The accident liability of car drivers

by G Maycock, C R Lockwood and Julia F Lester

Research Report 315

The Transport Research Laboratory is the largest and most comprehensive centre for the study of surface land transport in the United Kingdom. TRL provides research-based technical assistance to help set standards in all aspects of surface transport, improve safety and the environment, and encourage good transport practice for the benefit of both the public and private sectors, and hence ultimately for society as a whole.

As a national transport research laboratory, TRL operates world-wide and has developed close working links with many other international transport centres.

TRL is a non-profit distributing company whose customers include central government, local and regional authorities, consultants, industry (construction, infrastructure owners, operators, automotive), foreign governments and international aid agencies.

TRL employs over 500 people, including 380 technical specialists, many with higher degrees and Doctorates. Facilities include a state-of-the-art driving simulator, various indoor impact test facilities, a 3.8 km test track, a separate self-contained road network, a structures hall, an Intelligent Transport Systems Centre, an acoustics room and many other laboratories. Numerous mobile off-site test facilities, data collection devices and specialised vehicles are also employed to support research, consultancy and advice.

The Laboratory's primary objective is to work closely with its customers to carry out commissioned research, investigations, studies and tests to the highest levels of quality, reliability and impartiality. TRL carries out its work in such a way as to ensure that customers receive results that not only meet the project specification or requirement but are also geared to rapid and effective implementation.

Transport Research Foundation Group of Companies
Transport Research Foundation (a company limited by guarantee) trading as Transport Research Laboratory.
Registered in England, Number 3011746. TRL Limited. Registered in England, Number 3142272.
Registered Office: Old Wokingham Road, Crowthorne, Berkshire. RG45 6AU

TRANSPORT AND ROAD RESEARCH LABORATORY
Department of Transport

RESEARCH REPORT 315

THE ACCIDENT LIABILITY OF CAR DRIVERS

by G Maycock , C R Lockwood and Julia F Lester

Road User Group
Transport and Road Research Laboratory
Crowthorne, Berkshire, RG11 6AU
1991

ISSN 0266-5247

CONTENTS

THE ACCIDENT LIABILITY OF CAR DRIVERS

ABSTRACT

Data has been collected from a structured sample of just over 18,500 drivers using a postal questionnaire, to determine the relationship between the accident liability of these drivers, and factors such as age, driving experience, sex, Socio-economic group (SEG), and annual mileage by type of road. Accident liability is defined as the expected number of accident involvements per year. Generalised linear modelling techniques have been used to develop a statistical model which will predict the accident liability for an individual driver as a function of relevant explanatory variables.

The model suggests that accident liability is dependent mainly on exposure (total annual mileage), the driver's age *and* his or her driving experience measured as the number of years since passing the test. Predicted accident frequencies are not directly proportional to annual mileage, and are dependent on the proportion of driving done in the dark and on different types of road (built-up, rural and motorway). Accident liability falls with increasing age and driving experience. The form of the age and experience relation means that the proportional change in liability with increasing age or experience is larger for younger drivers than for older drivers; this is particularly marked for experience - i.e. the learning curve is steep. Men have higher accident liabilities than women at all ages. The differences between Socio-economic groups are relatively small.

1. INTRODUCTION

It is well established from UK national road accident statistics (Broughton, 1988, 1990) that the injury accident involvement rates of male car drivers per km driven falls by approximately a factor of 7 from the involvement rate of young drivers aged 17-20 to that of drivers aged about 50; after the age of 65 injury accident involvement rates begin to rise again. Female driver injury accident involvements show a similar pattern although young females exhibit only half the rate of their male counterparts, and involvement rates for older female drivers rise to a rate nearly twice that for men of similar age. The UK national accident database contains accidents involving personal injury which have been reported to the police.

The present study complements these findings by analysing the self-reported accident involvement rates of individual drivers. The purpose of this study is to establish the relationship between the accident liability of an individual driver (i.e. the expected frequency of his or her involvement in mainly non-injury accidents) and factors such as age, sex, exposure and driving experience. The study was undertaken as part of a programme of behavioural studies at TRRL and it is intended to provide a factual basis for more detailed psychological studies. The statistical models form a 'description' of the way accident liability varies between individual drivers which the psychological investigations would aim to 'explain' in terms which could be used as the basis of remedial interventions.

The effect of age and experience on accident liability are factors of particular interest since there are grounds for believing that they influence driving behaviour (and hence accidents) differently. Age effects clearly reflect growing maturity - perhaps changing perceptions of risk, or a growing sense of social responsibility. Experience on the other hand reflects a learning process within the driving task. The distinction between maturation and learning could have important implications for the application of road safety remedial treatments. If age is the dominant factor, then short of devising a method of accelerating the maturation process, raising the driving age might be the only practical countermeasure. If on the other hand, experience plays a crucial part, then improved safety might be achieved by devising better ways of imparting those skills necessary for safe driving to novice drivers - a matter of training.

The present report describes a study designed to model the factors affecting accident liability, and in particular to determine the relative importance of age and experience. Section 2 describes a literature survey reviewing what is known about the effects of age and experience on accidents. Section 3 details the methodology used in the present study and section 4 presents the data collected together with some basic tabulations of the characteristics of the sample of drivers and their accidents. Section 5 describes the multivariate modelling methodology, and section 6 presents the resulting models. Model predictions are discussed in section 7 and summarised in section 8

Four Appendices provide detailed accounts of various aspects of the work: Appendix A deals with the effect of memory loss, Appendix B describes the modelling process, Appendix C gives statistical details of the final model and Appendix D details of a simplified model and a model of injury accidents.

2. LITERATURE REVIEW

The literature covering the last 30 years has been reviewed (Barwick, 1990) to identify studies which have attempted to investigate the effect of age and experience on the accident rates of individual drivers. In order to be wholly satisfactory, it was considered that studies should be methodologically sound, they should include a valid measure of driving experience (usually - but not necessarily - the length of time since passing the driving test), they should include or control for exposure (miles travelled) and should treat male and female drivers separately.

A total of 12 studies were given detailed consideration, though many of them fell considerably short of the criteria outlined in the previous paragraph. Studies based on insurance policy data were regarded as unsatisfactory in that driving 'experience' was equated with the duration of the policy rather than the length of time the policy holder had been driving. A similar criticism can be levelled at studies based on the company records of professional drivers' accidents, where experience is equated with the length of employment with the company. Many studies did not satisfactorily control for exposure, and since exposure (miles travelled) does vary with age and sex, this omission seriously compromises the findings.

A study by Ferdun, Peck and Coppin (1967) - which was largely supported in a follow-up by Harrington (Harrington, 1972) - found an age effect among male drivers aged 16-19 when exposure was controlled for by means of a multivariate analytical method; older males have fewer accidents. Experience was defined both in terms of the total number of miles driven in an individual's lifetime and the number of months for which a licence had been held. Ferdun et al report that for males (in the 16-19 age group) experience factors did not affect accidents, whereas in the case of females, experience measured by the number of months a licence had been held did - as experience increased, accidents decreased. For female drivers, once experience factors were controlled for, age did not affect accident rates.

Kritz and Nilsson (1967) studied drivers aged from 18-50 in their first year of driving. They found a strong age effect for male drivers (at constant 'experience') whilst controlling for exposure by comparing the accident rates of drivers of different ages driving similar mileages; drivers less than 25 had a higher accident rate that older newly qualified drivers driving the same annual mileage. They found a significant age effect for female drivers although the magnitude was not as great as that of their male counterparts.

Pelz and Schuman (1971) - studying drivers up to the age of 44 -analysed the accident and violation records of drivers within specified annual mileage bands, as a function of the drivers age, and the age at which he or she learned to drive. The findings of this analysis are not easy to interpret, but they suggest that as far as young male drivers are concerned (aged under 25), age rather than driving experience is the dominant determinant of accidents and violations. Data from other driver groups (i.e. older drivers and female drivers) were not analysed in the same amount of detail.

It will be apparent from this brief survey that relevant studies of the accident effects of age and experience are not plentiful. In their review of the literature, the Organization for Economic Co-operation and Development (OECD, 1975) concluded: 'the situation as regards the relative effects of experience and age-related factors seems somewhat obscure'. More recently, the literature on this topic has been described as 'sparse and inconclusive' (Jonah, 1986). Since the literature survey mentioned above (Barwick, 1990) was completed, an extensive review of the subject has been published by the Traffic Injury Research Foundation of Canada (Mayhew and Simpson, 1990). As a result of this review, Mayhew and Simpson conclude 'while no clear picture emerges, the review ... suggests that both age-related factors and lack of driving experience account for some of the higher crash risk of young drivers - the relative contribution of these factors remains unknown'. The authors then go on to present an analysis of an Ontario Ministry of Transport database (the 'Trace' database) which clearly demonstrates an age and an experience (years licensed) effect for both men and women - though they conclude that the age effect is larger than the effect of driving experience.

Mayhew and Simpson point out that because the demographic characteristics of those applying for licences is changing - with more novice drivers being female and older than has previously been the case, it is of some practical importance for the design and implementation of remedial measures to identify the relative magnitude of age and experience on accident liability. The present study was designed to make a contribution to this topic.

3. METHODOLOGY

3.1 DEFINITIONS

The analysis described in this report aims to relate accident liability to a range of relevant explanatory variables or factors. It is necessary first to define what is meant by 'accident liability' in the context of this study. In the introduction, accident liability was defined briefly as the 'expected frequency of a driver's involvement in accidents'. The terms 'expected frequency', 'involvement' and 'accidents' need definition. It is appropriate to begin by defining an accident.

The types of accident recorded in the national accident database (STATS 19) are those involving injury (including fatalities) which have been reported to the police. Unfortunately, the national accident data is inadequate for the kind of analysis reported here - for the following reasons:

i) accident histories of individual drivers cannot be obtained from the national data,

ii) no information is available on exposure - the number of miles travelled per year, the types of road used or the times of year in which the journeys take place,

iii) the national accident records include the age of the driver but not the number of years driving experience.

Moreover, because the national records include only accidents involving personal injury reported to the police, the average frequency of such accidents per driver is very small - of the order of 0.01 accidents per driver per year. Such an average accident frequency is too low to enable multivariate analytical techniques to be used satisfactorily.

The data required for this study has therefore been obtained by means of a nationwide self-completion questionnaire survey. Details of the sampling strategy and the survey administration are given in the following sections. For the present it is sufficient to note that drivers were asked to provide details of 'all kinds of road accidents that they had been involved in as a driver over the last three years' - or for young drivers (under 23) accidents in which they had been involved since they started driving. A road accident was defined as any incident which occurred on a public road (not on private property) and which involved injury to the driver or another person, damage to property or to the vehicle being driven. Thus the majority of accidents included in this analysis are damage-only accidents. In fact, only about 11% of the accidents reported by respondents involved injury; for obvious reasons none involved the death of the driver.

The questionnaire approach described above, also makes clear that the information collected relates to accident 'involvements'. National accident tabulations normally provide information on the number of accidents, treating an accident as a single event regardless of the number of vehicles involved in each. Collecting accident data by questionnaire collects frequency data on individual driver involvements in accidents - a multiple vehicle accident could give rise to multiple involvements in the questionnaire database. This distinction needs to be borne in mind when comparing the results presented here with national accident tabulations.

It remains to define accident liability as the 'expected frequency' of accident involvements. The dependent variable used in the analysis is the number of accidents the driver has reported as having been involved in during a specific period (usually 3 years). The period over which the accidents have occurred is included in the analysis in such a way that the final predictive equation gives estimates of the expected values (the statistical expectation) of the number of accident involvements per year (frequency). So for example, the overall average value for the full data set (uncorrected for memory loss) is 0.12 accident involvements per year. This value (and all others estimated by the predictive models given in this report) are to be regarded as the underlying liability of the individual driver to become involved in accidents. The objective of the analysis is then to determine how this accident liability varies with age, experience, exposure and any other relevant variables.

One of the problems about retrospective surveys of accidents is that respondents are likely to forget some of the accidents they have experienced. This would mean that the number of accidents reported will be lower than the true number depending on the extent of forgetfulness. However, by comparing the apparent 'within individual' change in accident experience (accidents reported by an individual during the most recent year of driving compared with those reported in the previous year and the one before that) with the 'between individual' effect (derived from the age/experience relation) it is possible to estimate the likely magnitude of memory loss. It turns out that for all accidents, respondents forget about 30% per year (see Appendix A). Not surprisingly, the memory loss factor is lower for injury accidents. No corrections for memory loss have been applied to the tabulations of accident data reported in section 4, but the multivariate analysis has included corrections for this effect.

3.2 THE SAMPLE

A study carried out some years ago relating to 'accident involved' drivers (Quimby et al, 1986) confirmed that a strong relationship exists between the frequency of road accidents experienced by drivers and their age. Moreover, in the virtually random sample of drivers used in the accident involved driver study, a high correlation (0.9) was found between the drivers' age and their experience - measured as the number of years since passing the driving test. This is not surprising; young drivers of necessity have little experience whereas older drivers have usually been driving for some time. The high correlation between these two variables however, makes it difficult to determine reliably how much either factor alone affects accident liability.

Additionally, to ensure that the age effect is determined with a reasonable degree of accuracy at both ends of the age spectrum, it was thought necessary to over-sample both young drivers and older drivers. The sample of drivers was thus structured to have four components: for drivers aged 23 or over, a 'RANDOM' sub-sample - drivers selected at random, an 'OLD' sub-sample selected so that the combined numbers of the RANDOM+OLD sub-samples would be evenly distributed over the whole age range, and an 'INEXPERIENCED' sub-sample consisting of drivers of all ages who had less than 10 years driving experience at the time the sample was drawn; the fourth component was a 'YOUNG' sub-sample consisting of drivers less than 23 years of age. Each sub-sample included equal numbers of male and female drivers.

All samples were drawn by scanning the whole of the driver licence file maintained by the Driver Vehicle and Licensing Centre (DVLC) at Swansea selecting every nth driver for inclusion in the study. The value of n was adjusted in cells corresponding to the four sub-samples and 6 age bands to give target numbers for mail-out of questionnaires shown in Table 1.

In fact, the files were considerably over-sampled to provide a pool of reserve respondents because it was thought (by DVLC) that up to a half of the addresses in the driver file might be incorrect. In the event, the reserve respondents were not needed because the response to the first mailing was very good (see next section); the implied accuracy of the DVLC records was far better than expected - at most 20% of addresses were incorrect at the time of sampling.

TABLE 1

Numbers of questionnaires mailed by age in the four sub-samples

		Age ranges					
		< 30	30-39	40-49	50-59	60-69	> 69
RANDOM + OLD		2000	2000	2000	2000	2000	2000
INEXPERIENCED							
	3-4	400	400	400	400	400	
Experience:	5-6	400	400	400	400	400	
(years)	7-8	400	400	400	400	400	
	9-10	400	400	400	400	400	

	Age Ranges		
	17-18	19-20	21-22
YOUNG	5000	3000	2000

3.3 THE SURVEY

The survey was carried out on behalf of TRRL by NOP Market Research Ltd. At the request of TRRL, DVLC supplied a tape containing the names and addresses of respondents to NOP, who mailed out the 30,000 questionnaires and checked and coded the responses. The resulting data was returned to TRRL on 'floppy disk' for further validation and analysis.

The RANDOM, OLD and INEXPERIENCED sub-samples were surveyed in November 1987 and the YOUNG sub-sample in February 1988. Three mailouts were used, an initial mailing followed by two reminders to non-respondents. Each mailing included a copy of the questionnaire together with a letter on TRRL headed paper. The response rates are shown in Table 2.

It will be seen that the overall response rate for the survey was about 65%, which if the inaccurate addresses are taken into account could represent a 'real' response (ie a response from those actually receiving a copy of the questionnaire) of between 75 and 80%. Subsequent consistency and edit checks of the returned questionnaires eliminated a number of doubtful responses, so that the total number available for analysis was just over 18,500. Apart from a lower response than anticipated among drivers having only 3-4 years experience (probably a consequence of the experience bands used), the response rate was fairly uniform across all the cells shown in Table 1.

In the case of the first phase of the survey (ie excluding the YOUNG drivers), an attempt was made by NOP to contact a sample of non-responders. This was achieved

TABLE 2

Questionnaire survey response rates

	RANDOM + OLD + INEXPERIENCED Drivers	YOUNG Drivers
Number of questionnaires mailed out	19,972	10,000
Number completed and returned	12,700 (63.6%)	7,134 (71.3%)
Number returned from non-responders	1,185 (5.9%)	389 (3.9%)
No reply received	6,087 (30.5%)	2,477 (24.8%)

by means of a telephone survey of 472 drivers selected at random from the 6,087 who had not replied at all. Only 185 of these proved to have a traceable telephone number, which in view of the high level of telephone penetration among driving licence-holders, probably reflects inaccurate address information. Of those with traceable telephone numbers, 18% had either moved or were not personally contactable by telephone. Of those contacted, the main reasons for non-response were: objections to research (or questionnaires) - 19% (of the 185); lack of interest - 19%; personal reasons (too old, don't drive, too ill) - 8%. 8% who received the questionnaire claimed to have returned it, whilst a similar percentage claimed not to have received it; 10% were unsure whether they had received it or not. From this brief survey of non-responders there is no reason to believe that the drivers used in this analysis are unduly biassed in any way.

4. THE DATA

4.1 THE QUESTIONNAIRE

The following information relevant to the accident liability modelling to be reported here was sought from respondents by means of the questionnaire:

The number of accident involvements (as defined in section 3.1) in a car or a van in the last three years or, in the case of YOUNG drivers, since passing the driving test,

Age last birthday,

Sex,

Questions defining Socio-Economic Group of the respondent,

The year in which the driving test was passed - this enables the drivers 'experience' to be determined as the number of years since passing the test,

Exposure:

- an estimate of the number of miles driven in the last year,
- percentage of time spent driving on roads in built up areas, roads in rural areas, and motorways,
- amount of time spent driving during dark and light conditions.

Data on accidents and mileage was requested from each respondent in relation to any car or van driven during the sampling period -that is to say, the study was driver orientated, and did not collect information about the type of vehicle or vehicles being driven. Cars *and* vans were included in the survey because many drivers will use small vans for social as well as business mileage - the definition of a van was left to the respondents individual judgement. Some information was collected on the use of other types of vehicles - mainly motorcycles and HGVs - but the proportion of respondents driving other vehicles is relatively small (15%).

Detailed information was also collected about the first three accident involvements - did the accident involve injuries, and if so how severe? What other road user or roadside objects were involved in the accident? How much damage was done to the vehicles involved, and what were the costs of repair? Was the accident reported to the police or the subject of an insurance claim?

The information obtained on costs of repair has been published elsewhere (Taylor, 1990).

4.2 SUMMARY TABULATIONS

This section provides some simple tabulations illustrating the more important aspects of the data.

As has been already pointed out, the accidents reported by respondents are largely damage only accidents, and although they include accidents involving personal injury they do not include accidents in which the driver was killed. With this in mind this section includes tabulations of the mean accident involvement frequency (the number of involvements per year) by sex, age, SEG, driving experience and annual mileage of the driver. In all cases the accident frequency is calculated as $\Sigma A/\Sigma T$, where:

A is the number of accident involvements reported by a driver in time T, and,

T is the length of the period during which these accidents took place.

Only drivers who have been driving for more than 3 months have been included in the tables. Additionally, because of some uncertainty about the reporting of low annual mileages, only drivers who said they had travelled more than 400 miles per year have been included.

4.2.1 Annual Mileage

Tables 3 and 4 and Figure 1 show for men and women respectively, how accident frequencies increase as the mileage travelled per year increases. The accident frequency in each mileage band is the average of all respondents within the band, and the average mileage and number of respondents within each band is also given in the tables. The tables also show the average proportion of mileage driven on motorways, built up roads and rural roads as estimated by respondents.

It is clear that accident frequencies do not increase in proportion to the mileage travelled. For example in Table 3 (Males) from the lowest to the highest mileage band, the average mileage has increased by a factor of about 28 whilst the accident frequency has increased by only a factor of just over 2. It is also clear that the higher mileage drivers cover a higher proportion of their mileage

TABLE 3

Accident frequency and road-type exposure by annual mileage - Male

Annual Mileage:	0-2999	3-5999	6-9999	10-14999	15-29999	>30000
All accidents (per year)	0.119	0.102	0.129	0.163	0.186	0.260
Injury accidents (per year)	0.015	0.011	0.015	0.014	0.019	0.029
Average annual mileage	1490	4210	7420	11390	19130	42350
Proportion of miles on:						
Motorways	14.4	14.9	16.4	18.4	25.5	33.1
Built up roads	55.4	55.2	53.8	49.8	44.0	40.2
Rural roads	30.2	29.9	29.8	31.8	30.5	26.7
Number of respondents	1526	1530	1663	1669	1344	476

TABLE 4

Accident frequency and road-type exposure by annual mileage - Female

Annual Mileage:	0-2999	3-5999	6-9999	10-14999	15-29999	>30000
All accidents (per year)	0.072	0.080	0.128	0.144	0.212	0.220
Injury accidents (per year)	0.009	0.009	0.014	0.020	0.023	0.021
Average annual mileage	1390	4100	7200	11190	18020	43000
Proportion of miles on:						
Motorways	9.8	12.0	14.5	17.9	23.1	26.0
Built up roads	58.7	56.3	53.1	46.2	45.4	42.5
Rural roads	31.5	31.7	32.4	35.9	31.5	31.5
Number of respondents	3309	2161	1344	844	331	64

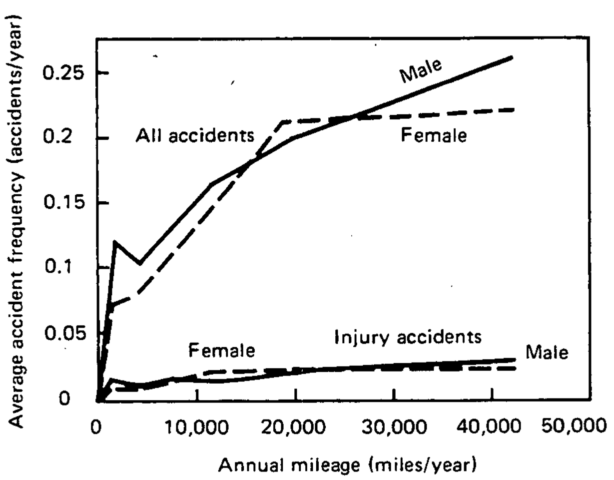

Fig.1 The observed effect of annual mileage on accident liability

on motorways. It way well be then, that the lack of proportionality between accident frequency and annual mileage is due to the fact that high mileage drivers cover a higher proportion of their mileage on motorways which are safer per mile travelled. The multivariate analysis should be able to shed some light on the extent to which a 'type-of road' effect of this kind explains this observed lack of proportionality.

It is also apparent (Figure 1) that the difference in accident frequencies between men and women drivers who travel similar distances per year is quite small; the magnitude of this effect will also be determined in the multivariate analysis.

From the data presented in Tables 3 and 4 on the proportion of mileage covered on the three types of road (built-up, rural and motorway), it will be seen that for both men and women, about 30% of mileage is travelled on rural roads independent of the drivers total annual mileage. However, as the annual mileage increases, the proportion of this mileage driven on motorways increases at the expense of mileage in built-up areas. About one third of the annual mileage of high mileage male drivers is on motorways.

4.2.2 Age and Driving Experience

Tables 5 and 6 show, for men and women respectively, how accident frequencies (accidents per year) and accident rates (accidents per million miles) for all accidents and for injury accidents, fall as a driver gets older and gains driving experience.

To obtain the figures in the tables, the data has been averaged over experience within age bands. This will bias the age effect somewhat since the distribution of experience within the age bands will be broader for the older ages than for the young drivers. Moreover, the figures have not been corrected for the effects of memory loss, so that the older drivers have accident rates which for this reason will be depressed relative to the younger drivers (accident periods less than 3 years mean a lower memory loss effect). Although the memory loss effect is probably dominant, these two effects tend to cancel and more detailed tabulations by age and experience suggest that tables 5 and 6 illustrate fairly realistically how accident frequencies and rates would vary over time for drivers who started to drive at age 18, and continued to drive until 70.

The tables clearly show the large age/experience effect as accident frequencies fall by a factor of 7 to 8 from young inexperienced drivers to older experienced drivers. The multivariate analysis will quantify the effects of age and driving experience separately, and correct for differential memory loss effects. The clear difference between the sexes is also apparent, with women experi-

TABLE 5

Accident frequency and accident rate by age - Male

Age (years):	<18	18	19-22	23-29	30-39	40-49	50-59	60-69	>70
All accidents (per year)	.462	.326	.226	.126	.092	.080	.078	.055	.056
All accidents (per million miles)	50.4	32.3	18.1	9.6	7.2	6.8	8.3	7.9	11.7
Injury accidents (per year)	.045	.031	.028	.015	.012	.008	.006	.006	.005
Injury accidents (per million miles)	5.0	3.1	2.3	1.1	1.0	0.7	0.6	0.8	0.9
Number of Respondents	1472	762	986	963	990	889	944	832	370
Annual mileage	9170	10080	12490	13140	12770	11760	9500	6930	4800

TABLE 6

Accident frequency and accident rate by age - Female

Age (years):	<18	18	19-22	23-29	30-39	40-49	50-59	60-69	>70
All accidents (per year)	.274	.199	.139	.092	.073	.062	.057	.046	.037
All accidents (per million miles)	52.2	38.1	23.2	15.2	13.5	11.8	14.4	12.7	11.9
Injury accidents (per year)	.030	.028	.020	.012	.007	.007	.007	.005	.005
Injury accidents (per million miles)	5.7	5.3	3.3	2.0	1.2	1.3	1.7	1.4	1.5
Number of Respondents	1357	765	1014	1067	1085	923	971	707	294
Annual mileage	5250	5220	5990	6010	5450	5260	3990	3600	3130

encing much lower accident frequencies than men. Overall, the ratio of 'all accident' frequency to injury accident frequency is 8.8.

Of course, as the tables show, there are considerable differences between the age groups and between men and women in the annual mileages covered. If accidents per million miles are calculated, the age/experience effect for men at least becomes U-shaped, with an upturn in accident rates for the oldest drivers. Moreover, whereas in terms of accident frequencies (per year) women have fewer accidents than men, in terms of accidents per million miles women have higher rates than men. It will be shown later that this result is due to the finding that accident frequencies are not proportional to annual mileage - as noted in the previous section - combined with the fact that women drive about half the annual mileage of their male counterparts. A meaningful comparison of the relative 'risk' of sub-groups in the population (for example women compared with men, or old compared with those in the middle of the age range) depends critically on the way accident frequencies are normalised for differences in exposure. This point will be considered further in section 7.1.

Table 7 shows for men and women combined the effects of age and experience for all accidents. The number of respondents in each cell of the table is given, but no attempt has been made to adjust accident frequencies for mileage. The table is included to show that both an age and an experience effect exist in the data; it will be the objective of the multivariate analysis to unravel these interactions.

4.2.3 Socio-Economic group

Tables 8 and 9 show for men and women respectively, the effects of socio-economic group. The grouping used in the tables are defined as follows:

SEG's A and B - Managerial, administrative or professional,

SEG C1 - Supervisory and clerical, junior managerial, administrative or professional,

SEG C2 - Skilled manual workers,

SEG's D and E - Semi-skilled and unskilled manual workers, pensioners, casual workers and the unemployed.

It will be seen from the tables that SEG does not appear to have a large effect on accident frequencies. There is a hint in the tables for both men and women, that SEG's A, B and C1 have a higher accident frequency than SEG's C2, D and E, though it is always possible that the true effect is masked by correlations within the data. The tables show that annual mileages and average ages do not vary greatly between the SEG groups. The effect of SEG will be examined again in the later multivariate analysis.

4.2.4 Accident Types

As indicated in section 4.1, questionnaire respondents were asked to provide some details of the accidents in

TABLE 7

Accident frequency by age and driving experience - Both sexes

Driving Experience (years)	Age (years) <18	18	19-22	23-29	30-39	40-49	50-59	60-69	>70
<1	0.385 (2733)	0.325 (902)	0.270 (410)	-	-	-	-	-	-
1 - 5	0.272 (96)	0.226 (625)	0.180 (1335)	0.114 (220)	0.100 (204)	0.074 (172)	0.078 (188)	0.085 (102)	-
5 - 10	-	-	0.152 (255)	0.108 (1390)	0.082 (676)	0.063 (442)	0.064 (465)	0.058 (282)	-
10 - 20	-	-	-	0.104 (420)	0.080 (988)	0.081 (432)	0.059 (315)	0.049 (191)	0.058 (40)
20 - 30	-	-	-	-	0.073 (207)	0.070 (673)	0.069 (472)	0.039 (254)	0.043 (101)
>30	-	-	-	-	-	0.057 (93)	0.072 (475)	0.047 (710)	0.049 (515)

( ) Figures in brackets are the number of respondents in each cell of the table.

TABLE 8

Accident frequency by socio-economic group - Male

Socio-Economic Group (SEG):	A/B	C1	C2	D/E
All accidents (per year)	0.165	0.157	0.141	0.137
Injury accidents (per year)	0.017	0.016	0.015	0.016
Number of respondents	1401	2146	2435	1475
Annual mileage	10970	11240	10440	10090
Average age	36.1	34.7	33.9	37.3

TABLE 9

Accident frequency by socio-economic group - Female

Socio-Economic Group (SEG):	A/B	C1	C2	D/E
All accidents (per year)	0.099	0.107	0.101	0.094
Injury accidents (per year)	0.013	0.012	0.013	0.014
Number of respondents	1590	2556	1853	1021
Annual mileage	5340	5330	4780	5090
Average age	35.5	34.5	29.9	35.2

which they had been involved. Table 10 gives the distribution of accident types by age and sex. Three types of accidents are identified in the table: 'moving' accidents - those involving another vehicle, a pedestrian or a cyclist; 'stationary' accidents - those involving a stationary or parked vehicle; and single vehicle accidents - in the main those involving roadside objects.

It will be seen from the table that young (and inexperienced) drivers of both sexes were involved in a higher proportion of single vehicle accidents than were mature drivers. There is a suggestion of a U-shaped relation between age and accidents involving stationery vehicles, the proportion of these accidents being at a minimum for middle aged drivers.

5. MULTIVARIATE MODELLING

5.1 INTRODUCTION

The tables given in section 4 of this report have shown that a driver's accident liability is dependent upon a number of variables - in particular, the annual mileage travelled and the sex, age and driving experience of the driver. However, these variables are themselves interrelated - for example, younger and older drivers travel fewer miles per year, higher mileage drivers travel more on motorways and less on rural roads, and age and experience are inevitably confounded to a significant extent.

TABLE 10

Accident type (percent) by age and sex

Age Group	<18	18	19-22	23-29	30-39	40-49	50-59	60-69	>70
					Male				
Moving	55	62	71	74	82	81	79	72	74
Stationary	20	17	13	13	11	11	16	21	18
Single Vehicle	25	21	16	13	7	8	5	7	8
					Female				
Moving	57	65	73	73	77	85	78	77	71
Stationary	24	19	15	15	14	6	15	18	24
Single Vehicle	19	16	12	12	9	9	7	5	5

In the tabulations of section 4, these interactions have not been separated, and in order to determine the relative magnitudes of the effects of the individual variables a multivariate method has to be used. Such a method will result in a statistical model of accident liability which will quantify the effects of the relevant explanatory variables, and will allow the true accident liability of individual drivers to be predicted from these variables. In addition, some loss of accident data has occurred as a result of memory lapses on the part of respondents. This loss depends on the accident period and potentially on other variables as well; it will be necessary therefore to make corrections for such effects within the modelling process.

The analysis has been carried out using multiple regression in the generalised linear modelling form provided by the statistical package GLIM (Numerical Algorithms Group, 1986). The dependent variable in the analysis was the number of accidents the individual driver had reported during the relevant period (3 years for the RANDOM, OLD and INEXPERIENCED groups and a variable period for the YOUNG drivers). The explanatory variables are annual mileage (as reported by the respondent), the proportion of this mileage driven in the dark, on built-up roads, rural roads and motorways, age (taken to be the drivers age at the mid-point of the accident reporting period) and the number of years of driving experience since passing the driving test (again calculated to the mid-point of the accident reporting period). Categories in the data such as the driver's sex and socio-economic group (SEG) also feature as explanatory variables in the model. The correction for memory loss (Appendix A) is included in the model as an offset.

5.2 THE FORM OF THE MODEL

The form of the main model is:

$$A_c/T = k(s,g)\ D\ R\ M^a \exp[b_1(g)/Ag + b_2(s)/(X + b_3)] \qquad (1)$$

where:

- A_c is the *expected* number of accidents in T years, corrected for memory loss effects as described in appendix A; A_c/T is thus the accident liability expressed as an expected accident frequency (accidents per year),
- D is a 'darkness' factor of the form $(1 + b_4p_d)$ where p_d is the proportion of driving undertaken in the dark,
- R is a 'road type' factor of the form $(p_b + b_5p_r + b_6p_m)$ where p_b, p_r, and p_m are the proportion of mileage driven on built-up roads, rural roads and motorways respectively $(p_b + p_r + p_m = 1)$,
- M is the estimated annual mileage reported by respondents,
- Ag is the age of the driver in years at the mid-point of the accident period,
- X is the number of years since passing the test -also determined from the mid-point of the accident period,

and k, a and b_1 - b_6 are coefficients to be determined.

The sex of the driver (denoted by s) and his or her socio-economic group (denoted by g) only enter the equation as interactions; in particular, the coefficient of the age effect (b_1) depends on the socio-economic group of the driver and the coefficient of the experience term (b_2) depends on the sex of the driver. Although the modelling clearly required interactions between age or experience and SEG or sex to be included, the most appropriate pairing of these interactions did not emerge clearly from the statistical analysis. The pairing selected - age(b_1)/SEG and experience(b_2)/sex - is thus largely arbitrary. This issue is discussed further in Appendix B.3.4 and B.3.5. With the interactions included in the model, the constant term (k) also has to be adjusted for SEG and sex.

This model form ensures that the number of accidents is directly proportional to the period during which they are reported to have occurred. It also ensures that zero accidents are predicted for zero miles travelled, and allows accident frequency to be proportional to mileage without forcing it to be so.

5.3 THE FITTING PROCEDURE

Appendix B presents a detailed description of the modelling process. In order to fit the multiplicative model of equation (1), the dependent variable (accident numbers) is subjected to a log (logarithm to the base e) transformation within GLIM; the error distribution is assumed to be Poisson. The explanatory variables are then introduced into a model one by one, and the effectiveness of each variable or combination of variables in explaining variations in the dependent variable is assessed by applying an appropriate statistical test. The explanatory variables may be tried in a variety of functional forms - for example, a simple additive term, or a power, or reciprocal or even an algebraic function. These alternatives are compared in terms of their explanatory power in the model, and the most appropriate selected. The technique is not automatic, but is a 'trial and error' process in which the best model is determined on the grounds of statistical significance, logical meaning and simplicity.

The 'null' model is simply $A_c/T = k$, where k is the mean *expected* number of accidents per year corrected for memory loss, for the data as a whole. The mileage term was next fitted to give the model $A_c/T = k\ M^a$. At this stage it was noted that the simple power function for annual mileage was deficient in the range 3000-6000 - possibly due to some bias in the self-reported annual mileages. An alternative mileage function described in Appendix B.3.3 was devised and fitted; since this function did not significantly affect the other model coefficients and may be an artefact of the data, it is not included in the main model. The age and experience variables were then introduced one at a time, followed by the SEG and sex interactions. Finally the effects of type of road, driving other vehicles and driving in the dark, were investigated.

The number of cases available for fitting the model varied because some respondents had 'missing' values for some but not all of the variables. In particular, when SEG was added to the model, a substantial proportion of drivers of unknown SEG had to be excluded. Comparisons between observed data and 'fitted' values (the values predicted by the latest version of the model) were continually made to check the appropriateness of the various model forms - in particular to check whether a more complex form than the simplest possible could be justified.

Each time a new variable was introduced into the model, 'interaction' terms between that variable and the other variables already included were also examined. The inclusion of an interaction term in a multiplicative model implies that the effect of one variable expressed as a multiplying factor or ratio, is dependent on the value of another. Higher order interactions (between more than two variables) were not tested.

Details of the fitting process, describing the effects of each variable tried and the outcome, is given in Appendix B. In the process of exploring these relationships, the effect of memory loss and its determinants was comprehensively investigated; Appendix A presents the results.

6. THE MODELS

6.1 THE MAIN MODEL: ALL ACCIDENTS

A car driver's accident liability - ie the *expected* annual accident frequency corrected for memory loss A_c/T, can be predicted from the following equation:

$$A_c/T = 0.00633 \exp\{s+g\}\ D\ R\ M^{0.279} \exp\{b_1/Ag + b_2/(X + 2.6)\} \qquad (2)$$

where:

s and g are adjustments to the constant associated with sex and SEG, the values of which are given in conjunction with b_1 and b_2 below,

$D = (1 + 1.6p_d)$ in which p_d is the proportion of driving undertaken in the dark,

$R = (p_b + 0.65p_r + 0.88p_m)$ in which p_b, p_r, and p_m are the proportion of driving in built-up areas, rural areas and motorways respectively,

M = distance driven annually (miles)

$b_1 = 13$, $g = 0$ for drivers in SEG groups A, B and C1,

$b_1 = 23$, $g = -0.72$ for drivers in SEG groups C2, D and E,

Ag is the drivers age in years,

$b_2 = 3.5$, $s = 0$ for males,

$b_2 = 2.3$, $s = -0.02$ for females.

X is the number of years since passing the driving test.

The model is based on 4198 accidents of all types experienced by the 13519 drivers in the database for whom complete data was available. Estimates of the parameter coefficients with their standard errors and significance levels are given in appendix C. This appendix also indicates that over 80% of extra-Poisson variability in the accident data (ie that part of the variation which could potentially be attributed to the systematic component) has been explained by the above model.

Some consideration has been given to the appropriateness of number of years driving (X) as a measure of driving experience. Among the young drivers in particular, distances driven within the first year or two of driving will vary considerably; it is of interest to attempt to determine whether driving experience could be better represented by total miles driven since passing the test. The results of such an investigation are presented in appendix B.4; it suggests that total miles driven could be a better measure of experience than number of years since passing the test. However, because of the practical difficulty of obtaining accurate estimates of total miles driven, it was decided not to use this measure of experience in the model presented above.

The main model relates to all accidents. The data allows accidents to be classified as moving vehicle accidents, accidents involving a stationary or parked vehicle, and accidents involving other road users or roadside objects. It is clear that the type of accident does change with age, and that a family of models relating to the different accident categories could be generated. Such models are however, outside the scope of this report.

6.2 A SIMPLIFIED MODEL: ALL ACCIDENTS

For the purpose of illustrating age and experience effects, it is convenient to use a simplified version of the main model given in equation (2) above. This simplified model omits the effects of sex and SEG to give a model which predicts age and experience effects averaged over men and women and over the various SEG categories. The detailed exposure variables are also omitted. The model is:

$$A_c/T = 0.00212\ M^{0.38} \exp\{20/Ag + 2.5/(X + 2.2)\} \qquad (3)$$

where the notation is as before.

This model is based on 5110 accidents experienced by the 17,130 drivers for whom mileage, age and experience data was available. Estimates of the parameter coefficients with their standard errors are given in appendix D.

6.3 INJURY ACCIDENTS

It is of interest to attempt to assess whether the model fitted to all accidents - most of which in the present study will be damage only accidents - is an adequate representation of injury accidents. In order to do this, a model of the form of equation (2) - including a memory loss correction - has been fitted to those accidents which were reported to have involved personal injury. The injury model is based on 463 injury accidents experienced by 13519 drivers. Estimates of the parameter coefficients of this injury accident model with their standard errors are given in appendix D with the coefficients of the main model for comparison.

Because there are about nine times as many damage-only accidents as injury accidents, the coefficients of the injury model have not been determined very precisely, and for this reason the model is not particularly satisfactory. However, a comparison of the coefficients enables the following conclusions to be drawn:

(i) The small reduction in the mileage exponent for injury accidents is far from significant.

(ii) The age effect is rather larger and the experience effect rather smaller for injury accidents than for all accidents, but neither of these differences is statistically significant.

(iii) There are only two differences between the models which approach the 5% level of significance. First, the memory coefficients are less, indicating that injury accidents are more difficult to forget than non-injury accidents. Second, the coefficient for rural driving (relative to urban driving) is larger for injury accidents. This is consistent with a higher severity rating of accidents on rural roads compared with urban roads.

There is therefore no strong evidence to suggest that injury accidents vary with annual mileage, age and experience differently from all (mainly damage-only) accidents.

7. MODEL PREDICTIONS

This section illustrates and summarises the effects of the main variables included in the models set out in section 6 above. A detailed discussion of these effects may be found in Appendix B.3.

7.1 EXPOSURE

7.1.1 Mileage Effects

Using equation (2) above and the equivalent injury accident model in Appendix D.2, figure 2 shows how annual mileage affects the expected frequency (accident

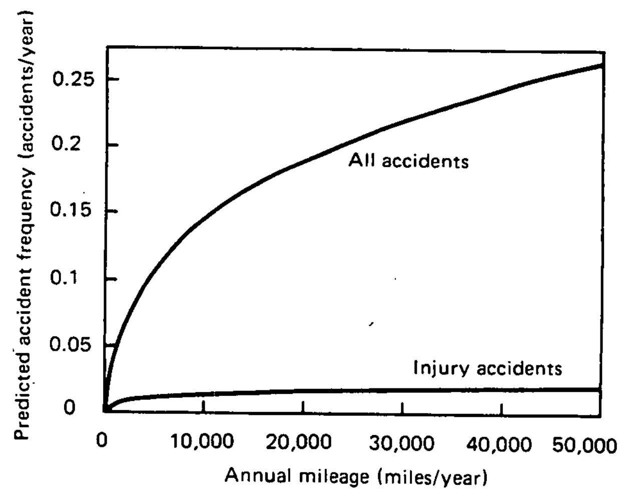

Fig.2 The predicted effect of annual mileage on accident liability

involvements per year corrected for memory loss) of all accidents and of those involving injury. The curves represent drivers aged 35 who started driving at 20 and show accident frequencies for male drivers in SEG groups A, B and C1.

7.1.2 Daylight/darkness effects

Equation (2) above includes the term D which is a linear function of the proportion of driving done in the dark. The questionnaire asked drivers to estimate the proportion of driving they do in the dark in summer and winter separately in four categories - up to a quarter, up to a half, up to threequarters and more than this. This data when included in the predictive equation in categorical form showed that driving in the dark in winter was a significant predictor of accident liability. There was no statistically significant difference in the coefficients of the darkness effect in summer or in winter, so the effect for the two seasons was assumed to be the same and the data combined. The category variables were converted to a continuous variable p_d, representing the proportion of driving done in the dark for the year as a whole.

The expression $D = 1 + 1.6p_d$ in equation (2) suggests that accidents in the dark are 2.6 times more likely to occur than accidents in daylight when all other variables included in equation 2 are properly controlled for. This does not of course mean that this increase in relative liability is due only to darkness; other factors which are associated with driving in the dark such as drinking and driving, will contribute to the increased risk of night-time driving.

There is some evidence of an age-darkness interaction such that the coefficient of p_d is larger for older drivers than for younger ones (see Appendix B.3.3). There is no evidence however for an upturn in total accident frequencies as drivers get older (see 7.2 below); it would seem therefore that any increase in the risk to older drivers of driving in the dark is compensated for by the fact (Appendix B, Fig. B3) that these drivers drive less in these conditions. The effect has not been included in the main model.

7.1.3 Type of road

Equation 2 also includes the term R which is a linear function of the proportion of driving on roads in built-up areas, on rural roads and on motorways. The expression $R = p_b + 0.65p_r + 0.88p_m$ suggests that the accident liabilities on built-up roads, rural roads and motorways are weighted in the ratio 1:0.65:0.88. This is unexpected; while it is not surprising that built-up roads have the highest accident weighting, motorways which have a low accident rate per Km would be expected to have a lower weighting than rural roads. Appendix B.3.3 includes a detailed discussion of this finding which is summarised below.

When a simple linear model of the form $A_c/T = (k_bp_b + k_rp_r + k_mp_m)M$ is fitted, the ratios of the coefficients k are 1:0.47:0.18; these values accord reasonably well with the ratios of national injury accident involvement rates per Km (STATS19) of 1:0.37:0.14. The unusual ratios obtained in equation (2) in all probability therefore, reflect the finding that accident liability is not proportional to annual mileage. The interpretation of this result is unclear. What is clear is that drivers who drive high mileages have a much lower accident rate per mile. These drivers also cover a higher proportion of their mileage on the motorway system. It is tempting to draw the conclusion that the low accident rate per mile of motorways is in part due to the fact that they are being used by the high mileage drivers who are safer per mile. It may additionally have something to do with the fact that young inexperienced drivers who have a relatively high accident rate per mile, use motorways rather less than older safer drivers.

However, the possibility that this type of road effect is an artefact of the model form needs further investigation.

7.1.4 Exposure effects - Discussion

The analysis confirms previous studies (Quimby et al, 1986, Taylor and Lockwood, 1990) in showing that accident liability is not proportional to annual mileage travelled. Including the effects of light and darkness, seasonality, and type of road in the model as part of the representation of exposure, makes little difference to the exponent of the annual mileage term. Thus the suggestion made in 4.2.1 that this lack of proportionality may be a consequence of the fact that drivers with high annual mileages drive a greater proportion of their mileage on motorways which are safer per Km, does not seem to be borne out.

The form of the mileage term implies a decreasing accident risk per Km travelled as the annual distance driven increases. This kind of effect could arise as a result of a number of mechanisms. It may be that the skills required for safe driving need regular 'practice' and that high mileage drivers are more practised; it could be that drivers who drive high mileages are different kinds of people from those who do not and the analysis has not included the appropriate 'explanatory' variables. It has recently been pointed out (Janke, 1991) that it is not unreasonable to hypothesise that drivers with a low level of competence (for whatever reason) will drive less; those that drive less will be less practised and it is unclear which way the causal link operates. The effect could even be regarded as evidence of 'risk compensation' - high mileage drivers adjust their level of risk to compensate for the higher levels of exposure.

Since a meaningful comparison of the accident 'risk' of different groups in the population depends on the way accident rates are normalised for exposure, further work aimed at obtaining an understanding of the accident-exposure relation would be very worthwhile. In the meantime, because accident frequencies are not proportional to annual mileage, between group comparisons of accident rates per mile or per Km need to be treated with caution.

7.2 AGE AND EXPERIENCE EFFECTS

Figures 3, 4 and 5 illustrate the age and experience effects (Equation 3 above) for drivers whose annual mileage is 7500, averaged over males and females and over SEG groups.

Figure 3 illustrates the sensitivity of accident liability to age by showing how predicted accident frequency would change with age, if it were possible for the driving experience of the drivers involved to remain constant at the four values shown on the figure (0, 5, 10 and 20 years). The left-hand end of the curves are truncated to represent the fact than driver cannot begin to drive before age 17 - it would not be possible, for example, for a driver with 20 years experience (curve D) to be younger than 37.

Similarly figure 4 illustrates the sensitivity of accident liability to driving experience by showing how predicted accident frequency would change with driving experience

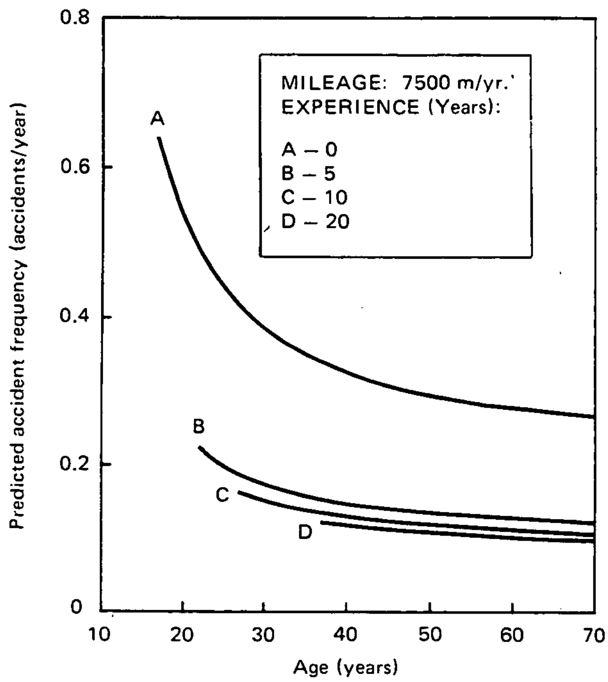

Fig.3 The predicted effect of age on accident liability

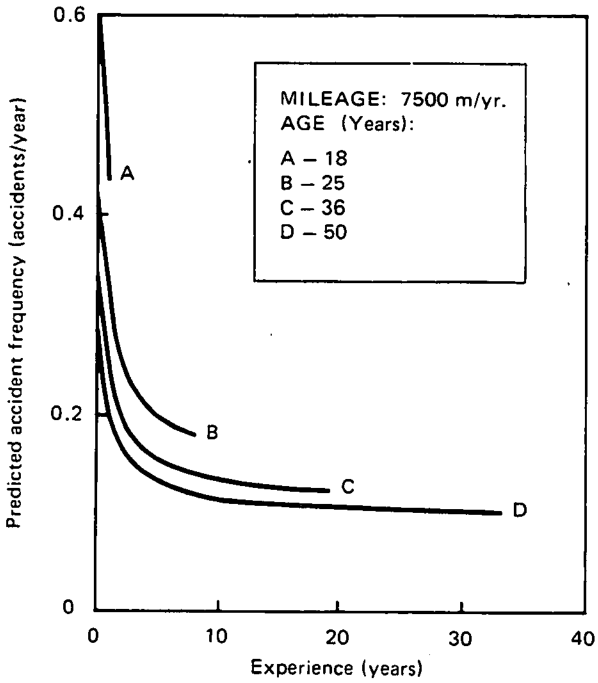

Fig.4 The predicted effect of driving experience on accident liability

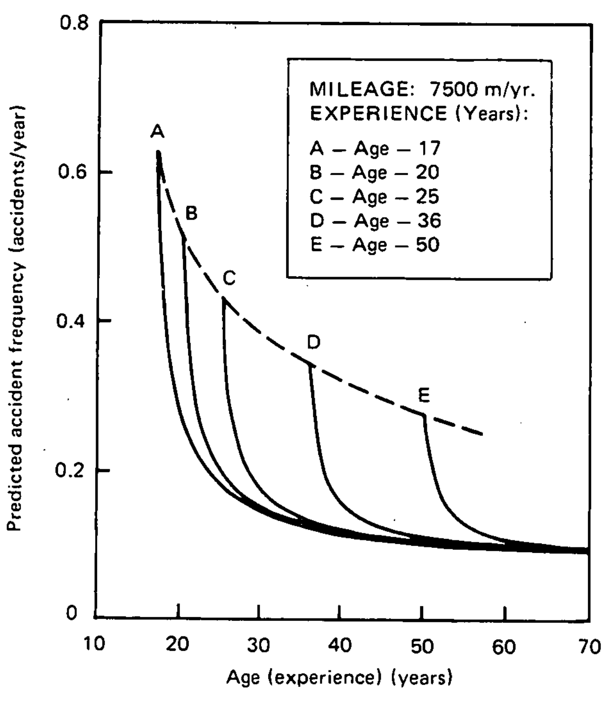

Fig.5 The predicted effect of age and driving experience on accident liability

for drivers whose ages remain constant at the four values shown on the figure (18, 25, 36 and 50 years). This time it is the right-hand sections of the curves which need truncating - for example, an 18 year old driver can only have been driving for 1 year.

In practice of course, age and experience increase together. Figure 5 shows how accident liability changes with age and experience combined, for drivers who start to drive at the ages shown on the figure (A-17, B-20, C-25, D-36 and E-50). The broken line on figure 5 is curve A in figure 3 and represents the effect of age on the accident liability of novice drivers.

It is clear that both age and experience contribute to the change in accident liability which occurs with passing time. The relative size of these effects will depend on the exact position of an individual on the age/experience continuum. Table 11 illustrates the effects for young and inexperienced drivers by presenting the percentage reduction in accident frequencies predicted by equation 3 for 1-yearly increments of age and experience.

The fist column in the table shows the effect of increasing driving experience: for example, in the first year of driving a driver's accident liability can be expected to decrease by 30% irrespective of his or her age; in the eighth year the corresponding reduction would be 3%. The second column shows the effect of age for drivers between 17 and 25. Due to age alone 17 year old drivers can expect that their accident liability will fall by 6% by the time they reach 18; the age reductions shown are independent of experience, though of course experience in these years will be limited. The third column shows the combined effects of age and experience: for example, the total reduction in accident liability due to the combined effect of a one year increase in age from 17 to 18 and the first year of driving experience will be 34%; a 24 year old who started driving at age 17 (7 years experience) could expect a 6% reduction in accident liability by the time he or she reaches 25 due to the additional year of both age and experience.

In terms of road safety remedial measures, it is obviously not possible to make direct use of the age/experience effects observed here - with the possible exception of raising the driving age. In order to explore the possibilities for the development of remedial measures related to age or experience, some understanding of the socio-psychological mechanisms underlying these effects is needed. Presumably, experience has to do with the process of learning those skills which matter for safe driving; age effects on the other hand are mediated through the process of maturation or changes in lifestyle or social factors with age. Work is in progress which it is hoped will shed some light on these issues.

In the complementary study of motorcyling (Taylor and Lockwood, 1990), motorcyclists accident liability was found to depend on the amount of car driving undertaken - experienced car drivers had fewer accidents when riding a motorcycle. In the present survey, no information was collected on motorcycle riding by car drivers, but drivers reported the amount of mileage they drove in vehicles (including motorcycles) other than the car or van included in the survey. Appendix B.3.6 shows that the effect of this additional driving experience on the car/van accident liability of the drivers involved was not statistically significant.

TABLE 11

The effect of age and experience on accident liability for young and inexperienced drivers

Percentage reduction in accident liability							
Experience Alone (Independent of Age)			Age Alone (Independent of Experience)				Age and Experience
During Year	1	30%	Between	17 and 18	6%		34%
	2	17%		18 and 19	6%		22%
	3	11%		19 and 20	5%		15%
	4	7%		20 and 21	5%		12%
	5	5%		21 and 22	4%		9%
	6	4%		22 and 23	4%		8%
	7	3%		23 and 24	4%		7%
	8	3%		24 and 25	3%		6%
OVERALL 8 years		59%	Between	17 and 25	31%		72%

7.3 SEX DIFFERENCES

Figure 6A illustrates the difference between the accident frequency experienced by men and women as a function of age, for a driver of Socio-economic group A, B or C1, who travels 7500 miles per year (close to the average mileage for all drivers), who starts to drive at 18 and continues to drive until age 70. The curves therefore combine the effects of increasing age and experience. The difference between the sexes is highlighted in Figure 6B which shows the percentage difference between male and female accident frequencies as a function of age. A similar pair of graphs would apply to SEG groups C2, D and E.

It will be seen that for young novice drivers, nominally driving the same annual mileage, women would expect to have 35% fewer accidents than male drivers covering the same mileage. This difference rapidly declines with increasing age and experience, until over the age of 30 the difference is only about 10%. This would seem to be the best estimate which can be made of the sex effect on accident involvements when mileage effects are corrected for. If this result is compared to the apparent sex differences in either the simple accident frequencies or accident rates per million miles in tables 5 and 6, it will be seen that comparisons based on either measure can be very misleading. The comparison shown in Figure 1 between men and women travelling similar distances is much more reliable.

7.4 SOCIO-ECONOMIC GROUP

Figure 7A similarly illustrates the effect of Socio-economic group by comparing how accident frequency changes with increasing age and experience for male drivers travelling 7500 miles per year belonging respectively to SEG groups A, B or C1 and groups C2, D and E. For descriptions of the SEG groups see section 4.2.3.

Figure 7B highlights the comparison by showing the percentage difference between the SEG groups as a function of age. Thus, the difference is about 15% for young and inexperienced drivers, but increases with age, until at age 70 drivers belonging to the lower SEG groups may be expected to have about 45% fewer accidents than drivers in the upper SEG groups. A similar result applies to women drivers.

8. SUMMARY AND CONCLUSIONS

1. Data has been collected from a structured sample of just over 18,500 drivers using a postal questionnaire, to determine the relationship between the accident liability of these drivers, their characteristics and driving experience. The sample was structured to have four components: for drivers aged 23 and over, a 'RANDOM' sub-sample - drivers selected at random, an 'OLD' sub-sample selected so that the combined numbers of the RANDOM+OLD sub-samples would be evenly distributed over the whole age range, and an 'INEXPERIENCED' sub-sample consisting of drivers of all ages who had less than 10 years driving experience at the time the sample was drawn: the fourth component was a 'YOUNG' sub-sample consisting of drivers less than 23 years of age. Each sub-sample included equal numbers of male and female drivers.

2. Basic tabulations of the data show the following:

(i) 'All accident' and 'injury' accident involvements fall markedly with increasing age; accident involvements also fall as driving experience - measured as the number of years since passing the test - increases.

(ii) Accident and injury frequencies do not increase in proportion to annual mileage travelled; the ratio of all reported accidents and injury accidents averages 8.8.

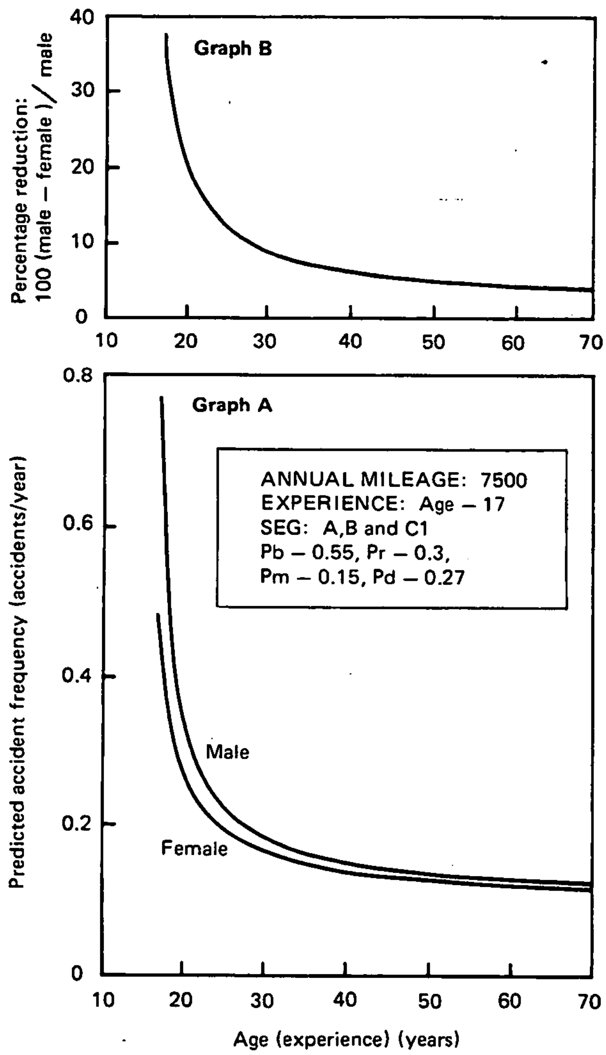

Fig.6 The predicted effect of sex on accident liability

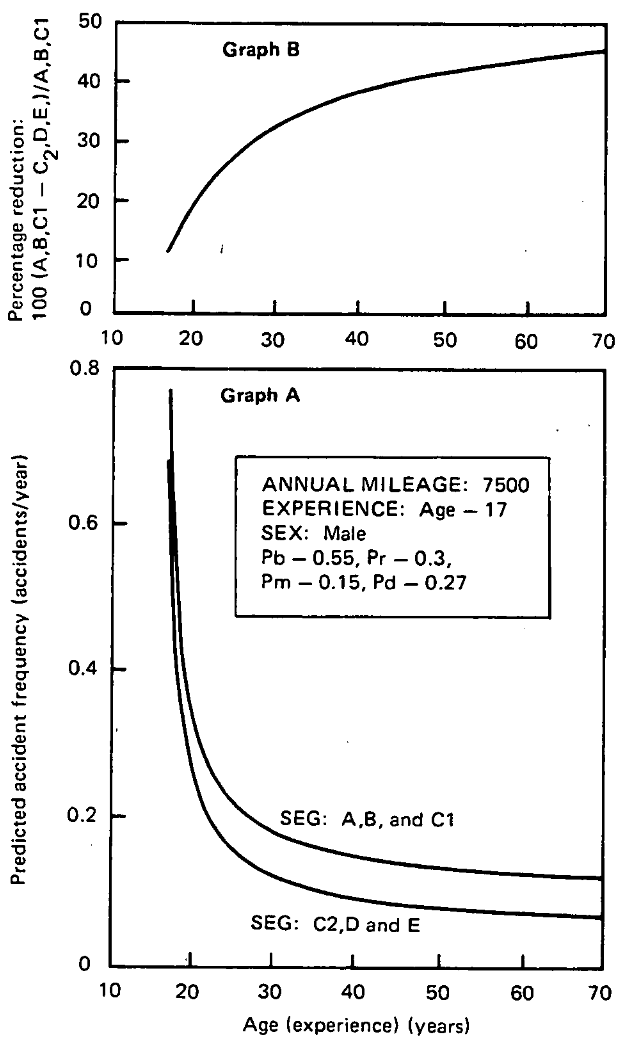

Fig.7 The predicted effect of socio-economic group on accident liability

(iii) For both men and women, about 30% of mileage is travelled on rural roads independent of the drivers total annual mileage. However, high mileage drivers use motorways to a much greater extent; about one third of the annual mileage of high mileage male drivers is on motorways.

(iv) In terms of accidents per million miles, accident rates for male drivers show a slight upturn for the oldest drivers. Whereas women have fewer accidents per year than men, they have higher rates in terms of accidents per million miles. These differences are however a consequence of the finding that accident frequencies are not proportional to annual mileage combined with the fact that women drive about half the annual mileage of their male counterparts.

(v) Young (and inexperienced) drivers of both sexes were involved in a higher proportion of single vehicle accidents than were mature drivers.

3. Generalised linear modelling techniques have been used to develop a statistical model which will predict the accident liability - the *expected* number of accidents per year corrected for memory loss (A_c/T) - for an individual driver as a function of relevant explanatory variables. In this context, accidents are all accidents on public roads including damage only accidents. The resulting relationship, with the notation of section 6.1, is:

$$A_c/T = 0.00633 \exp\{s + g\}\ D\ R\ M^{0.279} \exp\{b_1/Ag + b_2/(X + 2.6)\}$$

$$D = (1 + 1.6p_d)$$

$$R = (p_b + 0.65p_r + 0.88p_m)$$

where:

$b_1 = 13$, $g = 0$ for drivers in SEG groups 1 and 2 (A, B and C1),

$b_1 = 23$, $g = -0.72$ for drivers in SEG groups 3 and 4 (C2, D and E),

$b_2 = 3.5$, $s = 0$ for males,

$b_2 = 2.3$, $s = -0.02$ for females.

The correction for memory loss effects was achieved by modelling accidents as a function of the individual survey years - by comparing in effect, the reported accidents in the most recent year with accidents reported in the previous year and the one before that. Overall, respondents forgot about 30% of their accidents each year.

(4) The key implications arising from the analysis may be summarised as follows:

(i) Accident liability (expected accident frequencies) is dependent mainly on exposure (total annual mileage), the driver's age *and* his or her driving experience measured as the number of years since passing the test.

(ii) The proportion of driving done in the dark and on different types of road (built-up, rural and motorway) also affects accident liability, but to a lesser extent than age and experience. Interactions were found between age or experience and SEG or sex. The model has chosen to represent these as interactions between age and SEG and between experience and sex. Driving vehicles other than cars or vans had no detectable effect of the car accident liability.

(iii) Predicted accident frequencies are proportional to (annual mileage)$^{0.279}$. The reasons for the lack of direct proportionality are not known. Although accident frequencies are dependent to some extent on the proportion of mileage travelled on different types of road the type of road effect is not large, and including it in the model does not significantly increase the annual mileage exponent.

(iv) The form of the age effect means that the proportional change in liability with increasing age is larger for younger drivers than for older drivers. No upturn of accident frequency for older drivers could be detected, though there is some evidence of an age-darkness interaction such that the coefficient of p_d is larger for older drivers than for younger ones; this enhanced darkness effect for older drivers is probably compensated for by the fact that these drivers drive less in the dark.

(v) The form of the experience effect means that accident involvement falls rapidly after passing the test - ie the learning curve is steep. There is also an indication that experience could be better represented as total miles travelled rather than number of years since passing the test.

(vi) Young novice women drivers would expect to have 35% fewer accidents than male drivers covering the same annual mileage. This difference rapidly declines with increasing age and experience.

5. A model of the form given in (3) above but based on injury accidents only suggest that injury accidents vary with annual mileage, age and experience in a similar way to all (mainly damage-only) accidents. The memory loss effect is smaller for injury accidents than for all accidents; respondents forget about 18% of injury accidents each year.

9. ACKNOWLEDGEMENTS

The work described in this Report forms part of the research programme of the Road User Group (Head: Mr G Maycock) of TRRL, and is published by permission of the Director. The basic data collection was carried out by NOP Market Research Ltd.

The authors are indebted to those members of the public who took part in the survey by completing the questionnaire.

10. REFERENCES

BARWICK, J (1990) Age and Experience Factors in Accident Liability : A Review. Department of Transport, TRRL Working Paper WP/BSU/12: Transport and Road Research Laboratory, Crowthorne.

BROUGHTON, J (1990) Casualty rates among car occupants 1976-86. Department of Transport, TRRL Research Report RR244: Transport and Road Research Laboratory, Crowthorne.

BROUGHTON, J (1988) The variation of car drivers' accident risk with age. Department of Transport, TRRL Research Report RR135: Transport and Road Research Laboratory, Crowthorne.

FERDUN, G S, PECK, R.C. and COPPIN, R.S. (1967) The teen-aged driver. An evaluation of age, experience, driving exposure and driver training as they relate to driving record. Highway Research Record 163, 31-53.

HARRINGTON, D M (1972) The young driver follow-up study: an evaluation of the role of human factors in the first four years of driving. Accident Analysis and Prevention 4, 191-240.

JANKE, MARY K (1991) Accidents, mileage, and the exaggeration of risk. Accident Analysis and Prevention 23, Nos. 2/3, 183-188.

JONAH, B A (1986) Accident risk and risk-taking behaviour among young drivers. Accident Analysis and Prevention 4, 255-271.

KRITZ, L-B., and NILSSON, G (1967) Young drivers and road accidents. Official Swedish Council on Road Safety Research. Report from TRAG, Nr.100.

MAYHEW and SIMPSON (1990) New to the Road. Traffic Injury Research Foundation of Canada.

NUMERICAL ALGORITHMS GROUP (1986) The generalised linear interactive modelling system - release 3.77.

OECD (1975) Young driver accidents. Organisation for Economic Co-operation and Development Road Research Group.

PELZ, D C , and SCHUMAN, S H (1971) Are young drivers really more dangerous after controlling for exposure and experience? Journal of Safety Research 3(2), 68-79.

QUIMBY, A., MAYCOCK, G, CARTER, I D , DIXON, R and WALL, J G (1986)Perceptual abilities of accident involved drivers. Department of Transport, TRRL Research Report RR27: Transport and Road Research Laboratory, Crowthorne.

TAYLOR, M and LOCKWOOD, C R (1990) Factors affecting the accident liability of motorcyclists - A multivariate analysis of survey data. Department of Transport, TRRL Research Report RR270: Transport and Road Research Laboratory, Crowthorne.

TAYLOR, M (1990) The cost of vehicle damage resulting from road accidents. Department of Transport, TRRL Research Report RR256: Transport and Road Research Laboratory, Crowthorne.

APPENDIX A:THE EFFECT OF MEMORY LOSS ON REPORTED ACCIDENTS

A.1 BASIC ANALYSIS OF MEMORY LOSS

Drivers aged 23 or over were asked to report the accidents they had experienced during the previous 3 years and to give the date of the accident. However, respondents failed to provide a date for about one fifth of their accidents. For obvious reasons the analysis of memory loss effects in this appendix is based on dated accidents only. The appropriateness of applying the memory loss correction derived from dated accidents to the whole sample is considered in section A.5 below.

When aggregated over all drivers in the sample, it is to be expected that about the same number of accidents would occur each year - in fact rather more could be expected in the earlier years because of age/experience effects. As table A1 shows however, drivers were much more likely to report accidents that had happened recently.

This table excludes drivers under 23 years of age

Since studies of autobiographical memory (Rubin, 1976; Rubin and Baddeley, 1989) suggest that date errors tend to be normally or log-normally distributed around the true date with no systematic directional bias, the decline in the numbers of reported accidents with recall time is likely to be a memory loss effect. It is possible that some accidents from before the study period have been imported into the first reporting year (Wagenaar, 1986) but this effect is likely to be small.

Table A1 suggests that 28% of accidents are forgotten each year, and that age has no effect on the memory loss. However, a more thorough investigation of the memory loss problem than that of table A1 is desirable for two reasons. First, drivers below the age of 23 should be included in the analysis. In this case however, because age and experience have a large effect on the expected accident frequency particularly for teenagers, the memory loss effect will be confounded with these variables. Second, memory loss may vary depending on sex, SEG or other variables. For these reasons a multivariate analysis is needed to estimate the magnitude of the memory loss component and of any interactions between memory and other variables.

A.2 MULTIVARIATE ANALYSIS OF THE MEMORY EFFECT: METHOD

For every driver aged 23 or over, the number of dated accidents in each of the three annual periods (see table A1) was known from the accident dates supplied by them. The annual mileage (given only for the latest year) was assumed to be the same for each annual period. The values of age and experience for each driver varied from one year to the next in the obvious way. Each data year (three for each driver) was given a value of a variable 'MEM' corresponding to the recall period - the time from the completion of the questionnaire (November 1987 for most drivers) to the mid-point of the year for which accidents were being remembered; thus for an November '84 - October '85 data year, MEM = 2.5 yrs, and so on. Young drivers were dealt with by the same method, modified to allow for the fact that their accidents related to a variable number of years. The coefficient of MEM generated in the subsequent analysis represents the magnitude of the memory effect.

TABLE A1

Percentage of accidents reported by date of accident and age of driver

Age of driver	Date of Accident Nov84-Oct85	Nov85-Oct86	Nov86-Oct87	All dates
23-29	22%	35%	43%	100%
30-59	23%	32%	45%	100%
60+	25%	31%	44%	100%
All ages	23%	33%	44%	100%

GLIM was used to derive a model similar to that described in the main text, the difference being that the dependent variable was the number of dated accidents a driver reported during a one-year period, and MEM was the key additional explanatory variable.

A.3 RESULTS: THE FORM OF THE MEMORY LOSS EFFECT

It was found that the average memory effect was satisfactorily described by a simple exponential multiplier exp{u.MEM} where MEM is the recall period as defined in the previous section and u is the memory decay constant to be determined. A more complex term exp{a/(MEM+b)}, where a and b are constants was found to describe the effect slightly better (just significant at 5% level) but the extra complexity was not considered justified. The simple exponential form implies that on average the same percentage of accidents is forgotten each year. The percentage of accidents *remembered* each year is $100e^u$. From the initial analysis, the fitted value was:

$$u = -0.36 \pm 0.02$$

This value implies that accidents are forgotten at the rate of 30% per year in fairly close agreement with the results of the basic analysis in A1 above.

A similar analysis based on injury accidents only, gives a value $u = -0.20 \pm 0.06$, implying that injury accidents are forgotten at the rate of 18% per year. This value is significantly less than that for all accidents and significantly greater than zero (both significances at the 1% level). So not surprisingly, the more serious injury accidents are harder to forget than damage-only accidents, but nevertheless some injury accidents are forgotten.

A.4 RESULTS: THE FULL MEMORY LOSS MODEL

To explore the dependency of u on other variables, the interactions between the memory variable MEM and each of the other variables included in the model have been systematically investigated. As a result of this analysis, a car driver's annual accident frequency *based on those accidents for which dates have been reported* (A accidents in time T) can be predicted from the following equation - analogous to equation (2) in the main report:

$$A/T = 0.00359 \exp\{s + g\}\ D\ R\ M^{0.25} \exp\{b_1/Ag + b_2/(X + 2.5) + u\} \qquad [5]$$

where:

s and g are adjustments to the constant associated with sex and SEG, the values of which are given in conjunction with b_1 and b_2 below,

$D = (1 + 1.75p_d)$ in which p_d is the proportion of driving undertaken in the dark,

$R = (p_b + 0.68p_r + 0.95p_m)$ in which p_b, p_r, and p_m are the proportion of driving in built-up areas, rural areas and motor ways respectively,

M = distance driven annually (miles)

b_1 = 21, g = 0 for drivers in SEG groups 1 and 2 (A, B and C1),

= 30, g = - 0.79 for drivers in SEG groups 3 and 4 (C2, D and E),

Ag is the drivers age in years,

b_2 = 4.0, s = 0 for males,

= 2.4, s = 0.17 for females.

X is the number of years since passing the driving test.

u is the memory decay constant.

In particular, u, is given by:

$$u = u_{SEG} - 0.64(p_d - 0.27)$$

where u_{SEG} = -0.40 for drivers in SEG groups A,B,C1

-0.26 for drivers in SEG groups C2,D,E

(Note that an average value of p_d is 0.27, so that for a driver doing an average amount of driving in the dark, $u = u_{SEG}$).

All the coefficients in this model with the exception of the coefficients of age (Ag) are within two standard errors of the equivalent values given for equation (2) in the main report indicating that any differences between these model coefficients and those of the main model are unlikely to be statistically significant. The coefficients of age in the above equation are however significantly larger than those in equation (2) for reasons which will become apparent when the question of undated accidents is discussed in the next section.

Thus, memory loss was found to vary with SEG and with the proportion of driving done in the dark. Drivers in SEG groups A, B and C1 who do an average amount of driving in the dark forget 33% of their accidents each year whereas those in SEG groups C2, D and E forget only 23% each year. Drivers in SEG groups A, B and C1 who do not drive in the dark forget 20% of their accidents each year whereas those that do half their driving in the dark forget 42% each year. There was no evidence that age, experience, sex, type-of-road, or driving other vehicles, affected the memory loss constant. These results are purely empirical; possible explanations for these memory loss effects will not be attempted.

An increased memory loss effect might be expected in older drivers. This study has found no evidence for such an effect - at least as far as dated accidents are concerned. It is of course possible that there may be a genuine deterioration of memory with age which is too small to be detected in this dataset. On the basis of 95% confidence intervals, any non-significant age-memory interaction is likely to be smaller than is indicated by the following range: the minimum rate at which 20 year olds could forget accidents is 28% per year, whereas the maximum rate at which 60 year olds could forget accidents is 34% per year.

There was an interaction between mileage and memory significant at the 5% level, which implied that higher mileage drivers tended to forget more accidents. This interaction has been omitted from the above expression for u as it was not significant at the 1% level. It is worth pointing out that if the darkness/memory interaction were omitted, the mileage/memory interaction would be much stronger because of the correlation between driving high annual mileages and driving in the dark.

Tables A2, A3 and A4 show the statistical details relating to the MEM terms in equation (5) above.

Tables A2 and A3 show that the main terms and the interactions are all highly significant, so that no simpler representation of the effect of memory would be acceptable. Table A4 gives the relevant coefficients and standard errors.

Once the memory loss constant is known, a correcting factor P can be calculated for each driver to correct the number of accidents reported by individual respondents. It can be shown that for someone asked to recall any accidents over the last T years, the proportion of accidents that will be remembered is

$$P = \{\exp(uT) - 1\}/uT, \text{ so that } A_c = A_r/P,$$

TABLE A2

Significance levels of the memory variable (MEM), SEG and proportion of driving in the dark in equation (5)

Variable	Scaled deviance increase *	Degrees of freedom	Significance level
Memory (MEM)	332	3	< 0.01%
SEG	50	3	< 0.01%
Darkness	23	2	< 0.01%

* These are the scaled deviance increases that occur when the variable is removed from the final model.

TABLE A3

Significance levels of possible simplifications to equation (5)

Possible Simplification	Scaled deviance increase	Degrees of freedom	Significance level
Drop SEG/memory interaction	10.5	1	0.1%
Drop dark/memory interaction	15.3	1	0.01%
Drop SEG/age interaction	13.9	1	0.02%

TABLE A4

Coefficients and standard errors for the memory terms

Memory Effect	Model term	Fitted coefficient	Standard error
Memory loss for SEG groups A,B,C1	SEG(1).MEM	-0.40	±0.03
Memory loss for SEG groups C2,D,E	SEG(2).MEM	-0.26	± 0.03
Darkness/memory interaction	(d-0.27)*MEM	-0.64	± 0.16

where A_c is the corrected number of accidents and A_r the reported number. Note that for short time periods, P is almost equal to one. In the GLIM analysis, the memory correction is achieved by adding $\log_e P$ into the offset.

A.5 UNDATED ACCIDENTS

The memory correcting factor is derived from the model described in sections A2-A4. As noted earlier, about a fifth of the accidents had to be excluded from this model because they were undated. The following question needs now to be addressed: is it justifiable to use this memory correcting factor in the main model, which includes dated *and* undated accidents?

Using the same memory correction for both dated and undated accidents implies that the chance that the *date* of an accident will be omitted by a respondent is independent of the chance that the accident itself will be forgotten. The obvious alternative hypothesis is that respondents are more likely to forget the date of an accident if it happened a long time ago. It is clearly impossible to test these alternatives directly - for example by calculating the proportion of undated accidents in the three yearly periods - so the evidence has to be indirect.

Table A5 shows the proportion of undated accidents for the whole survey period as a function of age group. It will be seen that the proportion of undated accidents increases with age. This is in direct contrast to the results presented in table A1, which showed that for dated accidents the decay constant was roughly independent of both time and age.

TABLE A5

Undated accidents as a percentage of all accidents

'YOUNG' DRIVERS SAMPLE	
Drivers reporting on a period of at least 3 years	11%
'OLD' DRIVERS SAMPLE	
23-29 yr olds	26%
30-59 yr olds	36%
60 yr olds, or older	43%

The results of the multivariate analysis tell the same story. The simple exponential form of the decay constant implies a constant percentage memory loss effect from year to year, and the absence of an interaction between the magnitude of this memory effect and age confirms the absence of an age effect (Table A1). The larger value of the coefficients of age in equation (5) above which relate to dated accidents only, reflects the greater difficulty apparently experienced by older drivers (Table A5) in recalling or at least recording the date of the accident.

A consistent interpretation of these results would suggest therefore that older drivers are not having more difficulty than younger drivers in recalling the accidents as time passes, but they are not recalling (or at least recording) the dates so well. But the recall of dates would appear to be independent of the recall of the accidents. If this is the case, the total accident numbers (dated and undated) corrected for memory loss as suggested above, should be a sound basis for the analysis of accident liabilities as reported in the main report.

A.6 OTHER MEMORY LOSS STUDIES

In the early 70's a attempt was made to determine experimentally the effect of memory loss over a 2.4 year period using postal questionnaire and interview surveys respectively. The study related to accidents for which an insurance claim was made and compared the number of claims recorded by the insurance companies with the self-reports of subjects; 212 valid questionnaires and 193 interviews were involved.

The results reported in TRRL leaflet LF352 (1973) states that drivers missed about 15% of the accidents in the questionnaire survey, and about 33% in the interview survey. The 15% memory loss for the postal questionnaire is considerably less than that found for 'all accidents' in the present survey, but a part of this difference is undoubtedly due to the fact that accidents involving insurance claims are the more serious accidents and because they generate more post-accident activity - particularly filling in claim forms - they are more readily remembered.

LF 352 attributes the better recall performance of the postal questionnaire to the fact that the process is more anonymous than the face to face interview and it allows the respondent more time to check accident details and to think back over a period of time.

A.7 REFERENCES

TRANSPORT AND ROAD RESEARCH LABORATORY. Validation of self reporting of accidents. LF 352. Crowthorne 1973 (Transport and Road Research Laboratory).

RUBIN, D. C. (1976). Frequency of occurrence as a psychological continuum: Weber's fraction, Ekman's fraction, range effects, and the phi-gamma hypothesis. Perception and Psychophysics, 20, 327-330.

RUBIN, D. C. and BADDELEY, A. D. (1989). Telescoping is not time compression: a model of the dating of autobiographical events. Memory and Cognition, 17, 653-661.

WAGENAAR, W. A. (1986). My memory: a study of autobiographical memory over six years. Cognitive Psychology, 18, 225-252.

APPENDIX B: THE MODELLING PROCESS

B.1 THE STATISTICAL MODEL

The statistical model that forms the basis of the present work consists of three components:

(i) the 'systematic' component (the relationship between the dependent variable and a number of explanatory variables - given by equation 1 in the main text),

(ii) the 'sampling' or 'measurement' error associated with the dependent variable, and

(iii) residual error due to the lack of fit of the model. This component may be due to the use of incorrect functional forms for those terms included in the systematic component, or it may arise from the omission of explanatory variables (for example, behavioural characteristics of the drivers) or both.

Consider first of all the nature of the error component in (ii). It is reasonable to assume that accidents are random events generated at a rate given by the underlying accident liability represented by the systematic component of the model. This means that if the model fitted perfectly, predicting exactly the expected number of accidents a driver would have in a year, the actual occurrence of accidents would be represented by a Poisson process whose mean value was given by the predicted liability. For example, a driver with an expected accident frequency of 0.12 accidents per year (the uncorrected mean frequency for all drivers in the present survey) would in any year have a probability of 0.87 of having an accident free year, a probability of 0.11 of having one accident, a probability of 0.006 of having 2 accidents, and so on, according to the Poisson distribution. In this connection it is important to note that in the present analysis it is the *reported* number of accidents which is assumed to have a Poisson error structure. The memory correction is incorporated into the predictive equation as an offset - a case by case multiplier estimated as described in Appendix A. The analysis does not use corrected numbers of accidents as a dependent variable; only the predicted accident liabilities are corrected.

It is also possible to make assumptions about the form of the unexplained variation in (iii). A very convenient form for the distribution of the component is the Gamma distribution. This, when combined with the Poisson process assumed for (ii) above enables the accident data to be treated as Negative Binomial data, and the parameter of the underlying Gamma distribution which provides the best overall fit to the data estimated from the residual. This is of rather academic interest in the present study, and has not been attempted. The assumed error structure does however have implications for significance testing and overall goodness of fit which are described in B.2 below.

As regards the systematic component of the model ((i) above), the multiplicative form shown in equation (1) - bearing in mind that exp(a + b + c) is equivalent to exp(a).exp(b).exp(c) - has proved to be extremely robust over a range of studies. Not only has this form yielded superior data fits compared to alternatives, but the logarithmic transformation used in GLIM to fit the models is statistically the most appropriate one to use with Poisson data (McCullagh and Nelder, 1989). Once the most suitable form of the systematic component had been decided, much of the work in incorporating explanatory variables into this model is concerned with achieving the most appropriate functional form of these variables. The process is described briefly in section 5.3 of the main report and in greater detail below.

B.2 SIGNIFICANCE TESTING

The systematic component of the model is determined principally by establishing which variables make a significant contribution to explaining between-driver variation in accident frequency. Variables are introduced one at a time (exploring various functional forms in which they might be included) starting from the 'null' model $A_c/T = k$. At each stage the 'best-fit' model is selected. Note: The null model corresponds to the estimation of the mean accident frequency over all drivers and this is computed in GLIM as $\Sigma A/\Sigma T$. It is for this reason that the mean accident frequencies presented in Section 4.2 were also computed in this way.

The statistic calculated by GLIM which forms the basis for testing the significance of adding terms to the model is the 'scaled deviance'. This is a term analogous to the residual sum-of-squares in Normal regressions. For Poisson errors it is a maximum likelihood ratio statistic. Provided that the predicted mean value of the dependent variable is greater than about 0.5 the scaled deviance (with Poisson errors) is asymptotically distributed as a chi-squared variable with (n-p-1) degrees of freedom (where n is the number of data points and p the number of independent variables fitted). However, in the present study, the dependent variable has a mean value below 0.3. This has the effect of reducing the expected value of scaled deviance below that of a chi-squared variable and means that scaled deviance cannot be used to test the overall goodness-of-fit of the model. Maycock (1988) has shown however, that in these circumstances the generalised Pearson chi-squared statistic also calculated by GLIM as:

$$X^2 = \Sigma\ (\text{observed value-fitted value})^2/\text{variance function})$$

$$= \Sigma \frac{(y-\mu)^2}{\mu} \text{ for poisson errors,}$$

continues to be (asymptotically) reasonably distributed like chi-square with an expected value of 1 per degree of freedom. Thus *generalised chi-square* (X^2) was used in the present study as a measure of the overall goodness-of-fit.

The significance of adding each extra term to a model is usually assessed by looking at the difference between the scaled deviance for two nested models. If the models have df_1 and df_2 degrees of freedom respectively then the deviance difference is (asymptotically) distribution as a chi-squared variable with (df_1-df_2) degrees of freedom. Maycock (1988) demonstrated that this property holds for data means as low as 0.02 (for Poisson data) and in this study *scaled deviance difference* was thus used to test the significance of adding individual terms to the model.

Variables may be introduced into the model as continuous variables or as multi-level factors (variables with values in two or more categories). In the case of factors, the above test using deviance difference indicates the significance of the inclusion of the factor as a whole. If the models are fitted using GLIM and the assumption made that the error structure is Poisson, the standard errors calculated by GLIM assume a perfectly fitting model; component (iii) in the model structure (B.1 above) is ignored. To correct the calculated standard errors to allow for this 'extra-Poisson' component it is necessary to multiply them by the factor $\sqrt{X^2/df}$ before applying the usual t-test.

B.3 INVESTIGATING ALTERNATIVE MODEL FORMS

B.3.1 Introduction

Having accepted that the multiplicative form (Equation 1) for the systematic component of the statistical model is appropriate, the subsequent modelling process consists of determining which terms should be included in the model, and deciding the most appropriate functional form for these terms.

Because of the large number of variables and interactions which are potential candidates for inclusion in a statistical model of accident liability, it was decided to use the 1% level of significance as the criterion for the selection of model terms to reduce the risk of including inappropriate terms by chance.

The most appropriate functional form for the main terms was determined by trial and error; goodness of fit was judged on the basis of minimising the scaled deviance and on visual examination of a plot of residuals. In traditional multiple regression, a residual plot consisting of individual data points can be used. This kind of plot is unsuitable for accident data in which a high proportion of the observed number of accidents is zero; grouped data has to be used. In the present case, individual driver data was divided into about a dozen age groups. For each group, the 'observed' total number of accidents (O) is compared with the model prediction (E) by means of the statistic X - the square-root of the value of chi-square:

$$X = (O-E)/\sqrt{E}$$

X is plotted as a group residual against the mean age for the group. If the model is a good fit and the data has Poisson errors, the magnitude of X will be typically 1 and will exceed 2.5 in about 1% of cases.

With two exceptions, interactions between all possible pairs of variables were considered (interactions with the memory loss term are discussed in Appendix A); 34 possible pairs were tested. To avoid spurious effects, as indicated above, the 1% level of significance was used as a criterion for inclusion in the main model. However for completeness, a number of interactions significant only at the 5% level are recorded below.

B.3.2 Age and experience

At the outset it was well known that inexperienced young drivers had a far higher accident liability than experienced older drivers. The analysis needed to answer four questions:

(i) Are age and experience *both* relevant to the prediction of accident liability?

(ii) How should age be included in the model? Simple exponential forms for the age effect would predict a monotonically falling accident liability with increasing age. In contrast, analysis of national accident data (accident rates per Km travelled) suggests that accidents rates rise after age 65. In addition to determining the most appropriate form for this term, it was important therefore to check whether a rising curve for older drivers could be justified.

(iii) How should experience be included in the model? The important issue here is to ensure that the functional form chosen properly reflects the change of accident liability with increasing driving experience; does a driver's performance improve *steadily* as he gains experience, or does his performance improve rapidly in the first few months after passing the test but does not improve much thereafter?

(iv) Should the model contain an interaction between age and experience, or between age and mileage, or between experience and mileage?

The results of the analysis showed that:

(i) Both age and experience are separately very important, as can be seen from the table of significance levels in table C1, appendix C.

(ii) *Age.* There was no evidence from this study that accident liability (as defined in this report) increased for drivers over 65 years of age. However, a simple negative exponential age term proved not to be as good as the reciprocal form used. This reciprocal form means that the rate of decline of accident liability with age is not a simple multiplier, but falls more rapidly for young drivers than for older drivers.

Within this overall liability-age relation, there is some evidence that driving in the dark increases the accident liability of older drivers - see section B 3.3 (Darkness) below.

It is worth recalling that in the present paper accident liability relates largely to damage only accidents. There is strong evidence from national injury data, that *casualty* rates increase considerably with age presumably because older drivers are more vulnerable to injury.

(iii) *Experience.* As with age, a simple negative exponential representation of the experience effect proved to be inadequate. Driver accident involvement falls much more rapidly during the initial period after passing the test than the negative exponential form would predict; this is demonstrated in table B1, which is based on two pilot models which included only mileage, age and experience.

The table shows the incremental reductions in accident liability predicted by the two alternative models after 1 year of driving, a further 5 years, a further 10 years and a further and 30 years of driving experience. For example, according to the simple model a driver would be expected to have 5% fewer accidents if he has been driving for 5 years than he had when he had been driving for only a year, and so on.

TABLE B1

Reductions in liability caused by increments of experience, according to two different models of experience

Increment of experience (Years)	Simple model exp(ax)	Complex model exp(a/(x+b))
0 - 1	1.4%	30%
1 - 5	5%	35%
5 - 10	7%	13%
10-30	24%	12%

It is clear that the simple exponential function dramatically underestimates the learning effect in the early years.

(iv) **Interactions.** The final model was not improved by adding an age-experience or an age-mileage interaction. There was however an experience-mileage interaction which was significant at the 5% level. This interaction is related to the question of whether experience should be measured in years, or by the total number of miles a driver has driven since passing the test. This is in effect an alternative definition of experience and is examined more fully in B 4 below.

B.3.3 Exposure (mileage, type of road and darkness)

Mileage

In the main model, accident liability varies with $M^{0.279}$, where M is the annual mileage travelled. This expression is however a simplified representation of the actual data and overpredicts the liability of drivers in the 3000 to 6000 miles/year range. A better fitting model is produced if $M^{0.279}$ in the main model is replaced by:

$$M^{0.264}\exp\left(\frac{-7.0 \times 10^5}{(M-5100)^2+1500^2}\right)$$

The use of this rather artificial function to represent the annual mileage effect is a highly significant improvement in statistical terms; it produces a deviance decrease of 31.1 for the loss of 3 degrees of freedom (< 0.01%).

Fig B1 compares the above complex expression with the simple power function used in the main model. The main difference is the 'valley' in the region of 3-6000 miles. This gives the curious prediction that a driver doing 5000 miles/year is less likely to have an accident than one doing 3000 miles/year. The cause of this anomaly in unknown, but it may be associated with the way the annual mileages are reported by respondents.

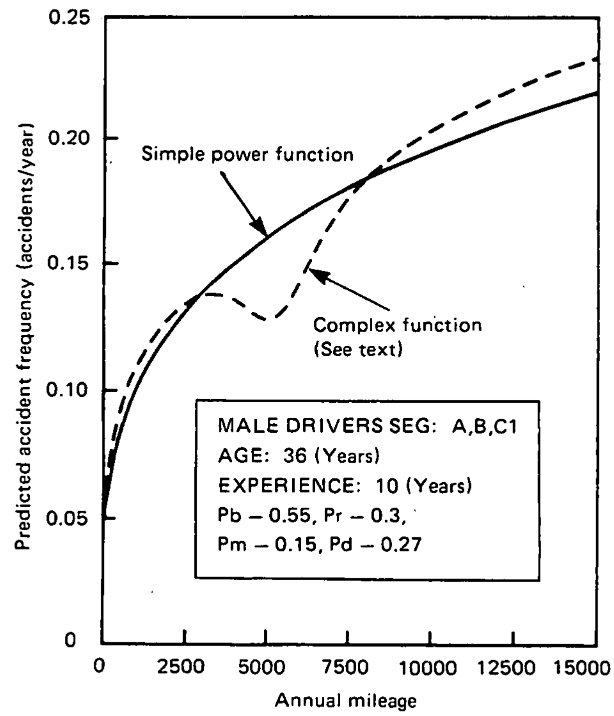

Fig.B1 Comparing a simple power function for annual mileage with a more complex function

The use of this more complex function in the main model is accompanied by little change in the coefficients of age, experience or any of the other variables. Accordingly for simplicity it has not been included in the main model.

Using the simplified mileage term, no interaction between mileage and other variables was significant at the 1% level. However, there were a number of interactions significant at the 5% level. These were between mileage and:-

Experience (see B.3.2 and B,4 below)

Memory (see appendix A).

SEG - the sense of the interaction is that the exponent of M for SEG groups A, B and C1 is 0.05 larger than for SEG groups C1, D and E.

The proportion of driving done in the dark (p_d) - the sense of the interaction was that drivers who do a high proportion of their mileage in the dark, have a lower exponent of M than those who do not.

Type of Road

Common sense suggests that accident liability should be proportional to annual mileage travelled. However, the modelling presented here and elsewhere, shows this not to be so - accident liability does not increase with increased annual mileage as much as would be predicted by a proportional relationship. In models in which the proportion of annual mileage by type of road has not been taken into account, a possible explanation for such an effect would be that drivers who drive long distances do so on safer roads. If this is the full explanation, the following model would have proved to be a good fit to the data:

$$A/T = (b_b p_b + b_r p_r + b_m p_m)M \qquad \text{Model M1}$$

where p_b is the proportion of driving on built-up roads,

p_r is the proportion of driving on rural roads,

p_m is the proportion of driving on motorways,

M is the annual mileage in Kms

b_b, b_r & b_m are factors to be estimated.

In fact, the following model is a far better fit:

$$A/T = (b_b p_b + b_r p_r + b_m p_m)M^a \qquad \text{Model M2}$$

where a is the exponent of mileage to be estimated.

Moreover, it turns out that the value of a is almost the same whether the proportion of driving by type-of-road is included in the model or not. It would seem to follow therefore, that the fact that drivers who drive high mileages tend to do a higher proportion of their mileage on motorways (see Tables 3 and 4 in the main text) is not even a partial explanation of the lack of proportionality between accident frequency and annual mileage.

Though M1 is not a particularly good statistical fit to the data, it is worth recording the estimates obtained by fitting this model since the coefficients will give estimates of accident rates per Km travelled. Table B2 gives the figures for all accidents and for injury accidents. The coefficients for injury accidents are subject to large errors but they can be compared with data reported in Road Accidents Great Britain (RAGB).

(These estimates of errors do not take account of the errors in the independent variables.)

Although there are differences between the two sets of injury figures, both indicate that motorways are safer per Km travelled than rural roads which in turn are safer than built-up roads. Incidentally, these figures also suggest that both the self-reported accident data and the data on proportion of mileage by type of road is not inconsistent with national accident data.

Table B3 gives the corresponding figures for the better fitting model M2, again fitted to all accidents and injury accidents. In this case RAGB does not contain data that can be used for comparison.

This gives a very different picture, all three types of road having much more nearly equal weights.

The interpretation of this result is unclear. What is clear is that drivers who drive high mileages have a much lower accident rate per mile (Km). Table 3 and 4 in the main text show that these drivers also cover a higher proportion of their mileage on the motorway system; figure B2 supplements this by showing the distribution of the proportion of driving by road type for all drivers. It is tempting to draw the conclusion that the low accident rate per Km of motorways has something to do with the fact that they are being used by the high mileage drivers who are safer per Km. It may additionally have something to do with the fact that young inexperienced drivers who have a relatively high accident rate per Km use motorways rather less than older safer drivers.

However, the possibility that this type of road effect is an artefact of the model form needs further investigation. Further light might have been shed on this matter had it been possible to disaggregate the accidents by type of road, and to fit separate power function models to the three road types.

Proportion of driving in the dark

The questionnaire asked drivers to estimate the proportion of driving they do in the dark in summer and winter separately in four categories - up to a quarter, up to a half, up to threequarters and more than this. This data

TABLE B2

Accident involvements per 100 million vehicle kilometres, for cars, on different types of road

	Model M1 (estimates of b_b, b_r & b_m) All Accidents	Model M1 Injury Accidents	RAGB 1986 Table 41 Injury Accidents
On built-up roads	2370 ±60	190 ±30	192
On rural roads	1120 ±70	140 ±40	71
On motorways	420 ±70	40 ±40	26

(These estimates of errors do not take account of the errors in the independent variables.)

TABLE B3

Type of road coefficients, using model M2

	All Accidents	Injury Accidents
Units:	$(\text{miles})^{-0.41}$	$(\text{miles})^{-0.3}$
Estimate of b_b (built up roads)	0.00669	0.00135
Estimate of b_r (rural roads)	0.0037	0.00124
Estimate of b_m (motorways)	0.0046	0.0014

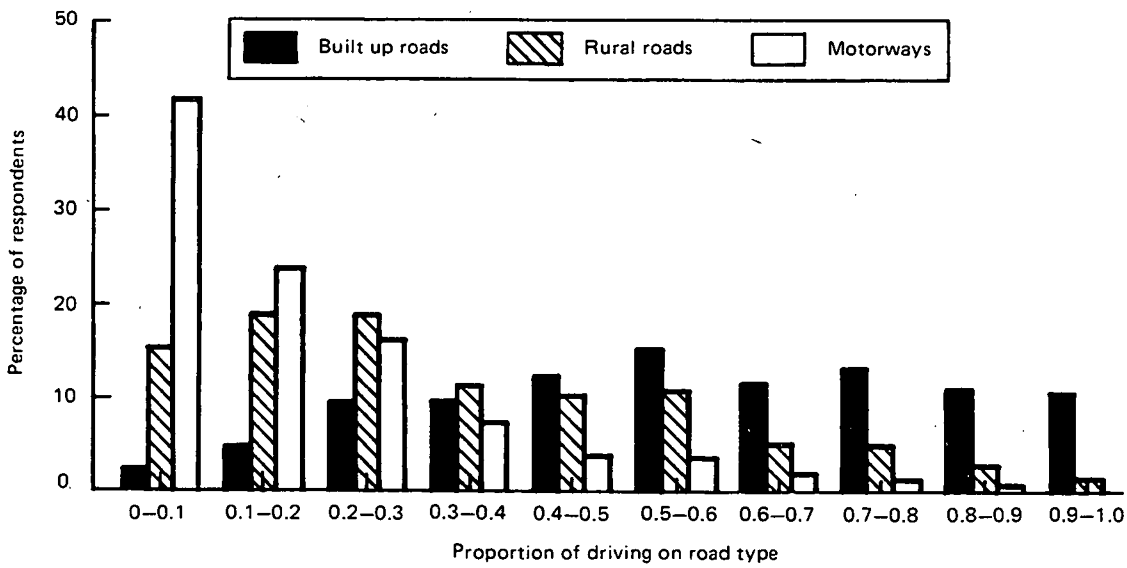

Fig.B2 Distribution of driving on various types of road

when included in the predictive equation in categorical form showed that driving in the dark in winter was a significant predictor of accident liability. There was no statistically significant difference in the coefficients of the darkness effect in summer or in winter, so the effect for the two seasons was assumed to be the same and the data combined.

The category variables were converted to a continuous variable p_d, representing the proportion of driving done in the dark for the year as a whole and the final model in the main report includes this term. Figure B3 shows how the distribution of p_d varies with age group. Not surprisingly, drivers under 20 years of age report that they undertake the highest proportion of their travel in the dark (an average of about one third) with respondents driving less in the dark as they get older.

As regards interaction terms, appendix A on memory, shows that drivers who do a lot of driving in the dark are more likely to forget accidents. No other interaction with the proportion of driving done in the dark proved significant at the 1% level. It is perhaps surprising given the differences between lighting standards in built-up areas and on rural roads, that no interaction between darkness and type-of-road was found even at the 5% level of significance. However, there were two interactions significant at this level. One was between mileage & darkness - an effect mentioned above in the section on mileage - and the other was the interaction between darkness & age mentioned in B 3.2 above.

It is reasonable to hypothesise that deteriorating visual performance with age , particularly at night, might make the older driver specially prone to accidents in the dark. To test this, a term was tried of the form:

$\exp\{d_w[a_1 + a_2(Ag - 50)]\}$ for drivers over 50

$\exp\{d_w a_1\}$ for drivers under 50

d_w is a variable representing amount of travel in the dark in winter; a_1 represents the age-independent effect and a_2 the age-dependent effect; 50 was an arbitrary age cut-off which was not optimised. a_2 would be significant in statistical terms only if there was an increasing effect of driving in the dark for the over 50's. The values of a_1 and a_2 with their standard errors were respectively, 0.14 + 0.02, and 0.009 + 0.003.

This result suggests that the effect of driving in the dark is a significant determinant of accident liability and that its

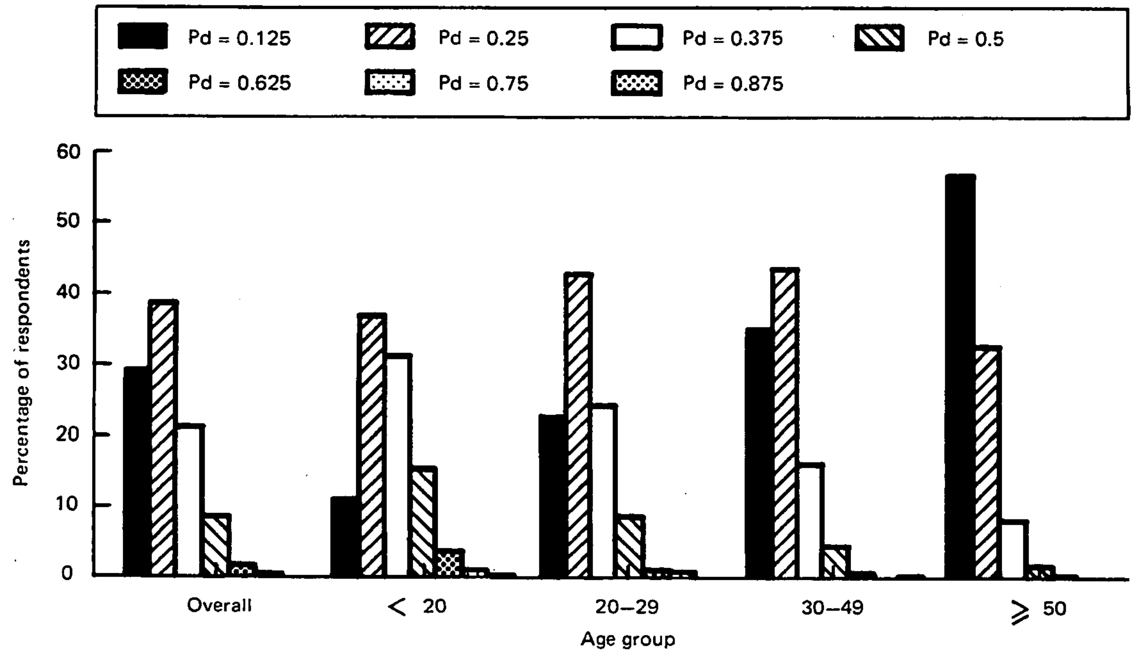

Fig.B3 Distribution of Pd by age group (Pd – Proportion of driving in the dark)

magnitude increases with age for the over 50's. The magnitude of the interaction is such that a driver aged 65 is affected by dark very roughly twice as strongly as a driver under 50. The data presented in Figure B3 on the changing pattern of driving in the dark with age may simply reflect changing patterns of lifestyle, but as far as the older age group is concerned probably also includes some deliberate reduction in dark travel to compensate for declining physical abilities. Since however, there is no evidence for an upturn in total accident involvement frequencies as drivers get older (3.2 above), it would seem that any increase in the risk to older drivers of driving in the dark is compensated for by the fact that these drivers drive less in these conditions.

B.3.4 Sex

It was found that there was *either* a strong experience-sex interaction, *or* a strong age-sex interaction, but not both. A model with an experience-sex interaction was a slightly better fit than a model with an age-sex interaction, but the difference was not significant. The former was selected for inclusion in the main model, but it must be realised that this was a fairly arbitrary choice.

B.3.5 Socio economic group (SEG)

From the questionnaire responses, it had been possible to classify drivers into four socio-economic groups. However, at an early stage in the analysis it was shown that the full SEG classification could not be justified statistically in the analysis; a simplified SEG classification with only two groups was therefore adopted.

As in the case of the sex category, there was *either* a strong experience-SEG interaction *or* a strong age-SEG interaction, but not both. A choice therefore had to be made on the limited evidence available. In this case, the age-SEG interaction was included in the main model, but again it must be remembered that this choice is somewhat arbitrary.

There was also an interaction between SEG and memory (see appendix A). There were no other interactions significant at the 1% level, but there was an interaction between mileage and SEG significant at the 5% level (see section B.3.3).

B.3.6 Driving other vehicles

About 1 in 7 drivers also drove vehicles (including motorcycles) other than the car/van mileage reported on in the questionnaire. Only about 15% of these covered higher mileages in the other vehicles than they did in the car/van; about half did under 10% of their total mileage in other vehicles. Experience of driving these other vehicles could have affected the respondent's accident liability when driving the car. A number of alternative ways of representing this extra driving experience in the accident liability model were examined; none of those tried was significant at the 1% level.

The maximum size of this non-significant effect based on the 95% confidence limits are as follows: driving another vehicle decreases car accident liability by at most 10%; driving another vehicle at least as far as you drive a car, decreases car accident liability by at most 30%. (Note: these figures apply to drivers who do not change their car mileage when they start driving another vehicle).

There were no significant interactions involving other-vehicle-driving, either at the 1% or the 5% level.

B.4 SHOULD EXPERIENCE BE MEASURED IN YEARS, OR BY TOTAL MILES DRIVEN SINCE PASSING THE TEST?

B.4.1 Modifying the experience term

A model was fitted in which X (experience in years) in the final model (equation (2)) was replaced by X_m, an estimate of the total number of miles the driver had driven since passing the test. X_m was estimated by multiplying the driver's current mileage by the number of years he had been driving. Despite the fact that this is a very crude estimate of X_m the resulting model was at least as good a fit as the main model - the scaled deviance of the model using X_m was in fact 0.2 smaller than the main model, a non-significant improvement. In B.3.3 above it was pointed out that a more complex mileage function represented the actual mileage data more effectively. If a comparison between the effectiveness of X and X_m is made using a model containing this complex mileage function the model containing X_m has a scaled deviance 4.1 lower than the model containing X. Presumably a model using accurate data for X_m would be a still better fit. This result suggests that X_m is a better representation of experience than X.

This evidence therefore indicates that total-miles-driven is at least as good, and probably a better measure of experience than the number of years since passing the driving test.

The revised experience term is:

$$\exp\left(\frac{b_{sex}}{X_m + b_m}\right)$$

where $b_m = 6000 \pm 2000$ mile

$b_{SEX} = 8000 \pm 900$ mile for males

$b_{SEX} = 4200 \pm 800$ mile for females

This is plotted in Figure B4 and tabulated incrementally in table B4. Table B4 shows for example, that a male driver's accident liability is 20% less when he has driven 10,000 miles than it was when he had driven 5000 miles. The changes are smaller for a woman driver, since it has been assumed that the experience term interacts with sex - see the warning in section B 3.4.

Substituting X_m for X in the main model causes two significant changes in the other model coefficients. First, the age effect becomes substantially larger. Second, and not surprisingly since mileage now features in the revised experience term, the mileage exponent increases - liability now being proportional to $M^{0.43}$. These changes suggest a further possible reason why accident liability is not proportional to annual mileage: drivers who do high mileages may gain experience (and become safer) more quickly which partly compensates for their high exposure. This would affect inexperienced drivers more than

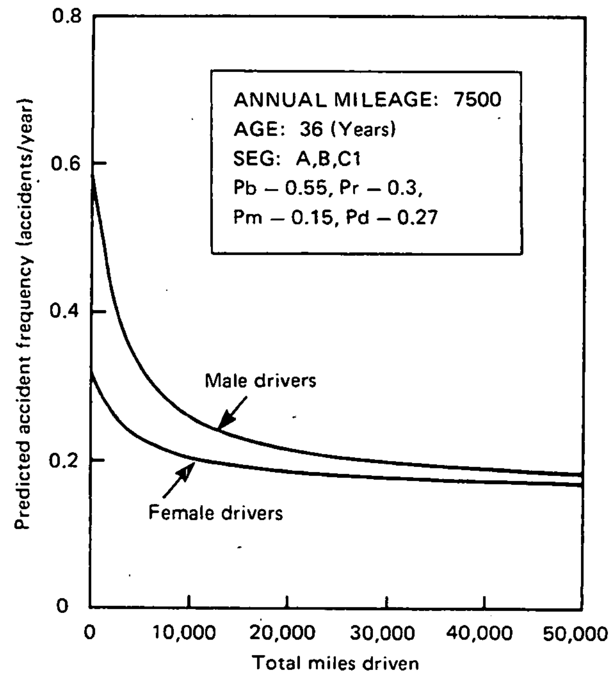

Fig.B4 The predicted effect of driving experience expressed as total miles driven since passing the driving test

experienced ones, and therefore there ought to be a mileage-experience interaction in the main model. Such an interaction was indeed found, significant at the 5% level, but it was omitted from the main model as it was not significant at the 1% level.

B.4.2 Implications for future surveys

The results given above suggest that total-miles-driven since passing the driving test might be a more appropriate measure of experience than years-since-passing-the-test. Clearly, the latter is much easier to determine accurately, whilst total-miles-driven will inevitably be associated with considerable error. The question arises: is it likely that measures of total-miles-driven can ever be obtained with sufficient accuracy to be useful in the modelling context? Two points can be made.

First, the total-miles-driven data used as described in section B 4.1 was quite crude, yet it gave as good a fit as the data relating to years-of-driving. This suggests that, in practice, the errors in total-miles-driven will be tolerable.

Second, for experienced drivers accident liability is insensitive to total mileage driven. For example, suppose a male driver claims he has driven 50,000 miles altogether, but that the true figure is 75,000 miles. If the inaccurate estimate is used to predict his accident liability, then it will be overestimated by only 4% (see table B4) - that is to say a 50% error in the estimate of miles has caused an error of only 4% in the accident liability. Hence when planning a study to relate driver

TABLE B4

Reductions in liability caused by increments in total-miles-driven

Increments in total miles driven since passing test	Reduction in accident liability MALES	FEMALES
0 - 5,000	45%	27%
5,000 - 10,000	20%	11%
10,000 - 15,000	11%	6%
15,000 - 20,000	7%	4%
20,000 - 25,000	5%	3%
25,000 - 50,000	11%	6%
50,000 - 75,000	4%	2%
75,000 - 100,000	2%	1%
100,000 - ∞	7%	4%

safety to total-miles-driven, it is only necessary to obtain approximate data from the more experienced drivers, and concentrate resources into obtaining accurate data from the less experienced drivers.

B.5 REFERENCES

MAYCOCK, G (1988) Significance tests in generalised linear models with sparse Poisson or Negative Binomial data sets. *Department of Transport TRRL Working Paper WP(BSU)8*: Transport and Road Research Laboratory, Crowthorne. (Unpublished - available on personal request only).

McCULLAGH, P and NELDER, J A (1989) Generalised linear models (2nd Edition). *Monographs on Statistics and Applied Probability, 37*: Chapman and Hall.

APPENDIX C: MODEL DETAILS: THE MAIN MODEL

C.1 COEFFICIENTS AND ERRORS

Table C1 lists the estimates of the parameter coefficients with their standard errors for the main model (equation (2) in the main text); the errors have been scaled up by a factor of 1.057 to allow for extra-Poisson variation as described in Appendix B.

A number of points can be made about the error estimates in table C1; most are of technical interest only.

The error of both the SEG and Sex (Constant corrections) terms are not meaningful and are not therefore given in the Table. In effect these terms are adjustments to the constant and take the values they do to ensure that the model prediction passes through the mean of the observed data sub-sets.

Because darkness interacts with memory, the estimate of the effect of darkness depends on the memory effect. As the memory loss coefficient is also subject to error, the error of the darkness term is higher than it would be if the memory correction was assumed to be error free; the error value in table C1 takes account of this effect.

The GLIM system itself only produces estimates of the log-linear coefficients; non-linear coefficients have to be estimated iteratively by determining that value which minimises the deviance. An error for each non-linear coeﬃcient (ie darkness, rural, motorway, and the parameter in the denominator of the experience term) was estimated by finding the value of the coefficient which gave a deviance 1 greater than the minimum.

C.2 SIGNIFICANCE LEVELS

Table C2 shows significance levels in terms of the increase in scaled deviance caused by omitting each variable in turn from the final model (equation (2)),

TABLE C1

Estimates and parameters for the main model

Effect	Estimated coefficient	Standard error
Mileage	0.279	0.016
Age (for drivers in SEGs A/B/C1)	13	3
Age (for drivers in SEGs C2/D/E)	23	3
Experience (for males)	3.5	0.4
Experience (for females)	2.3	0.4
(Parameter in the denominator of the experience term)	2.6	0.6
Darkness	1.6	0.5
Rural driving (relative to urban)	0.65	0.05
Motorway driving (relative to urban)	0.88	0.09
Constant (Ln)	-5.06	
SEG (Constant correction)	-0.72	
Sex (Constant correction)	-0.02	

without changing the memory correcting factor. In the case of SEG and darkness, this procedure is inadequate because the memory correction factor depends on these two variables; the model of equation (5) which takes memory into account explicitly (see appendix A) suggests that SEG and darkness are not quite as significant as table C2 suggests, but nevertheless both are still highly significant.

Table C2 confirms that all the variables in the table can be justified at the 0.01% level or better, but does not show if the variables have been included in the final model in the most appropriate form. Table C3 considers ways of simplifying the model without dropping any variable completely; it is seen that none of the simplifications can be justified.

C.3 VARIABILITY EXPLAINED BY THE MODEL

The final model is based on data from 13,519 respondents. Table C4 shows the initial and final values of both scaled deviance and generalised chi-square (see Appendix B) for two initial 'null' models - one without memory correction and the other with full memory correction.

It will be seen that the scaled deviance of the main model is considerably smaller than the number of degrees of freedom for reasons given in Appendix B.

An indication of the 'amount of variation' explained by the systematic component of the model cannot be obtained from the scaled deviance but can be estimated using the values of X^2. The Poisson component of the data will contribute to X^2 an amount equal to the number of degrees of freedom; in terms of the model postulated in this analysis, this variation cannot be eliminated. The variation which can potentially be explained by the systematic component of the model is thus the values of X^2 minus the number of degrees of freedom, ie 11,021, and 8672 respectively for the null models (without and with memory correction). The final model has reduced this variability to a residual of 15078-13507 = 1571. It is not entirely clear which of these null models is the most appropriate one on which to base an overall figure for the

TABLE C2

Significance levels for each variable

Variable	Scaled deviance increase (over 11241)	Degrees of freedom	Significance level
Mileage	342	1	< 0.01%
Age	70	1	< 0.01%
Experience	100	2	< 0.01%
Type of road	33	2	< 0.01%
Sex	64	1	< 0.01%
SEG	112	1	< 0.01%
Darkness	70	1	< 0.01%

TABLE C3

Significance levels for possible simplifications of the model

Possible simplification	Scaled deviance increase	Degrees of freedom	Significance level
Drop sex/experience interaction	14.7	1	0.01%
Drop SEG/age interaction	21.0	1	< 0.01%
Combine rural with motorway driving	7.1	1	0.8%
Combine urban with rural driving	32.9	1	< 0.01%

TABLE C4

Initial and final deviances and X^2 for the main model.

	Degrees of Freedom	Scaled Deviance	Generalised Chi-Square (X^2)
Null model without memory correction	13518	13605	24539
Null model with memory correction	13518	13539	22190
Final model	13507	11241	15078

explanatory power of the model, but from the values given above, the final model explains 85.7% of the variability of the uncorrected data, and 81.9% of the variability of the corrected data once the irreducible variability due to the Poisson process has been subtracted.

It has not been possible to explain the remaining variation in the data with the variables already investigated. Some of this residual variation is likely to remain unexplained, but some of it could be due to other measurable characteristics of individual drivers - characteristics such as attitudes, cognitive abilities, and social factors. However, the potential variation already explained by the demographic and exposure variables included in the final model is quite high. It may well be therefore that future studies of individual accident liability should concentrate on identifying those individual characteristics which mediate the age and experience effects, and which can help to explain why accident frequency is not proportional to annual mileage.

C.4 RANGE OF APPLICABILITY OF THE FINAL MODEL

Table C5 presents the (rounded) 5th and 95th percentile values of the distributions of the continuous variables, classified by four age groups, sex and two SEG groups. It also gives the proportion of respondents in each factor level and each age group. Together these provide a useful guide to the areas in which the model is particularly robust and therefore likely to be reliable.

The table shows that in all categories of the data the 95 percentile range covered by age is from 17 to at least 65 with the exception of high mileage drivers; older drivers have a more restricted range. The data for women drivers is rather more restricted than for men, in that the 95 percentile values of driving experience and annual mileage are smaller. Some caution is therefore needed in extrapolating the model to women drivers with more than 30 years of driving experience or who travel more than 14,000 miles each year.

C.5 OBSERVED VERSUS PREDICTED

Table C6 compares the average accident frequencies (accidents per year) reported by respondents (uncorrected for memory loss effects) by age/experience group with the corresponding prediction of the final model. In the table, the figures printed in normal type are the reported number of accidents; figures in italics are the corresponding model results based on the number of accidents predicted by equation (2) less the predicted memory loss effect. The figures below each pair (reported and predicted) are the number of drivers in each cell of the table.

TABLE C5

5 to 95 Percentile Range for the Continuous Variables

VARIABLE	Percentage of data	AGE (Years) 5th percentile	AGE (Years) 95th percentile	EXPERIENCE (Years) 5th percentile	EXPERIENCE (Years) 95th percentile	ANNUAL MILEAGE 5th percentile	ANNUAL MILEAGE 95th percentile
Overall	100.0	17.4	65.5	0.29	38.0	500	22,000
SEX:							
Male	48.7	17.3	68.5	0.29	44.0	800	30,000
Female	51.3	17.4	63.5	0.25	29.0	300	14,000
SEG:							
A, B & C1	54.7	17.4	65.5	0.29	38.0	500	23,000
C2 D & E	45.3	17.4	65.5	0.25	36.0	480	21,000
AGE:							
<20	34.6	-	-	0.21	2.2	324	20,000
20-29	18.7	-	-	0.84	11.0	500	30,000
30-49	24.7	-	-	4.0	27.0	500	25,000
≥50	22.0	-	-	4.0	52.0	500	17,000
ANNUAL MILEAGE:							
<6000	53.4	17.4	67.5	0.25	38.0	-	-
6-14,999	33.5	17.4	64.5	0.29	39.0	-	-
≥15,000	13.1	17.3	57.5	0.29	34.0	-	-

TABLE C6

Comparing reported accident liabilities with model predictions

		Age (years) <18	18-22	23-39	40-59	60+	All
Experience (years)	<1	0.37 *0.38* 2365	0.29 *0.29* 1267				0.336 *0.340* 3632
	1-4	0.28 *0.28* 115	0.183 *0.178* 1583	0.106 *0.111* 360	0.085 *0.080* 290	0.086 *0.072* 70	0.162 *0.158* 2418
	5-9		0.16 *0.14* 197	0.098 *0.102* 1745	0.063 *0.070* 737	0.063 *0.059* 206	0.091 *0.093* 2885
	10-19			0.085 *0.090* 1183	0.072 *0.063* 650	0.055 *0.049* 146	0.078 *0.078* 1979
	20+			0.072 *0.085* 180	0.070 *0.069* 1441	0.052 *0.050* 984	0.063 *0.063* 2605
	All	0.361 *0.368* 2480	0.205 *0.199* 3047	0.093 *0.098* 3468	0.070 *0.069* 3118	0.055 *0.053* 1406	0.125 *0.125* 13519

APPENDIX D: SIMPLIFIED AND INJURY MODEL DETAILS

D.1 THE SIMPLIFIED MODEL

For the purpose of illustrating age and experience effects, a simplified version of the final model was formulated. This model omits the effects of sex and SEG to give a model which predicts age and experience effects averaged over men and women and over the various SEG categories. The model is given as equation (3) in the main text and is based on 17,130 drivers who experienced 5110 accidents. Estimates of the parameter coefficients with their standard errors are given in Table D1.

D.2 THE INJURY ACCIDENT MODEL

A model was fitted to injury accidents only which had the same form as the model fitted to all accidents. First a model which included the memory loss term was fitted (as described in appendix A) to determine the memory decay constant for injury accidents; then the main model was fitted to data corrected for memory loss. The injury model is based on 463 injury accidents experienced by 13519 drivers. The injury model coefficients with their standard errors are given in Table D2; the table also gives the coefficients of the main model for comparison. The statistical errors in the injury model are about three times as large as those of the all-accident model because there are about nine times as many damage-only accidents as injury accidents. Because of these relatively

TABLE D1

Coefficients of the simplified model.

Model parameter	Model term	Coefficient	Standard Error
Mileage	ln M	0.38	± 0.02
Age	1/Ag	20	± 2
Experience	1/(x+b)	2.5	± 0.2
Parameter in denominator of experience term	b	2.2	± 0.5
MEMORY TERMS			
Memory	MEM	-0.33	
Mileage/memory interaction	(ln M - 9)*MEM	0.08	

Note. The memory function is also simplified, darkness and SEG being left out. As mentioned in section B.4, the mileage/memory interaction becomes significant if the darkness/memory interaction is omitted.

TABLE D2

Coefficients of models fitted to injury accident data

Model parameter Coefficients	Injury accident model Coefficients	Injury accident model Standard Errors	Main model (for reference)
Mileage	0.23	± 0.04	0.279
Age (for SEGs A/B/C1)	23	± 8	13
Age (for SEGs C2/D/E)	31	± 8	23
Experience (for males)	2.2	± 1.1	3.5
Experience (for females)	1.6	± 1.1	2.3
Darkness	1.4	± 2	1.6
Rural driving	1.1	± 0.2	0.65
Motorway driving	1.3	± 0.3	0.88
Constant (Ln)	-7.53		-5.06
SEG constant correction	-0.55		-0.72
Sex constant correction	0.05		-0.02
MEMORY TERMS			
Memory (for SEGs A/B/C1)	-0.20	± 0.08	-0.40
Memory (for SEGs C2/D/E)	-0.08	± 0.08	-0.26
Memory-darkness interaction	-0.34	± 0.3	-0.64

large errors, it cannot be claimed that the injury model is an adequate one. However, the purpose of fitting the injury model is to provide an indication of whether the age, experience and mileage effects determined for all accidents are similar to the corresponding effects for injury accidents or dramatically different.

As a rough rule of thumb, if the difference between the injury-accident model coefficient and the corresponding coefficient for the all-accident model is more than twice the standard error shown in the table, then the difference is likely to be statistically significant at the 5% level. Using this rule, it will be seen that the small reduction in the mileage exponent for injury accidents is far from significant. The age effect is rather larger and the experience effect rather smaller for injury accidents than for all accidents, but neither of these differences is statistically significant.

There are only two differences between the models which approach the 5% level of significance. First, the memory coefficients are less, indicating that injury accidents are more difficult to forget than non-injury accidents. Second, the coefficient for rural driving (relative to urban driving) is larger for injury accidents. This is consistent with a higher severity rating of accidents on rural roads compared with urban roads.