A MORE APPROPRIATE METHOD TO EVALUATE THE PASSIVE SAFETY OF VEHICLES

S. Achmus, R. Zobel
Volkswagen AG, Wolfsburg, Germany

ABSTRACT

Evaluation of the passive safety of vehicles from accident data has been attempted by different institutions. Researchers try to distinguish the relative importance of different factors associated with accidents often by means of regression models. But the decisive problem in evaluating the passive safety of vehicles from accident data by regression models is the assumption of linearity. Till now all applied procedures combine the crash parameters linearly either to calculate the injury severity (linear regression) or to calculate the logit of the injury risk (logistic regression). These methods are not an appropriate tool for evaluating the passive safety of vehicles since the influence of the crash parameters is obviously not linear. An increase in the change of velocity from 10 km/h to 40 km/h has a different effect than an increase from 80 km/h to 110 km/h. Most of the crash parameters exhibit nonlinear behaviour but it gets „linearized" by the above mentioned linear regression models. Therefore, the influence of the crash variables is described wrongly and one gets falsified results for the evaluation of passive safety of vehicles.

To avoid the linearity of these models we introduce a nonlinear nonparametric additive model. Here the crash parameters are combined additively after an appropriate transformation on each variable: $Y = c + f_1(x_1) + \ldots + f_k(x_k)$ with Y being the injury severity, $x_1, \ldots, x_k$ the crash parameters, c the average injury severity calculated from the accident data and $f_1, \ldots, f_k$ the unknown functions which have to be estimated nonparametrically from the data. Nonparametrically means that we do not assume the functions to be polynoms or to have another a priori given structure. The estimated functions show the effect of each of the variables, i.e. the function f_j represents the effect of the parameter x_j. It can be seen within which ranges of the parameter the contribution to the injury severity increases or decreases and within which areas the contribution to the injury severity remains constant.

This new approach is shown by an example from a real world data base.

WITHIN THE SCOPE of statistical procedures evaluation of injury severity is often based on restrictive assumptions which are not met by the data and the dependences between parameters and injury severity or injury risk. For example in a logistic regression model which is used for evaluation of passive safety of vehicles it is supposed that all non-categorical variables have a linear influence on the target variable, the logit of the risk for a certain injury level. Non-categorical variables are those parameters with continuous values which are not divided into certain categories. This linearity forced by the logistic regression model means that an increase (or decrease) of the values of any parameter implies an increase of the target variable. With this procedure it is not possible to find out ranges of the parameter where the risk remains constant or in return decreases. When such complicated dependences occur in the data they are inevitably described the wrong way, so the entire logistic regression model leads to a falsification of the dependences on other parameters (cf. Section 3). Such falsifications can also be caused by too rough classifications of the categorical parameters. If only few categories are defined for the values of a parameter this often leads to an incorrect description of the influence of this parameter and therefore it might falsify the dependence on other parameters. But in return many categories make the estimation worse (cf. Section 3).

In this paper an additive model is presented which does not need such restrictive assumptions regarding to the influence of crash parameters on the injury severity, rather any nonlinear relationship between the parameters and the injury severity is admissible. The aim is to show the problems that occur from linear assumptions and the arbitrariness of an evaluation of the passive safety when the influence of the parameters is not described correctly.

1. THE ADDITIVE MODEL

The additive model is applied to the database in order to examine the influence of different parameters on the injury severity and to demonstrate nonlinear behaviour of parameters. For better illustration the results of the two models, the additive model and the logistic regression model, are compared in an example of application.

The Volkswagen - database is a subset of the accident database of the Medizinische Hochschule Hannover. It covers crashes with at least one injured person that happened in or near Hanover. The analysis is performed on a dataset which is chosen according to the problem. A relative large amount of severe crashes is included in the dataset. Further just drivers of passenger cars which had a frontal collision are considered. Due to a limitation of the computing time the calculation is based on 300 cases. In a next step this number will be increased. Seven crash parameters are selected according to their significance:

(Δv)	velocity change caused by impact
(sb)	use of seat belt
(ang)	angle between involved vehicles
(opp)	type of opponent in collision
(vk)	velocity at the moment of collision
(hei)	height of the driver
(age)	age of the driver if velocity change $\Delta v \geq 25$ km/h

The complex definition of the variable age is explained below in Section 2, where the influence of this parameter is discussed. Variables like sex, weight of the driver, overlap and others proved to be not as significant as the above mentioned by means of improvement of the entire fit.

In the additive model the variables are combined additively after they have been transformed appropriately. Details about estimating these unknown transformations can be found in (Breiman, Friedman, 1985), (Buja, Hastie, Tibshirani 1989). The transformation, e.g. on the variable Δv

$$\Delta v \mapsto f_1(\Delta v),$$

is not supposed to have any special structure, i.e. f_1 is neither assumed to be linear ($a_0 + a_1 \cdot x$) nor to be polynomial ($a_0 + a_1 \cdot x + \ldots + a_k \cdot x^k$) nor is any other structure given a priori. f_1 is just thought of as any arbitrary unknown function and it is identified during the estimation of the entire additive model. Then in the additive model the injury severity is modelled by the sum of the transformed parameters:

$$Y = c + f_1(\Delta v) + f_2(sb) + f_3(ang) + f_4(opp) + f_5(vk) + f_6(hei) + f_7(age).$$

Here the target variable Y is the injury severity of the driver on the AIS - scale (MAIS: maximal AIS-value of all body regions) and the constant c represents an average injury level of the dataset. The functions $f_1,\ldots,f_7$, specify the influence of the parameters $\Delta v, sb,\ldots,age$ on the injury severity. This will be explained in detail below.

It must be mentioned that the additivity in the model is a restriction. If the injury severity depends on two variables x_1, x_2 through a function which depends on both variables at one time (e.g. $Y = x_1 \cdot x_2 + \ldots$), it could be necessary to model the influence of these parameters by a function which depends on these two variables $f(x_1, x_2)$ at one time, instead of modeling the influence by the sum $f_1(x_1) + f_2(x_2)$. So in these cases the above model can only approximate the real dependence by an additive term which may be an acceptable approximation within the range of values of the parameters x_1, x_2. Sometimes this problem can be avoided by combining two parameters appropriately (in the above example $x_{1,2} = x_1 \cdot x_2$ or cf. parameter age).

On the other hand the logistic regression model assumes the restriction of additivity and linearity ($Y = b_0 + b_1 \ \Delta v + b_2 \cdot sb + \ldots + b_7 \cdot age$). Here the logit of the risk of an injury level of at least 3 (MAIS 3+) is taken as the target variable, and the same parameters with coefficients $b_0, b_1,\ldots,b_7$ are used.

So the comparison of both models will show whether the assumption of linearity in the logistic model is acceptable to an evaluation of the passive safety of vehicles or whether more sophisticated dependences exist.

2. INFLUENCE OF DIFFERENT PARAMETERS ON THE INJURY SEVERITY

In this Section the influence of the seven parameters resulting from the additive model and from the logistic regression model is discussed.

In the additive model the influence of the velocity change on the injury severity is given by the function f_1. The value of the function $f_1(\Delta v)$ represents the expected increase (positive value) or decrease (negative value) of the injury severity when the velocity change is Δv (Fig.1).

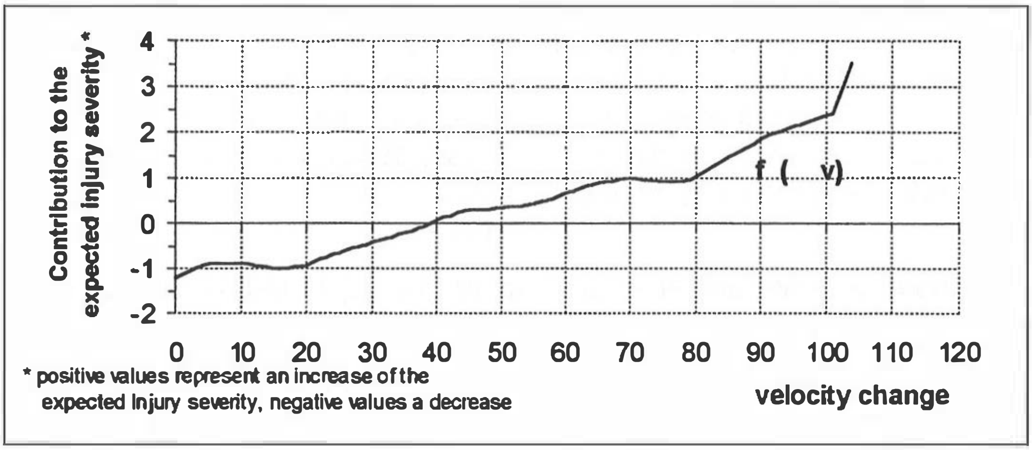

Fig.1 - Influence of the velocity change on the injury severity in the additive model

No contribution ($f_1(\Delta v) = 0$) to the injury severity is expected, i.e. the average injury severity which results from c and the contributions of the other parameters is expected, when the velocity change caused by the impact is $\Delta v = 39$ km/h. For minor values of the velocity change, for example $\Delta v = 15$ km/h, the expected injury severity is about 1 AIS-unit less than it would be expected for $\Delta v = 39$ km/h. Major values of the velocity change, for example $\Delta v = 70$ km/h or $\Delta v = 90$ km/h, lead to an increased expected injury level by 1 AIS-unit or 2 AIS-units respectively. It can be seen that for low velocity changes $\Delta v \leq 20$ km/h the influence $f_1(\Delta v) \approx -1$ of the velocity change on the injury severity is constant. Just values less than 5 km/h let this influence be $f_1(\Delta v) \approx -1.2$. For values between 20 km/h and 70 km/h the contribution to the injury severity varies linearly, and after a small constant range (70 - 80 km/h) it varies also linearly but a bit faster. That means that an increase of the velocity change by 10 km/h within the first range (up to 20 km/h) leads to no variation in the expected injury severity, whereas an increase by 10 km/h within the range 20 - 70 km/h raises the expected injury level by 0.4 AIS-units, and within the range 80 - 100 km/h it leads to an increased injury severity by 0.7 AIS-units. Consequently the influence of the velocity change is not linear.

Now the detailed result of the additive model is compared to the result of the logistic regression model (Fig.2). The logit of the risk for a certain injury severity can be considered as a characteristic quantity for a crash severity. The influence of the velocity change on this quantity is linear, this is forced by the logistic regression model. No constant ranges or rises of different intensity can be modeled.

It should be kept in mind that the function $f_1(\Delta v)$ represents the change of the expected injury severity and $b_1 \Delta v$ can be interpreted as the change of the logit of the risk of a certain injury level (here MAIS 3+). Obviously the additive model reflects the nonlinearity of the influence of the velocity change while the logistic regression model linearizes this influence.

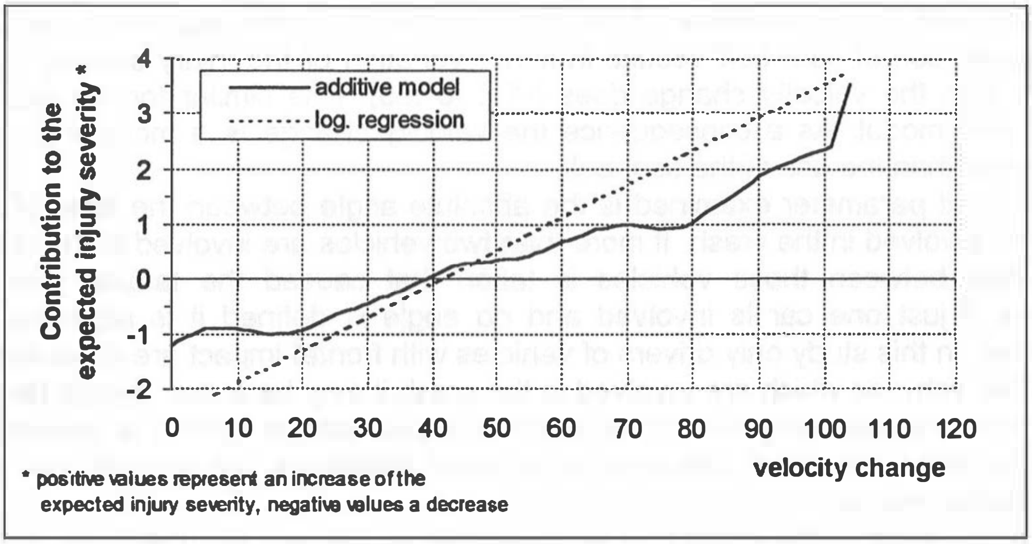

Fig.2 - Influence of the velocity change caused by the impact on the injury severity identified by the additive model ($f_1(\Delta v)$) and the logistic regression model ($b_1 \ \Delta v$)

The second parameter to look at is the use of the seat belt. This variable is categorical, that means it is devided into certain categories and cannot be considered as a parameter with continuous values. Three categories are to be distinguished: belted, not belted and unknown. The following dependence of the injury severity on the use of the seat belt results from the additive model (Fig.3).

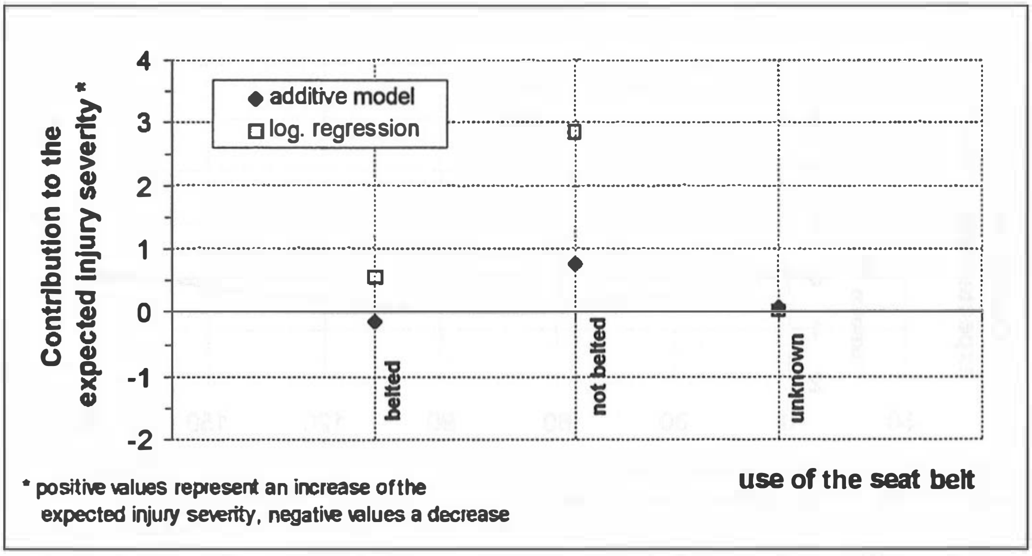

Fig.3 - Influence of the use of the seat belt on the injury severity identified by the additive model ($f_2(sb)$) and the logistic regression ($b_2 \cdot sb$)

According to the additive model unbelted drivers have to expect an increased injury severity by 1 AIS-level. If it is unknown whether the seat belt were used or not used there is no significant contribution to the injury severity.

In the logistic regression model the coefficients relating to the categories of the use of the belt show a similar trend but the difference between belted and unbelted drivers is larger. Further the category unknown seems less endangered.

Compared to the influence of the velocity change in the additive model the parameter 'use of seat belt' results in minor variation of the injury severity (-0.2 to 0.8) than the velocity change does (-1.2 to 2.5). It is similar for the logistic regression model. As a consequence the velocity change is a more important parameter than the use of the seat belt.

The next parameter examined is the absolute angle between the axes of the vehicles involved in the crash. If more than two vehicles are involved in the crash the angle between those vehicles is taken that caused the largest velocity change. If just one car is involved and no angle is defined it is regarded as unknown. In this study only drivers of vehicles with frontal impact are considered. For other vehicles which are involved in the crash it may be a rear impact (angle ≈ 0°), a side impact (angle ≈ 90°) or a frontal impact (angle ≈ 180°) or something between these situations. Influence of different directions, left or right, has not been accounted for.

The influence of the angle can be described as follows. For the driver in the considered vehicle the most dangerous impact situations are frontal vs rear impact and frontal vs frontal impact (Fig.4). An angle between 0° and 5° results in an increase of the injury severity by 0 - 0.4 AIS-units, while an angle between 170° and 180° leads to an increase by 0 - 0.2 units. On the other hand a frontal vs side impact leads to a decrease of the expected injury level, especially angles between 60° and 100° lead to a decrease by at least 0.5 AIS-units. This parameter is not as significant as the velocity change proves to be, but it leads to a variation of the injury severity by 1 AIS-unit (-0.7 to 0.4 units) which may not be neglected.

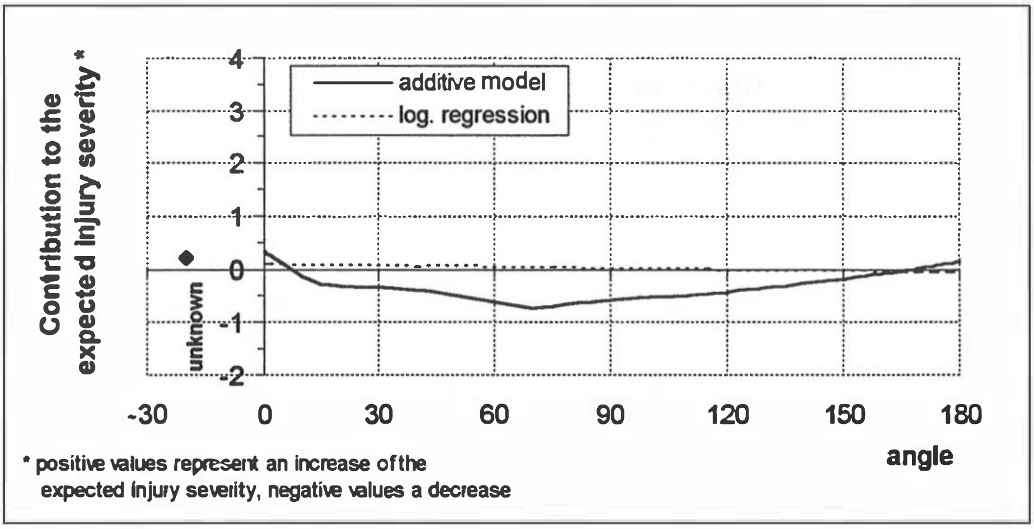

Fig.4 - Influence of the angle between the vehicles on the injury severity identified by the additive model ($f_3(ang)$) and the logistic regression ($b_3 \cdot ang$)

The logistic regression model on the other hand classifies the influence of the angle on the injury risk as not significant. The influence of this parameter on the logit of the risk for an injury of MAIS 3+ is almost constant. The logistic regression cannot find out that frontal vs side impacts are less dangerous because of the forced linear influence of the angle on the logit. The angle seems not significant for the risk of severe injury (MAIS 3+) although it can be seen from the additive model that there is a non-negligible influence.

The next parameter to look at is a categorical one. Four types of opponents are distinguished: car, truck, tree (and the like) and others.

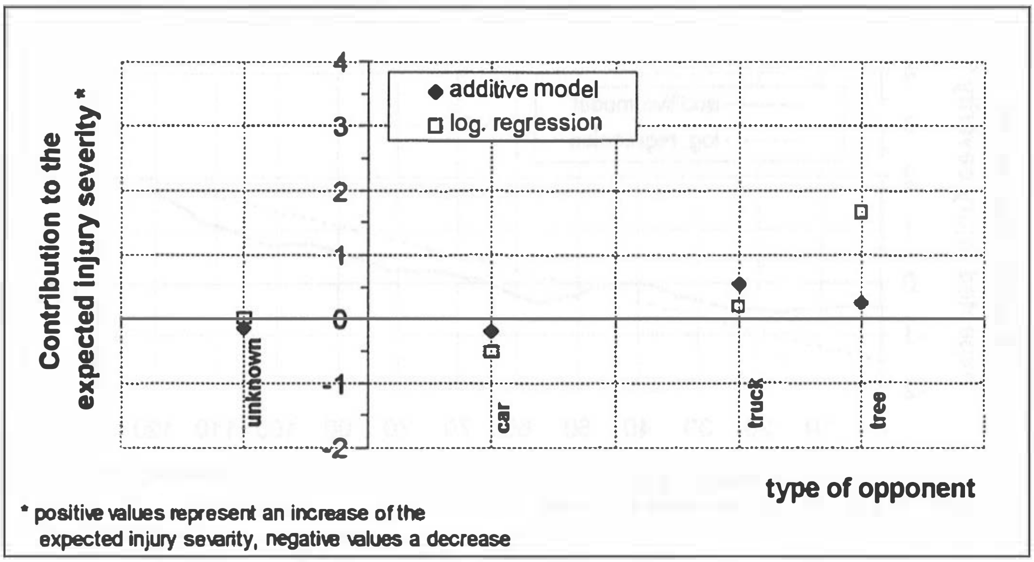

Fig.5 - Influence of the type of opponent on the injury severity identified by the additive model ($f_4(opp)$) and logistic regression ($b_4 \cdot opp$) respectively

The result of the additive model is plausible (Fig.5) and the logistic regression model shows a similar trend. Trucks as opponents are classified as not significant in the logistic model, which is not explainable.

The range of the contributions to the injury level which result from this parameter is from -0.2 to 0.5 in the additive model. Consequently the type of the opponent has less influence on the injury severity than the velocity change. This can be explained by the fact that to a certain degree this influence is taken into consideration by the parameter velocity change itself.

Another parameter to discuss is the velocity at the moment of the impact. It shows a very large influence on the injury severity in the additive model (Fig.6). A decrease of the injury severity by 0.5 AIS-units is expected for values of the velocity up to $vk = 30$ km/h. An increase of the injury level is to expect for values of at least $vk = 70$ km/h, for instance a velocity of 110 km/h leads to an increased injury severity by 1 AIS-unit. Within the range 40 km/h to 70 km/h the curve seems to oscillate. It must be checked by further investigation based on a larger data sample whether this is statistically significant or not.

Although the parameter velocity change has already been included in the additive model the velocity at the moment of impact is significant. It must be noted that these two variables are not completely correlated, e.g. a high velocity can result in a medium-sized velocity change depending on the velocity of the opponent, the angle, etc. Further for a fixed velocity change the higher the velocity the more energy has to be absorbed. Also the risk of a multiple collision increases with increasing velocity, so there are several reasons for including this parameter together with the parameter velocity change.

In comparison the linear influence of the velocity which is forced by the logistic regression model can be seen. The arguments to put forward are similar to those stated in the discussion of the parameter velocity change. Regions of constant influence and ranges of different rises in intensity cannot be modelled by the logistic regression, while the additive model shows these complex dependences.

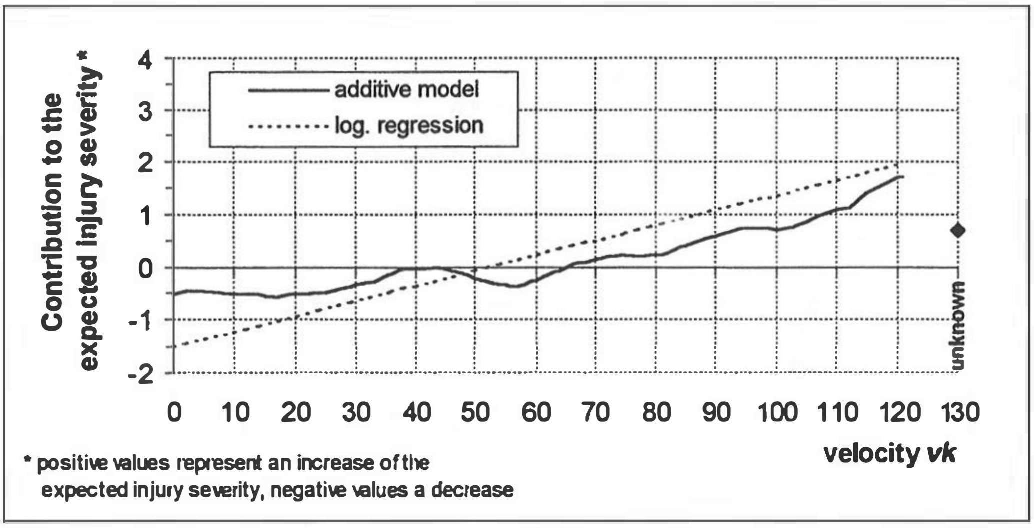

Fig.6 - The influence of the velocity at the moment of impact on the injury severity identified by the additive model ($f_5(vk)$) and logistic regression ($b_5 \cdot vk$)

The next two parameters are related to the driver: height and age. It can be seen that medium sized persons of 1.68 m to 1.90 m have to expect a minor injury level than persons with an extrem body height (Fig.7). For small persons of a height of 1.55 m - 1.65 m a major injury severity of 0.5 AIS-units to 1 AIS-unit is expected compared to persons who have an average height of 1.75 m.

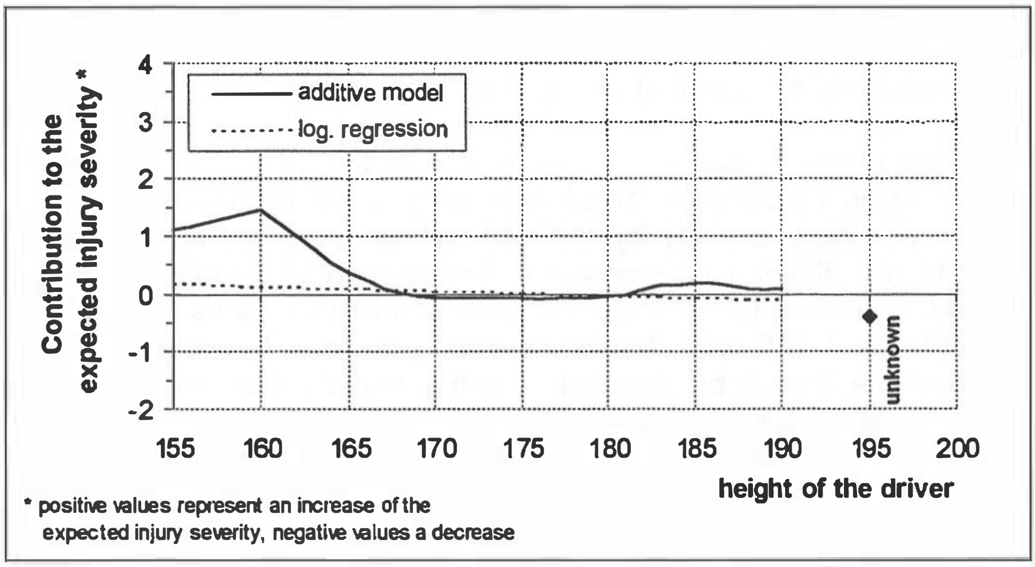

Fig.7 - The influence of the height of the driver on the injury severity identified by the additive model ($f_6(hei)$) and the logistic regression ($b_6 \cdot hei$)

In the logistic regression model the height of the driver doesn't seem to be significant. This can be explained by the special form of this influence identified by the additive model. A straight line cannot imitate the higher injury risk for extremely small persons and nearly constant risk for persons of height 1.68 m to 1.90 m. The result is an almost constant influence.

The last parameter to discuss is the age of the driver. It may be supposed that scratches and the like are independent of the age while fractures probably depend on the age. Therefore for the additive model a complex variable *age* is constructed. If the velocity change is less than 25 km/h the variable *age* is defined as -10. Otherwise if the age is unknown this variable is set to 99. In all other cases the variable *age* is defined as the age of the driver (Fig.8).

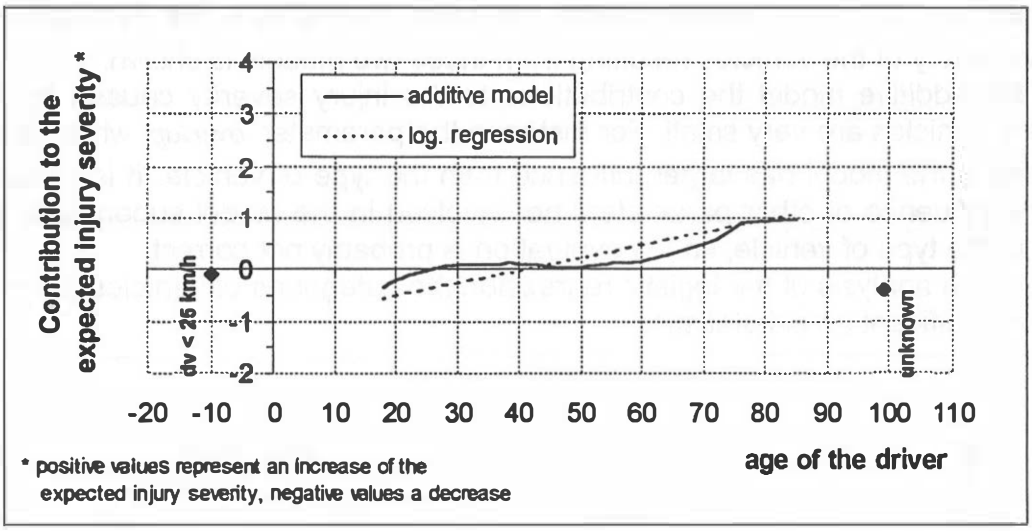

Fig.8 - The influence of the age of the driver on the injury severity identified by the additive model (complex definition of the variable *age*) and the logistic regression model (simply the age)

From biomechanical research a constant risk would be expected for 25 - 40 years old persons, whereas 40 - 70 years old persons are highly more endangered (cf. Zobel, Herrmann, Wittmüß, Zeidler (1994)). From the additive model the constant range for persons of an age between 30 and 55 years is larger than it would have been expected from biomechanics. Younger drivers (≤ 25 years) can expect a decreased injury severity by up to 0.3 AIS-units whereas 60 - 80 years old drivers must expect an increase of the injury severity by 0.2 AIS-units to 1 AIS-unit. When the age is unknown the contribution to the injury level is -0.4 units, that means less severely injured. This might be a systematical failure of the inquiry of the data, parameters related to persons more often remain unknown in less severe accidents.

In the logistic regression model the complex formulation of a variable which is defined as the age only for a certain range of the velocity change is not applicable. Here simply the age of the driver must be used for all values of the velocity change. The result is a linear dependence which roughly represents the influence resulted from the additive model, but the range where the influence remains constant, i.e. 30 - 55 years, cannot be reflected.

3. APPROACHES TO EVALUATE THE PASSIVE SAFETY OF VEHICLES BASED ON TWO STATISTICAL MODELS

From application of logistic regression models on accident data evaluation of the passive safety of vehicles is derived (e.g. Cameron (1995), Tapio, Ernvall (1995)). The additive model allows to make a similar approach but it takes into account the nonlinearities of the factors involved. How far the nonlinearities affect the evaluation algorithm is analysed. For this purpose the two statistical procedures described above are used. For evaluation of vehicles one parameter is added which is related to different car types, consequently then the models include eight variables. Nine types of vehicles are distinguished: A, B, ..., I and one group of others and unknown vehicles. The dependences of the other variables Δv, $sb,\ldots,age$ remain almost the same. In Figure 9 the evaluation of passive safety of the vehicles resulting from those two models is shown.

In the additive model the contributions to the injury severity caused by the different vehicles are very small. For instance the parameter *overlap* which is not included in the model has larger influence than the type of vehicle. It is possible that the influence of other parameters not involved in the model superposes the effect of the type of vehicle, so the evaluation is probably not correct.

From the analysis of the logistic regression the categories of vehicles C and D are not significant (Wald-statistic).

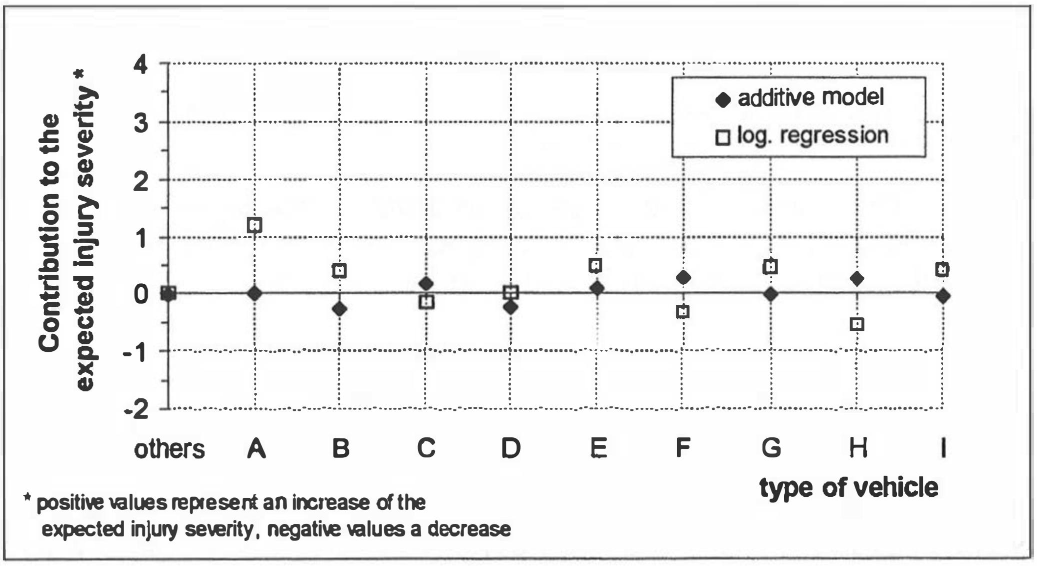

Fig.9 - Evaluation of passive safety of nine different vehicle types identified by the additive model ($f_8(veh)$) and the logistic regression ($b_8 \cdot veh$)

Again a negative value must be understood as a positive evaluation, it represents a decrease of the expected injury severity, while positive values reflect an increase of the injury severity. The classification of the vehicles regarding their passive safety according to the two models is completely different (Table 1). The underlined vehicles are the average with regard to passive safety. Vehicles are in decreasing order of safety from left to right. So the vehicles B and H are classified as the safest cars by the additive model and the logistic regression model respectively, and the vehicles F and A are estimated to be the most 'unsafe' cars.

additive model

	B	D	I	G	others	A	E	C	H	F	
good safety	----------	----------	----------	----------	----------	----------	----------	----------	----------	--------->	poor safety
	H	F	C	others	D	I	B	G	E	A	

logistic regression model

Table 1 - The succession of the vehicles in accordance with the evaluated passive safety.

One conspicuous circumstance is the classification of the vehicles C, F, and H. The logistic regression model classifies them as the safest cars, while the additive model places them worst. This is obviously a contradiction, but the reasons are as follows.

By fitting the logit of the risk for a certain injury level by a linear combination of the parameters, much information is given away at the beginning of calculation. No difference is made between MAIS = 0, 1, 2 and no difference is made between MAIS = 3, 4, 5, 6. For instance one can think of two vehicles 1, 2 with the same values of the parameters and the following MAIS-values (Table 2):

	MAIS											
vehicle 1	2	2	1	2	6	6	2	5	2	2	6	4
vehicle 2	3	0	0	1	3	4	0	3	3	0	4	3

Table 2 - Example for two vehicles with different classification resulting from the additive model and the logistic regression

Vehicle 2 would be expected to be safer by an analysis with the additive model because on an average it produces minor injury severities. But by an analysis with the logistic regression model with boundary at MAIS 3, vehicle 2 seems to be worse than vehicle 1 because vehicle 2 produces more often an injury severity of MAIS 3+ than vehicle 1 does. By consideration of MAIS 3+ all cases are classified equally besides the two hatched cases (Table 2). The logistic model uses much less information than the additive model, it does not look at differences within MAIS 3+ and MAIS 2-. If a logistic regression for the boundary MAIS 2+ is applied to this data example, vehicle 1 would seem much worse than vehicle 2. Based on our data set the evaluation by a logistic regression model for the risk of MAIS 2+ results in following succession regarding to passive safety

E C <u>others</u> D F A B G H I.

This is a competely different result than for an analysis for the risk of MAIS 3+. So one has to think of a valuation of these two different results and the others which are derived from calculations for the risk of MAIS 1+, MAIS 4+, etc.

Further the logistic regression can be affected by other falsifications. For example by the logistic regression model the influence of the height of the driver is not detected correctly (cf. Fig.7). If vehicle A is assumed to be a car which is mainly driven by small persons of a height of about 1.60 m, the increased injury risk of this group of persons is inadmissibly carried over to the vehicle A. This results in an increase of the coefficient of this vehicle type and the car A is evaluated with poor safety, although it possibly would be relatively safe if the driver population was on the average (cf. Zobel, 1995). This wrong classification results from the fact that the logistic regression cannot model the complex dependence as the additive model can do. A forced linear influence of parameters leads to falsifications, and in the above example it results in a shift of the inreased injury risk from the parameter height onto the variable car type A. Nevertheless such a wrong description can lead to an improvement of the entire fit of the model, although the coefficients do not reflect the correct influence.

To avoid the falsification caused by forced linear influence for the logistic regression exclusively categorical variables can be defined (cf. Cameron, 1995). But here other problems occur. One difficulty is to choose the number of categories for each variable. By defining only few classes it is nearly impossible to describe the influence of the parameters correctly as the dependences between the parameters and the injury severity are complicated. By few categories just a dependence consisting of few constant ranges can be imitated. (E.g. by defining categories of the age by 18-25, 26-59, 60-100 just three constant values for the influence of the age are possible although the additive model showed that within the ranges 18-25 and 60-100 the influence is not constant.) In return too many categories increase the error of estimation and therefore lead to increased confidence intervals for the coefficients in the logistic regression model. So in both cases the result has low reliability. When the number of classes is fixed, appropriate categories must be chosen and that means to choose boundaries for the categories. To avoid arbitrariness of this procedure a detailed analysis of the parameters is necessary.

The different evaluations of the two models show that the logistic regression model does not describe the influence of the parameters sufficiently. The nonlinearities seem to upset the result.

In contrast one has to ask how far the assumption of additivity in the additive model is admissible. This has to be shown by further investigation. Present investigations are on criteria for the goodness of the estimated functions, the construction of confidence bands and on statistical procedures for selection of significant parameters.

Another problem of logistic regression are missing values. Generally all cases with at least one missing value in the considered variables are kicked out of the calculation. A substitution by the means of the variables is not approriate that would strongly falsify the results.

The additive model can handle the missing values as has been described in detail for the variable *age* (Section 2). That offered a method to substitute the missing values for the logistic regression model by plausible values seen from the result of the additive model. The calculations shown in this paper are based on substitutions by such plausible values resulting from the analysis by the additive model.

CONCLUSION

It is agreed by most accident analysts that an identification of the inherent safety of a vehicle can be made by a comparison of the injury severity for cases of similar accident severity. This accident severity includes parameters with respect to the accident itself, i.e. collision mode, velocity, angle, and occupant related parameters (like age, gender, use of seat belt), etc. Based on most data bases this identification of inherent safety of vehicles as injury severity vs accident severity is not possible because many relevant parameters of accident severity, e.g. velocity change and velocity, are not available. Therefore statistical models are used that allow to substitute these parameters.

This paper shows that the assumptions of statistical models are restrictive, e.g. when a logistic regression model is used, the influence of the parameters is supposed to be linear. If categorical parameters are used, e.g. young drivers (yes/no), urban (yes/no), etc., this means that the linear approach of normal logistic regression is substituted by an approach of stepwise constant functions. The additive model assumes like the logistic regression that the influence of all parameters can be added. Therefore the difference of the hazard for a young occupant compared to an older occupant remains the same for all values of the other parameters (like velocity change, angle, ...). But the additive model allows more degrees of freedom than the logistic approach. It can be used to test whether the linear and stepwise constant approach of logistic regression is valid.

The result of this paper is that the assumption of linearity of the influence factors is highly questionable and falsifies the results. The simplification to linearity or stepwise constant functions influences the results of vehicle evaluation significantly. The sophisticated approximation of the accident severity by an additive model based on 300 in depth cases of frontal impacts clearly shows that nonlinearities of the dependences between parameters and injury severity exist.

Therefore a validation of the statistical model and the method of vehicle rating which is derived from this model is needed. This could be done as follows: An in depth accident data base is used which allows the calculation of accident severity including collision mode, velocity and individual parameters of the occupant, and which also includes all parameters needed for the statistical model. Then an evaluation of the inherent safety of a vehicle is made by the suggested statistical model. The result has to be compared with an evaluation of the inherent safety which is derived from the comparison of the injury severity for the different vehicles when the accident severity is fixed. If both results coincide the suggested method can be acceptable, otherwise not. Without such a validation no statistical procedure can be accepted. Statistical approaches do not always describe reality of an accident correctly as shown by the preceding results. Their assumptions like linearity etc. have to be checked carefully whether they are justified by the actual data.

Consequently the question *'How can the inherent safety of vehicles be identified from accident data ?'* is still open. Current statistical models derive an evaluation of the passive safety from accident data as a result of the vehicle, driver behaviour (distribution of age, sex, etc.) and environmental effects (distribution of mileage on rural or urban environments, etc.), and not exclusively as a consequence of the inherent safety of the vehicle itself. Before evaluation of the passive safety by any statistical model it must be ensured that every parameter with significant influence on the injury severity is available, included and correctly described in the model. Otherwise these external influences can superpose the inherent safety of the vehicle and consequently the real inherent passive safety of the vehicles remains undetected.

REFERENCES

Breiman, L., Friedman, J.H.: *Estimating optimal transformations for multiple regression and correlation,* J. Amer. Statist. Assoc., Vol.80 (1985), pp. 580-607

Buja, A., Hastie, T., Tibshirani, R.: *Linear smoothers and additive models,* Ann. Statist., Vol.17 (1989), pp. 453-555

Cameron, M. et al.: *Measuring crashworthiness: Make/model ratings and the influence of Australian design rules for motor vehicle safety,* IRCOBI (1995), pp. 297-310

Evans, L.: *Double pair comparison: A new method to determine how occupant characteristics affect fatality risk in traffic crashes,* General Motors Research Laboratories, Michigan 1985

Folksam Car Model Safety Rating 1991/92, Stockholm 1992

Tapio, J., Ernvall, T.: *Logistic regression in comparison of the passive safety of car models,* University of Oulu, Publications of Road and Transport Laboratory, Vol.32 (1995)

Zobel, R.: *Accident data and the passive safety of vehicles* or *Can you rate the passive safety of vehicles from accident data ?,* IRCOBI (1995), pp.375-388

Zobel, R., Herrmann, R., Wittmüß, A., Zeidler, F.: *Prediction of thoracic injuries by means of accelerations, deflections and the viscous criteria derived from full-scale side impacts,* Forschungsvereinigung Automobiltechnik, 14th International Technical Conference on Enhanced Safety of Vehicles, Munich 1994