TRANSPORT RESEARCH LABORATORY

TRL REPORT 219

THE ACCIDENT LIABILITY OF CAR DRIVERS: THE RELIABILITY OF SELF REPORT DATA

by G Maycock, Julia Lester and C R Lockwood

Prepared for: Mr H J Wootton, Acting Research Director, TRL
Project: Seedcorn Research Programme (SR52)

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 Offices: Old Wokingham Road, Crowthorne, Berkshire, RG45 6AU.

First Published 1996
ISSN 0968-4107

CONTENTS

EXECUTIVE SUMMARY

For accident studies which need to be based on the accident histories of *individual* drivers, it is necessary to collect data using self-report techniques such as questionnaires and interviews. However, in view of the errors and biases that self-reported data collection can involve, the users of these methods must be particularly concerned with the issue of data reliability. This report present a study aimed at exploring the reliability of self-reported data relevant to analyses of accident liability.

During 1987/88 a postal questionnaire survey was carried out to study the accident involvement of a large structured sample of car drivers (termed the 'original' survey). A 'follow-up' postal survey of the same drivers was carried out 3 years later. These surveys have provided a means of assessing aspects of data reliability by providing the opportunity to compare the data collected from the same respondents on two occasions three years apart to be compared.

In the original survey, in addition to basic demographic variables, drivers were asked to report the number of accidents in which they had been involved during the last three years, to estimate of their annual mileage and to estimate the time spent driving on different types of road. After doubtful cases and those who had not driven in the year preceding the survey had been removed, 8617 were available for analysis. In terms of the demographic details of the respondents at the time of the follow-up survey, just over half (51%) were women, and the average age was 38.

The annual mileage driven by male drivers had increased from 10,390 in the original survey to 12,190 in the later survey; the corresponding increase for female drivers was from 5,010 to 5,970 miles. For those drivers who had said that their mileage had not changed from the previous year (ie from 1989 to 1990), a plot of the reported annual mileage of male and female drivers in the original survey (1987) against the annual mileage reported by the same drivers in the later survey (1990) suggests that there is a considerable amount of inaccuracy in the estimates of annual mileage.

A comparison of the accident frequencies reported by the same drivers in the two surveys classified by sex and by 'old' and 'young' driver groupings shows that the high average accident frequency experienced by young male drivers in the original survey virtually halved in the three years between surveys. Although young women drivers did not have nearly as high an accident frequency, their accident liability also fell considerably between the surveys. The 'old' drivers had much lower accident frequencies than the younger drivers.

The types of accident and the circumstances in which they occurred in the follow up survey did not differ greatly from those reported in the original survey. The proportion of injury accidents remained virtually unchanged; 11 per cent of accidents in the original survey involved injury compared with 12% in the follow-up.

The Generalised Linear Modelling methodology was used to relate accident liability to dependent variables such as age, sex, annual mileage and driving experience. However, because in the case of the young drivers in the original survey, the period of time over which the accidents were being reported was variable - from a few months to several years - a comparison of the accident liability of this group with the other driver groups (all of which reported on accidents over the last 3 years) necessitates correcting for those accidents which are forgotten. This correction and its accuracy is a crucial element in the comparison, and is considered in detail in the report.

Table 12 from the report shows a comparison of the accident liability of three groups of drivers - original/young, follow-up/old and follow-up/young - in terms of a factor F which is the ratio of the accident liability of these groups to that of the original/old group when adjustments have been made for differences in age, sex, mileage and driving experience between the groups. Bearing in mind the fact that the errors shown in Table 12 are the estimated *standard errors* of these ratios (not the 95% confidence limits), it will be seen that none the four sub-samples differ significantly at the 5% level.

TABLE 12

Original and follow-up surveys - modelled accident comparisons in terms of the factor F (with standard errors)

Value of F for:	Original 1987/88 survey	Follow-up 1990/91 survey
OLD sub-set	1.0 (Reference Level)	1.009±0.05
YOUNG sub-set	1.04±0.10	0.89±0.07

The main findings of this study were as follows:

On average annual mileage increased by about 18 per cent between the surveys. Although in an overall statistical sense the annual mileages reported in the original survey (1987/88) by those drivers who claimed in the follow-up survey not to have changed their mileage 'from last year' were proportional to their reported mileages in the later survey (1990/91), there was a great deal of scatter in the individual mileage estimates. This suggests that individual driver's estimates of annual mileage cannot be regarded as accurate.

Tabulations by age and sex showed that accident frequencies decreased with age for both men and women, and were considerably lower for women than for men. Fitting a simple statistical model which took into account the effects of annual mileage, age, sex and driving experience, showed that the residual difference between the original and the follow-up survey for the same drivers in the 'old' subgroup was less than 1 per cent. A comparison of the young driver group in the two surveys proved to be considerably more difficult because of the need to adjust for memory loss effects. However, accident data collected by self-report is able to produce consistent data over a period of some years, providing the accident reporting period is also consistent and no correction for memory loss is needed.

An overall comparison of the average accident frequencies reported by the older respondents in both the original and the follow-up surveys with two other surveys carried out more recently shows that the self-reported accident frequencies obtained in the four surveys are comparable. This result confirms the reliability of the self-report method for use in accident studies.

THE ACCIDENT LIABILITY OF CAR DRIVERS: THE RELIABILITY OF SELF-REPORT DATA

ABSTRACT

This report presents a study aimed at exploring the reliability of self-reported data in analyses of accident liability. Data from two comparable postal questionnaire surveys of accidents carried out in 1987/8 and 1990/91 has provided the opportunity to compare accident and exposure data collected from the same respondents on two occasions three years apart. The report also includes some overall comparisons between the accident data collected in these surveys with similar data collected during the course of two later accident liability surveys. The study shows that although in an overall statistical sense the annual mileage reported by drivers in the first survey is proportional to that reported by the same drivers in the second survey, there is a great deal of variability in the individual mileage estimates. In contrast, providing the accident reporting period is constant from survey to survey and memory loss effects can thereby be avoided, the study shows that considerable consistency in the accident liabilities resulting from the surveys of this kind can be expected. The similarity of the average accident frequencies reported in the '87 and '91 surveys with those obtained in two more recent studies confirms the reliability of the self-report method for use in accident studies.

1. INTRODUCTION

During 1987/88 a postal questionnaire survey was carried out in order to obtain information relating to the driving experiences and accident involvement of a large structured sample of car drivers. The results of this survey - the first large scale survey of its kind - are reported in Maycock, Lockwood and Lester (1991) (RR315). The main aim of the RR315 study was to gain a better understanding of how accident liability - the expected frequency of accident involvements - relates to factors such as age, driving experience and annual mileage.

When drivers completed questionnaires in the 1987/88 survey, they were asked if they would be willing to help in further research into driving and road safety. About 70% agreed to help, and it is these volunteers who have provided the 'follow-up' data used in the reliability study which forms the subject of this report. The follow-up postal survey of those drivers who volunteered to help in further research was carried out in 1990/91, three years after the initial survey. The questionnaire used in both surveys was very similar, although some additional information was sought in the follow-up survey to provide data for other projects which had become of interest since the original survey. A copy of the questionnaire used in the follow-up survey is included as Appendix A.

The decision to undertake a 'follow-up' survey was dictated to a large extent by the need to assess the reliability of the self-reported data collected by this type of accident survey. For many research purposes, it is essential to base studies on the accident histories of individual drivers. In order to do this in the UK, it is necessary to obtain details of the accidents and the characteristics of the individual's driving from the drivers themselves. Although other accident data sources exist - notably the national accident data known as STATS19 and data held by insurance companies - these sources do not provide the necessary data for the analysis of individual driver's accident rates. Accident data collected routinely by the police using the STATS19 report procedure does not allow accidents to be linked to particular drivers nor does it contain information regarding some key personal characteristics of the drivers involved - in particular, driving experience and exposure to risk - which are important factors in accidents. Moreover, STATS19 is concerned only with accidents involving injury and generally speaking, there are too few injury accidents to enable analyses relating to individual drivers (as distinct from groups of drivers) to be carried out. Injury accidents though important from the point of view of economic cost, are a small proportion (roughly 11-12 per cent) of all accidents; the majority involve damage to vehicles or property only.

Information held by insurance companies about the accident involvement of individual drivers would seem to be a potentially valuable source of accident data. However, data collected by insurance companies is primarily collected to support the commercial operation of the company. Much of the detailed information needed for accident studies is stored on paper files rather than on a computer, and is not therefore readily accessible for research purposes. The data is mainly concerned with policies, claims and claimants rather than with accidents and drivers. Moreover, an insurance company's client base will not necessarily be representative of all drivers and accidents.

In order therefore, to gather the information needed to study drivers' accidents in relation to the characteristics of the individual drivers, it is necessary to collect the data using a self-reporting procedures such as questionnaires and interviews - with all the errors and biases that self reported data collection implies. Studies which use this method must therefore be particularly concerned with data reliability. The 'follow-up' survey reported here has provided a means of assessing the reliability of data collected by self completion questionnaire, by enabling the data collected from the same respondents on two occasions three years apart, to be compared.

If the self-report method for obtaining accident data is reliable, then other independent surveys using the same methods should give similar accident liability results. Since the original and follow-up self-report accident surveys were carried out, two other self-report surveys have been undertaken - one to study driver fatigue and the other to obtain driver reactions to police speed enforcement. The overall accident rate results for these surveys will be reported briefly in section 8 of this report.

2. BACKGROUND TO THE PRESENT STUDY

2.1 DATA ACCURACY AND RELIABILITY

When a driver is asked "How many accidents have you been involved in during the last 3 years?" or "How many miles have you driven during the last year?" - along with other questions about his or her driving experience - a number of motivational and cognitive influences come into play. In terms of motivation, some drivers may not wish to admit to having been involved in an accident. Others might feel that some questions - for example questions about age or socio-economic group - are not relevant to the issue of driving.

Even if the respondent is disposed to answer the questions asked, there remains the problem of the reliability of recall. The recall of events is subject to different kinds of error (Gaskell et al, 1992). Events may simply be forgotten - as will be illustrated later in relation to accidents. In many respects this is the simplest form of error to deal with. However, even if an event has been remembered, people are generally poor at estimating when an event occurred, and may tend to think that it occurred more recently than it actually did. This bias is known as forward telescoping and could occur for example either because the respondent genuinely thought that the event happened later than it did, or because the respondent wants or feels they should report the event even though it falls outside of the chosen recall period (Loftus et al, 1988). Giving the dates of accidents, or remembering how many accidents occurred in a specific period, could be subject to telescoping error for either or both of these reasons. Whether or not the date has been recalled accurately, there is evidence to suggest that the driver's recall of the details of an accident can be distorted (Chapman & Groeger, 1992; Diges, 1988).

With regard to questions about the individual's driving characteristics - such as annual mileage - the respondent is not so much being asked to remember how far they have driven, but to construct an estimate of the magnitude of a regular and routine activity. Here it is likely that respondents will use generic recall (in fact the questionnaires suggest they do just this); they will think about a typical day or week, for example their usual drive to work, to the shops, or to school, and will base their estimate on this typical set of daily or weekly trips with adjustments for special events such as holidays. In this case, errors could result from the generic memory not being fully representative of actual behaviour - unusual events could be left out. Simple computational errors can also creep in when scaling up to annual estimates from daily or weekly ones.

2.2 PREVIOUS STUDIES

Studies examining the validity of information gathered from drivers by comparing self-reported details with some independent source of the same measures are rare in the literature. In an unpublished study reported in a 1973 TRL leaflet an attempt was made to compare individuals' accident records as held by an insurance company with the same individuals recollection of those accidents which involved an insurance claim over the same two and a half year period. Not only was the reliability of the reporting of the accidents checked, but also the method of elicitation was investigated, comparing self-report postal questionnaires with face to face interviews. Response rates to both methods of elicitation were the same, but in an interview situation 67% of claims were recalled, whereas 85% were recalled in the postal questionnaire survey. This result suggests that not only are self-reported accidents a reasonably reliable measure, but also that for gathering certain types of information postal questionnaires are more successful than face to face interviews. The relative effectiveness of questionnaires may have arisen in this case either because respondents were less willing to admit their accidents to an interviewer, or because they have more time to think about and check the details of their accidents in the case of a postal questionnaire - or a combination of both.

In connection with the reliability of annual mileages, White (1976) has compared drivers' estimates of the annual mileage their vehicle has travelled with the vehicles' odometer readings. Large variation was found for individuals between their estimate and the instrument reading, but for groups of drivers over and under estimations of annual mileage tended to cancel each other out. In conclusion, White considered that although individual reports of estimated annual mileage were not valid, mean annual mileage for a group could be used satisfactorily.

2.3 OBJECTIVES AND METHODS USED IN THE PRESENT STUDY

The present study attempts to examine the reliability of annual mileage and self-reported accident frequency. The method to be used is to compare the mileages and accident frequencies obtained in the original survey of drivers (RR315) with the same data obtained in the follow-up survey three years later. It is important to note that this methodology does not compare the accuracy of recall for

the same event when asked at two different times, but rather compares information given by the same people about their mileage and accidents in two separate but contiguous time periods to examine whether the data is consistent over time.

The findings of this study will not therefore assess the accuracy of an individual's recall for the same event, but will investigate whether the techniques used to elicit information yield consistent and comparable results.

3. THE SURVEYS

3.1 THE ORIGINAL DRIVER SAMPLE

For the 1987/88 survey the names and addresses of a sample of drivers was supplied by the Driver and Vehicle Licensing Agency (DVLA) and the administrative aspects of the survey was carried out under contract by NOP Ltd. The questionnaires were sent out in two phases. For the first phase, approximately 20,000 drivers aged 23+ years (henceforth referred to as the 'old' driver sample) were sent a questionnaire in November 1987. For the second phase 10,000 questionnaires were sent out in February 1988 to drivers aged less than 23 years (henceforth referred to as the 'young' driver sample) - see RR315 for full details. The questionnaires for the two phases contained basically the same questions, however the old drivers were asked about their accident involvements over the last three years, whilst the young drivers were asked about their accidents since passing the driving test. Therefore, for young drivers in the initial survey the accident reporting period varies from driver to driver and account has to be taken of this when calculating accident frequencies (involvements per year).

The response rate to the initial survey was 63.6% for the old driver sample and 71.3% for the young driver sample - an overall response rate of 66.2%. Subsequent consistency and edit checks of the returned questionnaires eliminated a number of doubtful responses, so that the total number of questionnaire responses available for analysis - old and young samples combined - was approximately 19,400, representing a 64.7% response rate. In RR315 a reduced sample consisting of 18,500 drivers was used for the multivariate analysis, by excluding those who hadn't driven during the previous year.

In the case of the follow-up survey, only drivers who expressed a willingness to take part in further research into driving in the first survey were considered for inclusion. Therefore, it is necessary to check the characteristics of this 'volunteer' group to determine whether they are significantly different from those drivers who did not volunteer to help in further research. Table 1 compares these two groups for both the old driver sample and the young driver sample. Although volunteers appear to drive further than non-volunteers, and in the 'old driver' sample the volunteers are rather younger and include fewer women, the differences between the groups are not such as to invalidate the subsequent analysis.

3.2 THE FOLLOW-UP DRIVER SAMPLE

The follow-up survey was carried out in 1990/91. Questionnaires were sent to drivers who volunteered in the initial survey to help in further research, and who were still on record as available to help. A further 700 or so volunteers were excluded since they had recently been surveyed by a TRL contractor. This is not a trivial consideration, since the 700 were selected on the basis of their accident histories in the first survey. About a half of these drivers were accident involved drivers, and their omission from the follow-up survey is likely to distort comparisons between the accident rates of these drivers and those responding to the original survey based on all drivers. For this reason, this report will present analyses based only on data from the same drivers in the two surveys.

After the 700 drivers referred to in the previous paragraph had been excluded, 12,324 drivers were available for the follow-up survey. The survey was again carried out in two

TABLE 1

Drivers volunteering in the initial survey to help in further research compared with non-volunteers.

	Accident frequency per year[1]	Annual mileage[2]	Age (years)	Driving experience (years)	% Female
Old Driver Sample:					
Volunteers	0.08	8230	44.8	17.3	49.7
Non-volunteers	0.06	5737	50.3	17.7	59.1
Young driver Sample:					
Volunteers	0.26	7338	19.0	1.5	52.3
Non-volunteers	0.25	6368	19.1	1.6	54.5

1 Number of accident involvements per year in a car or a van (uncorrected for memory loss effects).
2 Mileage in a car or van

phases. In November 1990, 7,357 questionnaires were sent to those who had been in the old driver sample in the initial survey who were now aged over 26; and a further 4,967 questionnaires were sent in February 1991 to the young driver sample who were now aged under 26. The follow-up survey was carried out in-house at TRL.

3.3 THE QUESTIONNAIRE

The questionnaire used in the follow-up survey was similar to that used in the initial survey, containing in particular questions about:

- the number of accident involvements in a car or a van during the last three years and some details of the most recent accidents.
- age and sex of respondent;
- the year in which the driving test was passed - to determine driving experience as the number of years since passing;
- exposure measures - an estimate of annual mileage in a car or van, and the time spent driving on different types of road.

An accident was defined in the questionnaire as "any incident which involved injury to another person or yourself, damage to property, damage to another vehicle, or damage to the vehicle that you were driving". Respondents were asked to report only those accidents which had occurred on public property, in which they had been the driver and which had occurred during the last three years - or for the young drivers, those that had occurred since passing the test. Accident details obtained from the questionnaires included whether the accident occurred in daylight or darkness, what type of road the accident had occurred on, what other vehicles or objects were involved, the severity of the injuries and damage sustained and whether the accident was reported to the police or to an insurance company. Due to space limitations on the questionnaire this information was obtained for at most, the three most recent accidents.

Several variables in the initial survey were omitted from the follow-up - in particular, questions defining the Socio-economic group (SEG) of the driver and the proportion of driving in daylight and darkness. Although these variables were found to be significant contributory factors to accident liability in the multivariate model reported in RR315, they were omitted from the follow-up because they were not major determinants of accident liability and space was needed in the questionnaire to include new items of interest for current research. One of the new items added to the follow-up survey asked drivers whether the car or van driven most often by the respondent was company or privately owned, and if the accidents reported were in a company or privately owned car or van. Section 8 below will look briefly at the responses to this question in the context of self reported accident rates obtained in other recent accident liability surveys.

3.4 SURVEY ADMINISTRATION

A covering letter on TRL headed paper was sent with all questionnaires explaining the reasons for carrying out the survey. Two reminder letters were sent to non-responders, the first approximately six weeks after the initial contact and the second approximately six weeks after the first reminder. Both reminders also included another copy of the questionnaire. Approximately 41% of the whole sample were sent one reminder and 25% of the whole sample were sent a second reminder.

Returned questionnaires were edit checked for logical consistency, coded and the data entered onto a database. Computer records of the questionnaires were checked by the supervisor on a sample basis for accuracy against the written questionnaire. Since the purpose of the survey was to follow-up the driving experiences and accidents of the same drivers over time, it was of the utmost importance that each questionnaire returned was completed by the same person who had completed a questionnaire in the initial survey.

In order to achieve this, the sex and age of the respondent was cross-checked between the two surveys and compared also with demographic information provided by DVLA when the original sample was drawn. If the personal details given in the two questionnaires gave grounds for suspecting that they had not been filled in by the same person, an examination of the questionnaire itself usually in conjunction with the questionnaire from the 1987/88 survey was undertaken. If as a result of these investigations the suspicion that the questionnaire had been completed by different drivers was confirmed, then the data was discarded.

There were approximately 80 questionnaires from 'wrong' respondents identified in this way; these were removed from the database and included in the figure for null responses. 8,888 completed questionnaires were available for analysis. Table 2 shows response rates for the survey, and the characteristics of responders and non-responders are compared in Table 3.

As can be seen an overall response rate of 72 per cent was achieved, and the response rate did not differ markedly between the young and the old drivers. Table 3 shows that there was little difference in the average age of responders and non-responders (No reply + Null) and little difference between the proportion of men and women in the three categories of responders. However, the drivers who did not respond at all, (No reply) tended to be rather younger than the those that responded. Although all these age differences are statistically significant ($p<0.001$) - hardly surprising given the large numbers involved (N=12,324) - they are of little practical significance in the present context.

TABLE 2

Response rates for the follow-up survey

	Old Drivers	Young Drivers	Total
Number of questionnaires sent	7357	4967	12324
Number completed and returned	5402	3486	8888
Response rate	73.4%	70.2%	72.1%
Number of null returns	835	386	1221
Proportion	11.4%	7.8%	9.9%
No reply received	1119	1095	2214
Proportion	15.2%	22.1%	18.0%

TABLE 3

Average age and proportion of females by response category in follow-up survey

	Responders	No Reply	Null
Average Age (years)	37.9	33.5	40.4
Proportion Female (%)	52	47	51

The structure of the follow-up sample largely reflects that of the original survey sample which was structured by sex, age and driving experience, in order to investigate the relationships within the accident data; it was not intended to be a representative sample of the driving public. In terms of the demographic details of the respondents at the time of the 1990/91 follow-up survey, just over half (51%) of respondents were women. The average age of the whole sample of respondents was 38; the average age being 39 for males with a range from 20-92, and 37 for females with a range of 20-87 years. About a half of the respondents had passed their driving test in 1980 or before, the other half passing their test between 1981-1987.

4. RELIABILITY OF REPORTED AGE AND EXPERIENCE

4.1 AGE DISCREPANCIES

Even when the checking referred to in 3.3 above had confirmed that the questionnaires relating to the original and the follow up surveys had been completed by the same individual, there were still discrepancies in the personal information given by drivers - particularly in relation to age, and more often, driving experience. For example, a respondent might report his or her age as 38 years in the 1887/88 survey and as 39 years in the follow-up three years later, where according to DVLA records the respondent had their 43rd birthday during 1990. Such occurrences were not very common and a decision was made to deal with them in the following way:

> Age was primarily calculated from the drivers date of birth given in the DVLA records which were available for all drivers taking part in the survey. However, if there was a difference between self-reported age and 'DVLA' age of only ±1 year, self-reported age was used instead. This strategy was adopted because it was felt that such small discrepancies were most likely to arise as a result of the respondent's birthday being close to the date on which they filled in the questionnaire. In such cases the respondent's estimate of their age would probably be more accurate than that calculated using date of birth from DVLA records. Large discrepancies between self-reported age and 'DVLA' age (greater than ±3 years) had already been dealt with in the investigation of 'wrong' responders.

4.2 DRIVING EXPERIENCE DISCREPANCIES

Much more common were discrepancies between the two surveys about the year in which respondents passed their driving test. For the young driver sample this was relatively easy to deal with, since DVLA information with regard to

test pass dates was held for all drivers in this sample. Therefore, for the young driver sample, a variable for driving experience was calculated using DVLA records of when the respondent passed their driving test.

For the old driver sample test pass dates from DVLA records were not available for many drivers. In terms of the methodological aspects of carrying out longitudinal surveys it may be of interest to examine the extent of the problem with these driving experience data discrepancies. Approximately 430 respondents (6% of the old driver sample) gave test pass dates which were different by more than ±3 years in the two surveys; about a third of these were different by 10 years or more. Although this number may not seem very large given the total number of respondents to the survey, and may indeed not make much difference to large scale multivariate analyses of the data, it is as well to bear such findings in mind when considering the reliability of self-reported data.

It was not considered feasible to check all the experience discrepancies by reference to the questionnaires. However, all discrepancies of 10 years or more were checked using both the 1987/88 questionnaires and the 1990/91 questionnaires. In this way more than half of the discrepancies were resolved, usually by taking note of the respondents written comments alongside their answer. These comments often mentioned that the respondent had taken two driving tests, perhaps one in another country before taking a test in Britain, and the discrepancy had arisen when the respondent had recorded the date of one of these tests on the original questionnaire and the date of the other test in the follow-up survey. However, even after examining the questionnaires discrepancies remained in a significant number of cases; these cases were dealt with in the following way:

> If the two test dates varied by less than 10 years, the year given in the initial survey was taken for the experience variable. If the two dates differed by 10 years or more, then the experience variable was entered as missing data, so would not contribute to any analyses.

5. COMPARISONS BETWEEN SURVEYS: EXPOSURE

5.1 INTRODUCTION

This section provides an overview of how the responses to drivers' exposure to risk in the follow-up survey compare with the responses to the same questions by exactly the same drivers in the initial survey three years earlier. In both surveys 98 per cent of the drivers had driven a car or van in the last year. In comparing the driving experiences of drivers in the two surveys, it seems sensible to exclude those who did not drive in the year preceding the survey. Accordingly, the figures given in the sections that follow represent the responses made by 8,617 drivers - 5,226 from the 'old' driver sample and 3,391 from the 'young' drivers who responded to both surveys and had driven in the previous year. In some cases the number of valid responses are somewhat lower than the maximum possible because respondents occasionally failed to provide a response to particular questions. Percentages are given as a percentage of valid responses.

5.2 ANNUAL MILEAGE

The majority of drivers who drive at all, drive regularly. 93 per cent of men and 87 per cent of women reported driving more than once a week in the earlier survey; in the follow up 94 percent of men and 90 per cent of women still drive more than once a week. The annual mileage driven in a car or van reported by male drivers had increased from 10,390 in the original survey to 12,190 in the later survey. The corresponding increase for female drivers was from 5,010 to 5,970 miles. The average increase was thus 18 per cent. National data on mileage and the number of registered vehicles (Transport Statistics Great Britain, 1992) suggests that between 1987 and 1990 the distance driven per vehicle rose by about 4 per cent from 10,150 Km to 10,570 Km. Taken at face value therefore, the sample of drivers in the present surveys would seem to have increased their average mileage by an amount which is considerably greater than the national trend. The extent to which this is genuine or merely an artifact of the drivers estimates of mileage is impossible to tell. Table 4 shows the changes in annual mileage between the original and the follow-up survey by age group, for male and female drivers.

Figure 1 shows the same information presented as the difference in annual mileage reported by the same drivers in the different surveys. Here, a positive difference denotes an increase in annual mileage from the original to the follow-up survey, and a negative difference denotes a decrease in annual mileage from the original survey to the follow-up.

It will be clear from Figure 1 that most drivers aged under 50 years have experienced a considerable increase in their average annual mileage from the 1987/88 survey to the follow-up three years later. Although one would expect younger drivers to increase their annual mileage with increasing age and driving experience, it appears that most respondents under 50 years are reporting higher annual mileages than they were three years earlier. Drivers over 50 have reported decreases in their average annual mileage - this would of course be expected for the over 60 age group given the reduction in the need to drive after retirement.

5.3 ACCURACY OF MILEAGE ESTIMATES

The difficulty of obtaining accurate estimates of annual mileage has already been referred to in Section 2 and the

TABLE 4

Average annual car/van mileage for the same drivers in the two survey periods

Age at mid-point of accident period in original survey	MALE			FEMALE		
	Original survey	Follow-up survey	Number of drivers	Original survey	Follow-up survey	Number of drivers
Less than 18	8,870	12,120	861	4,950	6,380	892
18-18.9	9,050	13,240	424	4,840	7,450	462
19-23.9	11,930	15,320	506	5,640	7,320	554
24-29.9	13,440	15,180	371	5,380	6,310	502
30-39.9	12,470	13,800	529	5,290	5,890	603
40-49.9	12,230	12,710	480	5,400	5,340	480
50-59.9	10,090	10,040	460	4,340	4,320	439
60 and over	6,720	6,000	518	3,690	3,510	359
All ages	10,390	12,190	4149	5,010	5,970	4291

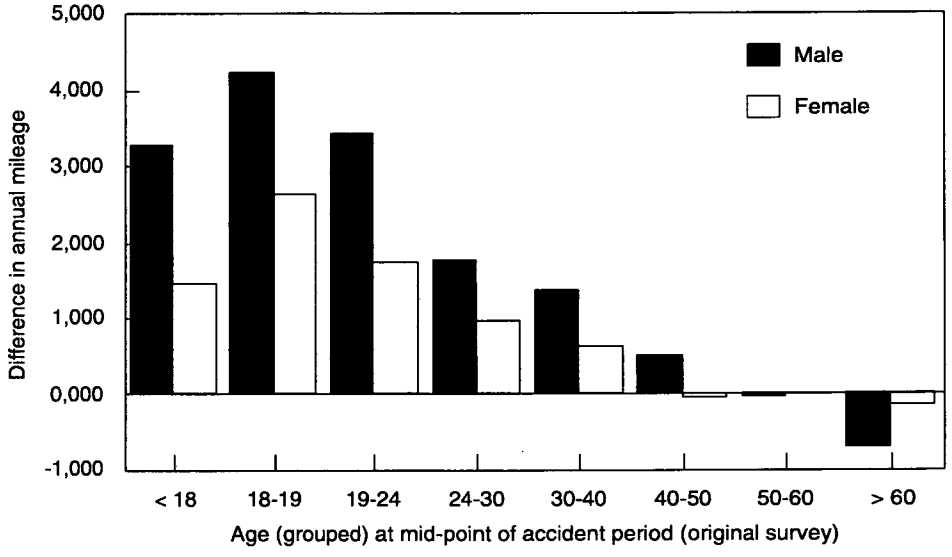

Fig.1 Differences in annual mileages (follow-up - original)

overall increases in mileage between the two surveys has been commented on in Section 5 above. It is of interest however, to see whether the data obtained in these two surveys can provide any evidence for the reliability of individual driver's estimates of their own annual mileage. Drivers in the follow-up survey were asked whether they had driven 'more this year than last year', 'less this year than last year', or 'about the same amount' as last year. In reply, 28 per cent reported that they had driven more in 90/91 than the previous year, 17 per cent reported that they had driven less, and the remainder (55%) reported no change. More drivers claiming an increase in annual mileage than are reporting a decrease is consistent with the increase in annual mileage between the surveys noted in 5.2 above.

Figure 2 shows the annual mileage of male and female drivers in the original survey (1987) plotted against the estimated annual mileage for the same drivers in the later survey (1990); the plot is limited to drivers who had said that their mileage had not changed from the previous year (ie from 1989 to 1990). To avoid extreme values, only those drivers whose reported annual mileage in the two surveys were between 400 and 20,000 have been included.

Figure 2 shows that there is a considerable amount of scatter - to put it mildly - although the general trend for annual miles travelled in 1987 to be proportional to those travelled in 1990 is evident. The lines on Figure 2 are regression lines though the origin for male and female drivers, the slopes of which are 0.892±0.011 for men and 0.843±0.013 for women, reflecting the general increases in annual mileage already noted.

Although the dataset is restricted to drivers whose annual mileage remained unchanged from 1989 to 1990, Figure 2

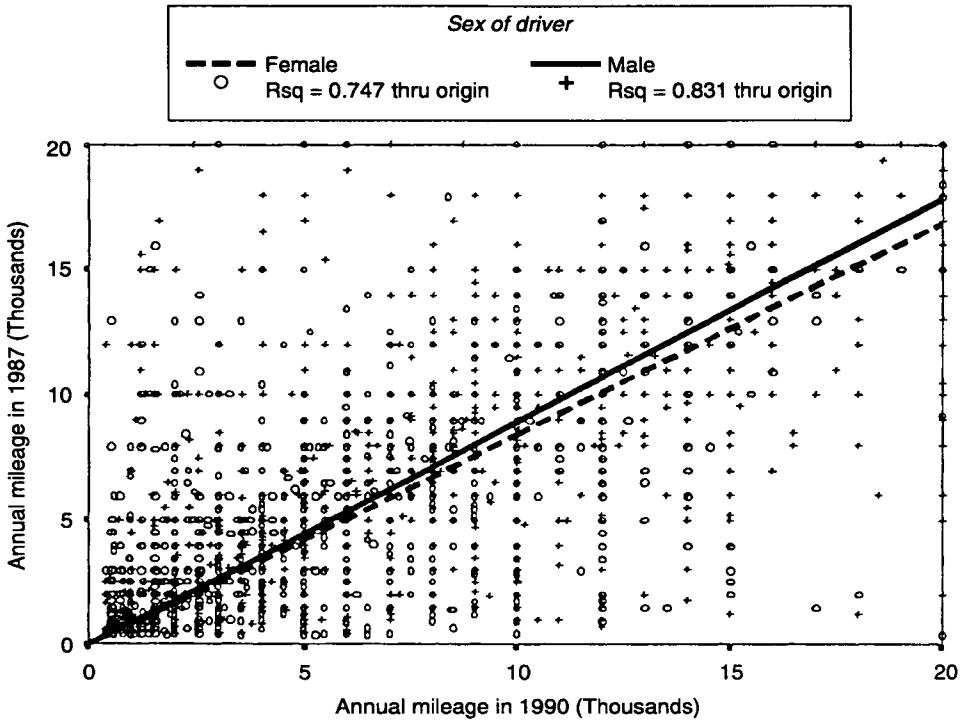

Fig. 2 Annual mileage reported in 1987 plotted against those in 1990 for selected drivers

relating as it does to annual mileages in 1987 and 1990 is likely to reflect some genuine changes in mileage over the three years between the surveys. As a grossly conservative approach, let us treat all the variation shown in Figure 2 as representing errors in mileage estimation. The residual variance about the regression line is 1.54×10^7 for male drivers and 1.02×10^7 for female drivers. If we assume that the errors are Normal and independent of mileage, this suggests an error standard deviation of about 3,900 miles for men (about 46 per cent of the mean value) and 3,200 miles for women (62 per cent of the mean value). It is unlikely that the distributions are Normal or that the error variance is independent of the mean. It is also likely - as has already been pointed out - that some of this variation will be due to genuine changes in mileage, but without introducing arbitrary judgements, it is impossible to tell how much. Nevertheless, the 'rough and ready' comparisons between the surveys illustrated in Figure 2 support the result of White, (1976), Section 3.3 above, in demonstrating that although individual drivers estimates of their annual mileages can be considerably in error, group averages can be regarded as reasonably reliable. The implications for the accident modelling of inaccuracies in the mileage estimates of the order suggested by this analysis need to be further evaluated.

5.4 ROAD TYPE

Table 5 shows the average proportion of time spent by men and women drivers on the three types of road - motorways, roads in built up areas and roads outside built up areas - in the two surveys.

The Table shows that women spent a rather smaller proportion of their driving time on motorways than men, but that the proportion of motorway driving increased for both

TABLE 5

Proportion of time spent driving on different types of road.

Road type	Original survey		Follow up survey	
	Male	Female	Male	Female
Motorways	18.2	12.4	21.6	15.1
Road in built-up areas	50.9	55.6	48.0	54.5
Roads outside built-up areas	30.9	32.0	30.4	30.4

sexes between the original and the follow up survey. These increases in motorway driving are consistent with the higher overall annual mileages reported in the later survey. The increases in motorway driving also mean, of course, that the proportion of time spent driving on the other types of road has fallen.

5.5 OTHER FEATURES

Approximately 25 per cent of men and 7 per cent of women in both surveys drove another type of vehicle regularly, e.g. a lorry or a motorcycle. The average annual mileage on these other vehicles had increased by about 10 per cent for men from 5,590 to 6,150, but decreased by nearly 30 per cent for women from 2,240 to 1,560. No account has been taken in the subsequent analysis of accidents of either driving/riding experience or accidents in these other vehicles.

Questions included for the first time in the follow-up survey showed that 85 per cent of drivers drove all of their car/van annual mileage in the same vehicle and that 15 per cent reported that the car or van they drove most often is company owned. The company car effect on accidents is considered further in section 8.1 below. It is important to remember that these drivers are not necessarily 'professional' drivers who drive as a part of their job, but are probably 'ordinary' drivers who happen to be driving a company car.

6. COMPARISONS BETWEEN SURVEYS: ACCIDENTS

6.1 INTRODUCTION

This section compares the accident types and the accident frequencies reported in the follow- up survey with those reported by the same drivers in the original survey 3 years earlier. In both surveys accident involved respondents gave details of the most recent accidents (up to a maximum of three) in which they had been involved during the accident reporting period. In considering the differences and similarities in the accident types, it should be borne in mind that the data refers only to the three most recent accidents and not all accidents during the reporting period.

6.2 OVERVIEW OF ACCIDENTS AND ACCIDENT FREQUENCIES

Table 6 shows the numbers of accidents reported by the same drivers in the two surveys classified by sex and by the 'old' and 'young' driver groupings.

The length of the accident reporting period was not the same for all respondents in both surveys because in the original survey the accident reporting period for young drivers was the time that had elapsed since they had passed the test. However, taking this into account, Table 6 also shows the average accident frequencies (accidents per year) for male and female drivers in the old and young categories.

TABLE 6

Numbers of accidents and average accident frequencies for the same drivers in the two surveys.

	MEN		WOMEN	
	Original survey	Follow-up survey	Original survey	Follow-up survey
Old drivers:				
Number of accidents	636	562	458	480
Average number of accidents per year[1]	0.082	0.072	0.058	0.061
Number of drivers	2593		2633	
Young drivers:				
Number of accidents	810	795	457	602
Average number of accidents per year[1]	0.317	0.164	0.176	0.113
Number of drivers	1619		1772	

1 Here and in later Tables 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. (For the follow-up survey and for the 'old' driver sample this is always three years; for young drivers in the original survey however the period varies.)

Table 6 clearly shows the high average accident frequency experienced by young male drivers in the original survey - a rate which has virtually halved in the three years between surveys. Although young women drivers do not have nearly as high an accident frequency, their accident rate too has reduced considerably between the surveys. The older drivers have much lower accident frequencies than the younger drivers. Even so, the number of accidents per year has fallen somewhat between the surveys for male drivers, though it has increased marginally for women drivers in the 'old' category. These accident frequencies make no allowance for differences between the groups in the levels of annual mileage. The accident changes between the two surveys in relation to the key determinants of accident liability - age driving experience and exposure - will be considered below.

The types of accident, and the circumstances in which they occurred, did not differ greatly between those reported in the initial survey and those reported in the follow up survey. 72% of the accidents reported in the initial survey compared with 78% in the follow-up happened in daylight. 64% of the accidents reported in the original survey involved a moving vehicle other than the respondent's own vehicle, compared with 67% in the follow-up. A larger difference was found in the proportion of accidents involving animals and roadside objects. In the original survey 20% of the accidents reported involved something other than another vehicle, whereas in the follow-up survey this figure had fallen to 13%. It seems likely that this result is reflecting the decrease in single vehicle accidents which is known to occur with increasing age and experience of the driver.

The proportion of injury accidents (i.e. accidents in which any road user involved in the accident was injured) remained virtually unchanged from the initial survey to the follow-up; 11 per cent of accidents in the original survey involved injury compared with 12% in the follow-up. In both surveys, the proportion of accidents involving slight injuries was 8-9 per cent, the remainder being serious injuries; no accidents were reported involving fatal injuries.

The location of the accidents was not obtained in the original survey. In the follow-up survey however, 4% of accidents took place on a motorway, 68% took place on roads in built-up areas and 28% on roads outside built-up areas.

There was little difference between the surveys in the extent of damage reported by respondents either to the driver's vehicle or to another driver's vehicle. In both surveys approximately 30% of accidents were reported to the police. However, whereas in the original survey 46% of the accidents reported involved the respondent making an insurance claim, this figure had increased to 57% in the follow-up survey three years later. 19% of accidents described in the follow-up survey were in a car or van which was company owned.

6.3 COMPARING DIFFERENT DRIVERS OF THE SAME AGE

The purpose of this section is to compare driver accident frequencies from the two surveys in as straightforward a way as possible. The difficulty in making valid comparisons in simple tabular form is that drivers' accident frequencies depend on their age, driving experience and annual mileage. All of these variables will change from the original survey to the follow up. Moreover if as is indicated in section 2.1 memory effects are important, then the number of accidents reported by respondents will also depend on the recall period. Individual drivers in the young driver category in particular, completed the questionnaire between 2 and 70 months after passing the test; the effect of memory loss on the number of accidents reported will be considerably greater for those recalling accidents over nearly 6 years compared with those who had been driving for only a few months.

In section 7 comparisons will be presented in terms of a multivariate model which takes account of these factors. However, for the purpose of presenting age for age comparisons in this section, only data from the 'old' driver category will be used (5226 drivers). This avoids having to make any corrections for potential memory loss among respondents since all respondents were recalling accidents for the full 3 years. It also has the advantage that for the older drivers accident frequencies are not changing rapidly with the passing years.

Table 7 and 8 shows for the 'old' driver category the accident data from the original and the follow up survey tabulated in 6-year age bands - for convenience the age bands correspond to driver ages at the end of 1987 and 1990 when the original and follow-up survey questionnaires were completed. For each age band shown in the tables, the average annual mileage, the average age at the mid-point of the accident period, and the average experience (number of years since passing the test) at the mid-point of the accident period are given together with the number of drivers in each group. Comparisons between the two surveys can be made for each age group by scanning the columns in the tables vertically. It will be seen that average age within a given age band differs by less than a year between the original and the follow-up surveys. Average driving experience is larger in the follow up survey by a little over a year, and annual mileage - as would be expected from the overall increases in mileage - is up to 18 per cent higher in the later survey. Since an increase in driving experience will reduce accidents whilst an increase in annual mileage will increase accidents, the changes in these two variables between the surveys will, in accident terms, tend to cancel.

Comparing the accident frequencies within age bands in Table 7 and 8 shows that although the follow up survey rates are perhaps somewhat lower for the 24-29 age group where the additional increase in age and experience are still

TABLE 7

Average accident frequencies (uncorrected for memory loss), annual mileages, mid point age and driving experience for male drivers grouped by age.

2593 MALE DRIVERS	Age at date of survey (grouped)								
	17-23	24-29	30-35	36-41	42-47	48-53	54-59	60 and over	All ages
Original Survey:									
Accidents per year[1]	0.154	0.134	0.099	0.065	0.066	0.079	0.080	0.047	0.082
Average annual mileage	13,960	13,940	12,640	12,160	12,520	11,170	10,490	7,080	11,080
Average mid-point age (years)	21.0	25.2	30.9	37.3	42.8	49.1	55.1	65.7	44.8
Average driving experience (years)	4.0	7.2	10.4	14.6	18.5	21.1	24.7	35.3	19.7
Number of respondents	106	369	296	343	282	308	272	617	2593
Follow up survey:									
Accidents per year[1]	-	0.102	0.102	0.064	0.083	0.089	0.069	0.041	0.072
Average annual mileage	-	15,890	14,810	13,210	14,000	12,680	10,750	7,030	11,700
Average mid-point age (years)	-	25.6	30.7	36.9	42.8	49.3	54.9	67.1	47.8
Average driving experience (years)	-	8.1	11.6	16.0	18.6	23.6	26.1	36.5	22.7
Number of respondents (N)	-	268	357	286	358	280	295	749	2593

1 The Standard Error of the accident frequencies is approximately 0.6x[Accidents per year/N]$^{0.5}$. Thus for example, for the 24-29 year old group the accident frequency is 0.134 ± 0.011. The 95% confidence interval for this value is thus approximately 0.112 to 0.156.

TABLE 8

Average accident frequencies (uncorrected for memory loss), annual mileages, mid point age and driving experience for female drivers grouped by age.

2633 FEMALE DRIVERS	Age at date of survey (grouped)								
	17-23	24-29	30-35	36-41	42-47	48-53	54-59	60 and over	All ages
Original Survey:									
Accidents per year[1]	0.087	0.092	0.064	0.061	0.055	0.050	0.043	0.023	0.058
Average annual mileage	5,460	5,610	5,250	5,140	5,630	4,920	4,470	3,710	4,970
Average mid-point age (years)	21.1	25.5	31.0	37.3	42.8	49.1	54.9	64.8	42.2
Average driving experience (years)	3.8	7.0	9.1	12.0	14.4	15.0	16.8	25.2	13.7
Number of respondents	84	514	339	386	287	299	254	470	2633
Follow up survey:									
Accidents per year[1]	-	0.082	0.074	0.083	0.057	0.041	0.041	0.049	0.061
Average annual mileage	-	6,631	6,230	5,860	5,790	5,210	4,890	3,610	5,330
Average mid-point age (years)	-	25.8	30.5	36.9	42.7	49.2	54.8	66.1	45.2
Average driving experience (years)	-	8.0	11.1	13.6	15.9	18.4	18.2	26.6	16.7
Number of respondents	-	265	506	323	387	274	291	587	2633

1 See footnote to Table 7

having a significant influence, overall, the frequencies are very similar. Statistically speaking, with the exception of the '60 and over' age group for women (see comment below), none of the differences between surveys are significant at the 5 per cent level - though due to the relatively small number of drivers in each group the confidence limits of accident frequencies (see footnote to Table 7) are relatively large. More importantly, there is no obvious systematic bias for one survey to give higher accident frequencies than the other over the age range.

In terms of the overall accident frequencies, the difference for male drivers between an accident frequency of 0.082 in the original survey and 0.072 in the follow-up is just about statistically significant at the 5 per cent level, and also in the expected direction since all drivers are 3 years older and have 3 years additional driving experience. The increase in the effect of age and experience for this group of drivers is likely to outweigh the effect of a 6 per cent increase in annual mileage. It is surprising that the same effect is not observed in the case of women where in effect, the overall accident frequency remains unchanged. Women's accident frequencies are of course generally lower than those for men, and the age effects - especially for the younger drivers - are correspondingly less. However, examination of the within age band comparisons suggests that the accident frequency for the '60 and over' age group in the original survey is anomalously low compared with the follow-up result. However, bringing this value up to the level of the 54-59 year old age band would make little difference to the fact that overall, women's accident frequency appears to have changed little between the surveys.

6.4 COMPARING THE SAME DRIVERS OVER TIME

To illustrate the accident changes which have occurred to the same individuals between the two surveys, Table 9 shows accident frequencies in both surveys for drivers grouped by their age at the mid-point of the accident period for the original survey; the changes are illustrated in Figure 3. Thus for example, male drivers whose age was between 19 and 23.9 at the mid-point of the original survey accident

TABLE 9

Average accident frequencies (uncorrected for memory loss) for the same drivers in the two survey periods

Age at mid-point of accident period in original survey	MALE			FEMALE		
	Original survey	Follow-up survey	Number of drivers	Original survey	Follow-up survey	Number of drivers
Less than 18	0.440	0.161	882	0.227	0.121	924
18-18.9	0.306	0.172	430	0.170	0.115	483
19-23.9	0.200	0.139	513	0.127	0.083	567
24-29.9	0.113	0.100	376	0.083	0.078	508
30-39.9	0.077	0.076	532	0.061	0.071	613
40-49.9	0.076	0.079	487	0.052	0.050	489
50-59.9	0.065	0.062	466	0.041	0.043	451
60 and over	0.052	0.038	526	0.023	0.052	370
All ages	0.140	0.107	4212	0.087	0.082	4405

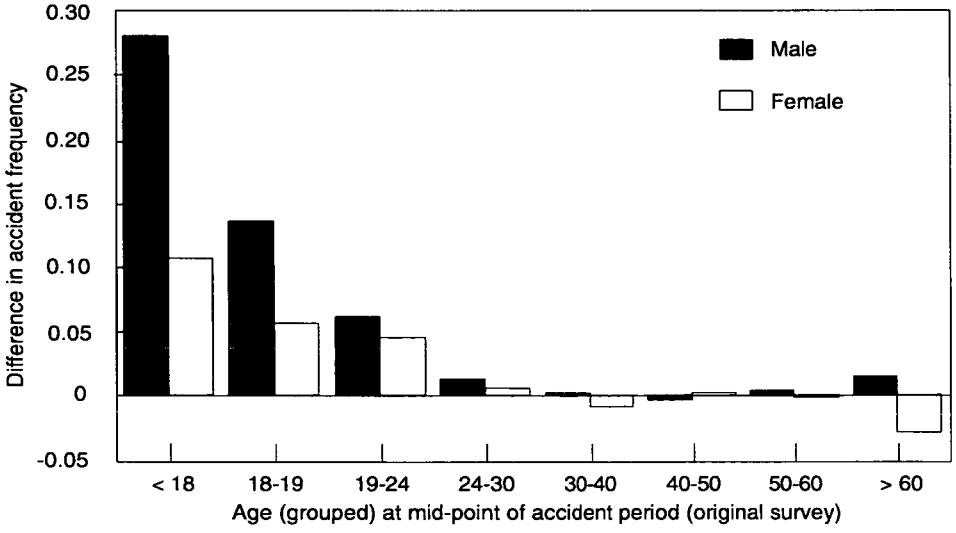

Fig.3 Differences in accident frequencies (original - follow-up)

period, had an accident frequency of 0.200 accidents per year. The accident frequency reported by the same drivers three years later was 0.139 - a 30 per cent reduction. It is clear from Table 9 and Figure 3 that the younger drivers experienced the largest reduction in accidents due to the additional 3 years of age and driving experience whereas the effect for the older drivers is small. The increase in accident rate of women drivers in the '60 or over' age group from the original to the follow up survey is evident here as in Table 8.

The accident data shown in Table 9 and Figure 3 does not make any correction for accidents which are forgotten. In the original survey, the first three age groups involve drivers for whom the accident period was variable. In a majority of cases, the relevant accident period is less than the three years appropriate to the older drivers. The memory loss effect means that for these drivers fewer accidents would be forgotten than would be the case for the older drivers, and the apparent reduction of accidents with age in Table 9 will accordingly be over stated. Thus for example, the reduction of a factor of over 7 from the 0.440 accidents per year appropriate to the youngest group of male drivers to 0.06 or thereabouts for the older drivers will be inflated by the fact that lost accidents have not been taken into account. The modelling approach to be described in section 7 will make some allowance for memory loss effects.

Though the differences shown in Table 9 and Figure 3 are exaggerated by the fact that memory loss corrections have not been made it can be seen that young drivers considerably 'improve' their accident frequency in the three years between surveys, reflecting their increasing maturity and driving experience. With the exception of the over 60 women drivers - an exception already noted - there is little change in drivers' accident frequencies between the surveys for those aged over 30.

Table 10 parallels Table 9 in format but shows the differences in injury accident frequencies between the original survey and the follow-up for the same drivers. Although the general trends apparent in Table 9 are discernable for injury accidents in Table 10, the small numbers of accidents involved makes it difficult to identify any statistically robust trends.

7. COMPARISONS BETWEEN SURVEYS: ACCIDENT MODELLING

7.1 THE MODEL

One of the aims of the follow-up survey was to test the robustness of the multivariate modelling technique developed using all of the initial survey data and reported in Maycock et al (RR315, 1991). In this section the observed follow-up accident frequency will be compared with the accident frequency for the same drivers using a slightly modified version of the 'simple' accident model reported in RR315. The simple form of the model was used in preference to the main model, because data relating to the proportion of time drivers spent driving in the dark, and the driver's socio-economic group - both of which featured in the main RR315 model - were not collected in the follow-up survey.

The form of the model fitted to drivers' accident frequency data was as follows:

$$A/T = k\, M^{\alpha} \exp \{b_1/Ag + b_2(s)/(X+2.2) + [U\ MEM]\} \quad (1)$$

TABLE 10

Average injury accident frequencies (uncorrected for memory loss) for the same drivers in the two survey periods

Age at mid-point of accident period in original survey	MALE			FEMALE		
	Original survey	Follow-up survey	Number of drivers	Original survey	Follow-up survey	Number of drivers
Less than 18	0.049	0.020	882	0.019	0.017	924
18-18.9	0.026	0.012	430	0.014	0.015	483
19-23.9	0.020	0.018	513	0.024	0.012	567
24-29.9	0.014	0.013	376	0.011	0.009	508
30-39.9	0.012	0.009	532	0.008	0.004	613
40-49.9	0.008	0.008	487	0.003	0.007	489
50-59.9	0.006	0.004	466	0.007	0.007	451
60 and over	0.006	0.006	526	0.005	0.005	370
All ages	0.015	0.012	4212	0.087	0.011	4405

Where: A/T = Accident frequency (accidents per year), and T is the period over which the accidents are being reported,

k is a constant

M = Annual mileage, and α the exponent to be determined,

Ag = the driver's age in years at the mid-point of the accident period, and b_1 the coefficient to be determined,

X = the driver's driving experience measured as the number of years between passing the driving test and the mid-point of the accident period, and $b_2(s)$ is the coefficient to be determined; s is the sex of the driver - there is a different value of b_1 for male and female drivers.

MEM is the length in years of the period over which accidents are being recalled, with U the memory loss coefficient; this term in the model applies only when memory loss effects are being estimated from dated accidents.

A brief summary of the principles of the Generalised Linear Modelling methodology and the methods used for fitting the explanatory variables and testing their significance is included as Appendix B.

Annual mileage is fitted as a power term with α as the exponent. This has proved to be an extremely robust form of the accident-mileage relation, though of course it means that accidents are not proportional to mileage driven. High mileage drivers are involved in fewer accidents per mile travelled than are low mileage drivers. The reason for this is not known.

The accident-age relation is adequately represented by a reciprocal 1/Age term. This relation means that accidents fall with increasing age at a higher rate for young drivers than for older drivers. A similar but more dramatic relation is that between accidents and the number of years driving experience. The constant 2.2 included in the experience term in the above model is determined interactively as explained in Appendix B, to give the best functional form for the accident-experience relation; it was the value used in the RR315 simple model. The shapes of these functional forms are illustrated in RR315.

The driver's sex was not included in the simple model in RR315. However, the previous tabulations of accident frequencies have shown that sex is an important determinant of a driver's accident liability. The most effective way of incorporating the sex of the driver into the accident model was as an interaction with the experience term - just as in fact it was included in the main RR315 model. The above model therefore has a coefficient of experience $b_2(s)$, which indicates that there are different coefficients for the driving experience effect for men and women.

For the young drivers in the original survey, because the period of time over which the accidents were being reported varied from a few months to several years, the comparison of the accidents of this driver subset with the other driver groups (all of which reported on accidents over the last 3 years) necessitates correcting for those accidents which are forgotten. Since this correction and its accuracy is a crucial element in the comparison, it will be considered in detail in the next section before the main results are presented.

7.2 MEMORY LOSS EFFECTS

The effects of memory loss can most readily be illustrated by determining for each of the four sub-sets of the data being considered in this report - original/follow-up surveys by old/young grouping - the number of accidents reported in each month of the survey, and plotting these figures against the time in months from the date of completion of the questionnaire. Of course, this can only be done if the accidents have been dated. In the case of the young drivers in the original survey - since some drivers had been only driving (and having accidents) for a few months whilst others had been driving (and having accidents) for some years, it is also necessary to adjust the monthly accident numbers according to the number of drivers responding in each month of the survey. Figure 4 shows the results for the old drivers and Figure 5 for the young drivers.

In Figures 4 and 5 it will be seen that the numbers of accidents reported in each month decline as respondents try to recall accidents which occurred earlier in the 3 year period. The small number of accidents recalled by the relatively few young drivers who had been driving for more than three years makes this data unreliable; it is not therefore shown in Figure 5. If for the moment the assumption is made that the monthly number of accidents for these drivers would, in the absence of recall problems, remain roughly constant over the three year period, exponential decay curves fitted to the data in Figures 4 and 5 will estimate a memory loss constant representing a constant proportion of accidents being forgotten each month. If A_t is the number of accidents recalled in month t, and A_1 the number recalled in the first month, then the monthly memory loss constant u (corresponding to the annual value U in equation 1) is given by:

$$A_t = A_1 e^{ut}$$

The correction that has to be applied to the total number of accidents reported in a time period T (in months) is then:

$$\frac{uT}{e^{(uT-1)}}$$

This correction is treated in the GLM analysis as an offset.

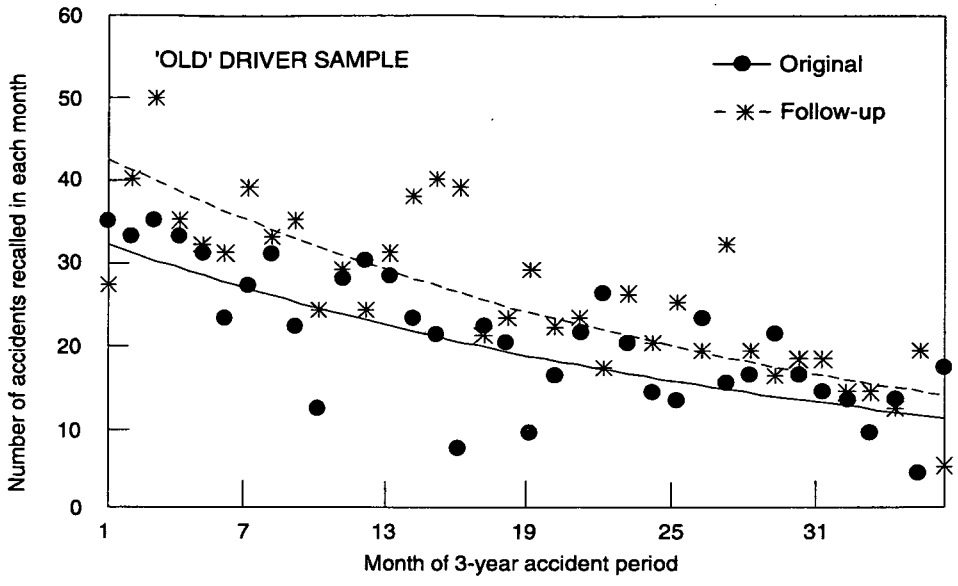

Fig. 4 Dated accidents in each month for 'old' drivers plotted against month of occurrance

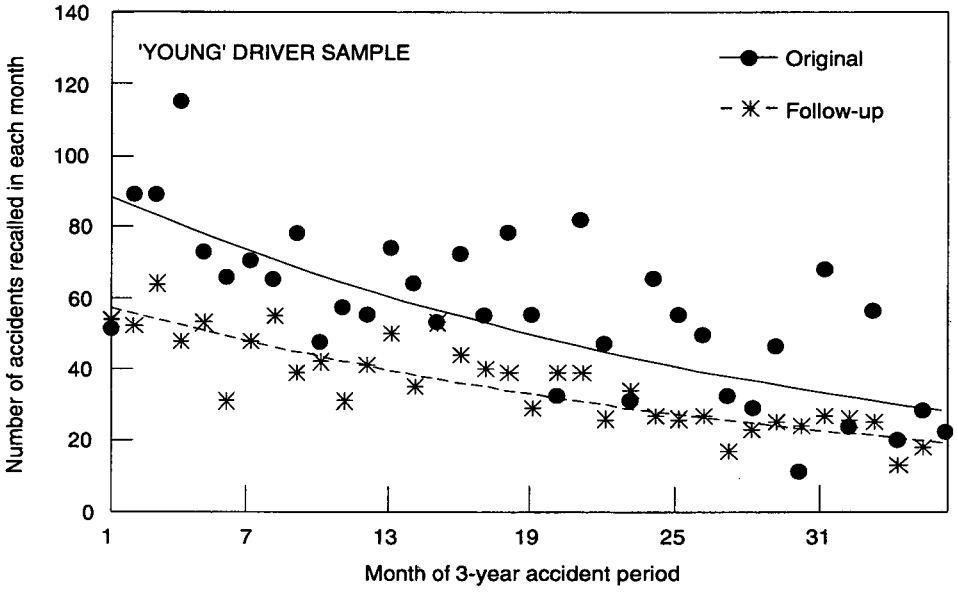

Fig. 5 Dated accidents in each month for 'young' drivers plotted against month of occurrance

The memory loss effect determined from Figures 4 and 5 with their standard errors, together with the proportion of accidents in each sub-sample which were dated, are shown in Table11. Of course, the memory loss constants determined from the data shown in Figures 4 and 5 take no account of the fact that accidents would have declined anyway during the three year period due to the effects of age and driving experience (counteracted to some extent by an increase in mileage); the constants derived from Figures 4 and 5 will thus tend to over-estimate the effect of memory loss. The alternative (and in principle better) way of determining the memory loss effect, is to model the accidents from each year (or part of a year) of the survey separately, and to include the memory loss effect as an additional term in the model as indicated in equation (1) above. Because of inter-correlations between the variables - notably, the memory recall period and the length of driving experience for young drivers in the original survey - the results obtained by the multivariate method depend to some extent on the way the model is structured. Table 11 shows the memory loss effects obtained by incorporating the MEM term (U - the annual effect as shown in equation (1)) into a separate model for each of the four sub-sets of the data. This means that the coefficients for the other terms included in

TABLE 11

The proportion of accidents undated and the annual memory loss constants (U) for the four data sub-group

	Original Survey	S.E.	Follow-up survey	S.E.
Old Driver Sample:				
Proportion of accidents undated	31%		9.7%	
Memory loss constant from Figure 4 (per year)	-0.37	0.07	-0.39	0.05
From GLM (per year)	-0.34	0.05	-0.28	0.04
Young driver Sample:				
Proportion of undated accidents	8.8%		7.4%	
Memory loss constant from Figure 5 (per year)	-0.39	0.07	-0.26	0.07
From GLM (per year)	0.11	0.06	-0.26	0.04

the four models will be different, but the coefficients of the memory loss term so determined may be regarded as being the 'best' estimate for the individual sub-sets of data when the other factors have been taken into account.

Table 11 shows that for the young driver sample in the original survey and for both samples in the follow-up survey, less than 10 per cent of accidents were undated. By contrast nearly a third of the accidents in the old sub-sample of the original survey were undated. This is probably due to inadequate questionnaire layout for the earlier phases of the original survey - too little space was provided for respondents to report the dates of their accidents. The later questionnaires were improved considerably in this respect. Despite however, the lack of dated accidents in the 'old' sub-set of the original survey, the estimate of the memory loss constant for this group of drivers is very similar for the original and the follow-up surveys, suggesting that the omission of the undated accidents is not significantly distorting the memory loss effects.

Taken at face value, Table 11 suggests that the memory loss effects are similar for the three data sub-sets involving a full three year recall period - original/old and both age groups in the follow-up survey; an average value of -0.29 would be appropriate for these data sets. In the original analysis (Maycock et al, 1991) there was some indication that the memory loss coefficient was dependent to some extent on occupational group, but there was no evidence that it interacted with age, experience, or the sex of the driver. Table 11 shows that the memory loss constant for the original/young driver sub-set is however very different. Part of the reason for this is the high correlation between memory recall period and the length of the driver's driving experience. However, even when a model with common coefficients for mileage, age and experience is fitted to all the data or to a subset involving the young drivers (original plus follow-up) only, with separate memory effects for the four sub groups, the original/young driver subgroup has a memory loss constant which is at least negative, but is never larger than -0.06±0.04.

The estimation of the memory loss effect - which unfortunately is crucial to a comparison of the original young driver subset with the other sub-sets - is therefore problematic. It is clearly desirable in planning accident liability surveys of this kind to avoid the need for memory corrections as far as possible, by using a common recall period. For the analysis presented in the following section, a common memory loss constant of -0.29 per year will be used.

7.3 COMPARISONS OF MODELLED ACCIDENT FREQUENCIES

Table 12 shows a comparison of the four sub-sets of data using the common memory loss constant u=-0.29, a common exponent for mileage (0.3), and common coefficients for age and experience, and a factor F to distinguish the four data sets.

$$A/T = 0.0038\ F\ M^{0.3} \exp\{17/Ag + b_2/(X + 2.2)\} \quad (2)$$

where b_2=3.3 for male drivers and 2.3 for female drivers.

The factor F in equation (2) and Table 8 indicates in effect, the ratios of the absolute values of accident liability of the three sub-samples follow-up/old, original/young and follow-up/old in relation to the original/old sub-sample as a reference. Bearing in mind the fact that the errors shown in Table 12 are the estimated *standard errors* of these ratios (not the 95% confidence limits), it will be seen that none of the four sub-samples differs significantly at the 5% level. In view of the quite large differences in accident liabilities

TABLE 12

Original and follow-up surveys - modelled accident comparisons in terms of the factor F (with standard errors)

Value of F for:	Original 1987/88 survey	Follow-up 1990/91 survey
OLD sub-set	1.0 (Reference Level)	1.009±0.05
YOUNG sub-set	1.04±0.10	0.89±0.07

apparent in Tables 5 and 6, this result confirms the appropriateness of the average mileage, age and exposure terms in 'correcting' the data for variations in these parameters between the sub-samples.

The comparison in the first row of Table 12 between the old driver sub-sample in the original survey with the same drivers in the follow-up survey is particularly important because all drivers reported on a full 3 years of accident data; this comparison is therefore unaffected by the memory loss correction. It will be seen that the accident liabilities reported by these drivers in the two 3-year periods when corrected for mileage, age and driving experience, differ by less than 1 per cent.

In the case of the young drivers, a memory loss correction has had to be applied. As was indicated earlier, establishing the most appropriate value of the memory correction factor is not straightforward and an overall value of -0.29 has been used. A less negative value (say -0.2) would produce a ratio between the two surveys which was closer to 1, but since it is not possible to determine with any confidence what the most appropriate value should be, the comparison between the two surveys for the young drivers does not enable the reproducibility of accident liability data from this particular sub-sample to be assessed with any degree of reliability.

Table 13 shows the coefficients of mileage, age and experience (both in their reciprocal forms) for men and women separately obtained when the four sub-sets of data summarised in Table 12 are modelled separately (again with the overall memory loss correction term included).

Table 13 shows that the mileage exponent is determined robustly and consistently in all four data sub-sets. The experience effects, although not determined particularly robustly within the individual data sub-sets, are also reasonably consistent. By comparison, the age effects are considerably less consistent between the data sub-sets, and are particularly poorly determined for the young drivers. This is mainly because the age range for these young drivers is limited to 5 years at most (20-25 in 1990 when the follow-up survey was carried out) compared with 35 years (25-80) for the older drivers. Since the age effect in accident liability studies of this kind is relatively large, when designing studies of this kind it is important to decide whether age effects need to be taken into account, and if they do, to select a sample with an adequate age range.

8. OTHER ASPECTS OF ACCIDENT LIABILITY SURVEYS

8.1 COMPANY CAR DRIVERS IN THE 1990/91 FOLLOW-UP SURVEY

The follow up survey (1990/91 - including the young driver sub-set) included the question: 'Is the car you drive most, company or business owned or privately owned?' Table 14 shows the average accident frequencies, mileages and ages for the sample of 8,617 drivers who responded to this

TABLE 13

Model coefficients for the four sub-sets of data separately with their standard errors.

Model terms	Young sub-set		Old sub-set	
	Original	Follow-up	Original	Follow-up
Log constant	-5.0 ± 0.7	-4.4 ± 0.6	-6.2 ± 0.3	-5.1 ± 0.3
Mileage exponent (α)	0.31 ± 0.03	0.30 ± 0.03	0.36 ± 0.04	0.26 ± 0.04
1/Age	6 ± 14	-17 ± 15	22 ± 4	12 ± 5
1/(Experience+2.2) M	3.4 ± 0.7	5.6 ± 1.3	3.2 ± 0.9	3.7 ± 1.5
F	2.2 ± 0.7	4.9 ± 1.3	2.6 ± 0.9	3.4 ± 1.5

TABLE 14

Company car drivers compared with private car drivers in the 1990/91 follow up survey (including young drivers)

	Drivers of company owned vehicles		Driver of privately owned vehicles		Overall averages	
	Male	Female	Male	Female	Male	Female
Accidents in 3 years	0.427	0.342	0.293	0.237	0.323	0.245
Annual mileage	20,790	11,470	9,660	5,520	12,230	5970
Age (November 1990)	35.2	33.8	39.9	37.0	38.8	36.8
Number of respondents	937	333	3,234	4,041	4,171	4,374

question disaggregated by company car drivers and private car drivers. In order to avoid any difficulties arising from differential memory loss effects, only accident data referring to a complete 3 year period will be used in this section, and no corrections will be made for memory loss effects.

It will be seen from Table 14 that 22 per cent of men and 8 per cent of women said they drive company owned cars. Moreover, the drivers of company owned cars drove nearly twice the annual mileage of those driving privately owned cars.

Comparing the accident frequencies (per 3 years) of company car drivers with private car drivers, it is clear from Table 14 that the average accident frequency of company car drivers is considerably higher than that of private car drivers for both sexes. Of course, some of this difference will be due to the higher mileage covered by the company car drivers combined with the fact that they are younger and less experienced as drivers. When the effects of age, experience and mileage are adjusted for using multivariate analysis (not reported here) - in which it turns out that the sex differences are largely accounted for by the differences in mileage and age - company car drivers have a 29 per cent higher accident frequency than the drivers of privately owned cars. The survey did not collect sufficient details of the company car drivers or the mileage they did to provide a 'profile' of their driving characteristics. In view of the average mileages covered it seems likely, however, that many of these drivers were driving as part of their job.

8.2 COMPARING ACCIDENT LIABILITY RESULTS FROM THE ORIGINAL AND FOLLOW-UP SURVEYS WITH OTHER SURVEYS.

The main purpose of this report is to explore the reliability of self-reported accident data. Sections 6 and 7 in this report have concentrated on comparing the reported accident data from the same drivers in two successive time periods each of 3 years duration. However, another way of assessing data reliability would be to take repeat independent samples, and to compare the accident rate reported by these groups of different drivers. Although it is not possible here to make such a comparison using samples of drivers selected according to an identical sampling strategy, two other self report surveys have been conducted since the follow-up survey of 1987-90, and in the context of data reliability and repeatability it is of interest to compare the overall accident frequencies obtained in these surveys with those obtained in the original and follow-up surveys reported above.

The most appropriate basis for this comparison is the 'old' sample of drivers obtained for the original and the follow-up surveys (remember, this was a sample roughly evenly distributed across the age range with equal numbers of male and female drivers). As a basis for comparison, Table 15 shows the overall accident frequencies (accidents in 3 years - no corrections for memory loss), annual mileage and age for the original survey sample and those that responded to the follow-up survey.

It will be seen from Table 15 that the sub-set of drivers who responded to the follow up survey were somewhat younger and covered slightly higher annual mileages than the drivers who responded in the original survey. Despite these differences (both of which would be expected to increase accident liability), the accident liability of the follow-up sub-set was a few per cent lower than that of the original sample - though these differences are not statistically significant.

Comparing the change in the same group of drivers from 1987 to 1990 shows that over this 3-year period annual mileage has increased (as already discussed in section 5 above); accidents involving men have fallen by about 11 per cent - an amount consistent with the extra 3 years of age and driving experience between the surveys - and accidents involving women drivers appear to have increased by a non-significant 5 per cent.

The accident data in Table 15 may be compared to data collected using similar methods in a recent survey under-

TABLE 15

Overall accident frequencies, annual mileage and age for the original and follow-up respondents.

	Drivers who responded to the original survey 1987 data		Drivers who responded to the follow-up survey 1987 data		1990 data	
	Male	Female	Male	Female	Male	Female
Accidents in 3 years	0.260	0.180	0.245	0.174	0.217	0.182
Annual mileage	10,450	4,530	11,080	4,970	11,700	5,325
Age (November 1987/90)	47.5	46.7	46.3	43.7	49.3	46.7
Number of respondents	5,881	6,605	2,593	2,633	2,593	2,633

taken in August 1994 related to fatigue and driving and a survey related to speed enforcement undertaken in March 1993. In the case of the fatigue driver sample, the sample consisted of male drivers only, sampled to include drivers of all ages approximately uniformly across the age range (as in the case of the original 1987 survey); the drivers in the fatigue survey sample were asked whether or not they drove a company car. Table 16 shows the average accident frequencies, annual mileages and ages of those that drove company cars and those that didn't.

Table 16 shows that about 17 per cent of male drivers drove a company car - a figure not dissimilar to that given in Table 14 (section 8.1) for the 1990 follow-up survey. Also like that survey, the company car drivers drove just over 20,000 annually, and were somewhat younger than the drivers of privately owned cars. The accident frequencies given in Table 16 confirm the findings of the earlier survey in showing that drivers of company cars have a considerably higher accident frequency than the drivers of privately owned vehicles. The final column in the Table provides overall average values for comparison with the 1987/90 sample results given in Table 15. Bearing in mind the inevitable differences between the samples, the averages presented in Table 16 and those given in Table 15 are very similar.

Table 17 shows the average accident liabilities of a sample of drivers involved in a survey related to the enforcement of speed. In the course of this study, a random sample of just over 10,000 drivers in different parts of the country responded to a postal questionnaire which asked them to report the number of accidents in which they had been involved in the last three years. They also reported their annual mileage and age (among other things). Although these drivers were asked whether they were professional drivers and about the use of their car for business purposes, the questions and responses relating to the use of company cars were not compatible with the other surveys. Accordingly, only the overall averages are given in Table17.

Comparing the accident rates in Tables 15 and 16 with Table 17 for male drivers shows a remarkable degree of consistency both in annual mileage, and in average accident frequencies, bearing in mind that the surveys were conducted over a period of 7 years. For female drivers, a comparison of Tables 15 and 17 suggests that the average accident frequency obtained in the speed survey data is a little higher than that obtained in the 1987/90 survey though the difference is not particularly large and is consistent with the higher mileage and younger age of the drivers in the speed survey.

TABLE 16

The average accident frequencies of a structured sample of male drivers obtained during a study of fatigue and driving.

	Drivers of company owned vehicles	Driver of privately owned vehicles	Overall averages
Accidents in 3 years	0.310	0.198	0.218
Annual mileage	21,030	9,380	11,400
Age (August 1994)	42.6	48.8	47.7
Number of respondents	797	3,800	4,597

TABLE 17

The average accident rates of a random sample of drivers obtained during a study of speed enforcement.

	Male	Female	Overall average
Accidents in 3 years	0.245	0.229	0.240
Annual mileage	11,160	6,200	9,630
Age (March 1993)	47.5	42.1	45.5
Number of respondents	7,422	3,333	10,775

These results support the contention of this report that self-reported accident data can be regarded as a reliable source of data for studying the relationships between accidents and a range of other variables.

9. SUMMARY AND CONCLUSIONS

Drivers who had taken part in a survey in 1987/88 about their driving experiences and accidents were followed up in another survey three years later in order to gather information about their current driving and any accidents that they may have had in the intervening three years. Such a survey was carried out for the following reasons:

1. To examine the methodology of carrying out longitudinal postal surveys of large samples of drivers, and to investigate the reliability of measures taken.

2. In particular, to examine the changes in driving exposure (annual mileage) and accidents for the same drivers over time.

A comparison of the data collected from the same drivers in these two surveys (termed here the 'original' and the 'follow-up' survey) has provided an opportunity to compare the reproducibility of the reporting of mileage and accidents over these two periods. This comparison will not assess the accuracy of an individual's recall for the same event, but will provide some indication of whether the techniques used to elicit information yield consistent and comparable results over time. The opportunity has also been taken to compare the overall accident frequencies reported in the original and the follow-up surveys with two more recent self-report surveys - one carried out for a study of fatigue and driving (1994) and the other in a study of speed enforcement (1993).

The main findings from the follow-up can be summarized as follows:

(i) On average annual mileage increased by about 18 per cent between the surveys. This increase seems to be rather higher than expected from national data, but there is no way of determining whether it is real or an artifact of the survey methodology.

(ii) The annual mileages reported in the original survey (1987/88) by those drivers who claimed in the follow-up survey not to have changed their mileage 'from last year', were compared with their reported mileages in the later survey (1990/91). Although in a statistical sense the two mileages were proportional to one another, there was a great deal of scatter in the individual data points. If this scatter was solely due to errors in estimating mileages (rather than genuine changes), it is estimated that the standard deviation of the mileage error is between 46 per cent and 62 per cent of its mean value; in other words individual mileages may well be rather poorly estimated.

(iii) The proportion of time spent by drivers driving on three types of road - motorways, roads in built-up areas, and roads outside built-up areas - was reasonably consistent from survey to survey.

(iv) Average accident frequencies (reported accidents per year) ranged from over 0.3 for 'young' male drivers in the original survey to 0.06 for 'old' women drivers in both surveys. Tabulations by age and sex showed that accident frequencies decreased with age for both men and women, and were considerably lower for women than for men. Apart from the youngest age group there was little difference between the accident frequencies observed in the two surveys.

(v) Fitting a simple statistical model which took into account the effects of annual mileage, age, sex and driving experience, showed that the residual difference between the original and the follow-up survey for the same drivers in the 'old' subgroup was less than 1 per cent. Accident data collected by this self-report method would therefore seem to have been highly consistent over the two 3 year periods involved.

(vi) The young driver accident data in the original survey involved a range of reporting periods. In order therefore to make comparisons of the accident frequencies reported by these drivers between the surveys, a correction has to be made for the fact that a variable proportion of accidents will be forgotten. Although a memory loss coefficient could be determined satisfactorily for the data as a whole, it was not possible to determine a satisfactory coefficient for the young driver sample on its own. In carrying out such studies therefore it is strongly suggested that drivers are asked to recall accidents over the same fixed period. A practical maximum would be the 3 year period used in these surveys.

(vii) A comparison of drivers claiming to be driving a company car with those driving a privately owned car shows that accident frequencies of the company car drivers are considerably higher - a difference which remains when the effects of annual mileage and driver age have been allowed for.

(vii) An overall comparison of the average accident frequencies reported by the older respondents in both the original and the follow-up survey with two other surveys carried out more recently shows that the overall self-reported accident frequencies are very similar in the four surveys, confirming the reliability of the self-report method for use in accident studies.

10. REFERENCES

AITKIN, M, ANDERSON, D, FRANCIS, B and J HINDE (1992) Statistical modelling in GLIM. Oxford Science Publications, OUP, Oxford.

CHAPMAN, PR & GROEGER, JA (1992). Subjective risk and subsequent memory. In GB GRAYSON & JF LESTER (Eds) Behavioural Research in Road Safety II. Proceedings of a seminar held at Manchester University, Sept 1991. *Department of Transport. PA2193/92.* Transport Research Laboratory, Crowthorne.

DIGES, M (1988). Stereotypes and memory for real traffic accidents. In MM GRUNEBERG, PE MORRIS & RN SYKES (Eds.) Practical Aspects of Memory: Current Research and Issues Vol 1: Memory in everyday life. John Wiley & Sons, Chichester.

GASKELL, G WRIGHT, D O'MUIRCHEARTAIGH, C (1992). Psychological processes in retrospective reporting. Paper presented at the JCSM Survey Methods Seminar, 'How well can survey respondents remember events and behaviour', *JCSM Newsletter 12(2).*

LOFTUS, E SMITH, KD JOHNSON, DA FIEDLER, J (1988). Remembering "when": errors in the dating of autobiographical memories. In MM GRUNEBERG, PE MORRIS & RN SYKES (Eds.) Practical Aspects of Memory: Current Research and Issues Vol 1: Memory in everyday life. John Wiley & Sons, Chichester.

MAYCOCK, G, LOCKWOOD, C R and Julia LESTER (1991). The accident liability of car drivers. *Department of Transport TRL report RR 315:* Transport Research Laboratory, Crowthorne.

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

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

WHITE, S B (1976). On the use of annual vehicle miles of travel estimates from vehicle owners. *Accident Analysis & Prevention 8,* pp257-261

APPENDIX A: THE QUESTIONNAIRE USED IN THE FOLLOW-UP SURVEY

TRRL/RUB/90-91

DRIVING SURVEY

Please answer all the questions in the three sections of this questionnaire by either ticking the appropriate box(es) or writing in the required details. Only give details of journeys where you were a ***driver***, not a passenger. Any information that you give will of course be treated in the strictest confidence.

SECTION 1

1. Have you driven a car or a van at any time during the last year?

TICK ONE BOX ONLY

YES ..... ❑ 1
NO ..... ❑ 2

If YES go on to the next question
If NO go to section 2 on page 2

2. On average, how often have you driven a ***car*** or ***van*** during this time?

TICK ONE BOX ONLY

everyday ..... ❑ 1
two or three times a week ..... ❑ 2
about once a week ..... ❑ 3
less often ..... ❑ 4

3a. About how many miles have you driven in a car or a van during the last year? Add mileage in different cars or vans together. (It may help to think of the number of miles you drive in a typical week, multiply this number by 50, and then add the mileage of any extra journeys, e.g. driving on holiday, or special occasions).

WRITE IN THE NUMBER

..... miles

3b. Was all or nearly all of this mileage done in the ***same*** car or van?

TICK ONE BOX ONLY

YES ..... ❑ 1
NO ..... ❑ 2

If YES go to question 3c
If NO go on to question 4

3c. What is the make and model of this vehicle?

WRITE IN THE DETAILS

make .....

.....

model .....

3d. Is this vehicle 12 years old or less?

TICK ONE BOX ONLY

YES ..... ❑ 1
NO ..... ❑ 2

If YES what is the year of manufacture or registration letter?

Write in the details .....

4. Is the car or van you drive most?

TICK ONE BOX ONLY

company or business owned ..... ❑ 1
privately owned ..... ❑ 2

5. Thinking about how far you have driven in a car or van this year in comparison with last year, do you think you have driven...

TICK ONE BOX ONLY

more this year than last year ..... ❑ 1
less this year than last year ..... ❑ 2
about the same amount ..... ❑ 3

6. Have you driven any other vehicles, e.g. a lorry or a motor cycle, at any time during the last year?

TICK ONE BOX ONLY

YES ..... ❑ 1
NO ..... ❑ 2

If YES go on to the next question
If NO go on to question 8

7. About how many miles have you driven in these ***other*** vehicles during this time? Add mileage in different vehicles together

WRITE IN THE NUMBER ..... miles

8. Please estimate how much of your driving ***time*** you have spent, in ***all kinds of vehicles***, on the following types of roads in the last year.

WRITE IN THE AMOUNT AS A PERCENTAGE

motorways ..... %

roads in built up areas ..... %

roads outside built-up areas ..... %

TOTAL 100%

Page 1

SECTION 2

We would now like you to tell us about all kinds of road accidents that you have been involved in, *as a driver*, over **the last three years**. By 'accident' we mean any incident which involved injury to another person or yourself, damage to property, damage to another vehicle, or damage to the vehicle that you were driving.
Please mention only those incidents which occurred on public roads, not on private property, and in which you were involved *as a driver*, not as a passenger; please include *all* accidents, regardless of how they were caused, or how slight they were.

9a. How many accidents have you been involved in during the last three years?

WRITE IN THE NUMBER..... (if none write zero)

If NONE go to question 20 on page 4

9b. How many of these accidents were in a car or van?

WRITE IN THE NUMBER..... (if none write zero)

9c. And of these car or van accidents, how many involved physical injury to yourself or others?

WRITE IN THE NUMBER ..... (if none write zero)

We would now like you to give further details of your LAST THREE ACCIDENTS IN A CAR OR VAN DURING THIS PERIOD using the following table.

Please answer questions 10 - 19 for your most recent accident by ticking the boxes in the appropriate columns below. Then answer questions 10 - 19 again for any earlier accidents.

If you have not been involved in an car or van accident at all during the last three years, go straight to question 20 at the beginning of section 3 on page 4.

	MOST RECENT ACCIDENT	NEXT MOST RECENT	ONE BEFORE THAT
Date of accident WRITE MONTH/YEAR	/	/	/

10. Did the accident occur during...

TICK ONE BOX ONLY

daylight	❑ 1	❑ 1	❑ 1
hours of darkness	❑ 2	❑ 2	❑ 2

11. What type of vehicle were you driving?

TICK ONE BOX ONLY

company or business owned car or van	❑ 1	❑ 1	❑ 1
privately owned car or van	❑ 2	❑ 2	❑ 2

12. What type of road were you driving on?

TICK ONE BOX ONLY

motorway	❑ 1	❑ 1	❑ 1
road in a built up area	❑ 2	❑ 2	❑ 2
road not in a built up area	❑ 3	❑ 3	❑ 3

13. Apart from your vehicle, what else was involved?

TICK MORE THAN ONE BOX IF NECESSARY

moving vehicle(s)	❑ 1	❑ 1	❑ 1
stationary vehicle(s)	❑ 1	❑ 1	❑ 1
parked vehicle(s)	❑ 1	❑ 1	❑ 1
pedestrian(s)	❑ 1	❑ 1	❑ 1
motor cyclist/cyclist(s)	❑ 1	❑ 1	❑ 1
other e.g. lamp-post, road sign, kerb, island, tree, hedge, animals etc.	❑ 1	❑ 1	❑ 1

	MOST RECENT ACCIDENT	NEXT MOST RECENT	ONE BEFORE THAT

14. Where there any injuries to yourself or others?

TICK MORE THAN ONE BOX IF NECESSARY

none	❑ 1	❑ 1	❑ 1
slight injuries (e.g. cuts and bruises)	❑ 1	❑ 1	❑ 1
serious injuries (needing hospital care)	❑ 1	❑ 1	❑ 1
fatal injuries	❑ 1	❑ 1	❑ 1

15. How much damage was done to your car or van?

TICK ONE BOX ONLY

no damage	❑ 1	❑ 1	❑ 1
slight damage (dents and scratches)	❑ 2	❑ 2	❑ 2
serious damage	❑ 3	❑ 3	❑ 3
total loss/write off	❑ 4	❑ 4	❑ 4

16. Which of the following best describes the worst harm caused to any vehicle other than your own?

TICK ONE BOX ONLY

no other vehicle involved	❑ 1	❑ 1	❑ 1
no damage	❑ 2	❑ 2	❑ 2
slight damage (dents and scratches)	❑ 3	❑ 3	❑ 3
serious damage	❑ 4	❑ 4	❑ 4
total loss/write off	❑ 5	❑ 5	❑ 5

17. Was the accident reported to the police?

TICK ONE BOX ONLY

YES	❑ 1	❑ 1	❑ 1
NO	❑ 2	❑ 2	❑ 2
don't know	❑ 3	❑ 3	❑ 3

18. Did you make a claim from your insurance company?

TICK ONE BOX ONLY

YES	❑ 1	❑ 1	❑ 1
NO	❑ 2	❑ 2	❑ 2

19. Please describe briefly how and where each accident happened

Most recent _____

_____

_____

_____

Next most recent _____

_____

_____

_____

One before that _____

_____

_____

_____

SECTION 3

20. In what year did you pass your driving test?

WRITE IN THE YEAR

19 .....

If you are unsure about the exact date, please write in an approximate date, or write in the space below the approximate number of years you have been driving.

..... Years

21. What was your age last birthday?

WRITE IN THE NUMBER

..... Years

22. Are you...

male ..... ❑ 1

female ..... ❑ 2

23. Thank you for your help in the past; are you still willing to be contacted in any future research into driving?

YES ..... ❑ 1

NO ..... ❑ 2

Finally are there any comments you would like to make about driving, particularly with regard to changes over the past three years?

Please check that you have answered all the questions as fully as possible.
Please return your completed questionnaire in the envelope provided.

THANK YOU VERY MUCH FOR YOUR HELP

APPENDIX B: THE STATISTICAL MODELS

B.1 MODEL STRUCTURE

A statistical model of the kind reported in 7 above has three components: (i) a systematic component - the relationship between the dependent variables (accident frequency) and the significant explanatory variables, (ii) the sampling error associated with the dependent variable - in this case assumed to have a Poisson error structure, and (iii) the errors due to the lack of fit of the model. The lack of fit component may arise either because incorrect functional forms have been used for those variables included in the model, or because some key variables have not been included at all.

As far as the Poisson error structure is concerned, it is reasonable to treat accidents as though they were random events generated at an average rate corresponding to the individual driver's accident liability. This means that if the model fitted perfectly, the actual number of accidents a driver would experience in a year would be represented by a Poisson process whose mean value is given by the model prediction. In fact, the type of model used in this report does not explain all the non-Poisson variability in the data with the result that the residuals are over-dispersed compared with a pure Poisson process. This over-dispersion can be handled in a number of ways. Since in the present case the over-dispersion is generally not large as a proportion of the total residual variability, no corrections have been made for it.

The basic form of the models is multiplicative - that is to say, all the elements of the model can be regarded as factors which are multiplied together to obtain the final estimate of accident liability. This type of multiplicative model has proved to be more satisfactory than additive alternatives.

Decisions about what terms to include and the functional form of the variables included are made using the statistic 'scaled deviance' as the indicator of goodness of fit. The process is described briefly in the following section.

B.2 FITTING THE MODELS

The models have been fitted using the Generalised Linear Modelling package GLIM4 (National Algorithms Group, Oxford).

To assess whether a new explanatory variable is worth including in the model, or whether a term is being included in the most appropriate functional form, goodness of fit is judged using a likelihood ratio statistic called 'scaled deviance'. Providing that the 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 included in the model). When the average mean value of the dependent variable falls below 0.5 (as is the case with accident frequencies in the present study) the deviance of the final model cannot be used as an overall measure of goodness of fit since under these conditions it ceases to be a chi-squared variable. Instead, the generalised Pearson chi-squared statistic X^2 is appropriate. Fortunately, the deviance *differences* obtained when new terms are added to the model are still chi-squared variables so a comparison of deviance difference with the appropriate point of the chi-squared distribution can be used to assess the significance of adding terms or modifying functional forms of the terms already included. Thus, if only one additional explanatory variable is being added to a model, the change in deviance has to reach 3.84 (the $p = 0.05$ point of the χ_1^2 distribution) to be significant at the 5% level and 6.64 to be significant at the 1% level.

Variables may be introduced into the models as continuous variables or as multi-level factors which are available in the form of categories within the data. In the case of factors, deviance difference is used to assess the usefulness of the factor as a whole - including all the levels; the significance of the individual levels has to be assessed using the standard errors computed by GLIM for the individual categories.

If a parameter is included in the model in a form which doesn't allow the GLIM fitting routine to calculate a coefficient and standard error directly - such as the constant 2.2 included in the experience term - the statistical significance of the addition of this parameter can still be assessed using deviance difference as for any other term. However, estimates of the optimum value for the constant and appropriate confidence intervals have to be obtained using a process which is called 'profile deviance' (see Aitkin et al, 1992). The process requires the model to be re-fitted for a range of values of the parameter of interest so that the value of the parameter can be optimised and its confidence limits estimated from plots of residual deviance.

MORE INFORMATION

The Transport Research Laboratory has published the following other reports on this area of research:

RR270 Factors affecting the accident liability of motorcyclists - a multivariate analysis of survey data. M C Taylor and C R Lockwood. Price Code C.

RR3I5 The accident liability of car drivers by G Maycock, C R Lockwood and Julia F Lester. Price Code C.

PR111 Cohort study of learner and novice drivers: Part 3 Accidents, offences and driving experience in the first three years of driving by B Forsyth, G Maycock and B Sexton. Price Code J.

TRLl69 Driver sleepiness as a factor in car and HGV accidents by G Maycock. Price Code H.

If you would like copies, photocopy and fill in the slip below. There is a 20% discount if you take all the reports listed above. Prices include postage and are correct at the time of publication. Please see the enclosed letter for current price code values and handling charge. Enquiries to TRL Library Services, Tel: 01344 770784, Fax: 01344 770193.

To: Publication Sales, TRL Library, PO Box 304, CROWTHORNE, Berkshire, RG45 6YU.
Please send me the following Transport Research Laboratory reports (state report Nos and quantity)

Report no ..... *Quantity* .....
Report no ..... *Quantity* .....
Report no ..... *Quantity* .....

Report no ..... *Quantity* .....
Report no ..... *Quantity* .....
Report no ..... *Quantity* .....

Name .....
Address .....
.....
.....
Postcode .....
Telephone .....
.....
Credit card address (if different from above).....
.....

PAYMENT:
• I enclose a cheque for £ .....
payable to TRL Ltd.
• Please debit my Deposit Account
no .....
• Please debit my Credit Card by £.....
• Credit card no .....
Expiry date .....
Signature .....