involved was partial, mainly the vehicle's registration plate and year. Since the main goal of this study was to analyze vehicle characteristics effects on the crash severity outcomes, the vehicle technical features were requested from the Portuguese central administration responsible for the coordination of inland transport, Institute for the Mobility and Inland Transportation (IMTT). An integrated database was developed where each crash record and technical characteristics for the vehicles involved in the collision were combined into a unique crash observation.

3.2.1 Crash Information provided from Police Records

For each crash event, information was gathered including road type and location, weather conditions, driver's alcohol and/or drugs test results, crash type, vehicles' registration plate and registration year, and crash outcomes, namely vehicle occupant's injuries and/or fatalities. Further, each vehicle's information in the crash dataset was recorded following the order stated at the Police record. As an example, the first Vehicle (V_1) in a collision report tends to be associated with the one that initially collided with the second vehicle (V_2) and/or causes the crash collision with it and possibly with a third vehicle (V_3). However the vehicle order in the police records does not follow this uniform protocol.

3.2.2 Vehicles Characteristics provided from IMTT Database

For all vehicles involved in the crash the vehicle registration plate was linked to the IMTT database. This database includes a specification sheet which contains the technical data attribute of vehicle ID and vehicle's registration plate. An example of the vehicle's characteristics acquired from those databases is listed below:
Brand (Toyota), Model (Corolla), Wheelbase (2465 mm), Size (4095 mm), Weight (1045), Engine Size (1332 cm^3), Fuel (Gasoline). These detailed vehicle specifications have not been detailed and considered in previous safety research.

3.3 Statistical Analysis

The crash data set was organized in order to be imported to Statistical Analysis Software, SAS® v9.1 and SAS®Enterprise Minner™5.2 software (SAS Institute Inc., 2009a, 2009b). Table II lists the original crash data set variables that were incorporated for further analyses of their significance to the crash severity indices.

The data set includes several classes and categories of variables, such as vehicle weight, vehicle engine size, vehicle age, crash type, weather conditions. CART methodology was selected for the following reasons. First, traditional statistics have limited utility in the task of variable selection for multiple variable comparisons (Lewis, 2000). Second, predictable variables are rarely satisfactorily distributed (Lewis, 2000). Third, in the crash data set complex interactions may exist amongst the explanatory variables, such as vehicle engine size, vehicle weight, vehicle age, crash type and weather conditions. CART has the potential to "uncover complex interaction between predictors which may be impossible to uncover using traditional multivariate techniques" (Lewis, 2000). In the next section, results for CART analysis are presented for the targets (dependent variables) and explanatory (independent variables).

Table II – Description of original crash data set variables

Dataset section	Variable name	Variable output/response
Crash safety outcomes	# Occupants injured and/or killed	# Light injuries # Seriously injuries # Fatalities
Driver information	Illegal alcohol level	Yes No
Weather conditions	Good conditions Poor conditions	Dry surface Wet surface
Road information	Road class	Two-lane Multi-lane Motorway
	Median type	Divided Undivided
	Speed limit (km/h)	50 90 100 120
Crash description	Crash location	Road name
	Crash type	Ran off road Rollover Collision with a fixed object Collision with a parked vehicle Head-on collision Rear-end collision Sideswipe collision Unknown/other
	# Vehicles involved	One Two Three
Vehicle information	Vehicle characteristics	Make Model Weight Engine size (cm^3) Wheelbase Size (length) (mm) Fuel type Year of 1st registration (age) (Yr) Mileage (VKT) (Vehicle kms traveled)

Table III identifies 4 categories for the dependent variables and depicts the independent variables found to be related to the variable of interest (target) based on the CART methodology. The dependent variable categories were defined by performing calculations and aggregations with the original crash outcomes, namely the number of light injuries (LI), serious injuries (SI) and killed (K). As an example: the dependent variable entitled "SIK" was created to signify the sum of the number of serious injuries and fatalities in a crash.
To test the significance between the target variable and the predictor categories (terminal leafs at the tree structure) the Fisher´s Exact Test was used since the Chi-Square test is not valid at the 5 % significance level for those cells that had expected counts less than 5 (SAS Institute Inc., 2009b).

Table III – Description of variables used as inputs and targets in CART methodology

Variable Category	Description	Name @SAS	Metadata Status @ Trees		
			Input	Rejected	Target
Age of Vehicle 1	$AgeV_1$ (yr) was calculated based on the year of the crash event minus the year of the first vehicle registration.	$AgeV_1$	T1,T2		
Age of Vehicle 2	$AgeV_2$ (yr) was calculated based on the year of the crash event minus the year of the first vehicle registration	$AgeV_2$	T1,T2		
Age Difference between vehicles (V_2) and (V_1)	AgeV2V1 (yr) stands for age of vehicle V_2 minus the age of vehicle $V_{1,}$ crash observation.	AgeV2V1	T1,T2		
Alcohol and/or Drugs	The Driver´s test for alcohol and or drugs is presented as: Code=0, legal; Code=1, illegal	AlcoholDrugs	T1,T2		
Crash Type	Single vehicle collisions types are: Ran off road Rollover Fixed object Multi-vehicles collision types are: Parked vehicle Rear End Head-On Sideswipe Unknown/other	CrashCode	T1,T2		
Divided/ undivided	Existence or absence of physical median: Code=0, undivided Code=1, divided	DivisionCode	T1,T2		
Number of Killed (K) plus Serious Injured (SI)	SIK stands for the sum of occupants serious injured (sum SI) plus the sum of occupants killed (sum K) in a crash SIK	SIK		T1,T2	
Serious and/or Fatal SIK	FatalSIK is a categorical response for a crash outcome used to predict either a serious injury, or fatality in a crash event. FatalSIK=1, if SI>0 and/or K>0, else, FatalSIK=0	FatalSIK		T1,T2	
Serious and/or Fatal SIK for vehicle 1 (V_1) occupants	FatalSIKV1 is a categorical response for a crash outcome used to predict either a serious injury, or fatality or both for occupants in vehicle 1 in a crash event. FatalSIKV1=1, if SI>0 and/or K>0, else, FatalSIKV1=0	FatalSIKV1		T2	T1

Serious and/or Fatal SIK at vehicle 2 (V_2)	FatalSIKV2 is a categorical response for crash outcome for a crash outcome used to predict either a serious injury, or fatality or to both for occupants in vehicle 2 in a crash event. FatalSIKV2=1, if SI>0 and/or K>0, else, FatalSIKV2=0	FatalSIKV2		T1	T2
Lanes Same Direction	Number of lanes	LanesSD	T1,T2		
Lanes Opposite Direction	Number of lanes	LanesOD	T1,T2		
Road Class	Roads is based in the number of lanes and coded as follows: Code=0, two lanes Code=1, multi-lanes Code=2, motorway	RoadClass	T1,T2		
Speed Limit	The legal speed limit are as follows: 50, 90, 100, and 120 km/h	SpeedLimite	T1,T2		
Wheelbase of Vehicle 1	Wheelbase of vehicle (V_1) (mm)	WBV_1	T1,T2		
Wheelbase of Vehicle 2	Wheelbase of vehicle (V_2) (mm)	WBV_2	T1,T2		
Wheelbase Difference between vehicles (V_2) and (V_1)	WBV2V1 stands for wheelbase of vehicle V_2 minus the wheelbase of vehicle $V_{1,}$ at crash observation, (mm).	WBV2V1	T1,T2		
Weight of Vehicle 1	Weight of vehicle 1 (V_1) (kg)	WTV_1	T1,T2		
Weight of Vehicle 2	Weight of vehicle 2 (V_2) (kg)	WTV_2	T1,T2		
Weight Difference between vehicles (V_2) and (V_1)	WTV2V1 stands for weight of vehicle V_2 minus the engine size of vehicle $V_{1,}$ at crash observation (kg).	WTV2V1	T1,T2		
Weather Conditions	Weather conditions at the moment of the crash: Code=0, Clear and/or dry pavement Code=1, rain and/or wet pavement	WeatherCode	T1,T2		
Engine Size of Vehicle 1	Engine size of vehicle (V_1) (cm^3)	ccV_1	T1,T2		
Engine Size of Vehicle 2	Engine size of vehicle (V_2) (cm^3)	ccV_2	T1,T2		
Engine Size Difference between vehicles (V_2) and (V_1)	ccV2V1 stands for engine size of vehicle V_2 minus the engine size of vehicle $V_{1,}$ at crash observation, (cm^3).	ccV2V1	T1,T2		

T1 Indicates the variables included at CART methodology to develop the Tree presented in Figure 1.

T2 Indicates the variables included at CART methodology to develop the Tree presented in Figure 2.

4. RESULTS AND DISCUSSION

At the present stage of the research the main objective was to identify and categorise those vehicle characteristics that are strongly correlated with the crash severity. The discussion in this paper is focused on the results obtained for the two-vehicle collision crashes. This chapter presents the initial results including descriptive statistics for dependent and independent variables, CART methodology selection of independent variables, and significance test for those variables categories.

4.1 Descriptive Statistics

The descriptive statistics analysis for the crash data set which includes a total of 314 crash observations is presented in this section. Data were arranged into the following categories according to the number of vehicles involved. "All" crashes refers to multi and single vehicle crash (314 crashes observations). "Two" refers to the sub set involving two vehicles crashes (186 crash observations). From those, 59% involved collision between two vehicles, 31% involved a single vehicle, and 10% involved three vehicles. The total numbers of injuries and fatalities are: 25 Killed (K), 59 serious injuries (SI), and 431 Light Injured. Per crash event, the overall crash output as grouped was follows: sum of light injuries (SUMLI), sum of serious injuries (SUMSI), and sum of killed (SUMK). Subsequent analysis showed that 24 crashes were categorized as fatal (sum K >0), and 67 crashes were categorized as serious and/or fatal (FatalSIK coded "1"). Table V shows that the maximum number of seriously injured per crash was three. On the other hand, the maximum number of fatalities in a crash event was two. For crashes involving two vehicles, (186 observations), the average differential values for weight, engine size and wheelbase are: 53 kg, 74 cm^3, and 15 mm, respectively.

Table IV – Descriptive Statistics for selected variables in all crashes (N= 314) and two-vehicle collision (N= 186)

Variable name @SAS	N	Mean	Std.	Minimum	Maximum
SUMLI[1]	314	1.373	1.05	0	8
SUMSI[2]	314	0.188	0.479	0	3
SUMK[3]	314	0.0796	0.282	0	2
SIK[4]	314	0.268	0.575	0	3
VehicleInvolved[5]	314	1.790	0.604	1	3
WTV_1[6]	186	1215 kg	315	640 kg	1957 kg
ccV_1[7]	186	1649 cm^3	446	1998 cm^3	3222 cm^3
WBV_1[8]	186	2598 mm	273	1998 mm	4025 mm
$AgeV_1$[9]	186	10 yr	5	<1yr	21 yr
WTV_2[10]	186	1267 kg	375	715 kg	3500 kg
ccV_2[11]	186	1744 cm^3	550	995 cm^3	3824 cm^3
WBV_2[12]	186	2613 mm	309	2159 mm	4100 mm
$AgeV_2$[13]	186	9 yr	5	<1yr	23 yr
WTV2V1[14]	186	53 kg	489	-1115 kg	2860 kg
ccV2V1[15]	186	74 cm^3	698	-1597 cm^3	2716 cm^3
WBV2V1[16]	186	15 mm	386	-1640 mm	1674 mm
AgeV2V1[17]	186	<1 yr	8	-17 yr	19 yr

1 Sum of LI; **2** Sum of SI; **3** Sum of K; **4** Sum of SI and K; **5** Number of vehicles involved in the crash; **6** Weight of Vehicle V_1; **7** Engine size of Vehicle V_1; **8** Wheelbase of Vehicle V_1; **9** Age of vehicle V_1; **10** Weight of Vehicle V_2; **11** Engine size of Vehicle V_2; **12** Wheelbase of Vehicle V_2; **13** Age of vehicle V_2; **14** Weight Differential between V_2 and V_1, at two vehicle collision; **15** Engine size differential between V_2 and V_1 at two vehicle collision; **16** Wheelbase Differential between V_2 and V_1 at two vehicle collision; **17** Age Differential between V_2 and V_1 at two vehicle collision.

4.2 Classification and Regression Tree Analysis

The classification trees were developed for the following targets variables: FatalSIK, FatalSIKV, FatalSIKV1 and FatalSIKV2, as explained in Table IV. CART analysis was performed using SAS® Enterprise Miner™ 5.2 software. Trees discussion is presented for two vehicles crashes, only two vehicles were involved in the collision.

The CART methodology was first applied to select the vehicles exploratory variables associated to the overall crash severity that that is expressed by the target FatalSIK. The results showed that the age of the vehicle V_1 was the first explanatory variable selected by CART methodology to separate the two-vehicle crashes. Subsequently for the V_1 vehicles category AgeV_1 < 6.5 yr old, the wheelbase differential between the two vehicles involved (WBV2V1) was used for the branch spilt. The highest FatalSIK average (75%) was found for vehicles involving these categories: V_1 newer that than 6.5 yr old, and WBV2V1 >= 216 mm. When one vehicle involved had a wheelbase larger than 216 mm than the other vehicle, this differential might increase the risk of having a crash outcome with FatalSIK being "1". For the older vehicle V_1 category, AgeV_1 >=6.5 yr old, the age of the other vehicle involved was also an important factor, the category of AgeV_2< 1.5 yr old. The fact that newer models include features that offer a better protection to its occupants, but they might contribute to a higher risk for the occupants in the older vehicle, age of V_1 > 6 yr old. For the newer V_2 category, the weight differential between the two vehicles involved (WTV2V1) lighter than 290.5 kg was associated to a higher average of FatalSIK (67%) than the average at the WTV2V1 category >=290.5 kg (0%). Fisher's Exact Test results for those vehicles categories and FatalSIK risk showed a p-value equal to 9.32E-8. Hence the target, FatalSIK and the exploratory variables presented above cannot be considered independent at the 0.05 significance level.

The higher probability of FatalSIK associated with the category of dissimilar vehicle weights lighter than 290.5 kg, needs more clarification through more careful analysis of crash data. When targeted FatalSIK was analyzed the injured and killed could be in V1 or V2, or at both vehicles involved.

An advances strategy that was developed by targeting the variables FatalSIKV1 and FatalSIKV2 is discussed next.

4.2.1 Classification Tree for FatalSIKV1 for Two-vehicle Crashes

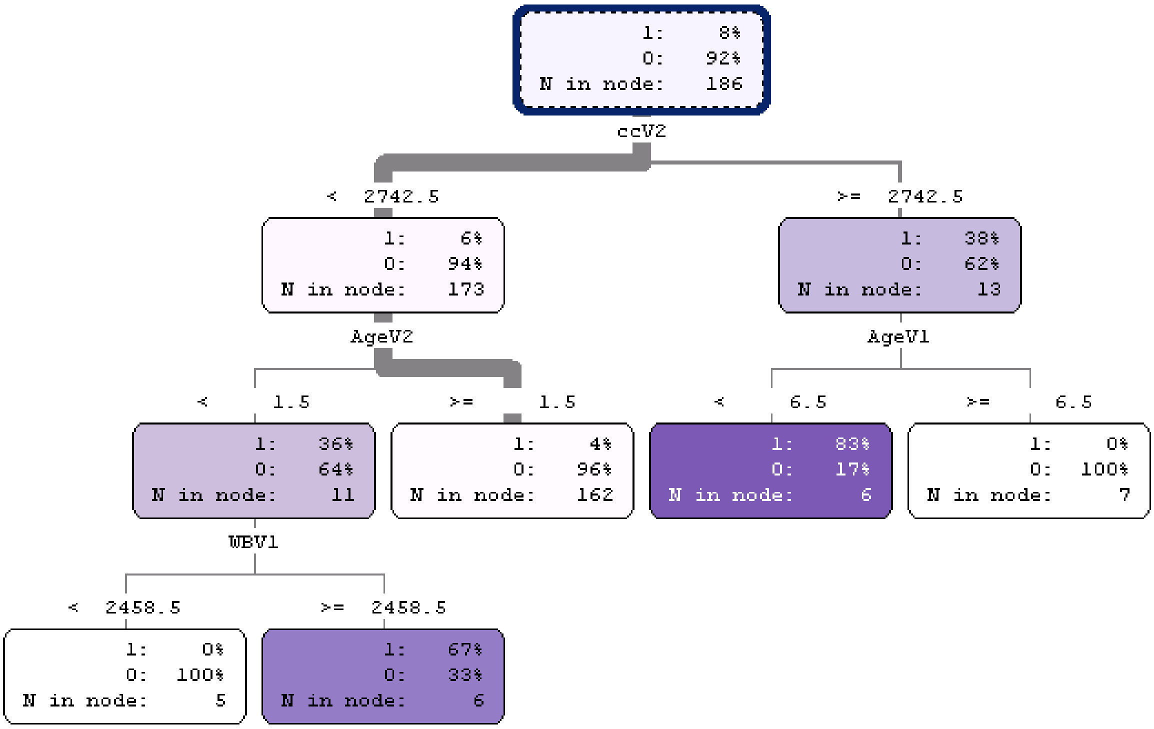

Figure 1 – Classification Tree for FatalSIKV1 in a two-vehicle crash data.

In this section, the analysis focuses the crash outcome in vehicle V_1. The input, as well as the target variables is indicated in table III. Figure 1 shows the output of CART methodology for the target variable FatalSIKV1, which is represented as a binary variable, “1” or “0”. The engine size of vehicle V_2 was the first explanatory variable selected by CART for the branch first split. The V_2 vehicles category ccV_2 >= 2742.5 cm^3 showed a higher average FatalSIKV1 (38%) when ccV_2 < 2742.5 cm^3 with a lower percent of FatalSIKV1 (6%). Subsequently, this category was split by the Age of V_2, showing that when vehicles with $AgeV_2$ <1.5 yr are involved, the crash results in a higher FatalSIKV1 average (36%) for occupants of vehicle V_1. This may be explained by the fact that newer vehicles models tend to include features that provide a better protection to its occupants, (for example deformation zones). Hence newer vehicles are more protective to its occupants; however they are might contribute to a higher risk for the vehicles of the other vehicle involved in the collision, V_1. Next, the tree was split by the wheelbase factor of V_1 (vehicle linked to the target variable SIKV1) leading to two terminals leaves. For V_1 in the category WBV1 >= 2458.5 mm the FATALSIKV1 average was higher (67%) than for the opposite category, WBV1 < 2458.5 mm with all observations of FatalSIKV1 being “0”. Usually, higher wheelbase values are associated with higher safety. However for the total of 186 crash observations involving two-vehicle collision the WBV1 average was 2598 mm (Table IV). The higher average values for FATALSIKV1 that were found at the category WB>= 2458.5 mm it may be relate to the fact that more V_1 vehicles were found in this category (closer to the average WB value). The highest FATALSIKV1 average (83%) was found for the collisions involving vehicles V_2 with ccV2 >=2742.5 cm^3 and vehicles V_1 less than 6.5 yrs old.

A real world Portuguese crash example is presented for those categories; it refers to a crash record at the motorway A4, involving two vehicles in a rear end collision. Vehicle V_1 was a Ford model Transit Kombi with the following characteristics: weight=1699 kg, engine size= 1998 cm^3, wheelbase=2933 mm, and one year old. This vehicle hit a second vehicle, V2, which was a Toyota model Dyna 250, with the following characteristics: weight=1800 kg, engine size= 2977cm^3, wheelbase=3215 mm, and is 20 yr old. Despite V_1 being a newer vehicle, which is one year old, all the injuries were produced at the Ford Transit Kombi with two occupants suffering light injuries and two others occupants seriously injured. No injuries occurred in the Toyota Dyna 250 model. As mentioned previously, this category of ccV_2 could be more aggressive to the occupants of the other vehicle involved in the collision, passengers at V_1. Even though it would be expected that the drivers of older vehicles tend to be at a higher risk than the driver of the newer vehicle, as Broughton (2008) showed, the impact of the collision velocity may explain the results for vehicle V_1 occupants' injuries severity. For the V1 age category (Age<6.5 yr), the crashes where found in roads classes (that include multi-lane and motorway facilities) with an average legal speed limit was 113 km/h. However, for the older V1 category, AgeV1>=6.5 yr, the average road speed limit was lower, 105 km/h. This inspection may contribute the highest average of FatalSIKV1 for the newer vehicles category, AgeV1<6.5 yr old, since the crashes occur at higher velocity impact, leading to a higher severity. These two factors may explain the highest average for FatalSIKV1. Fisher's Exact Test results for the tree terminal leafs categories (four categories corresponding to four terminals leaf) showed a p-value equal to 2.657E-8. Hence the target, FatalSIKV1 and the exploratory variables explained above cannot be considered independent at the 0.05 significance level.

4.2.2 Classification Tree for the Target FatalSIKV2 for Two-vehicle Crashes

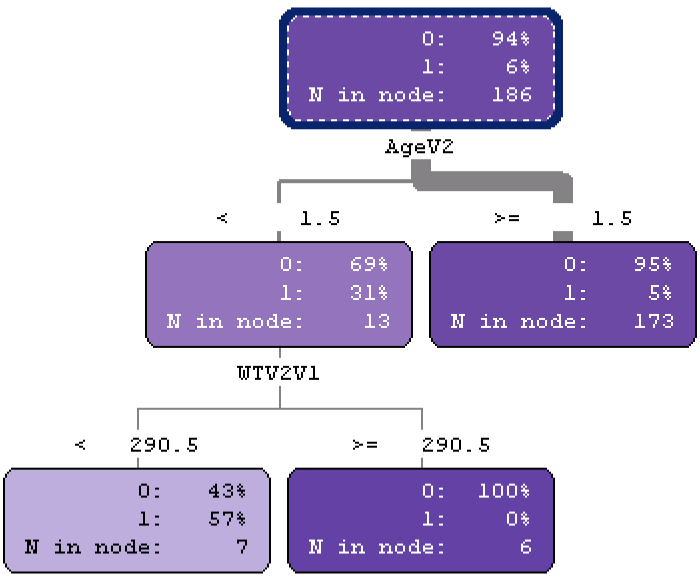

Figure 2 – Classification Tree for FatalSIKV2 in two-vehicle crash data.

In this section, the analysis focuses on the probability of a fatality or severe injury in vehicle V_2. The input variables as well as the target variable (FatalSIKV2) used in this tree are shown in Table IV. The output is shown in Figure 2. AgeV_2 was the first input selected by the CART methodology in splitting the two-vehicle crashes. This finding is consistent with Broughton (2008) work that showed that the driver of the older car tends to be at a higher risk than the drivers of the newer car. For the newer V_2 vehicles, AgeV2 < 1.5 yr old, the WTV2V1 appears to act as an important factor that may impact the serious injuries and/or fatalities in V_2. The category WTV2V1 < 290.5 kg shows the highest average of FatalSIKV2 (57%). This result is intuitive since the lower range of weight differential between the vehicles V_2 and V_1 is associated with an increase of the effect to WTV_1 during the collision leading to a higher risk of potential of injuries and/or fatalities for V_2 occupants. On the other hand, for the category WTV2V1 >=290.5 kg, the effect of WTV_1 is less aggressive. Further, the heavier WTV_2 provides a better protection during the impact of the collision, decreasing the risk of serious and/or fatal injuries for V_2 occupants (0%). These research findings are consistent with Woods and Simms (1997) study for collisions between dissimilar sized vehicles (they showed that weight and structural energy absorption were fundamental parameters to determine the injury risk).

To illustrate the effect of those vehicle characteristics a Portuguese real world crash is presented. A crash recorded at a multi-lane lane road, EN105, involving two vehicles in a head on collision resulted in severe outcomes. Vehicle V_1 was a Renault Clio model with the following characteristics: weight=900 kg, engine size=1870 cm^3, wheelbase=2472 mm. This vehicle was 18 years old. Vehicle V_2 was a Toyota Yaris weighting with 1155 kg, 1364 cm^3 2460 mm, and one yr old. Despite this second vehicle being 255 kg heavier than V_1 (WTV2V1=255 kg) and 17 yr newer than V_1 (it included air bags and other safety features), V_2 had one serious injury and one killed (two occupants as total), and V_1 had one killed (the driver was the single occupant). Fisher's Exact Test results for the tree terminal leafs categories (three categories corresponding to four terminals leaves) showed a p-value equal to 0.0012. Hence the target, FatalSIKV2, and the predictor explained above cannot be considered independent at the 0.05 significance level.

5. CONCLUSION

The preliminary results of this research showed that vehicle characteristics have an effect on passenger injuries and fatalities. The CART decision trees presented in this paper showed that engine size of the vehicle 1, wheelbase of vehicle 1, Age of vehicle 1 and weight differential amongst the two vehicles involved in the collision (WTV2V1) are important explanatory variables for crash severity response within each vehicle.

The CART methodology applied to the analysis of vehicle 1, V_1, response to the collision with the other vehicle, V_2, and expressed as FatalSIKV1, showed that the Age of V_1, along with the engine size of the second vehicle, ccV_2, impact the crash severity outcomes in V_1. The V_1 category "AgeV_1<6.5 yr", and vehicles V_2 category "ccV_2 >= 2742.5 cm^3" correlated with the highest FATALSIKV1 average (83%). One explanation for this highest average FatalSIKV1 may be the fact the vehicle category V_1 newer than 7 yr was involved in crashes that were reported at road facilities with higher average speed (113 km/h). Hence, when these vehicles are involved in a crash where the other vehicles have a higher engine size (ccV_2>2743 cm^3), the occupants of vehicle V_1 are exposed to a higher risk of suffering serious injuries and/or fatalities.

The CART methodology applied to the analysis of serious injuries and fatalities in vehicle V_2 using FatalSIKV2 as target variable. It showed that for those vehicles newer than 1.5 yr, the weight differential between the two vehicles involved in the collision (WTV2V1) was an important factor in vehicle V_2 crash severity outcomes. For newer V_2 category, a weight differential WTV2V1 <290.5 kg, occupants of this vehicles category have highest average of FatalSIKV2 (57%). CART output for the target FatalSIKV2 showed that when V_2 is heavier than V_1 but with an increment less than 290.5 kg, its occupants are exposed to a higher risk of serious and fatal injuries after the collision. On the other hand, the trees also indicated that when V_2 is 290.5 kg heavier than V_1, its occupants appear to be more protected and only light injuries were observed, leading to 0% average of FatalSIKV2.

Following the CART methodology results the variables: engine size of the vehicle ($ccV_{1,2}$), wheelbase of vehicle ($WBV_{1,2}$), Age of Vehicle ($AgeV_{1,2}$) and weight differential amongst the two vehicles involved (WTV2V1) were selected for further analysis.

The main finding from this initial phase of the research is that in a two-vehicle collision, the crash severity outcomes not only depend on the injured occupant's vehicle characteristics, but also the severity outcomes are affected by the characteristics of other vehicle involved. The presented research supports the conclusion that crash severity and vehicles characteristics differential, mainly weight differential, cannot be considered independent at the 0.05 significance level. Hence, the vehicle attributes that showed a strongest effect on crash severity will be considered in the development of vehicle-based crash severity prediction models in the subsequent phase of the research.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Portuguese Science and Technology Foundation (SFRH/BD/4135/2007 and SFRH/BPD/21317/2005), Luso-American Foundation (Project 02-03/2008, within the Program FLAD/NSF - "Portugal-USA: Networks and Partnerships for Research") and United States National Science Foundation (NSF Grant No. CBET-0756263). The collaboration between Drs. Coelho and Rouphail was under the auspices of the Luso-American Transportation Impacts Study Group (LATIS-G). Any opinions, findings, conclusions or recommendations expressed in this document are those of the author(s) and do not necessarily reflect the views of FCT, FLAD, NSF, or LATIS-G. In addition, the authors would like to thank the collaboration and expertise of Dr. Joy Smith and Dr. David Dickey, from the Department of Statistics of North Carolina State University.

REFERENCES

Al-Ghamdi, Al. (2002). Using logistic regression to estimate the influence of accident factors on accident severity. Accident Anal Prev. 34 (6), 729-741.

Bedard, M., Guyatt, G.H., Stones, M.J. and Hirdes, J.P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. Accident Anal Prev. 34 (6), 717-727.

Blincoe, L.; Seay, A.; Zaloshnja, E.; Miller, T.; Romano, E.; Luchter, S.; Spicer, R. (2002). The Economic Impact of Motor Vehicle Crashes 2000. NHTSA Technical Report: Report No. DOT HS 809446. http://www-nrd.nhtsa.dot.gov/Pubs/809446.PDF

Boufous, S., Finch, C., Hayen, A. and Williamson, A. (2008). The impact of environmental, vehicle and driver characteristics on injury severity in older drivers hospitalized as a result of a traffic crash. J. Safety Res. 39 (1), 65-72.

Broughton, J. (2008). Car driver casualty rates in Great Britain by type of car. Accident Anal Prev. 40 (4), 1543-1552.

Chang, L. and Wang, H. (2006). Analysis of traffic severity: an application of non-parametric classification tree techniques. Accident Anal Prev. 38 (5), 1019-1027.

Das, A., Abdel-Aty, M. and Pande, A. (2009). Using conditional inference forest to indentify the factors affecting crash severity on arterial corridors. J. Safety Res. 40 (4), 317-327.

Euro NCAP (2009a). Frontal Impact. European New Car Assessment Program. http://www.euroncap.com/tests/frontimpact.aspx Accessed 11-23-2009.

Euro NCAP (2009b). Euro NCAP - For safer cars | comparable Cars. European New Car Assessment Program. http://www.euroncap.com/Content-Web-Page/0f3bec79-828b-4e0c-8030-9fa8314ff342/comparable-cars.aspx

Evans, L. (2004). How to make a car lighter and safer. Vehicle mass and size in Traffic Safety. SAE International.

Hermans, E., Brijs, T., Wets, G. and Vanhoof, K. (2009). Benchmarking road safety: lessons to learn from data envelopment analysis. Accident Anal Prev. 41 (1), 174-182.

IIHS-HLDI (2008). Offset barrier crash test protocol (version XIII). Insurance Institute for Highway Safety, Highway Loss Data Institute. http://www.iihs.org/ratings/protocols/default.html Accessed on 11-23-2009.

Insurance Institute For Highway Safety (IIHS) (2009). News release, April 14 New crash tests demonstrate the influence of vehicle size and weight on safety in crashes; results are relevant to fuel economy polices. 1005 N. Glebe Rd, Arlington, VA 22201. http://www.iihs.org/externaldata/srdata/docs/sr4404.pdf

Koppel, S., Charlton, J., Fildes, B. and Fitzharris, M. (2008). How important is vehicle safety in the new vehicle purchase process? Accident Anal Prev. 40 (3), 994-1004.

Lewis, R. (2000). An introduction to classification and Regression Tree (Cart) Analysis. Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California. http://www.google.com/search?hl=en&rls=com.microsoft:en-us:IE-SearchBox&ei=i1vIS9qZBpGS8QTD_eSfCw&sa=X&oi=spell&resnum=0&ct=result&cd=1&ved=0CAgQBSgA&q=how+interpret+a+tree+CART&spell=1

Li, Y. and Bai, Y. (2008). Development of crash-severity-index models for the measurement of work zone risk levels. Accident Anal Prev. 40 (5), 1724-1731.

Portuguese National Authority for Road Safety, Autoridade National Seguranca Rodoviaria (ANSR) (2010). Annual Report for the Year 2009. Road Casualties. Department for Road Safety. "(Document written in Portuguese)".

http://www.ansr.pt/Default.aspx?tabid=273&language=pt-PT

SAS Institute Inc. (2009a). Statistics I: Introduction to Anova, regression, and Logistic Regression. Cary, NC, USA. ISBN 978-1-59994-914-7.

SAS Institute Inc. (2009b). Applied Analytics Using SAS®Enterrprise MinerTM 5. Instructor-based training. Cary, NC, USA. ISBN 978-1-59994-515-6.

Tolouei, R. and Titheridge, H. (2009). Vehicle mass as a determinant of fuel consumption and secondary safety performance. Transport Res D, 14 (6), 385-399.

Wenzel, T. and Ross, M. (2005). The effects of vehicle model and driver behavior on risk. Accid Anal Prev. 37 (3), 479-494.

Wood, D. and Simms, C. (1997). Safety and the car size effect: a fundamental explanation. Accid Anal Prev. 29, 139-151.

World Health Organization (WHO), (2009a). Global Status Report on Road Safety. Time for Action. http://whqlibdoc.who.int/publications/2009/9789241563840_eng.pdf

World Health Organization (WHO), (2009b). European Status Report on Road Safety. Towards safer roads and healthier transport choices. http://www.euro.who.int/document/e92789.pdf