C. Pastor
Bundesanstalt für Straßenwesen,
Bergisch Gladbach, Germany

Paired-Comparison Study: Correlation between Euro NCAP Star Ratings and Accident Data from the National German Police Road Accident Statistics

Abstract

Today, Euro NCAP is a well established rating system for passive car safety. The significance of the ratings must however be evaluated by comparison with national accident data. For this purpose accidents with involvement of two passenger cars have been taken from the German National Road Accident Register (record years 1998 to 2004) to evaluate the results of the NCAP frontal impact test configuration.

Injury data from both drivers involved in frontal car to car collisions have been sampled and have been compared, using a "Bradley Terry Model" which is well established in the area of paired comparisons. Confounders – like mass ratio of the cars involved, gender of the driver, etc. – have been accounted for in the statistical model.

Applying the Bradley Terry Model to the national accident data the safety ranking from Euro NCAP has been validated (safety level: 1star <2 star <3 star <4 star). Significant safety differences are found between cars of the 1 and 2 star category as compared to cars of the 3 and 4 star category. The impact of the mass ratio was highly significant and most influential. Changing the mass ratio by an amount of 10% will raise the chance for the driver of the heavier car to get better off by about 18%. The impact of driver gender was again highly significant, showing a nearly 2 times lower injury risk for male drivers. With regard to the NCAP rating drivers of a high rated car are more than 2 times more probable (70% chance) to get off less injured in a frontal collision as compared to the driver of a low rated car.

Introduction

Today, Euro NCAP is a well established rating system for passive car safety. The significance of the ratings must however be evaluated by comparison with accident data. The variety of real world crash events raises the question, whether everyday scenarios can be covered by a small number of crash tests, conducted under artificial circumstances. Recent studies have already shown a positive correlation between the NCAP rating and real world crash performance. However, none of the studies have shown the size of the effect as compared to other effects, generated by confounders like mass difference of the cars in a two car crash or gender of the driver, etc. In addition several models have been used which do not account for the multilevel structure of traffic accidents but just compare the relative performance of cars using accident databases.

Hence, the central objective of my study was (1) to detect any correlation between NCAP rating and real world crash behaviour by means of sound statistical models which account for the multilevel structure of accidents as well as (2) to compare the effect attributed to the NCAP rating with other effects attributed to covariates like mass difference of cars or age and gender of the drivers involved.

The Bradley Terry Model

There is considerable existing work in the statistics literature concerning the repeated ranking of members of a group of individuals. The most fundamental model in this field is generally attributed to BRADLEY and TERRY. The Bradley Terry Model deals with the area of paired comparisons, where ranking takes place between members drawn from a group two at a time. In the model each member is assigned a real-valued positive number η. Thus for a group of m individuals, with $\eta=(\eta_1,....,\eta_i,....\eta_m)^T$, where η_i is associated with individual i, the probability p_{ij} of individual i being superior to individual j is given by

$p_{ij}=\eta_i/(\eta_i+\eta_j)$; $Odds_{ij}=\eta_i/\eta_j$.

The standard model can alternatively be expressed in the logit-linear form

Logit $[p_{ij}]=\lambda_i-\lambda_j$, where $\lambda_i=\log(\lambda_i)$ for all i.

Thus, assuming independence of all contests, the parameters λ_i, λ_j, etc. can be estimated by maximum likelihood standard methods.

To put the model in a crashworthiness context, consider the individuals to be passenger cars, being involved in two car collisions. Here η can be thought of as representing the crashworthiness of car i and p_{ij} as the probability of the passengers of car i getting away less injured than the passengers of car j, in a collision between the two. Clearly this model lends itself to the case of one-on-one collisions (1:1 matching or pair matching).

From a statistical point of view the restriction to two-car accidents (1:1 matching) has the advantage that all observed and unobserved characteristics of the accident itself (time, location, weather conditions, severity etc.) are the same for both accident-involved cars and, therefore, these characteristics cannot account for differences in the injury risk of the two drivers involved in the accident. Consequently, the 'pure' effect of NCAP rating on the probability of car driver injury can be measured more precisely. However, on the other hand, a number of accidents (single car accidents, accident between cars with identical rating or accidents with cars not rated) are ruled out and can not be used in the context of this model.

The Bradley Terry Model has been widely studied and has many generalisations and applications in a broad range of areas. In-depth explanations of which can be found in many sources [DAV, HUN, FIR].

When using the Bradley Terry Model in connection with crashworthiness ratings it is fundamental to choose a reasonable "winner function", a function which decides what car is the 'winner' in a car to car competition – or in other words in a two-car accident. There are several possibilities, e.g. defining a Severity Score S

$S = f$ (#fatalities, #serious, #slight)

for each car declaring the car with the lower severity score to be the winner. However, in my study only the injury severity of the driver of each car was used as severity score S, thus the car with the less injured driver won the competition. Having a more sophisticated "winner function" can be beneficial as this can reduce the number of ties, where both parties show the same injury severity score. Clearly those kind of accidents (ties) do not contain any information about the relative crash performance of the two cars involved in the crash.

Data

For purposes of this study a sample of the German police recorded accident register with car to car accidents between 1998 and 2002 was available. These dataset contains NCAP tested vehicles and non-NCPA tested vehicles.

There are 235,047 (out of 981,627) vehicles which could be considered NCAP tested. The classification of car type and model has been done on the basis of the German type- and vehicle-manufacturer code, which distinguishes cars by their motorization, chassis, kind of propulsion, cubic volume to name a few. It was decided that these variables can sufficiently specify a certain car in order to decide whether it is similar to the NCAP tested variant or not.

The sample dataset was supposed to be a 70% sample of the German Official Police Traffic Accident Statistics of car to car collisions where just two cars have been involved in the accident.

Figure 1.1 summarizes the number of accidents with fatal, serious and slight consequences for any car passengers in car to car collisions. The sample dataset contains 53% of all cars where passengers had fatal injuries, 56% of all cars where the driver sustained serious injuries and 57% of all cars with slight injury consequences. If one reduces the sample set to NCAP tested vehicles only, the data contains 9% of all cars with fatal consequences to passengers, 12% of all cars with serious and 14% of all cars with slight injury consequences. Thus it could be estimated that 18% of all fatalities happen in cars which are NCAP tested.

It becomes obvious, that NCAP vehicles are under-represented in the group of fatal and serious

	Fatalities		Serious Inj.		Slight Inj	
	Number	%	Number	%	Number	%
ALL police recorded accident data (1998-2002)	5,327	100	101,265	100	699,700	100
Sample database	2,807	53	56,266	56	401,666	57
Sample database with NCAP tested vehicles only	501	9	11,761	12	97,666	14

Figure 1.1: Sample size of GERDAT and GERDAT_NCAP as compared to the national police data (for accidents involving two cars)

Ctyp	0		1		2		3		4		5		6		7		8	
Atyp	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng
1	1,2	0,2	0,0	0,2	0,2	0,2	0,0	0,2	**17,8**	**9,0**	0,0	0,0	0,0	0,2	3,0	0,4	3,0	1,0
2	0,2	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,8	0,2	1,2	0,6	0,0	0,0	0,0	0,4	0,0	0,0
3	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,0	0,0	0,0	**15,4**	**3,0**	0,0	0,0	0,0	0,2	0,0	0,0
4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
6	0,6	0,8	0,2	0,0	0,6	0,6	1,4	1,0	**16,6**	**10,4**	0,0	0,0	0,0	0,2	1,0	0,6	1,2	0,2
7	0,4	0,0	0,4	0,2	0,0	0,0	0,0	0,0	1,8	0,6	0,8	0,0	0,0	0,6	0,6	0,0	0,0	0,2

Figure 1.2: Distribution [%] of combinations of Ctyp (kind of accident) and Atyp (type of accident) for all fatal cases with NCAP tested cars; separated into guilty (g) and not guilty (ng) party (n=501)

Ctyp	0		1		2		3		4		5		6		7		8	
Atyp	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng
1	0,4	0,4	0,3	0,2	0,3	0,4	0,3	0,2	**5,5**	**8,5**	0,1	0,1	0,0	0,0	1,3	0,7	1,3	0,9
2	0,1	0,2	0,1	0,1	0,5	0,9	0,1	0,1	1,6	2,2	3,0	3,9	0,0	0,0	0,0	0,1	0,0	0,0
3	0,1	0,2	0,1	0,1	0,0	0,0	0,0	0,0	0,1	0,2	**13,3**	**14,2**	0,0	0,0	0,0	0,0	0,0	0,0
4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	0,1	0,0	0,5	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
6	0,3	1,0	0,4	1,0	2,9	4,5	1,0	1,4	**6,6**	**10,1**	0,1	0,2	0,0	0,0	0,4	0,7	0,4	0,6
7	0,7	0,5	0,5	0,2	0,3	0,1	0,1	0,1	0,6	0,7	0,2	0,2	0,1	0,1	0,2	0,1	0,3	0,1

Figure 1.3: Distribution [%] of combinations of Ctyp (kind of accident) and Atyp (type of accident) for all serious cases with NCAP tested cars; separated into guilty (g) and not guilty (ng) party (n=11,761)

Ctyp	0		1		2		3		4		5		6		7		8	
Atyp	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng	g	ng
1	0,1	0,3	0,2	0,5	0,2	1,1	0,1	0,2	**1,2**	**3,4**	0,1	0,2	0,0	0,0	0,3	0,4	0,3	0,4
2	0,1	0,2	0,1	0,7	0,6	3,2	0,1	0,2	1,2	2,1	3,0	5,6	0,0	0,0	0,0	0,0	0,0	0,0
3	0,1	0,2	0,1	0,2	0,0	0,1	0,0	0,0	0,1	0,2	**9,1**	**18,6**	0,0	0,0	0,0	0,0	0,0	0,0
4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
5	0,0	0,1	0,5	0,7	0,0	0,2	0,0	0,0	0,0	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0
6	0,2	0,8	0,6	4,4	*3,1*	*21,3*	0,6	1,6	**1,4**	**3,7**	0,1	0,1	0,0	0,0	0,1	0,4	0,1	0,4
7	0,4	0,7	0,3	0,6	0,1	0,6	0,1	0,1	0,2	0,4	0,2	0,4	0,0	0,0	0,1	0,1	0,1	0,1

Figure 1.4: Distribution [%] of combinations of Ctyp (kind of accident) and Atyp (type of accident) for all serious cases with NCAP tested cars; separated into guilty (g) and not guilty (ng) party (n=97,666)

casualties. Although they make up 24% of the sample dataset, they make just 17% of the fatal cases and 21% of the serious cases [fatal: 1/0,53*0,09=0,17; serious: 1/0,56*0,12=0,21].

Looking in more detail at the structure of the accidents described by “kind” and “type” of accident there are some interesting facts becoming obvious (see Figure 1.2 to 1.4).

Certain combinations of the kind of accident (Ctyp), which describes the entire course of events in an accident, and the type of accident (Atyp), describing the conflict situation, turn out to represent most of the fatal casualties. Details on the exact definitions of CTyp and ATyp can be found elsewhere [STA].

The combinations (Atyp/Ctyp), (1/4), (6/4) do represent front to front crashes, (6/4) being the “classical” example for a front to front collision, whereas in (1/4) one party looses control on its car and wherefore receives more often serious consequences. Both types of accidents are responsible for nearly 54% of all fatal casualties in NCAP tested vehicles and nearly 31% of all serious cases in NCAP tested vehicles.

The combination (3/5) mainly builds up of front to side collisions, having the fatal consequences for the guilty party (Figure 1.2). This is usually the vehicle crossing the street and getting the side impact.

Database	Numbers
NCAP vehicles in the SAMPLE DATABASE	235,047
Accidents with at least one NCAP vehicle	212,672
“Front-Front“ accidents with at least one NCAP vehicle	17,962
“Front-Front“ accidents with both cars NCAP-tested	1,963
“Front-Front“ + “NCAP-NCAP“ + accident with at least one person sev. or fatally inj.	495

Figure 2.1: Relevant case numbers in the database

Looking at the slight injury cases (Figure 1.4) the combination (6/2) becomes obvious. These are rear end collisions with a stopping vehicle and the slightly injured person is most often the person in the stopping vehicle.

With regard to the matched pairs design of the study front to front collisions have been considered to produce the most ideal “pair” for two car collisions whereas both cars sustain more or less the same crash severity. This is most truly given in a front to front accident configuration. As described these accidents can be identified by proper combinations of kind and type of accident, here Atyp/Ctyp combinations (1/4) and (6/4). Comparison with the NCAP star rating was thereby restricted to comparisons with the frontal star rating only.

For this study it was furthermore decided that accidents with fatal and serious outcomes should be most meaningful with regard to the NCAP star rating. Here only the injury status of the drivers have been taken into account.

However, the numbers of available accident cases decreases rapidly when the data is restricted to this interesting group of accidents. This is quite natural taking into account the great number of low severity accidents which are excluded from the analysis. Furthermore NCAP tested cars do not often collide with another NCAP tested car, but more often with other not NCAP tested cars, which can be seen by comparing row 1 and 2 of Figure 2.1.

Within the remaining 495 car to car accidents there have been 64 fatalities, 771 severely injured and 453 slightly injured persons.

Data Analysis

The study focuses on frontal car to car collisions of NCAP tested cars and compares the real world performance to the frontal offset test results of the NCAP assessment. The official frontal rating is calculated using the crash dummy readings of both frontal passengers. This rating has been recalculated to assess merely the driver readings, thus giving a frontal rating for assessing the drivers risk of getting injured. This procedure was carried out by the Monash University in the course of the SARAC2 project. Using this kind of assessment complies with the “winner function” as described in the paragraph, describing the Bradley Terry Model.

Winner	Loser	Frequency	Reduced Frequency
1-STAR	1-STAR	14	-
2-STAR	1-STAR	22	22
3-STAR	1-STAR	28	28
4-STAR	1-STAR	6	6
1-STAR	2-STAR	16	16
2-STAR	2-STAR	48	-
3-STAR	2-STAR	62	62
4-STAR	2-STAR	12	12
1-STAR	3-STAR	12	12
2-STAR	3-STAR	37	37
3-STAR	3-STAR	36	-
4-STAR	3-STAR	10	10
1-STAR	4-STAR	2	2
2-STAR	4-STAR	4	4
3-STAR	4-STAR	9	9
4-STAR	4-STAR	3	-
SUM		321	220

Figure 3.1: Winner-Loser-Frequency Matrix of frontal car to car collisions of NCAP tested cars

Before starting the analysis the only 0-STAR car in the database, a Chrysler Voyager, was decided to be taken out. Setting up the data into a winner/loser shown matrix produces the result in Figure 3.1.

Figure 3.1 contains information on 321 car to car frontal collisions. This means that there have been 174 ties out of 495 accidents, where the injury status of both colliding cars have been identical and which have for this reason been of no value for the analysis. However, more cases of the table do not contain valuable information for the model, because winner and loser are of the same category. Referring to the second frequency column in Figure 3.1, the number of valuable accidents reduces to 220. Looking at the matrix the NCAP star rating seems to make sense, since higher rated cars do more often win against lower rated cars and vice versa.

	λ_i Estimate	Std.Error	Signif
1-STAR	0,000	0,000	-
2-STAR	0,306	0,247	0,2145
3-STAR	0,855	0,249	0,0006
4-STAR	1,133	0,378	0,0027
Null deviance: 445 on 321 degrees of freedom Residual deviance: 425 on 318 degrees of freedom AIC: 431			

Figure 3.2: Crashworthiness estimation by using a simple Bradley Terry Model

	λ_i Estimate	Std.Error	Signif
1-STAR	0,000	0,000	-
2-STAR	0,439	0,274	0,1086
3-STAR	0,714	0,274	0,0092
4-STAR	0,900	0,412	0,0291
MR	1,692	0,312	5,9e-08
FEMALE	-0,590	0,191	0,0021
Null deviance: 442 on 319 degrees of freedom Residual deviance: 362 on 314 degrees of freedom AIC: 372			

Figure 3.3: Crash-worthiness estimation by using a Bradley Terry Model with covariates adjustment

Fitting the data to a Bradley Terry Model produces the result, shown in Figure 3.2. Again, there is a clear tendency for better rated cars being more crashworthy than lower rated cars. The standard deviation does not allow for a separation of each class category. Thus it is possible to distinguish between a 1-star and a 3-star car, and also between a 1-star and a 4-star car. All other confidence intervals overlap on a 5% significance level.

By extending the model one can include confounding variables. Mass ratio (MR) defined by

MR=max.weight(winner)/max.weight(loser)

and gender of the driver have shown to be of significance.

Figure 3.3 shows the results. The mass ratio is significantly marked to play an important role in front to front car collisions.

Results

The main results of the study can be summarized as follows:

(1) After adjusting for confounding factors there remains a significant safety difference between cars of the 1- and 2- star category and cars of the 3- and 4- star cars, thus the NCAP star rating seems to be reflected by real world accidents.

(2) The mass ratio of the cars involved in a frontal car to car accident is the most powerful covariate. A 10% change in mass ratio results in a 20% increase in probability to get better off in an accident, or in other words the odds to get better off change from 50/50 to 60/40 for the driver of an equally rated but heavier car.

(3) There is a gender effect. Female drivers show an almost 2 times higher injury risk. The odds change from a 50/50 chance to get better off to a 35/65 chance for a female driver of an equally rated car of similar weight.

(4) The NCAP rating can have at a maximum chance the odds from 50/50 to 30/70, thus the driver of the lower rated car is 2 times more probable to get more injured in a crash.

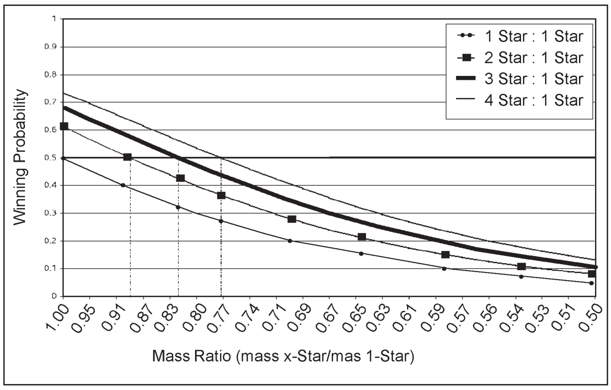

Figure 4.1: Comparison of star rating and mass ratio influence in frontal car to car accidents

The influence of the star rating as compared to the impact of mass ratio is depicted in Figure 4.1. The baseline car for all curves is an impact with a 1-star rated car.

Accidents between two 1 star rated cars with mass ratio 1 show a winning probability of 50% which is trivial. A collision between a 2-star and a 1-star rated car of equal weight shows a winning probability for the 2-star car of 62% and so on. It could be seen that the advantage of a 4-star car as compared to a 1-star car is compensated when the mass ratio reaches a value of 0,77; thus a 2,000kg 1-star car hitting a 1,500kg star 4-star car both will have a 50% probability of getting better off. Similar comparisons could be done with the other covariate which was the gender of the driver. The advantage of driving a 4-star car is thus compensated by the fact of a female driver when hitting a 1-star car with a male driver.

In further analysis it would be desirable to include further covariates like age of driver which has not been considered here. The special dataset of this study did not contain enough information for this purpose. It would furthermore of interest to extend the Bradley Terry Model to make more use of ties (accidents between cars with the same injury severity score).

Acknowledgements

This study was sponsored by the European Commission and carried out in context of the SARAC2 project. The author is especially graceful to the Monash University Accident Research Center (Prof. Max Cameron, Amanda Delaney) for providing data of the recalculated NCAP scores based on the driver dummy readings only.

References

DAV: H.A. DAVID: The Method of Paired Comparisons, 2nd edition, Oxford University Press, New York (1988)

FIR: D. FIRTH: Bradley Terry Models, Journal of Statistical Software, 12, 1-12 (2005)

HUN: D.R. Hunter: MM algorithms for generalized Bradley Terry Models, Annals of Statistics, 32, 384-406 (2004)

STA: Verkehrsunfälle – Fachserie 8, Reihe 7, download via http://www-ec.destatis.de/