S. Rohrer, L. Hannawald
Accident Research Unit, University of Technology (TU) Dresden, Germany

R. Koch
TU Dresden, Medical Faculty, Department of Informatics and Biometry, Dresden, Germany

H. Zwipp
TU Dresden, Medical Faculty, Department of Accident and Reconstructive Surgery, Dresden, Germany

Mortality of Occupants in Real Crash Scenarios – A Multivariate Analysis of Influences Regarding Mortality

Abstract

While the number of fatal accidents is diminishing every year, there is still a need of improvement and action to prevent these deaths. Basis for this purpose has to be an analysis about the factors influencing the car crash mortality. There are various studies describing the univariate influence of several factors, but crash scenarios are too complex to be described by a single variable. The multivariate analysis respects the interference of the variables and gets so to more detailed and representative results.

This multivariate analysis is based on about 2,600 cases (the data have been collected by the accident research units Hannover and Dresden (during the years 1999-2003). This paper presents a multivariate model (containing ten different variables) which detects 93% of these cases properly. This means it detects the cases as truly survived and truly death.

Notation

AUC	Area under the curve
CI	Confidence Interval
EES	Energy Equivalent Speed
MAIS	Maximal Abbreviated Injury Scale Value
OR	Odds Ratio
PROCAM	Prospective Cardiovascular Münster Study
ROC	Receiver Operating Characteristic

Introduction

There is no discussion about the necessity of traffic for our society. But there is also absolutely no doubt about the dangers which lies herein. Thanks to various improvements in different fields, especially in traffic safety (car design, road construction) we achieved an enormous decline in traffic death and injury severity. Major campaigns to increase safety belt use and to reduce impaired driving and the efforts of legislation were major contributors to the reduction in fatalities and in the fatality rate.

The highest number of fatal accidents in Germany was seen in 1970; more than 21,000 people lost their lives (see Figure 1). The German government took action and passed a law which obliged the automobile industry to install seat-belts in cars. This law was followed by the compulsion for passengers to use this restraint system (1976). Other major points, responsible for this steady decline, were the improvements in the emergency treatment at the scene of the accident and perhaps more important, the implementation of an emergency call system and the realization of an emergency telephone system along the road (1971, invented by Björn STEIGER). Till then it was absolutely normal to wait up to an hour before the ambulance car arrived. Also in the year 1970 the first rescue helicopter started his work. New findings concerning the pathophysiology of the shock and new imaging methods in the emergency room, e.g. the 16 scan computer tomography, led to a reduction of mortality. Nowadays the number of traffic deaths is as low as never before, although the quantity of

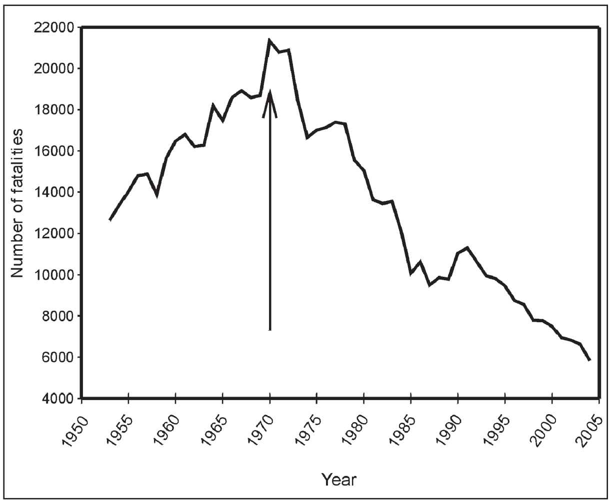

Figure 1: Fatalities in Germany from 1953 until 2004 (until 1989 numbers refer to Federal Republic of Germany and German Democratic Republic, since 1990 numbers represent the reunited Germany) [1, 2]

cars is three times higher than in 1970. So today the drivers on our roads are safer than they have ever been, in part because of the safer cars, higher safety belt use and stronger safety laws.

Despite these improvements, there are about 40,000 fatalities every year in Europe, which corresponds to the population of a town. Considering the age of these victims this number is dramatic, because most of them are young with full capacity for work. Concerning the economical feature of car crashes the direct costs amount to 45 billion Euro per year; the indirect costs are three to four times higher. Analogously to this Figure the annual amount of 160 billion Euros is assessed, this sum corresponds to 2% of the gross national product of the European Union [3].

But as long as the number of fatalities remains as high as it is, the efforts to reduce the mortality have to be continued. Therefore it is still of enormous importance to determine the factors associated with a higher risk of mortality. The analysis of data pools of traffic accident research institutions can detect negative trends. Only by this procedure it is possible to take early countermeasures.

Numerous studies described the univariate influence of different variables (such as seat-belt use, gender, age,…) on the mortality. But crash scenarios are too complex to be described by only one variable. The univariat analysis also often neglects the existence of confounders or covariates. With a multivariate analysis it is possible to examine the influence of several factors on the same outcome (mortality). Further, the multivariate analysis respects the interference between the different influencing factors and gets so to more detailed and more representative results. The odds ratio of the multivariate analysis reflects a more realistic probability.

Material and methods

Accident team structure, equipment and dataset

In Hannover (since 1973) and Dresden (since 1999) local traffic accident research units exist. They collect, code and report prospective data on motor vehicle crashes with at least one injured person. The team consists of two technicians, one medical student and one coordinator who stays at place to answer the phone and to guide the team quickly to the scene of the accident. The team is provided with two emergency vehicles and after being notified by police dispatchers they arrive often simultaneously with the rescue personnel on scene. In the field they record among other things the following data:

- Number, type and construction year of involved automobiles.
- Photographs for measurement of distance and deformations.
- Energy Equivalent Speed (EES), v, collision speed.
- Traffic control, road condition, weather condition.
- Reason for the accident.
- Deformities of the vehicle.
- Number of involved persons, number of injured persons.
- Personal data of the involved drivers or passengers (date of birth, height, weight,…).
- Transport of the injured.
- Initial assessment of injury type and severity [X-ray, ISS, AIS, incidence of multiple injury (polytrauma)].
- Treatment in the hospital.

All the collected information is supplied to a special data pool called GIDAS (German In-Depth Accident Study). To achieve representative results, it is necessary to record as many accidents as possible. But with the increasing quantity of recorded car crashes the quality of registered variables diminishes. The GIDAS data pool contains a healthy combination of 2,000 reported accidents per year with more than 2,000 single variables to code for each accident.

Definition of some important statistical terms

Each statistical test has an associated null hypothesis, the p-value is the probability that the sample could have been drawn from the population being tested (or fictive samples which are still more different from the null hypothesis), given the assumption that the null hypothesis is true. Null hypotheses are typically statements of no difference or effect. A p-value close to zero signals that the null hypothesis is false and that a difference is very likely to exist. Large p-values closer to 1.0 imply that there is no detectable difference for the sample size used.

Odds is the ratio of the probability something is true divided by the probability that it is not. An odds ratio is the ratio of two odds.

$$OR = \frac{\text{Odd for death by exposure}}{\text{Odd for death by non-exposure}} \qquad (1)$$

An odds ratio of 1.0 means, the independent (here exposure) has no effect on the dependent (here death). The larger the difference between the observed odds ratio and 1.0, the stronger the relationship. An odds ratio below 1.0 indicates that the independent is associated with a decrease in the odds. An odds ratio above 1.0 indicates an increase, thus the independent leads to an increase in odds and risk. Odds ratios are to be understood like relative risks. Multivariately, they can be estimated by logistic regression models.

If the 95% confidence interval on the odds ratio includes the value 1.0, the variable is not considered as a useful predictor variable. Thus, the independent variable is not associated with a change in the odds.

Cross-validation can be used to estimate the generalization error of the model. It is one approach estimating how well the model of the training data is going to perform on future as-yet-unseen data. Leave-one-out cross validation means that the function approximator is trained on all the data except for one point and that a prediction is made for that point. The average error is computed and used to evaluate the model.

An ROC curve is a graphical representation of the trade between sensitivity and specifity for every possible cut off of the estimated probability for the considered event (here death) estimated by a logistic regression model. By tradition, the plot shows the false positive rate (1-specifity) on the x-axis and the sensitivity on the y-axis. A ROC curve is good when it climbs rapidly towards upper left hand corner of the graph. This means that the sensitivity is high and the false positive rate is low. When the ROC curve follows a diagonal path from the lower left hand corner to the upper right hand corner it means that this test is no better than flipping a coin. Cause every improvement in false positive rate is matched by a corresponding decline in sensitivity. Quantifying how quickly the ROC curve rises is possible by measuring the area under the curve (AUC). The larger the area the better the prognosis. If the area is 1.0 the test is ideal; if the area is 0.5 then the probability for right prognosis is not more than chance.

So the closer the area is to 1.0, the better the test is, and the closer the area is to 0.5 the worser the test is.

Statistical methods

When the dependent variable is a dichotomy (here is the considered event the accidental death: yes/no) and the independent variable is of any type, a logistic regression is used. The regression equation reads as follows [4, 5]:

Log it (considered event)=

$$b_0 + b_1X_1 + b_2X_2 + \ldots + b_jX_j \qquad (2)$$

with the definition of logit:

log it (event)=log [odds(event)]=

$$\log \frac{\text{probability(event)}}{1 - \text{probability(event}} \qquad (3)$$

For the multiple regression it is necessary to determine the parameters b_0, b_1, b_2,..., b_J so that the likelihood of the sample is maximal.

The estimation of parameters and the validation of the resulting regression model are an object of statistical software, e.g. SAS® or SPSS®.

We included cases of the years 1999-2003, but only accidents with car drivers or passengers who sustained an injury. Condition for consideration was further the existence of the EES and the MAIS; if theses two variables were missing or coded “unknown” the case was excluded.

The multivariate analysis requires complete cases; the missing of the value of one variable leads to an exclusion of the whole case. Therefore this form of analysis demands a very high number of cases. The univariate analysis was carried out with 3,418 cases (with 77 fatalities), for the multivariate analysis this number was reduced to 2,609 cases (with 50 fatalities). Univariate estimates of the odds ratio with a 95% Confidence Interval (CI) were the starting point to find multivariate predictors. The development of different models to predict the mortality was carried out with manual and automatic steps. We used the so called forward stepwise method in which the likelihood ratio test determines which variables are added to the model. Starting with the constant, one variable is added after another in the order they predict the mortality as the best. Criteria for the comparison of different models were adequate validity (tested with the Hosmer-Lemeshow goodness of fit test, ROC- and AUC-analysis, Wald test of the coefficients), good interpretation and numeric stability (SPSS-results tested in SAS). After cross-validation the final model detects 92,62% of the cases in the right manner, which means it declares properly death or survival (see Figure 2).

Proceeding a univariate as well as a multivariate analysis several combinations are possible (see Table 1).

In conclusion:

- Univariate assessments can be falsified by bias; this effect can be uncovered by multivariate models.

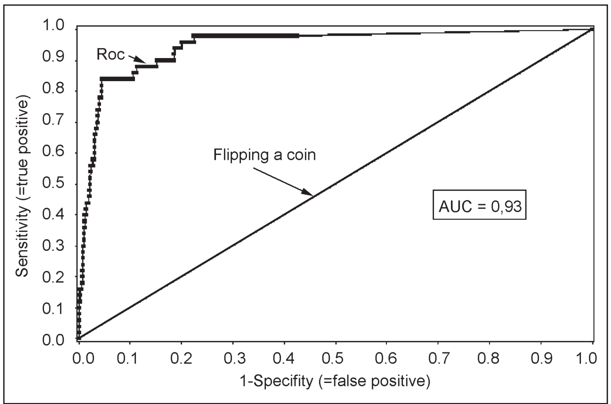

Figure 2: The ROC curve and the AUC of the multivariate analysis. The area under the curve represents the cases which are properly detected by this model

Univariate	Multivariate	Conclusion
significant	significant	Confirmation of the significance of the covariate
not significant	dispensable	Confirmation of the nonsignificance of the covariate
significant	dispensable	Multivariate analysis leads to devaluation. Reason either: a) Univariate pretended significance or b) The diagnostic/prognostic information of the covariate is already described by another covariate which is also contained in the model (e.g. interference/correlation of the covariates)
not significant	significant	Multivariate analysis leads to revaluation. Reason is an univariate covered significance (e.g. mixed population with opposite correlation in the different groups of the population)

Table 1: Comparison of the uni- and multivariate assessment [6]

- The assessment of the diagnostic/ prognostic significance of potential covariates can never finally answer the question of causal connection.
- Even with the same data pool it is possible to create different optimal models, which are equivalent in their quality. This means that the multivariate assessment of the covariates can differ. This is no contradiction, because the models have only a describing character for the relation with the objective variable (here: mortality).

Results

The univariate analysis considers only the influence of one variable on the mortality. The following table (Table 2) contains 25 different variables, whose influence is analysed independently from the others, and which are the basis for the multivariate choice of the optimal logistic regression model.

The resulting optimal multivariate model (see Table 3) contains ten different variables, all these variables are included simultaneously into the analysis. It is objective to determine the influence of the independent variable X (for example gender) on the dependent variable Y (here: death) in consideration of the fact that a third variable Z (for example EES) also affects Y.

The majority of the variables which are univariately significant shows also significant results in the multivariate analysis (see Table 4). The variable "passenger" shows uni- and multivariately no significance, therefore this variable can be neglected in further analysis. The variable "gender", which demonstrates in the univariate analysis an increased mortality risk for the males (OR=1,78) is devaluated in the multivariate model. By contrast, the characteristic older than 60 years shows a significantly increased mortality risk in the multivariate analysis. Therefore the variable "age" is revaluated.

It is important to detect which of the variables have confounders. These variables which show univariately higher or lower values of OR, as in the multivariate analysis must have covariates (see Table 5). Because the interactions between the different variables are detected and corrected in the multivariate analysis, therefore this OR is more realistic and closer to reality.

	n	Regression coefficient	s.e.	p-value	OR	95% CI
Energy Equivalent Speed (EES) in kph	3418					
0 –15 (ref.)	1036					
16 – 30	1459	1,742	1,061	0,101	5,71	0,71...45,69
31 – 45	624	3,168	1,036	0,002**	23,80	3,12...181,09
46 – 60	216	4,765	1,025	0,000***	117,37	15,73...875,87
>60	83	6,476	1,026	0,000***	649,41	87,00...4847,30
Gender	3347					
female (ref.)	1582					
male	1765	0,577	0,244	0,018*	1,78	1,11...2,87
Age in years	3323					
26-60 (ref.)	1702					
children up to 17	279	0,341	0,391	0,473	1,41	0,65...3,06
18-25	1027	0,042	0,879	0,766	1,04	0,61...1,79
>60	315	0,446	0,221	0,249	1,56	0,77...3,19
BMI °	2734					
normal (ref.)	1555					
overweight	774	0,367	0,341	0,282	1,44	0,74...2,82
obese	273	0,496	0,468	0,289	1,64	0,66...4,11
underweight	132	0,117	0,746	0,876	1,12	0,26...4,85
Age of driving license	1877					
>3 years (ref.)	1373					
≤3 years	504	0,139	0,425	0,743	1,15	0,50...2,64
Accident location	1763					
known (ref.)	1234					
unknown	529	-0,334	0,574	0,560	0,72	0,23...2,21
Blood alcohol concentration^	3282					
<0,03mg/l (ref.)	3226					
≥0,03mg/l	56	1,643	0,486	0,001***	5,17	1,99...13,42
Use of seeing aid	3032					
no (ref.)	2295					
yes	737	-1,900	0,725	0,009**	0,15	0,04...0,62
Swerve to avoid hitting	1582					
no (ref.)	872					
yes	710	-0,064	0,372	0,863	0,94	0,45..1,95
Braking	1618					
yes (ref.)	859					
no	759	1,036	0,399	0,009**	2,82	1,29...6,16
Illness	3158					
no (ref.)	2849					
at least one	309	0,414	0,412	0,315	1,51	0,67...3,39
Medicaments	3210					
no (ref.)	2917					
at least one	293	-1,779	1,010	0,078	0,17	0,02...1,22
Passenger	3418					
no (ref.)	2788					
at least one	630	-0,545	0,357	0,127	0,58	0,28...1,17
Belt usage	2925					
yes (ref.)	2682					
no	243	1,091	0,331	0,001***	2,98	1,56...5,70
Front airbag	3124					
yes (ref.)	1151					
no	1973	0,032	0,248	0,896	1,03	0,64...1,68
Airbag deployment	1151					
yes (ref.)	539					
no	612	-1,867	0,547	0,001***	0,16	0,05...0,45
Site	3418					
urban (ref.)	1898					
rural	1520	1,949	0,316	0,000***	7,02	3,78...13,05
Light conditions	3414					
daylight (ref.)	2130					
darkness, dawn	1284	0,705	0,232	0,002**	2,02	1,29...3,19
Locality of accident	3277					
within built up area (ref.)	2126					
outside built up area	1151	0,711	0,234	0,002**	2,04	1,29... 3,22

Table 2: Univariate analysis

	n	Regression coefficient	s.e.	p-value	OR	95% CI
Scene of the accident	3404					
junction/crossroad (ref.)	1470					
curve	604	2,435	0,454	0,000***	11,42	4,69...27,80
straight line	1330	2,122	0,437	0,000***	8,35	3,55...19,65
Rainfall	3392					
no (ref.)	2559					
yes	833	-0,221	0,284	0,437	0,80	0,46...1,40
Condition of the road	3404					
good (ref.)	3243					
affected	161	-0,40	1,014	0,694	0,67	0,09...4,90
Surface of the road	3400					
dry (ref.)	2115					
wet, frozen, snow covered	1285	-0,179	0,244	0,461	0,84	0,52...1,35
Power of the car in kW	2843					
<50 (ref.)	853					
50-69	1077	0,308	0,309	0,319	1,36	0,74...2,49
70-89	528	0,045	0,391	0,908	1,05	0,49...2,25
>90	385	-0,250	0,479	0,601	0,78	0,31...1,99
Seating position	3418					
front passenger (ref.)	2984					
rear seat passenger	437	-0,096	0,358	0,788	0,91	0,45...1,83

° BMI is adjusted to age and gender of the children (according to the definitions of the WHO)
^ The blood alcohol concentration is only ascertained for drivers and only for the autopsied dead
* (p≤0,05), ** (p≤0,01), *** (p≤0,001)

Table 2: Continuation

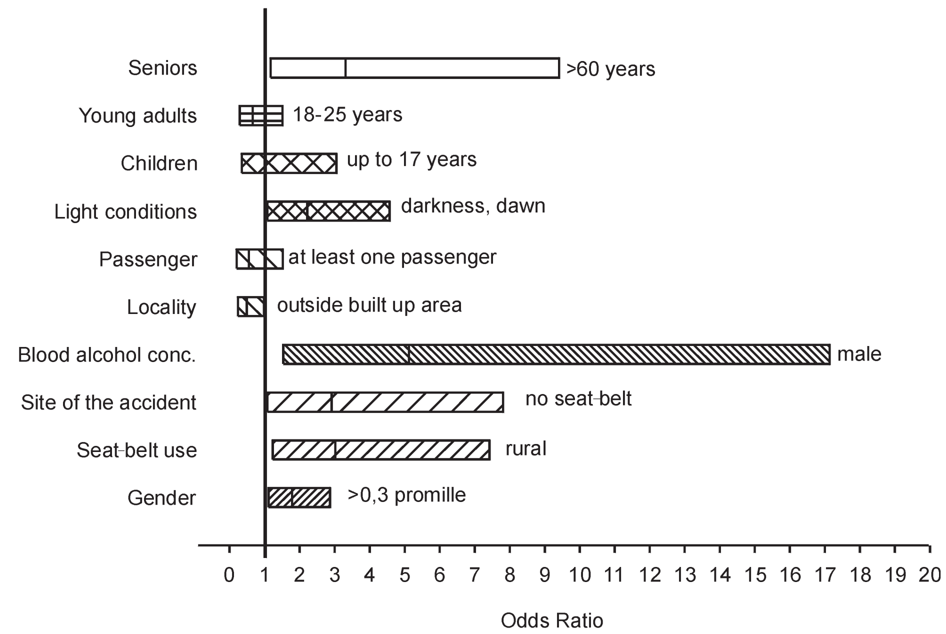

Figure 3: The different variables of the multivariate analysis with an increase (OR >1) or a decrease in mortality risk (OR <1). Significant results do not include 1.0

The variables “belt usage”, “light conditions”, “passenger” and “blood alcohol concentration” show uni- and multivariately the same OR. The variable “locality of the accident” demonstrates univariate an increased mortality risk (OR=2,04) for the characteristic “outside built up area”. In the multivariately model there is seen an inverse effect: “outside built up area” seems to have a protective effect (OR=0,49) on the mortality risk. The variables “site”, “EES”, “scene of the accident” and “gender” have univariately a higher OR as multivariately and the variable “age” has univariately a lower OR than in the multivariate model, therefore they have confounders.

	n	Regression coefficient	s.e.	p-value	OR	95% CI
Constant factor		-8.7761				
Belt usage						
yes (ref.)	2399					
no	210	1,1045	0,4594	0,0162*	3,02	1,23 ... 7,42
Site						
urban (ref.)	1450					
rural	1159	1,0671	0,5043	0,0343*	2,91	1,08 ... 7,81
Blood alcohol concentration^						
<0,03mg/l (ref.)	2567					
≥0,03mg/l	42	1,6326	0,6168	0,0081**	5,12	1,53 … 17,14
Energy equivalent speed(EES)						
0 –15kph (ref.)	794					
16 – 30kph	1119	0,7868	1,1285	0,4857	2,20	0,24 ... 20,06
31 – 45kph	478	2,2277	1,0657	0,0366*	9,28	1,15 ... 74,93
46 – 60kph	156	3,4933	1,0675	0,0011**	32,90	4,06 ... 266,58
>60kph	62	5,5121	1,0724	<,0001***	247,67	30,27...>999,9
Light conditions						
daylight(ref.)	1640					
darkness, dawn	969	0,7993	0,3673	0,0295*	2,22	1,08 ... 4,57
Locality of accident						
within built up area (ref.)	1657					
outside built up area	952	-0,7214	0,3582	0,0440*	0,49	0,24 ... 0,98
Gender						
female (ref.)	1279					
male	1330	0,1426	0,3660	0,6969	1,15	0,56 ... 2,36
Age						
26-60 years (ref.)	1356					
children up to 17 years	219	0,0305	0,5528	0,9560	1,03	0,35 ... 3,05
18-25 years >	779	-0,4124	0,4217	0,3281	0,66	0,29 ... 1,51
60 years	255	1,1977	0,5329	0,0246*	3,31	1,17 ... 9,41
Passenger						
no (ref.)	2127					
at least one	482	-0,6044	0,5206	0,2456	0,55	0,20 ...1,52
Scene of the accident						
junction/crossroad (ref.)	1147					
curve	467	2,0285	0,7953	0,0108*	7,60	1,60 ... 36,13
straight line	995	1,6204	0,7886	0,0399*	5,05	1,08 ... 23,71

^ The blood alcohol concentration is only ascertained for drivers and only for the autopsied dead
* (p≤0,05), ** (p≤0,01), *** (p≤0,001)

Table 3: Multivariate analysis

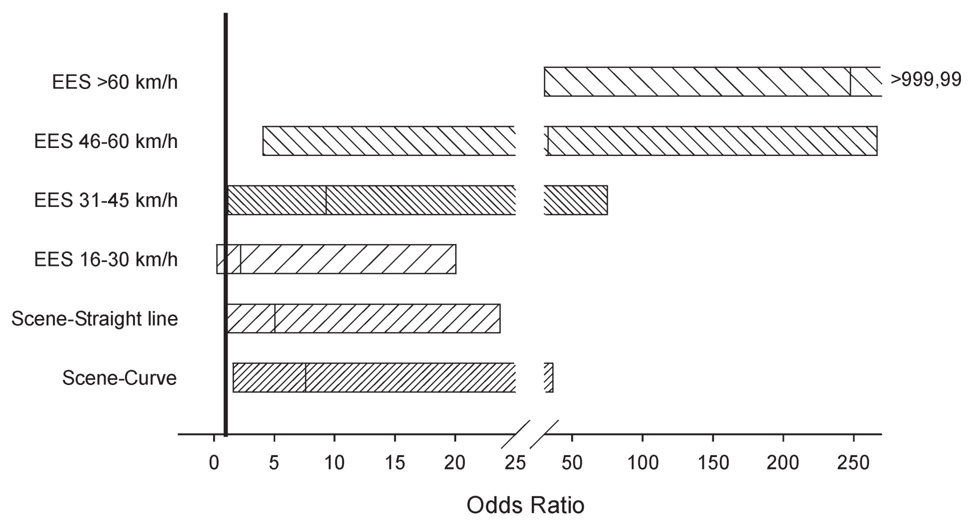

Figure 4: The different variables of the multivariate analysis with an increase in mortality risk (OR >1) or a decrease in mortality risk (OR <1). Significant results do not include 1.0. Break between 25 and 30

Univariate	Multivariate	Variables
significant	significant	• Belt usage • Site • Blood alcohol concentration • EES (31-45 kph, 46-60kph and >60kph) • Light conditions • Locality of the accident • Scene of the accident
not significant	dispensable	• Passenger
significant	dispensable	• Gender
not significant	significant	• Age (>60 years)

Table 4: Overview of the results of the uni- and multivariate analysis concerning the significance

Univariate higher OR as multivariate	• Site • EES • Scene of the accident • Gender
Univariate lower OR as multivariate	• Age
Uni- and multivariate same OR	• Belt usage • Light conditions • Passenger • Blood alcohol concentration
Opposite effect in uni- and multivariate analysis	• Locality of the accident

Table 5: Overview of the different OR of the uni- and multivariate analysis (by variables with several characteristics, it was referred to the significant results)

	EES in kph					Total
	0-15	16-30	31-45	46-60	>60	
female	562	676	245	74	25	1582
male	454	756	365	134	56	1765
Total	1016	1432	610	208	81	3347

Table 6: Overview of the different EES-values in dependence on the gender

Discussion

Which confounders these are can never finally be answered.

Site: By calculation of the OR of the variable “site” the multivariate model takes all the other nine variables also into consideration and respects the interference of the variables among themselves. One potential covariate is the EES. In rural areas higher driven speeds are possible which in the case of an accident ends in higher EES. The multivariate OR is therefore lower than the univariate, because the variable “site” is included in the variable “EES”. It is not possible to look only at the site; the driven speeds have also to be considered.

Scene: The scene of the accident depends mostly on the site and the locality. In rural or outside built up areas more curves and straight lines are existing than in urban areas where junctions and crossroads predominate. Therefore the scene has to be seen in context with the other variables.

Gender: The significant OR for the males in the univariate analysis will not last the multivariate model. One explanation is the observation that males drive with higher speed (see Table 6) and their cars show higher EES after a crash. The information of the variable “gender” is also incorporated in the variable “EES”.

Age: In the univariate analysis the characteristic “older than 60 years” shows no significance. The multivariate model results in a three times higher mortality risk than the reference group with persons aged 26 to 60 years. This characteristic was revaluated. But it is not easy to answer which covariates these are.

EES: As seen above this variable affects others, but it is also influenced by others. The EES is probably dependent on the site, the locality, the gender, the light conditions and the scene of the accident.

Locality: This variable shows for “outside built up area” like forest, fields, meadow and so on, a two times higher risk. In the multivariate model a protective result (OR=0,5) is illustrated. This apparent contradiction is caused by the limited validity of the univariate model: considering solely the locality, it is true that outside built up areas more fatal accidents occur. The multivariate analysis demonstrates that rather the relation that roads outside built up areas are situated in the rural area and allow higher speed is decisive than the locality of the accident. Hence only in relation with the variables “site” and “EES” it is possible to interpret the variable “locality” properly.

Limits of the multivariate model

For each of the 50 deaths it is possible to determine the prediction quality/validity to know how precise the multivariate model predicted properly the death of this case. The analysis of cases which have not been predicted in a good manner can give detailed information about the limits of this multivariate model.

- Case number 1010825 (Figure 5) is not well predicted by this model (0,95%). Three variables show significant values (rural, straight line, night time driving) which are associated with a higher death risk. A car crashes on the highway into the rear of a truck with projected cargo. The front seat passenger dies. As the victim was a front seat passenger the blood alcohol concentration has not been determined. The seat-belt use is not so relevant as well, because the massive head injuries caused by the cargo cannot be prevented by a seat-belt. This example also shows that accidents with low EES-values (16-30kph) can lead to death, when the collision object is a massive object which penetrates into the interior and causes due to its height massive head injuries.
- Case number 1990343 (Figure 6) was not well identified (7,57%). After overtaking a truck, the car starts skidding and stops standing crosswise on the highway. The following truck crashes into the side of the car. Three variables show significance (straight line, darkness, EES 55kph). The dead person was the front seat passenger, so the blood alcohol was not determined.

The limits of this multivariate model are the detection of collisions with massive objects if the front seat passenger or other passengers lost their lives. The variable blood alcohol concentration is not suitable to determine the risk of the front seat passengers or the other occupants. Special accident scenarios, for example accidents due to projected cargo, are predicted insufficiently.

The Tables 2 and 3 contain the p-value, which serves as evaluation for the prediction validity. But how suitable are these variables to be accepted in the multivariate model? In which way do they express the death rate? The SPSS®-Program lists for each case and each variable the predicted probability.

The variable “site of the accident” for example distinguishes very well between survival or death

Figure 5: The truck with its cargo and the car

Figure 6: The arrow shows the direction of the truck

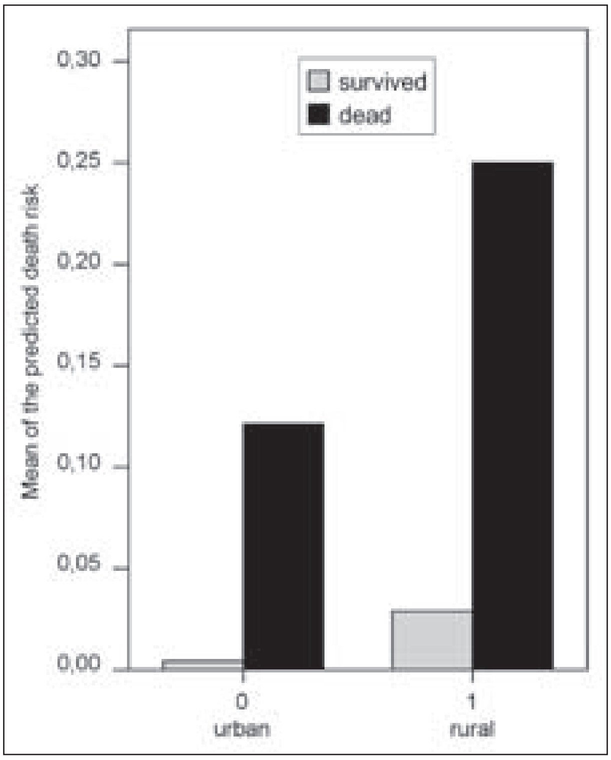

Figure 7: Mean of the predicted death risk for the variable “site of the accident”

(see Figure 7). By the characteristic value “rural” 25% of the cases with fatalities are properly

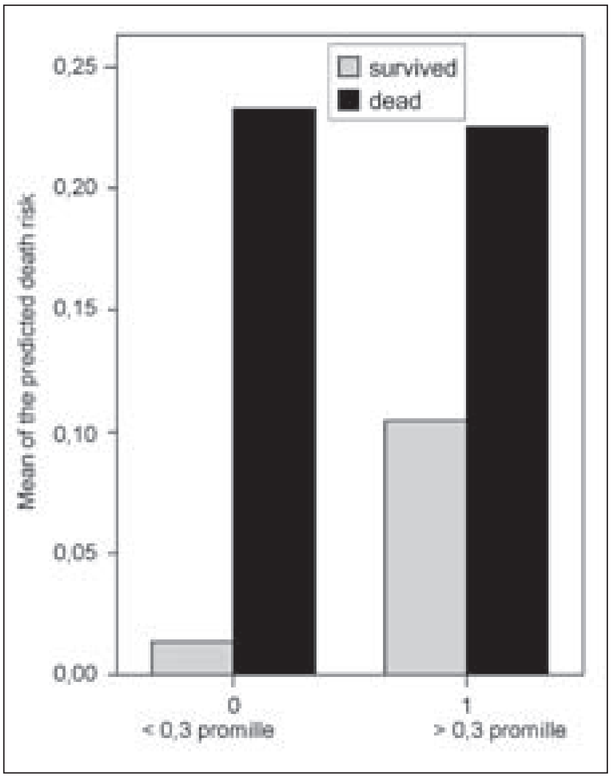

Figure 8: Mean of the predicted death risk for the variable "blood alcohol concentration"

Significant?		survived	dead
	<0,03mg/l	3166 (98,1%)	60 (1,9%)
	>0,03mg/l	51 (91,1%)	5 (8,9%)

Table 7: Overview of the number (and percentage) of the survived and dead persons in dependence on the "blood alcohol concentration"

detected as dead and about 12,5% by the characteristic "urban". The cases with the survivors show low mean predicted death rate. Therefore the variable "site of the accident" is very suited for this multivariat model.

The variable "blood alcohol concentration" is of less quality. Figure 8 shows that death is better predicted for those who have a blood alcohol concentration <0,03mg/l but the univariate analysis illustrates a 5 times higher risk for those who have a blood alcohol concentration >0,03mg/l. In the univariate analysis it is only considered if the difference between persons with <0,03 mg/l and persons with >0,03mg/l is significant. As seen in Table 7 the difference between 1,9% and 8,9% is significant. But this significance is favoured by the grouping: in the univariate analysis only 56 persons are in the group >0,03mg/l. Further this variable is not representative, because it does not consider the occupants. Usually the blood alcohol concentration is only considered for the drivers and if they are dead only for the autopsied persons.

For the graph (Figure 8) of the "mean of the predicted death risk" the predicted mortality is summarized. As more persons (60 victims) are dead with <0,03 mg/l, they are better predicted. Solely this variable is not suitable for predicting mortality, because there exists a lot of scenarios, where the driver or the occupants die without drinking any alcohol (e.g. after losing control, a truck crosses the halfway line and crashes into a oncoming car).

Just the entirety of all variables contained in the model determines the prediction validity of the model, which can be estimated by simple reclassification and cross validation. How precise a single variable serves to the prediction validity of the whole model, can only be answered by the p-value of all variables included to the model. By the combination of these variables are 93% of the cases detected as "truly survived" and "truly dead".

Conclusion

It was shown how critical it is to consider only the influence of one single variable on the mortality. Univariate analyses are simple to understand but their results are not realistic. The multivariate model allows more detailed results which are closer to reality. With this kind of analysis it is possible to consider several variables simultaneously and to respect their interferences. The multivariate analysis is based on a model which detects 93% of the cases properly. For the future it would be very interesting to develop a calculator for the car crash mortality based on a model which contains much more cases than considered here. Such a risk assessment calculator was developed to estimate the individual risk of a heart attack within the next 10 years based upon data of the PROCAM study [7]. The person has to give some information about gender, age, blood pressure, smoking of cigarettes and so on; the computer calculates than the risk to suffer a heart attack.

As a future vision the potential driver or occupant could answer some questions like:

- Light conditions during the drive (darkness, dawn vs. daylight).
- Seat-belt usage (yes vs. no).

- Existence of airbags (yes vs no).
- Age of the driver.
- Gender.
- Construction year of the vehicle.
- Predominating road type (highway vs roads in rural areas vs roads in urban areas).
- Estimated average speed.

Afterwards the calculator estimates the mortality risk if an accident happens. Especially for young drivers this calculator would be very interesting. By seeing the increased mortality risk for driving without a seat-belt or under the influence of alcohol they would perhaps drive more carefully and occupants would deny driving with an alcoholised driver.

References

[1] Bundesministerium für Verkehr- Bau und Wohnungswesen: Verkehr in Zahlen 2005/2006, Berlin, Deutscher Verkehrs-Verlag, page 180, 2005

[2] Ministerium des Inneren- Hauptabteilung Verkehrspolizei: Statistische Analyse des Verkehrsunfallgeschehens in der DDR, 1953-1989

[3] European Community: White Paper, European transport policy for 2010: time to decide, page 66, 2001

[4] R. KOCH: Schätzung von Risiken: Eine kurze Einführung in Ziele und Methoden, www.imib.med.tu-dresden.de/imib/biometrie/risikoschaetzungen.pdf

[5] L. KREIENBROCK, S. SCHACH: Epidemiologische Methoden, Jena, Gustav Fischer Verlag, 1995

[6] R. KOCH: Diagnostische bzw. prognostische Bedeutsamkeit einer Kovariablen-Vergleich uni- und multivariate Bewertung, www.imib.tu-dresden.de/imib/biometrie/Folien_logitmodelle_quer.pdf, 2003

[7] International Task Force for Prevention of Coronary Heart disease: PROCAM Risiko-Rechner, http://chdrisk.uni-muenster.de/calculator. php