METHODS OF STANDARDIZATION AND EVALUATION OF ACCIDENT DATA

Schlichting, K.-D.
Ingenieur-Büro Schlichting
3300 Braunschweig

and

Dörschlag, S. and Arand, W.
Institut für Straßenwesen
Technische Universität Braunschweig
3300 Braunschweig

ABSTRACT

Methods of standardization of 'raw' data from official accident reports that were used in the study of car accidents in the West German state of Nordrhein-Westfalen in 1980 are described.

The larger study aimed to identify technical car parameters which contributed to accidents and required the identification of groups of drivers who have an average risk of causing accidents.

The analysis involved the collection of additional data, that would allow the standardization of 'raw' driver-age accident frequency and car-model accident frequency data to account for the relative 'exposure' or degree of participation of driver and car populations in normal traffic. In the case of the car models, this involved the collection of data on the abundance of car models, and a calculation of the average annual mileage for different models. For drivers, the age distribution in normal traffic was approximated by the age distribution of drivers who were involved in, but not the cause of accidents, according to the official accident reports. The exposure-corrected age distribution of drivers who caused accidents reveals a much higher involvement of older drivers (above 54 years) in accidents than would be expected from their contribution to the total driver population in normal traffic. Using this approach for different accident types clearly shows the extremly high involvement of 1) the 18-22 year age group in causing longitudinal and single-car type accidents and 2) drivers above 55 years of age in accidents occuring at intersections.

The distribution showing the accident involvement of specific car models when exposure to traffic is taken into account is significantly different from the raw accident frequency distributions and shows that many car types with low accident frequencies have a high probability of accident involvement.

Speed data were also collected in a roadside study to examine the relationship between driver age, car type and choice of speed. Data for separate observation days showed the existence of significantly different speed distributions thereby precluding the analysis of the data for all days as a single group. However, by calculating a daily 85-percentile speed and the relative numbers of each age group exceeding it, the data could be combined. Drivers aged 25 - 45 years chose the highest speeds and speeds decreased for higher ages.

INTRODUCTION

This report describes methods that were used to evaluate 'raw' accident data in a study of passenger car accidents in Nordrhein-Westfalen, West Germany, in 1980, conducted by the Institut für Straßenwesen of the Technische Universität Braunschweig (Ref. 1). The aim of this study was to examine the relationship between accident occurrence and technical car parameters in accidents attributed to human rather than mechanical failure. Of special interest was to examine the possibility that certain car design features may contribute to an over-confidence in the driver of his ability to handle certain driving situations or reduce the drivers' awareness of speed or changing road conditions. An example of the former would be a car that undergoes an abrupt change in handling characteristics at a particular speed whilst cornering.

In order to separate the influence of car design from driver characteristics it was necessary to identify a group of drivers that has an average probability of causing accidents. These so-called 'normal' drivers could then be used in further studies to determine the type of car design parameters that contribute to accidents.

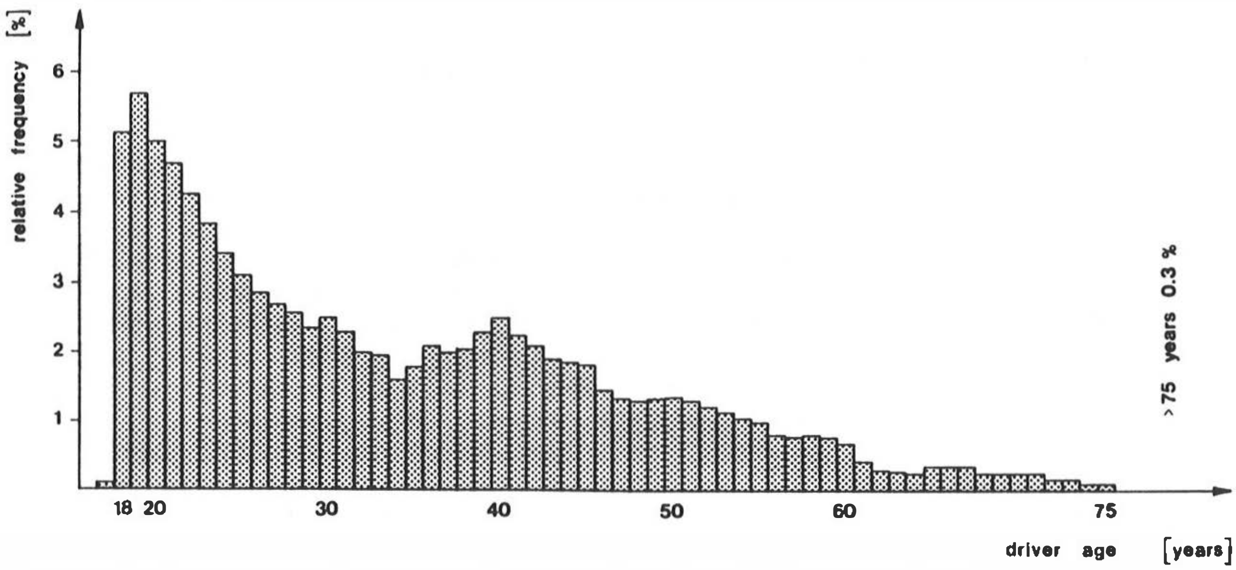

Figure 1: Frequency histogram showing the age of drivers who caused passenger car accidents. Data for the state of Nordrhein-Westfalen in 1980.

Many studies of road accidents are limited to an analysis of the raw data, for example, the age of drivers that have caused an accident (Fig. 1) or the types of cars involved (Fig. 2). However, it is important to consider the degree of participation of the different driver-age, or car-type populations in traffic and therefore exposure to accident risk. Also, the age distribution of drivers gives no information about the age-related behaviour that may contribute to accidents. Therefore from this data alone it is not possible to draw conclusions about the risk of accidents for special groups. The approach described here aims to taken account of exposure for drivers of different ages and car types, and of age groups whose behaviour is riskier than normal by the use of supplementary data and standardization techniques.

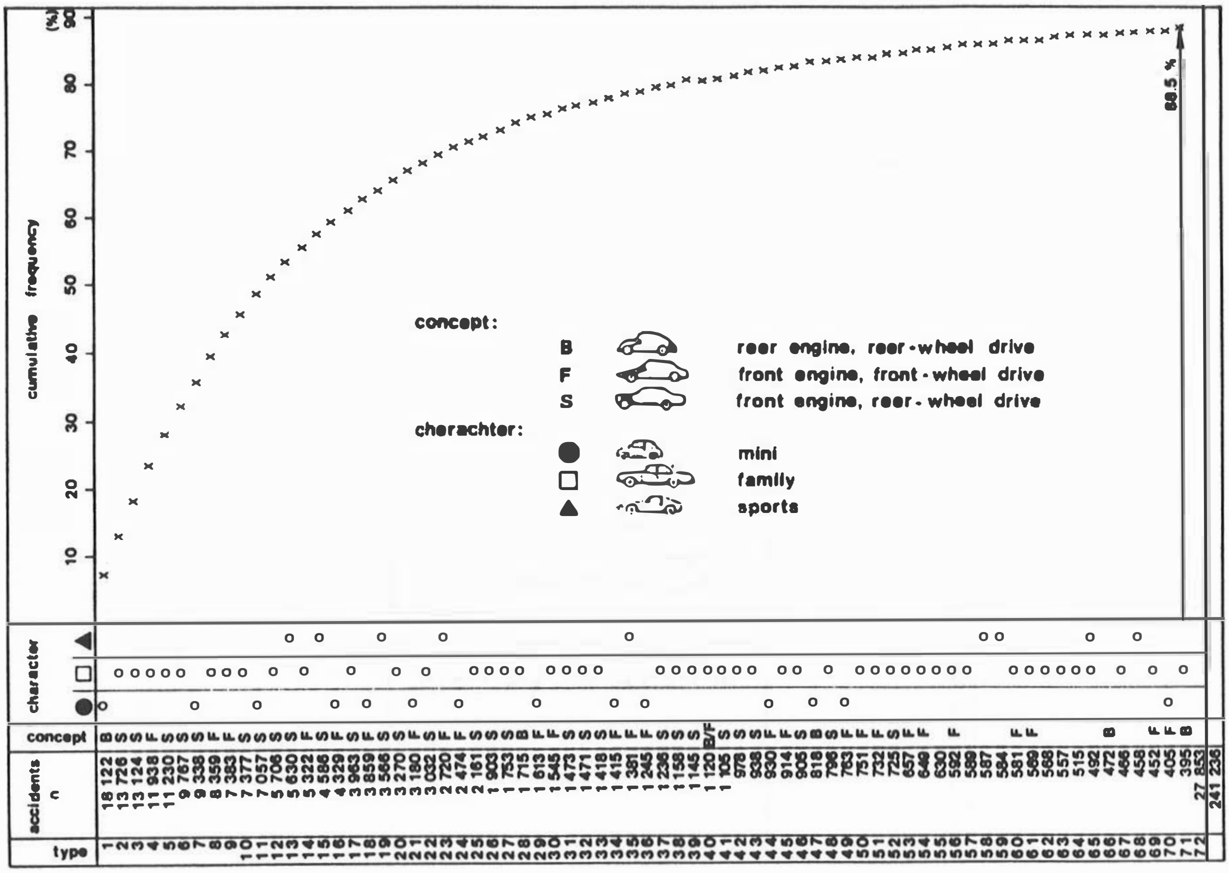

Figure 2: Cumulative frequency of types of passenger cars involved in accidents. The cars are listed in descending order of the numbers of accidents.

DATA SETS USED

In order to standardize the accident data several additional sets of data were collected. Figure 3 summarizes the way in which these were combined to determine driver and vehicle properties in relation to accidents. The data sets are described in turn.

The accident data set comprises 241,236 records describing passenger car accidents from a larger collection of 440,000 records covering all road accidents in Nordrhein-Westfalen in 1980. Each of these records contains information from the official accident report supplemented by data stored by the Federal Government Motor Vehicle Licensing Agency (Kraftfahrtbundesamt, KBA) such as the age and sex of the drivers and car parameters such as weight and power.

The fleet data consists of the numbers of cars registered in West Germany in 1980 for each of 171 different car models. It is possible to relate this information to the accident data because both data sets employ the same official KBA-Code to identify the make and model of a car.

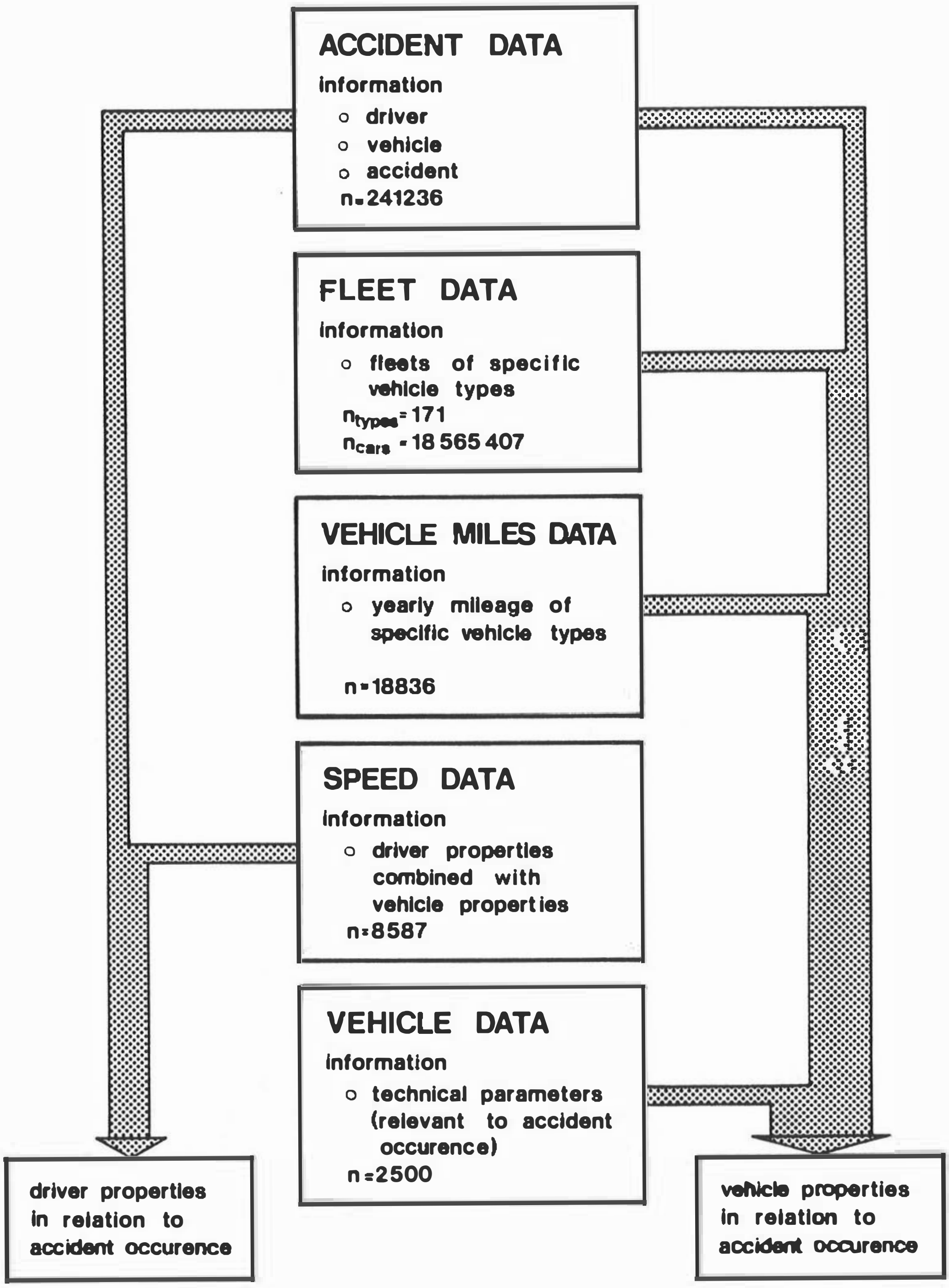

Figure 3: Data sets used for standardization and evaluation

In the third set the average annual mileage of different car models was stored. The information was obtained from 18,836 used-car advertisements published in German national daily newspapers. The average annual mileage was calculated from the total mileage quoted in the advertisement, the date the car was manufactured and the date of advertisement.

As the official KBA-Code for car type was not part of the advertisement, another code had be to used. This new code is based on the external appearance of the cars and is called TU-Code (Technical University Code). The TU-Code was added to all other data sets with the aim of having the possibility of connecting the different sets.

As part of the aim of identifying groups of drivers with an average accident involvement, it was important to obtain information about the age-related behaviour of non-accident drivers i.e. those in freely-flowing traffic. In particular, it was of interest to investigate whether drivers who chose high speeds corresponded to groups with a high accident involvement. Thus the speeds of 8,587 cars were observed in a roadside study. The TU-Code for passenger cars, their paths in bends, and the age and sex of drivers were included in this data set.

The vehicle data set contains information about the design of the car that may be relevant to accident occurrence, for example, the type of suspension geometry or whether the car has front, rear, or four-wheel drive. As this data is still being gathered, we will not discuss it in this report.

STANDARDIZATION OF DRIVER CHARACTERISTICS

To calculate the relative risk of causing accidents for various age groups, ratios were calculated that take account of the exposure in traffic for those groups. This is done by comparing the percentage of drivers of a certain age who caused accidents with the percentage of the total driver population (participating in normal traffic) of the same age. However, the problem in this study was to estimate the age dependent exposure of drivers in normal traffic (i.e. without accidents) because of the lack of basic data. It was postulated that the age distribution of drivers who were involved in an accident, but not the cause of the accident, closely approximates the age distribution of drivers in normal traffic and can therefore be used to show the relative degree of participation in traffic of the various age groups.

The official analysis of accident data distinguishes two groups of drivers; those who caused the accident and those who were involved but not the cause. The age-dependant relative frequencies of these two groups are shown in figure 4. The relative frequency of every age group totals 100 %.

This evaluation shows a totally different distribution in comparison with figure 1 where the age distribution of drivers who caused accidents was given. There, the frequencies of accidents decrease with increasing driver age. In contrast, figure 4 shows an increasing ratio of drivers above 50 years of age have caused the accidents in which they were involved.

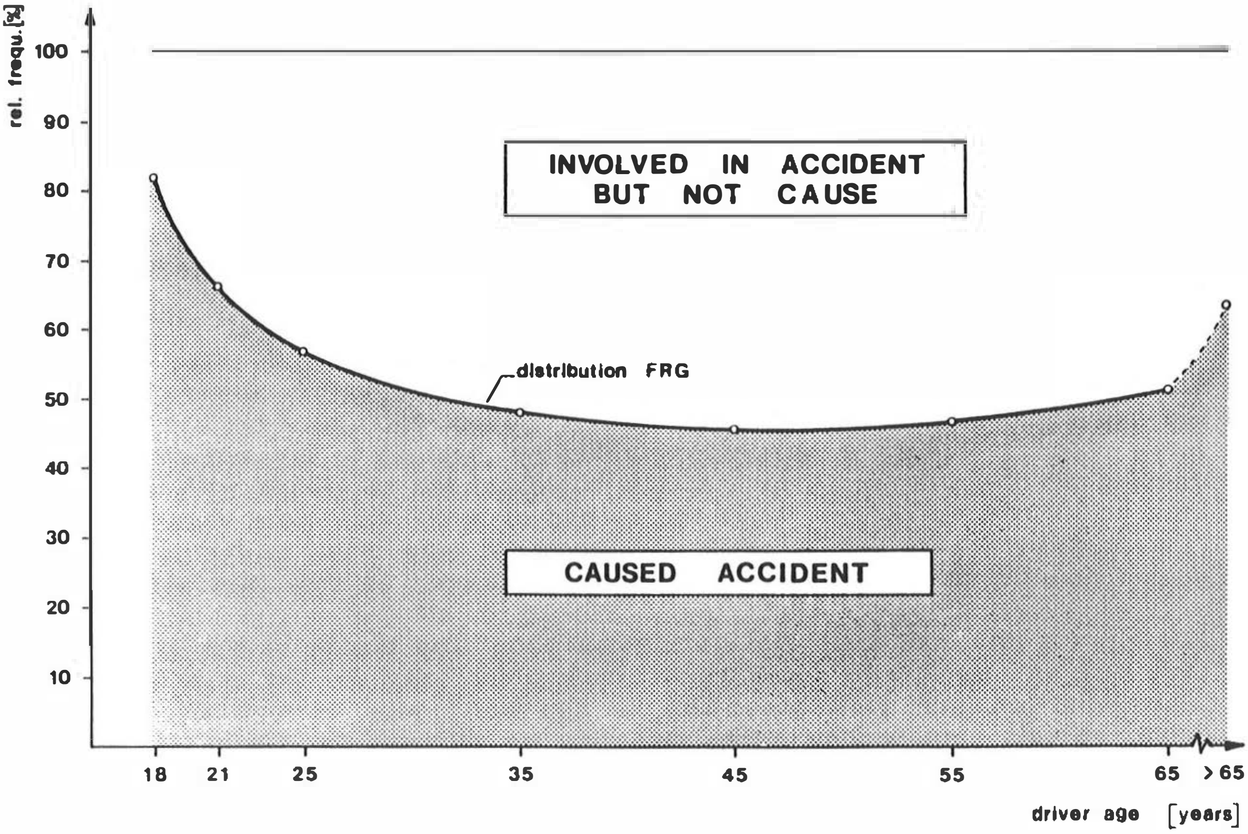

Figure 4: Percentage of drivers in each age group who
a) were involved in an accident but not considered to have caused it (top portion) and
b) caused an accident (lower shaded portion) versus age.
Data refers to West Germany in 1980.

The relative frequencies of drivers who caused accidents were standardized by calculating a 'cause-index', S, for each age group as follows:

$$S_{(age)} = \frac{n_{\%\,age,\ cause}}{n_{\%\,age,\ non\text{-}cause}}$$

where $n_{\%\,age,\ cause}$ = percentage of age group who caused an accident

and $n_{\%\,age,\ non\text{-}cause}$ = percentage of age group who were involved but who did not cause the accident,

and where the age distribution of 'non-cause' drivers (see upper part of fig. 4) is assumed to approximate normal traffic.

This index is plotted against drivers' ages in figure 5. If the value of the 'cause index', S, exceeds '1' then the corresponding group of drivers cause accidents more than would be expected from their representation in the total driver population.

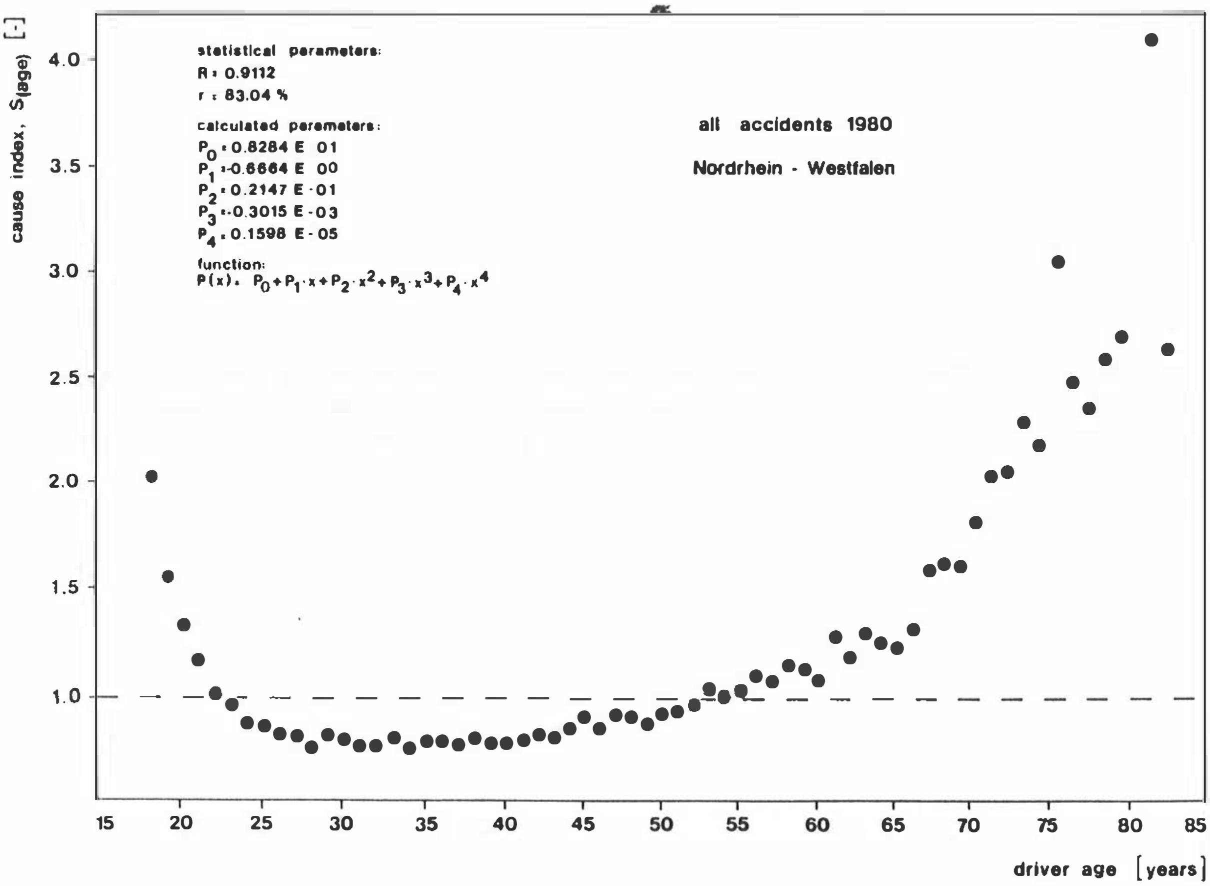

Figure 5: Age dependent 'cause index' S versus driver age

A 'cause index' of less than '1' indicates these drivers seldom cause accidents and an index of '1' shows that these groups cause accidents corresponding to their participation in traffic.

While the risk for younger (up to 22 years) as well as for older drivers (above 54 years) is extraordinarily high, the index is extemely low for the drivers within the age group 25 and 40 years.

STANDARDIZATION OF CAR CHARACTERISTICS

A standardization method similar to the one above was used to calculate a ratio for the involvement of specific car types in accidents. In this case, the exposure to traffic was estimated using the fleet and milage data sets. For every vehicle type (TU-Code) a ratio, M, was calculated as follows:

$$M_{(car\ type)} = \frac{n_{accidents,\ car\ type}}{n_{total\ accident}} \times \frac{av.\ km/yr._{all\ car\ types}}{av.\ km/yr._{car\ type}} \times \frac{n_{total\ cars}}{n_{car\ type}}$$

The first factor of this product is based on the accident data set, the second was obtained from the mileage data set and the third was obtained from the fleet data set which takes account of the relative abundance of specific models.

When car types are plotted in descending order of their ratio, M, (Fig. 6) a very different sequence is obtained compared with Figure 2, which shows cars in descending order of accident frequency. The lower part of figure 6 shows the change of position for each car type in the new sequence compared with figure 2.

The interpretation of the value of the car type ratio, M, is the same as for the cause index shown in figure 5 above. A ratio, M, exceeding '1' indicates that the car type is involved in more accidents than would be expected from its exposure to traffic.

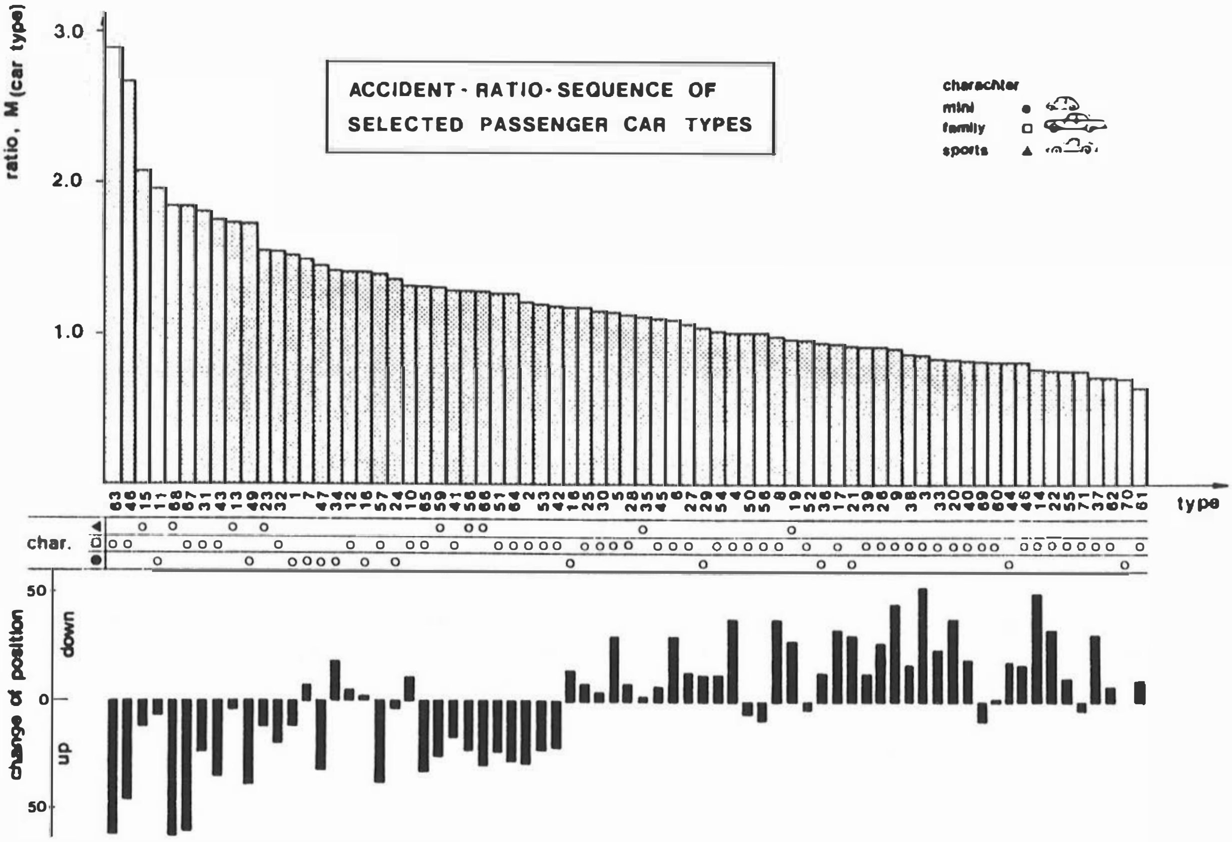

Figure 6: Upper part:
Car type (TU-Code) plotted in descending order of car type ratio M.
Lower part:
Change of position in sequence for each car type compared with Fig. 2 where the actual number of accidents was shown.

After this analysis it became evident that even though many types of cars have low accident frequencies, they must be considered as types with a high probability of accident involvement.

STANDARDIZATION OF SPEED DATA

This data set was established to examine relationships between the age of the driver and the car type used on one hand and the choice of speed in freely-flowing traffic on the other. 8,587 roadside observations were made over 58 days. It was intended to combine all observations together since the number of cars observed on a single day was too small for obtaining statistically significant conclusions. However, this was not possible since the use of X-tests for the daily speed distributions revealed that distinctly different populations existed.

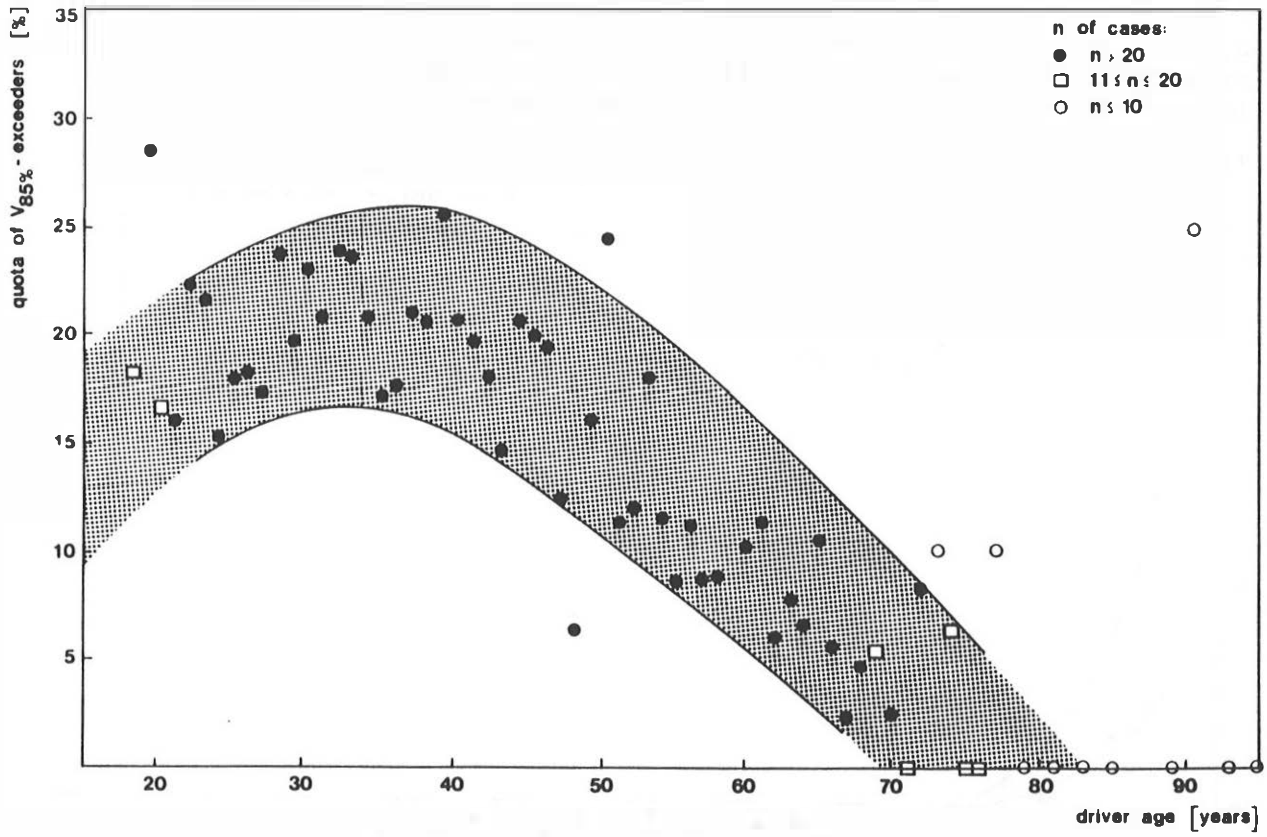

Figure 7: Quota of V 85%-exceeders versus age of drivers

The use of a parameter independent of the actual speeds (quota of exceeders of the 85-percentile speed) enabled observations for seperate days to be combined. First, the 85-percentile speed (V-85%, i.e. the speed that 85 % of the drivers do not exceed) was calculated for each of the 58 days. Using this value, the daily percentages of each driver-age and car-type population exceeding the V-85%-speed (quotas of V-85 % exceeders) were calculated. Figure 7 shows that drivers between 25 and 45 years chose the highest speeds and that the speed chosen decreases with increasing age. It is interesting to note that young drivers with ages between 18 and 20 years drove significantly slower than the 25 to 45 year age group.

METHOD OF DETERMINATION OF A 'NORMAL DRIVER'

When the cause index, S, for drivers was plotted for different types of accidents, the existence of two different types of functions was found. For drivers up to the age of about 45 years, the curves of best fit for the data have a hyperbolic shape for both 'single car' accidents (dashed curve in fig. 8) and accidents between participants going the same or opposite direction (solid line). For both these accident types, the cause index reaches a value of '1' for drivers between 22 and 23 years of age. It shows that drivers in their first four years of driving have a very high risk for these types of accidents.

For driver ages greater than 45 years, the curves of best fit have the shape of a polynomial of third degree (fig. 9). Figure 9 highlights the significantly higher risk of accidents at junctions for drivers in the older than 55 year age group than for younger drivers.

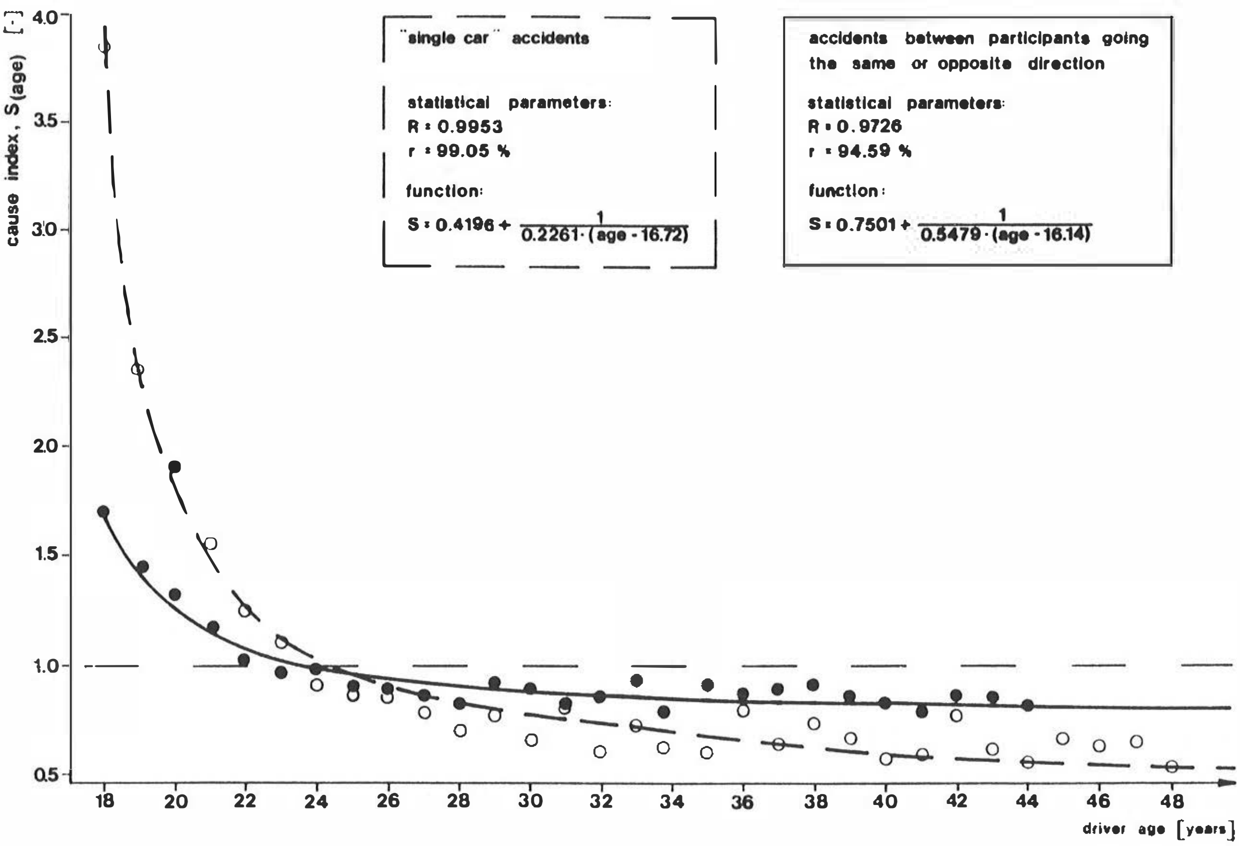

Figure 8: Age-dependent cause index, S, versus age for drivers aged between 18 and 48. Open circles show 'single car' accidents (line of best fit details in dashed box). Solid circles show accidents between participants going the same or opposite direction (line of best fit details in solid box).

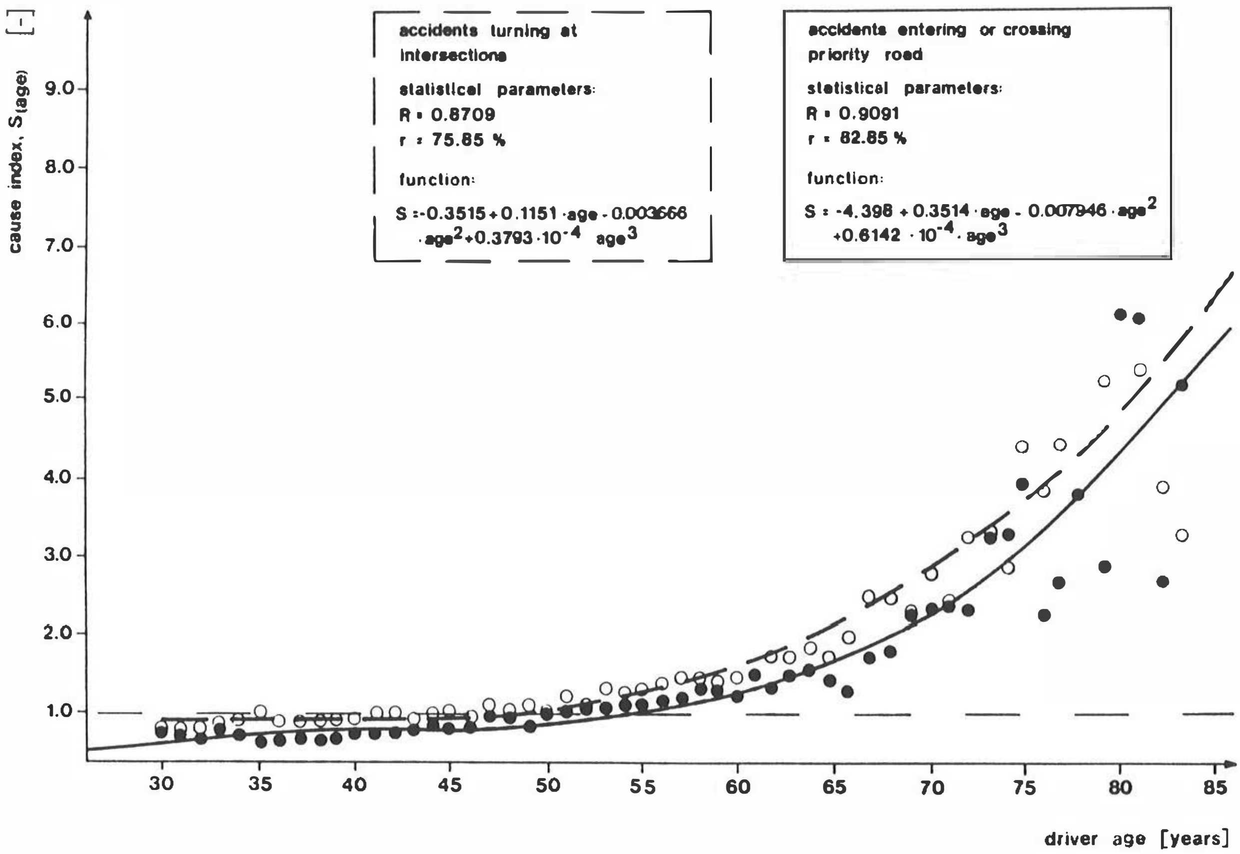

Figure 9: Age dependent cause index, S, versus age for drivers older than 30 years.
Open circles show accidents for cars turning at intersections (line of best fit details in dashed box).
Solid circles show accidents for cars entering or crossing a priority road (line of best fit details in solid box).

Based on this form of statistical analysis (as illustrated in Fig. 8 and 9), the larger study (Ref. 1) identified two distinctly different kinds of driver failure. Young drivers often lose control of the car because of an inappropriate speed. Significantly, this may not simply mean a high speed in absolute terms, but a speed which is too high in relation to the inexperienced drivers' skills or the road conditions or both (refer to fig. 7). The older drivers appear to cause accidents mainly because of their slower reactions and reduced decision-making capabilities.

REFERENCES

1 Dörschlag, S.; Schlichting, K.-D. Fahrer- und Fahrzeugeigenschaften und Unfallgeschehen Forschungsbericht des Instituts für Straßenwesen der Technischen Universität Braunschweig 1987 im Auftrage der Bundesanstalt für Straßenwesen, Bergisch Gladbach

2 Straßenverkehrsunfälle 1980 Hrsg.: Statistisches Bundesamt Wiesbaden