DEPARTMENT OF TRANSPORT

FEDERAL OFFICE OF ROAD SAFETY

DOCUMENT RETRIEVAL INFORMATION

Report No.	Date	Pages	ISBN	ISSN
CR 50	April,1987	i - xii 1 - 101	0642 513546	OR = 0158-3077 CR = 0810-770 X

Title and Subtitle EXAMINATION OF THE RELATIONSHIP BETWEEN ROAD TRAFFIC CRASHES AND VIOLATIONS IN THE ACT

Author (s) DOUGLAS E SHAW
CHRISTOPHER D ROWLES

Performing Organisation (Name and Address)

SIROMATH Pty Ltd,
Level 5, 156 Pacific Highway,
ST.LEONARDS NSW 2065

Sponsor (Name and Address) Federal Office of Road Safety
G.P.O. Box 594
CANBERRA A.C.T. 2601

Available from (Name and Address) **Price/Availability/Format**

Federal Office of Road Safety

Abstract Data from the Australian Federal Police on traffic offences in the ACT and data from the National Capital Development Commission on road crashes in the ACT were combined to form a data base linking offence histories and crash histories of drivers. Various summaries of this data are examined to determine relationships between offence histories and crashes. Two statistical techniques were applied to the data to obtain relationships which might be used to identify drivers who are in the future more likely to be involved in a crash. Neither of these techniques (linear discriminant analysis and generalised linear models) could establish a relationship.

KEYWORDS: Crashes, Violations, Drivers, Motor Cyclists, Statistical Analysis, Statistical Models, Correlation

NOTES:

(1) FORS research reports are disseminated in the interests of information exchange.

(2) The views expressed are those of the author(s) and do not necessarily represent those of the Commonwealth Government.

(3) The Federal Office of Road Safety publishes two series of research report
  (a) reports generated as a result of research done within the FORS are published in the OR series;
  (b) reports of research conducted by other organisations on behalf of the FORS are published in the CR series.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance received from The Australian Federal Police and from the National Capital Development Commission in the provision of the data for this study,and the assistance provided by those organisations in the interpretation of the data.

The authors would particularly like to thank Mr Bill Danaher of the Federal Office of Road Safety for his assistance,encouragement and understanding during the course of the study.

CONTENTS

LIST OF TABLES

Table

Table

Table

LIST OF FIGURES

EXECUTIVE SUMMARY

This study has analysed data from the Australian Federal Police on traffic offences recorded against drivers in the ACT in the period 1980–86, and data from the National Capital Development Commission on road crashes in the ACT in the period 1982–mid 1985. Any information on direct driver identification was removed from the data files before they were supplied to the authors. From these extensive data files a data base was created which links the offence histories and crash histories of drivers. The aim of the study was to examine relationships between offence histories and crashes, with a view to determining whether offence history could be used to predict the driver's propensity to be involved in a crash.

The offence data and crash data were summarised in various ways, leading to the following observations.

- Proportions of offences in various categories were consistent through the years for which data was available, with speeding offences accounting for 43% of all offences.

- Drivers against whom only a single offence had been recorded had more speeding offences and offences of failing to obey lights, signals, markings and directions than did drivers against whom more than one offence had been recorded. Conversely, drivers with multiple offences had more offences in the category of failure to obey licencing and registration regulations and in the category of equipment offences.

About three-quarters of drivers involved in crashes had no prior recorded offence history. However, in examining drivers with offence histories, the number of crashes per 100 drivers increased with the number of offences. That is, the more offences a driver has the more likely he/she is to be involved in a crash.

In looking at various sub-groupings of the data, the following were observed.

- Drivers not involved in crashes had more speeding offences than did drivers who were involved in crashes. However, the drivers involved in crashes had a higher proportion of offences relating to driving, overtaking, passing and pulling out, and a higher proportion of offences relating to equipment.

- Drivers had fewer speeding offences and fewer offences of failing to obey lights, markings, signals and directions after their first crash. However, after their first crash they had a higher proportion of equipment offences and offences relating to licencing and registration.

- There was no discernible difference as regard the proportions of different categories of offences between car drivers involved in crashes and motor cycle riders involved in crashes. Similarly, there was no discernible difference in the proportions of offence types between young and old drivers involved in crashes.

Two statistical techniques were applied in order to obtain a relationship between crash involvement and offence history which might be used to identify drivers who are in the future more likely to be involved in a crash. Neither of the techniques,linear discriminant analysis and generalised linear models,were successful in establishing a usable relationship. This inability to establish a relationship was due,at least in part,to the absence from the data available of any information about drivers with neither an offence history or a crash history. A more conclusive analysis could be performed if offence and crash data were available from a cohort of licenced drivers observed over a period of years.

INTRODUCTION

The cost to the community of road crashes is difficult to assess; in addition to the direct economic costs, there is the incalculable cost of human suffering and disruption. The economic costs of road crashes are substantial in their own right; in 1983 in Australia they were estimated conservatively at $2.7 billion, or about 2% of Gross National Product (Office of Road Safety, 1984). In view of these costs, road safety programmes aimed at reducing the incidence and /or severity of road crashes are clearly an appropriate response.

One avenue of approach to road safety is to identify a group of drivers as having a "high risk" of subsequent crash involvement, and to take steps to reduce the likelihood of these drivers having subsequent crashes. Such steps could include driver behaviour modification through education or licence restriction, or removal from the at risk group by licence suspension or revocation. There is, however, no established method of identifying the group of drivers having a high risk of crash involvement, even though there is general agreement that such a group does exist. (Although the high risk group is generally acknowledged as existing, it is also generally thought that they are involved in only a small percentage of crashes, i.e most crashes involve the drivers who would not generally be seen as having high risk of crash involvement).

Many jurisdictions approach the identification of high risk drivers by a "demerit" points system applied to offences. In such a system, the drivers accumulate "points" for traffic offences recorded against them, with some offences rating higher points than others. When the driver's accumulated points reach a certain level, action is taken to reduce the driver's risk of crash involvement. (There are also various schemes to "retire" demerit points after the expiry of a suitable time, etc). The rationale for such a "demerit" points system is that some offences are more likely to be associated with crashes than others, and that the accumulation of points identifies a high risk driver. The assumptions underlying a "demerit" points system are not fully supported by the research that has been carried out. (see Review of Related Studies for more details of studies based on large samples of drivers in the USA and Canada).

Most studies have been unable to find a "useful" relationship between specific types of offences and crashes. By "useful" in this context we mean a realtionship with predictive ability; many of the relationships studied led to statistically significant measures of association, but fell short of predictive ability. (Again, this is canvassed in Review of Related Studies). However, there are many possible reasons for this failure to find useful relationships, lying in the data sets used to carry out the analyses and in the offences selected for the study in the analyses.

Data on offences and crashes for the Australian Capital Territory have recently become available in computer-accessible form, and it is believed that this data set will be a reliable one for a study of offence/crash relationships. There are sufficient licenced drivers in the ACT (140 000 in 1982) to provide a data set of adequate size and the proportion of all crashes that are reported in the ACT is high by international standards. (7.1 crashes per hundred licenced drivers are reported in the ACT versus 3.5 in Calfornia (McConnell and Hagen, 1980)). Further, the geographical extent of the area is small, so that relatively uniform enforcement and reporting will apply throughout the area.

In this Report, we will discuss the data that was made available and the steps taken to create, from the data, a data base appropriate to the study of relationships between offences and crashes. We will then proceed to analyse the data to examine in some detail the differences in offence histories between drivers involved in crashes and drivers not involved in crashes. Where feasible, we obtain separate analyses for drivers and motor cycle riders. Attention will also be paid to the crash and offence patterns of "recidivists" - repeat offenders. Finally, we will develop models to relate the propensity of drivers to be involved in crashes to their driving offence history, and examine the implications of such a model for identification of "high risk" drivers and riders.

REVIEW OF RELATED STUDIES

There have been a number of studies examining the relationship between traffic offences recorded against drivers and the crash record of that group of drivers. Some of these studies relate to jurisdictions where some form of point-count system operates, so that attention is directed to the points awarded for offences rather than the offences themselves; this tends to result in a very broad classification of offences, e.g. "no points, one point, two points" as in California. Many of the studies have, as at least part of their aim, the assessment of an existing point-count system or the formulation of one.

The studies which relate to this Report can conveniently be divided into three groups, based on the data sets that they analyse. These data sets are very substantial, and hence provide some measure of confidence in the results from them. The three groups of studies are based respectively on data from California, North Carolina and Ontario.

1. CALIFORNIAN STUDIES

The California Driver Record Study was based on a sample of approximately 148,000 drivers entered into the study in 1964. A number of analyses were published as the Study proceeded. The 4th Part (Coppin, Lew and Peck, 1965) examines briefly the relationship between numbers of offences and numbers of crashes, without distinguishing between the types of offences. The offences in this study were those carrying points in the California system, and offences related to crashes were excluded on the grounds that they would bias the relationships. The analysis is concurrent, i.e. offences and crashes in the same time period are analysed. Over a 3 year period, it was found that increased offence frequency was accompanied by a significant increase in crash involvement.

The figures in Table 1 below show the relationship.

Table 1

Crash rates versus numbers of offences - California Driver Record Study.

Offences	Crashes per 100 drivers
0	14
1	25
2	35
3	44
4	57
5	62
6	64
7	79
8	69
>8	89

Coppin et al (ibid) quote a correlation between crash numbers and offence numbers of 0.23. When crash-related offences are included, this correlation rises to 0.27, which is an indication of the bias introduced by including the crash-related offences.

In Part 7 of the California Driver Record Study (Lew, Coppin and Peck, 1966), the same data set is used to examine the relationship between crashes and offences of different types. Using a sample of 48006, they calculated correlations between number of offences and numbers of crashes for various types of offences. (Again, crash-related offences have not been included). The results are

Table 2

Correlation between number of offences and number of crashes for various offence types - California Driver Record Study.

Offence Type	Crash Correlation
Speed	.161
Signs/Signals/Markings	.137
Overtaking/Passing	.106
Equipment	.100
Turning/Stopping/Signalling	.087
Right of way	.072
"Majors"	.045

These correlations are disappointingly low, and the order is perhaps surprising. Lew *et al*(ibid) comment that

a. the size of the correlation can be expected to increase with the size of the conviction category.

b. there is a very low correlation with "Majors", i.e. offences that come from the criminal code. They surmise that this low correlation is a result of the low number of such convictions recorded, and the skewed distribution of the data.

c. there is a surprisingly high correlation with "equipment" offences, which the authors ascribe to the recording of multiple equipment offences when a more substantial offence is recorded.

This Report comments further on these remarks by Lew et al when correlations are calculated.

Lew et al (ibid) concluded that "total offences" was a better predictor than any individual offence, and that multiple regression analysis was required to see whether a differential weighting system was justified. They cautioned that interpretation of such a regression was not simple, and may not give a useful prediction for individual drivers even if it behaved well for a large group of drivers.

Peck and Coppin (1967) did carry out a multiple regression analysis on the California Driver Record Study. They looked at three sets of predictors

(a) total numbers of one-count and two-count offences under the California system.

(b & c) total numbers of one-count and two-count offences under the California system, but adding information on specific violations and on convictions not attracting point counts:

The more complex equations were found not to provide any useful improvement in the predictive power of the equations. That is, simple counts of numbers of one-count and two-count offences give the best prediction of crashes. Further, one-count offences contributed most in the regression, and the regression equation gave equal weight to one and two count offences. Although this result casts some doubt on the Californian point count system, it is far from conclusive because of selectivity regarding more serious offences in the sample.

In a later study on Californian data, Peck,Mc Bride and Coppin (1971) extended the analysis to include cross-validation in both concurrent and non-concurrent time periods. (Cross-validation is a technique whereby a predictor is developed on one part of the sample and the characteristics of the predictor evaluated by its application to an independent part of the sample. This avoids the common occurrence that the predictor works better than it should on the data for which it was developed). For concurrent data, ie offences and crashes in the same period, prediction of crash status is poor - only 33.2% of males involved in crashes and 22.3% of females are correctly predicted. Non-concurrent prediction - predicting crash involvement from offences in a previous period - is even worse. In predicting one year's crashes from two proceeding years of offences, only 13.4% of males and 7.1% of females were correctly predicted. They still found one-count offences the best predictors, and found non-countable offences to contribute significantly in the prediction of male driver involvement in crashes. The cross-validation R^2 for **males** was 5.15% for concurrent time periods, 1.42% for non-concurrent time periods. Driver exposure data was available for a small sample (536) of males, and inclusion of this data in the regression analysis improved R^2 to 13%.

Kwong, Kuan & Peck (1976) used subset selection procedures on a suite of 109 possible predictors of crash involvement. Using 14 year histories of 5400 Californian drivers, they only achieved an R^2 of 7.3%.

McConnell & Hagen (1980) carried out discriminant analysis on California data, using age, sex and offence history as variables. The first discriminating function accounted for 86 to 88% of total variability. They identified 5 high risk driver groups from the discriminant analysis, and compared the accident rates in equal sized groups of "high risk" drivers selected

- by the discriminant analysis approach
- by a synthesis of the Californian point count system
- by regression analysis using prior offences and crashes
- by regression analysis using prior offences and crashes, plus age and sex.

The average numbers of accidents per driver in these four "high risk" groups were

.5007

.4516

.5421

.5342

The two regression methods were significantly better than the point count system; the "risk group" method was not significantly better.

Peck & Kuan (1982) re-analyzed the data used in Kwang et al (1976) and McConnell & Hagen (1980). For a sample of 92999 drivers, they found that the relationship between the number of offences and the number of crashes was as shown in Table 3.

Table 3

Crash rates for various numbers of offences in a prior period - California Driver Record Study

Prior Offences (1969-71)	Crashes/100 drivers (1972-74)
0	12.2
1	18.3
2	23.8
3	29.6
4	35.6
5	36.3
>5	51.7

They also carried out a non-concurrent regression on a subset of the drivers (10259) for whom exposure and personal data were obtained by a survey. Using nine predictor variables in the regression, they were able to make 27% correct predictions of crash involvement (which is better than the non-concurrent predictions achieved by Peck, et al (1971).

2. NORTH CAROLINA STUDIES

Campbell (1958) reported a study of 40467 North Carolina drivers, and concluded that there was a "stable and substantial contingency between non-accident violations and accidents when dealing with the average record of a large group of drivers" (p. 18). The results in Table 4 were obtained for non-crash related offences:-

Table 4

Crash rates for various offence frequencies - North Carolina.

Offences ("Moving")	Average Crashes per Driver
0	.167
1	.391
2	.560
3	.699
4	.857
5	1.001

Driving exposure was noted as relevant, but data was not available.

A statistical analysis of 1200 of the driver records showed that offences of particular types were significantly correlated with crashes, but the correlations were small (0.01 to 0.1). Campbell (ibid) comments that such correlations may be spurious - in the North Carolina data drunk-driving is negatively correlated with crashes, but faulty mufflers and no-licence offences are positively correlated. The multiple correlation of individual offence counts with crashes is 0.27 (R^2 = 7%) - about the same as for correlation with a count of total offences. This study is also limited to concurrent periods, i.e. offences and crashes in the same time period; for true prediction the correlations would be even lower.

Stewart and Campbell (1972) tabulated 2502240 North Carolina drivers of 22 or more years of age for two 2-year periods. They find that prior offence record is a poor predictor of crash involvement - 71% of crashes were for drivers who had no offences in the prior 2 - year period. Previous crash record is a better predictor. Age of the driver appears to have little effect, although the youngest group is least predictable.

3. ONTARIO STUDIES

Chipman (undated copy of report) selected 500 - 600 drivers from each of five levels of demerit points in the Ontario system as at 1/7/70, and their associated offences and crashes in the succeeding 2.5 years. This amounted to 3030 offences and 1105 crashes in 2650 drivers in 2.5 years. Chipman reported that

a. women at all levels had fewer subsequent convictions than men

b. men with chauffeurs' licences had higher subsequent collision rates

c. the age of men affected offences but not crashes

d. only demerit points distinguish between drivers involved in crashes and drivers not involved in crashes.

Chipman and Morgan (1976) extended Chipman's 1974 study over 4.5 years, looking at successive 6 month periods. Those who have an offence in the previous 6 month period have about twice the risk of a crash in the next 6 month period when compared with those who do not have an offence. They also looked at types of offences (as was done in California), and tested for association with crashes within initial demerit point strata. The occurrence of any conviction was significantly correlated in all strata, while occurrences of demerit point offences (about half of all offences) was not so well correlated. Of the offence types, speeding was significantly correlated in all but one strata, and the remaining offences significantly correlated with crash numbers only for drivers with high initial demerit points. "Major" offences were not correlated at all - the same experience as in California.

Chipman (1982) analysed data for 3084 Toronto drivers for whom exposure data was also available. The results of this study were as shown in Table 5.

Table 5

Crash rates versus demerit points for women and men - Ontario

	Demerit Points	Crash Rate per 1000 drivers	Crash Rate per 10^6 driver kms
Women	0-2	31.2	12.2
	3-5	50.5	13.6
Men	0-2	76.9	11.3
	3-5	94.9	11.6
	6-8	135.6	15.9
	9+	153.8	14.6

4. COMMENTS

In assessing the three groups of studies, we can make the following comments

a. It would appear that including the crash-related offences , i.e. offences recorded at the crash scene, imparts a slight positive bias to the relationship between offences and crashes.

b. Relationships between offences and crashes are stronger for concurrent periods (i.e. for offences end crashes in the same time period.) It is claimed that consecutive periods give a more realistic picture for evaluation of ability to predict crashes; in such analyses the correlations, etc are very much worse. The true picture would seem to be somewhere between the two - in the non-concurrent analyses, we are deprived, in predicting a crash, of the offence record of the driver during the second period. This offence record would be available to us in a predictive situation.

c. The number of offences, without regard to their type, seems to be generally the best predictor of crash involvement. This runs counter to the point-count systems in use in many jurisdictions.

d. Drivers with higher numbers of offences are more likely to be involved in crashes, although the numerical rates differ widely from study to study (See, e.g. Tables 1,3 and 4). The discrepancies reflect differences between concurrent and non-concurrent analyses, and between definitions of the offences included in the studies.

e. Relationships between specific classes of offences and driver involvement in crashes sometimes appear to be contrary to intuition. For instance, occurrences of driving offences which might be thought of as quite serious, e.g. drunken driving and driving offences which are violations of the criminal code, are reported as negatively correlated or uncorrelated with crash involvement. On the other hand, offences related to equipment and licencing irregularities are found to be positively correlated with crash involvement. Various explanations for this have been put forward, e.g.

- serious offences provide a very small sample from a skewed distribution;
- equipment offences are often recorded at the same time as other, more serious, offences.

However, these recurrent anomalies do not seem to be adequately resolved.

f. Regression modelling, when attempted, does not appear to have been very successful; the R^2 values to assess the fit have generally been less than 10% of the variation observed in the crash records. Even when extra explanatory variables are introduced, e.g. age, sex, driving exposure, the fit as measured by R^2 is far from adequate.

g. The regression analyses carried out have used a direct least squares (with weighting in some cases) approach, despite the fact that the data do not appear to satisfy the assumptions of such an analysis.

h. Because of the very large sample sizes usually employed in these studies, the statistical analyses usually find the measure of association (correlation, R^2, etc.) to be statistically significant. This means that, although the measures of association are small, the data suggest that it is highly unlikely that the sample come from a population with no association. One must distinguish, however, between a small but significant measure of association and one that is useful as a basis for prediction.

i. The various assessments made of the predictive value of associations established by the authors (whether established by correlation, regression or discriminant analysis) indicate that the ability of the associations to predict crashes amongst the driver population is not high.

DATA SOURCES FOR THE STUDY

The data for the proposed study of relationships between traffic offences and road crashes in the Australian Capital Territory (ACT) comes from two sources - the Australian Federal Police Force, which maintains data on traffic offences recorded in the ACT jurisdiction, and the National Capital Development Commission (NCDC), which maintains data on all reported crashes in the ACT. Further, the Australian Federal Police Force maintains a data file which enables the offence files to be linked to the crash files. We will discuss each of the main data files in turn.

1. THE OFFENCE DATA

The offence data file provided by the Australian Federal Police contained 290186 records, and consisted of extracts of all Traffic Infringements records kept by the Australian Federal Police. Any information relating directly to the identity of the person against whom the offence was recorded (eg, name, driver's licence number, etc.) was deleted before the records were supplied.

The Australian Federal Police offence data commenced being recorded in the form in which the records now exist in January 1980. However, a scan of the data showed that the file included offences committed on dates earlier than 1980. These were excluded from the data before any subsequent analyses were performed.

The other items considered relevant on the Australian Federal Police offence data files were

- 0 personal identification number (essential for linking with crash data)
- 0 offence data (restricted to post 1/1/80, as discussed above)
- 0 the offence description
- 0 the State of vehicle registration

The field showing the State of vehicle registration of the driver against whom an offence was recorded was used to reduce the offence data file. Offences for drivers of vehicles registered in States remote from the Australian Capital Territory were removed from the records, as it was considered likely that such offences would relate to visiting drivers who would

- 0 possibly exhibit different driving behaviour from those regularly driving in the ACT, and
- 0 not remain in the ACT for long, and hence not contribute proportionally to the crash records.

We retained records of offences for drivers of vehicles registered in the ACT, vehicles registered in the State of New South Wales (which surrounds the ACT) and vehicles for which the state of registration field contained a blank or some version of "unknown".

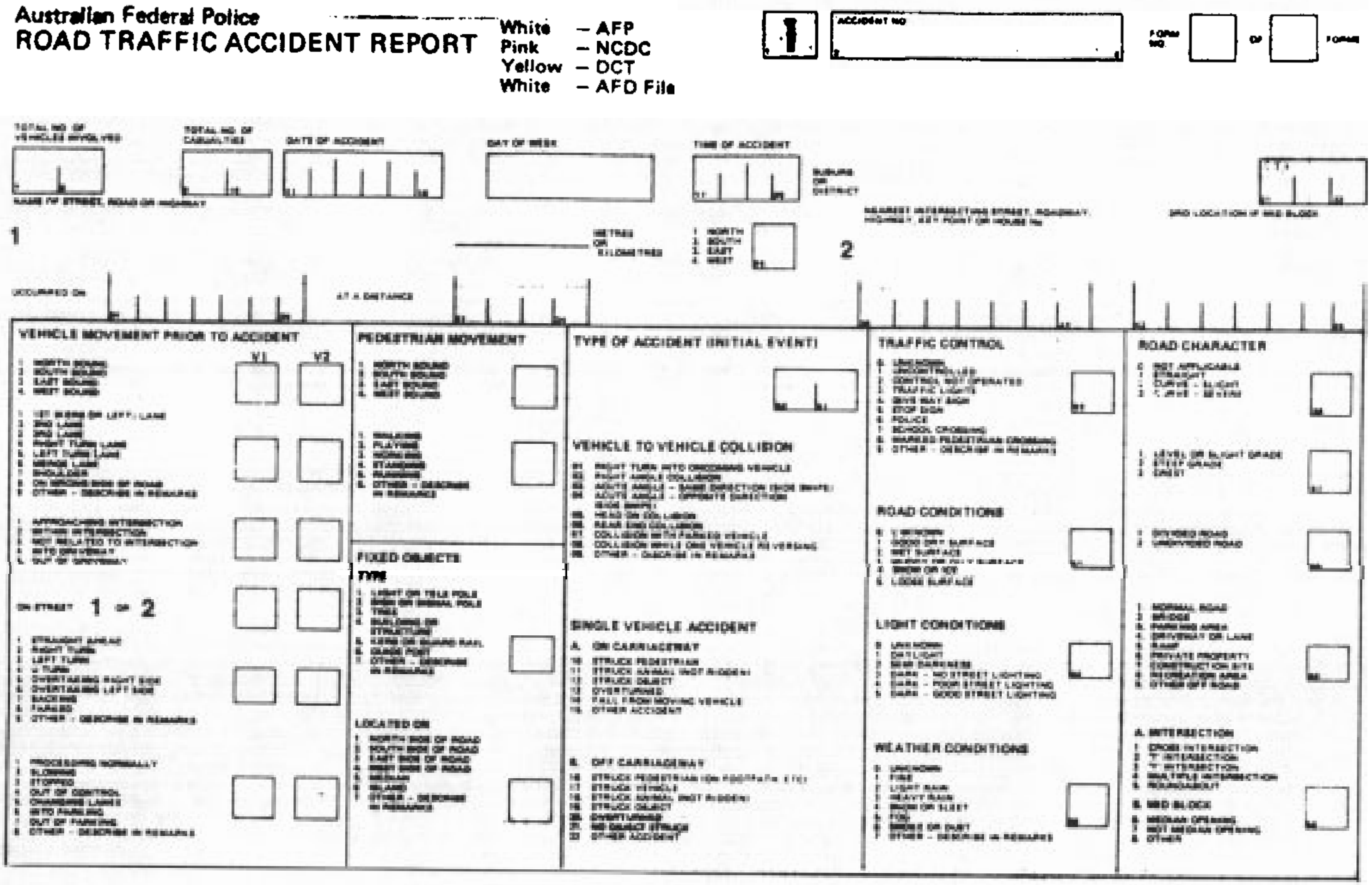
The "offence description" field represented the major data handling problem in dealing with the offence records. The procedures used to reduce the offence description to a managable set of offence classes are discussed in a separate section (see "Categorisation of Offences"). In this categorisation process, some further records were deleted; these were records of offences where the offence was apparently unrelated to an actual driving or vehicle maintenance behaviour (eg various offences of "aiding and abetting") and these instances where an offence was not formally recorded against the driver (because of special provisions such as "first offender" regulations). The number of offence records excluded on this basis was small.

We retained as a basis for our study of traffic offences a total of 123311 offence records. The broad characteristics of these offences are examined in the section "Summaries of the Data".

2. THE OFFENCE - CRASH LINKAGE

The Australian Federal Police Force officer attending a crash completes, either on the spot or subsequently, a Road Traffic Accident Report (Figure 1 is an example of the Report). This report is forwarded to the National Capital Development Commission for processing. However, before forwarding, the Australian Federal Police record a personal identification number for each driver involved in the crash. This personal identification number, together with the file number of the accident report, form the basis of a record in the second file maintained by the Australian Federal Police. The second file, containing 96788 records, provides the means of linking the offence data held by the Australian Federal Police (where identification is by the personal identification number) to the crash data held by the National Capital Development Commission (where identification is by crash report file number).

Figure 1 (a) : Australian Federal Police Road Traffic Accident Report - Upper Section



Australian Federal Police
ROAD TRAFFIC ACCIDENT REPORT

White – AFP
Pink – NCDC
Yellow – DCT
White – AFP File

ACCIDENT NO

VEHICLE MOVEMENT PRIOR TO ACCIDENT

V1 V2

PEDESTRIAN MOVEMENT

FIXED OBJECTS

TYPE

LOCATED ON

TYPE OF ACCIDENT (INITIAL EVENT)

VEHICLE TO VEHICLE COLLISION

SINGLE VEHICLE ACCIDENT

A. ON CARRIAGEWAY

B. OFF CARRIAGEWAY

TRAFFIC CONTROL

ROAD CONDITIONS

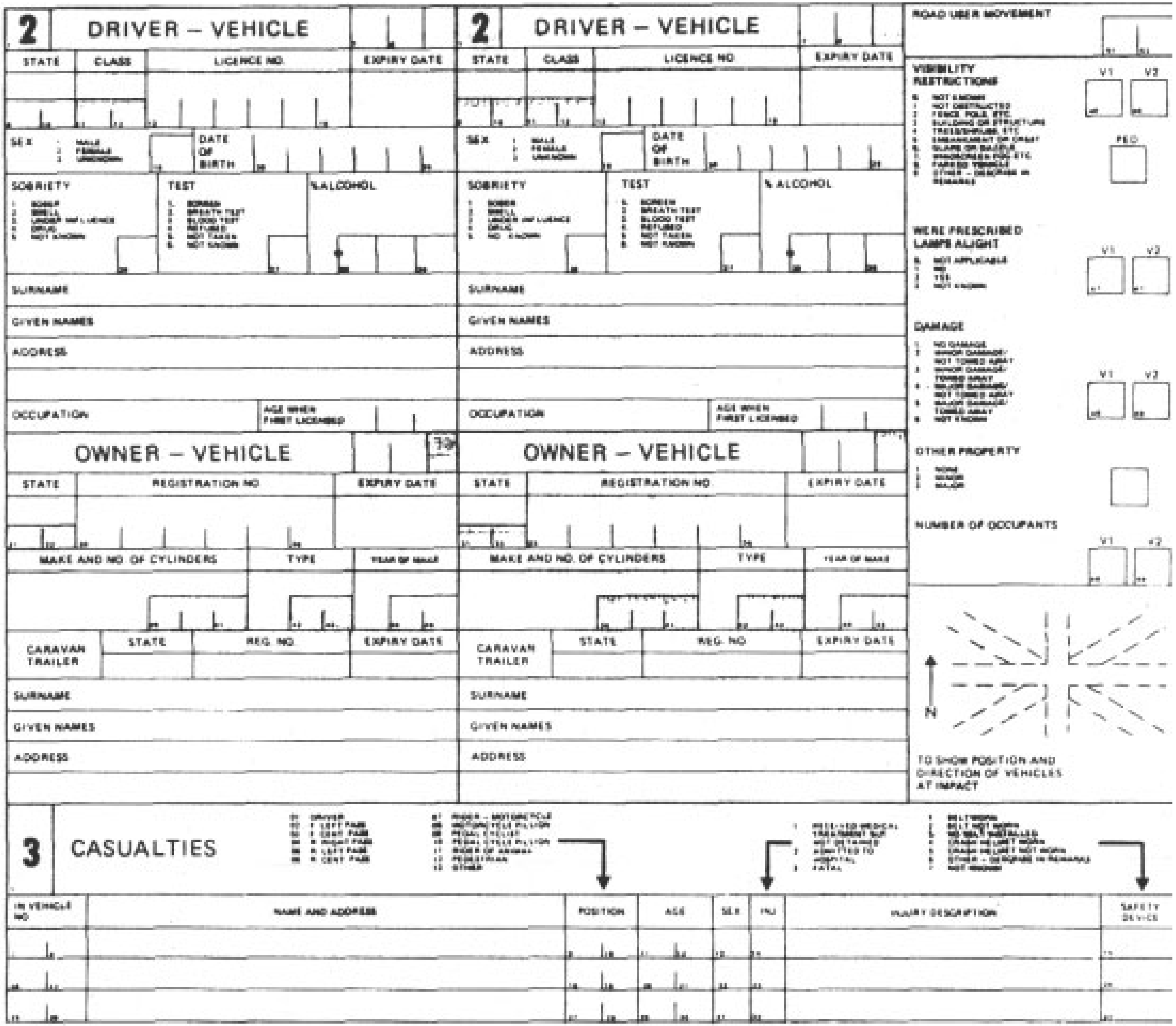
LIGHT CONDITIONS

WEATHER CONDITIONS

ROAD CHARACTER

A. INTERSECTION

B. MID BLOCK



2 DRIVER – VEHICLE

STATE | CLASS | LICENCE NO. | EXPIRY DATE

SEX: 1 MALE, 2 FEMALE, 3 UNKNOWN | DATE OF BIRTH

SOBRIETY | TEST | % ALCOHOL

SURNAME

GIVEN NAMES

ADDRESS

OCCUPATION | AGE WHEN FIRST LICENSED

OWNER – VEHICLE

STATE | REGISTRATION NO | EXPIRY DATE

MAKE AND NO. OF CYLINDERS | TYPE | YEAR OF MAKE

CARAVAN TRAILER | STATE | REG. NO | EXPIRY DATE

SURNAME

GIVEN NAMES

ADDRESS

2 DRIVER – VEHICLE

STATE | CLASS | LICENCE NO | EXPIRY DATE

SEX: 1 MALE, 2 FEMALE, 3 UNKNOWN | DATE OF BIRTH

SOBRIETY | TEST | % ALCOHOL

SURNAME

GIVEN NAMES

ADDRESS

OCCUPATION | AGE WHEN FIRST LICENSED

OWNER – VEHICLE

STATE | REGISTRATION NO | EXPIRY DATE

MAKE AND NO. OF CYLINDERS | TYPE | YEAR OF MAKE

CARAVAN TRAILER | STATE | REG. NO | EXPIRY DATE

SURNAME

GIVEN NAMES

ADDRESS

ROAD USER MOVEMENT

VISIBILITY RESTRICTIONS | V1 | V2

PED

WERE PRESCRIBED LAMPS ALIGHT | V1 | V2

DAMAGE | V1 | V2

OTHER PROPERTY

NUMBER OF OCCUPANTS | V1 | V2

N

TO SHOW POSITION AND DIRECTION OF VEHICLES AT IMPACT

3 CASUALTIES

IN VEHICLE NO | NAME AND ADDRESS | POSITION | AGE | SEX | INJ | INJURY DESCRIPTION | SAFETY DEVICE

Figure 1 (b) : Australian Federal Police Road Traffic Accident Report - Lower Section

3. THE CRASH DATA

As described above, the crash data is maintained by the National Capital Development Commision, and is based on the Road Traffic Accident Report completed by the attending Australian Federal Police officer. The NCDC is in the process of converting its crash records to an in-house computer system, and the conversion is complete for the years 1982, 1983 and 1984, and has been carried out up to about the middle of 1985. This was the data supplied for the study of relationships between offences and crashes.

The complete records of each reported crash, with the names, licence numbers and registration field deleted, were provided on tape. This amounted to some 50000 records in all. As can be seen from the Road Traffic Accident Report form, a wealth of data is available for each crash, not all of it relevant to the study reported here. We have extracted from the available data information about the crash and the conditions which it occurred (but not the extensive information about its street location), and information about the drivers involved in the crash.

Again, some broad summaries of this crash data are provided in the section "Summaries of the Data".

CATEGORIZATION OF OFFENCES

As discussed in the section on 'Data Sources for the Study', the Australian Federal Police records of traffic offences contain a field for "offence description". This is the police officer's designation of the offence recorded against the driver, which the police officer writes in English. That is, the offence is recorded by an English language description (possibly in shortened form) and not by an offence number or other uniform offence designation. The officer's description of the offence is transcribed directly to the computer file, without editing.

Although the Australian Federal Police stated that there was some attempt to impose uniformity in the offence descriptions, there was still a great diversity in modes of description. The matter was even further complicated by the inability of the computer to make reasonable judgements about minor disimilarities in description. For example, the computer sees "EXCEED 08" and "EXCEED.08" as different offence descriptions; it also sees every misspelling as creating a new offence. For small amounts of data this problem can be overcome manually, but for the very large data files involved in this study correction of each offence description to a standard form was not feasible.

A scan of the offence records to be studied showed over 2200 individual offence descriptions present. Even after going over the offence descriptions and imposing some of the obvious equivalences, more than 650 offence descriptions remained. This was clearly far too many for any rational analysis based on offence types, so it was decided that the offences should be assigned to a manageable number (ten or so) of broad categories.

Such a categorization had been carried out in order to analyse offence/ crash relationships in the California Driver Record Study (Lew, Coppin and Peck, 1966). The categories chosen there were

speeding
traffic signs, signals and markings
turning, stopping and signalling
driving, overtaking and passing
equipment
right-of-way
major (driving while suspended, hit-run,
DUI, reckless driving, drugs, manslaughter)

Note that the categories used in the California study do not include those offences relating to the licencing and registration of the vehicle.

In a study in South Australia (private communication from a member of the study team), a different categorization of offences was used. Here the categories were

dangerous driving
driving under influence
alcohol concentration
exceed speed limit
incorrect position on road

fail to give way

improper turns

disobey lights/crossings

disobey signs

lighting offences

L&P permit offences

fail to stop after accident

demerit disqualifications

In terms of providing a categorization of offences to reflect patterns of driver behaviour, the above list seems to make unnecessarily fine distinctions in some cases, e.g. between "driving under influence" and "alcohol concentration". It also chooses to disregard most offences relating to the state of the vehicle or its accessories and fittings.

Guided by the two examples of categories discussed above, and mindful of the need for categories to fairly represent patterns of driver behaviour, we assigned the offences as described in the "offence description" field of the Australian Police Force data to one of twelve categories.

1. Speeding - all offences related to driving at excessive speed

2. DUI/PCA - all offences related to driving under the influence of alcohol, drugs, etc.

3. Dangerous/reckless/culpable - offences related to dangerous, reckless or culpable driving (including manslaughter).

4. Driving/overtaking/passing/pulling out - offences related to driver behaviour in any of these aspects of driving

5. Failure to obey lights, markings, signals and directions.

6. Turning, stopping and signalling - offences related to driver behaviour in any of these aspects of driving

7. Equipment - all offences related to the state of the vehicle, its equipment and accessories

8. Driving while licence suspended

9. Failure to obey licencing and registration regulations

10. Unsafe behaviour in or on a vehicle

11. Failure to stop at or report an accident

12. Others

The categories set out here were not designed to equalize numbers within categories, but rather to group and identify offences which, in our opinion, represented similar patterns of driver behaviour.

We can make the following comments on the categories we chose to represent the offences.

a. Category 12 - "Others" contains a very few offences for which the offence description was uninterpretable.

b. We have separated offences relating to "driving while licence suspended" from other offences relating to licencing and registration. Driving with a suspended licence indicates a different history of driving behaviour from that of a driver committing a more technical offence relating to licencing or registration rules.

c. Category 10, for "Unsafe behaviour in or on a vehicle", covers offences not directly related to the driving of the vehicle, for example having protruding limbs or having an excessive number of persons in the vehicle.

d. Some thought was given to a further sub-division of Category 5, with the possibility of extracting offences relating to failure to obey red lights. Offences relating to not obeying red lights could be thought of as an obvious contributor to crashes. However, we did not pursue this finer sub-division because

   i. it was not clear from the records that all offences of disobeying red lights could be unambiguously identified, and

   ii. it seemed that disobeying red lights was but one of a number of similar driver behaviours (eg improper right turns) which would all be likely to be strongly related to crashes.

e. The categories we have chosen to use divide into those which relate to behaviour in the actual control of the vehicle, ie Categories 1-6, and those which relate to the vehicle itself and the regulations governing licensing, registration, the vehicle and its equipment, and non- driving behaviour in the vehicle, ie Categories 7-11. The 12th category of unidentified offences is so small as to be negligible.

For the remainder of this Study, we will identify offences by the categories to which we have assigned them.

SUMMARIES OF THE DATA

1. THE OFFENCE DATA

As discussed in the section "Data Sources for the Study", the offence data from the Australian Federal Police Force was reduced by excluding offences prior to 1980, and by excluding offences committed by drivers identified as being from Victoria, Queensland, Tasmania, South Australia or Western Australia. A total of 123311 offence records were retained for further study.

The 123311 offences were committed by 56167 individual drivers, which corresponds to an average of 2.20 offences per driver. A more informative picture is given by the frequency distribution among drivers of the number of offences; this information is shown in Table 6 and in Figure 2. Note that two drivers has in excess of of forty-nine offences in the period of just over five years covered by the study.

Table 6

Distribution of Numbers of Offences

No of Offences	No of Drivers
1	32690
2	10483
3	4974
4	2684
5	1505
6	1005
7	679
8	499
9	361
10	273
11	213
12	149
13	123
14	104
15	74
16	71
17	46
18	42
19	31
20	32
21	21
22	16
23	9
24	8
25	9
>25	66

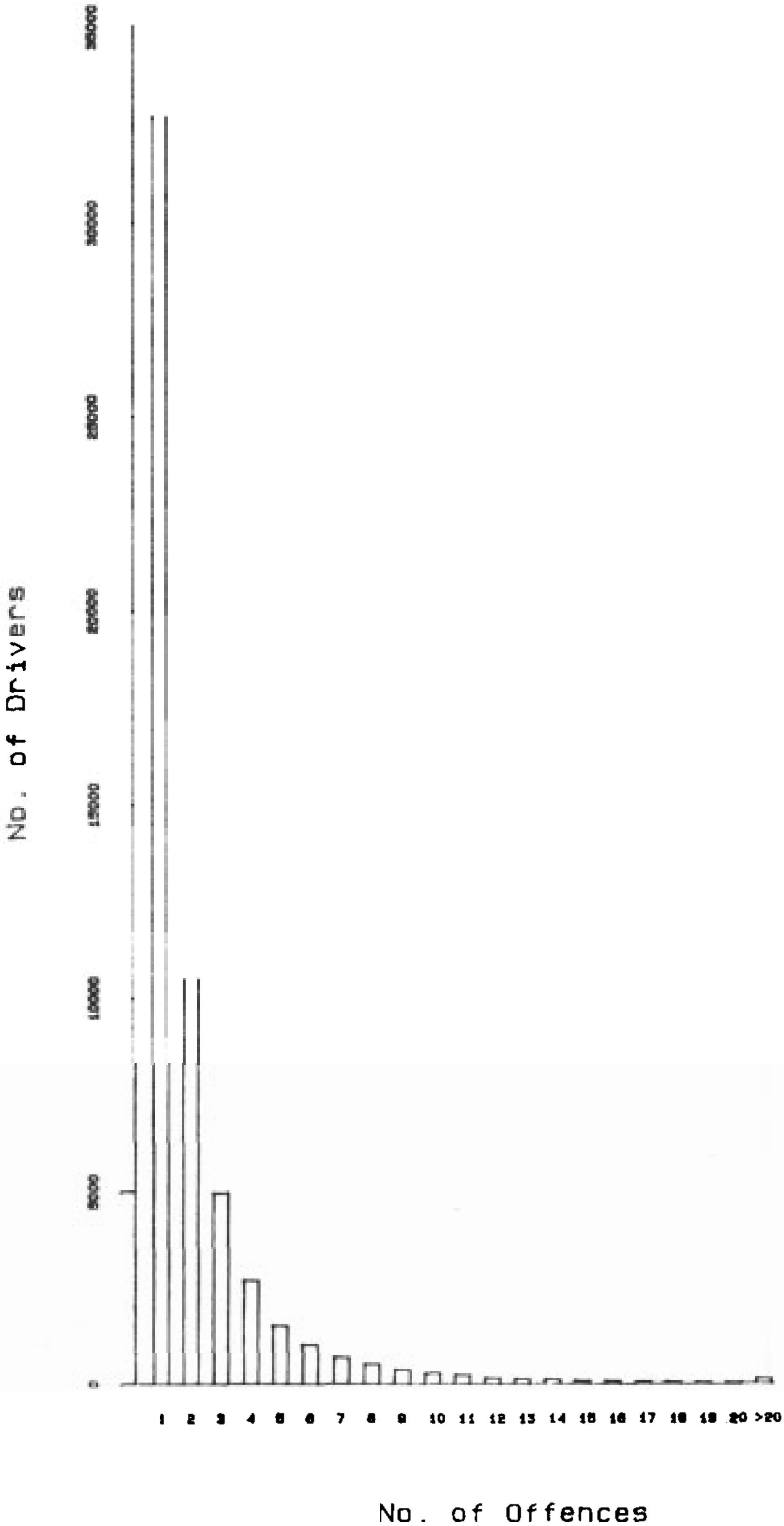


Figure 2: Distribution of numbers of offences

Another distribution of interest is that of the number of offences simultaneously recorded against a driver. It has been suggested that many of the offences in the file may be "technical" offences,regarding equipment, licencing,registration,etc.,recorded against drivers who have been stopped for offences relating to the driving of a vehicle. Table 7 shows the number of occasions on which a driver has had 1,2,3,.... offences recorded on the same date.

Table 7

Distribution of the Number of Offences at the Same Date

No of Offences on the Same Date	No of Occasions
1	76250
2	13304
3	4302
4	1187
5	314
6	108
7	36
8	22
9	5
10	5
11	2
12	3

Regarding the results in Table 7,we note the following.

a. Only dates relating to offences,and not the time of day of the offence were available. Thus it is not possible to distinguish between the driver who is stopped twice on the same day and the driver who is stopped once and has two offences recorded.

b. Table 7 shows that,on 80% of days on which an offence is recorded,a single offence is recorded against the driver;62% of all offences arise as single offences.

c. It is beyond the scope of this study to speculate on the relative worth, as indicators of driver behaviour,of various combinations of offences recorded on the same date.

In the light of the results in Table 7 and the comments above,we will not attempt any analysis relating to the effects of the so-called "add-on" offences - offences of a technical nature recorded against a driver already stopped for an offence related to driving behaviour.

Turning to the distribution of these offences over the classes of offence defined in the section "Categorization of Offences", the results for all the offences, and the offences for each year, are shown in Table 8. The distribution of all offences is shown in Figure 3.

Recall that the categories are, briefly

1. Speeding
2. DUI/PCA
3. Dangerous/reckless/culpable driving
4. Driving/overtaking/passing/pulling out offences
5. Failure to obey lights, markings, signals and directions
6. Turning, stopping and signalling offences
7. Equipment
8. Driving while licence suspended
9. Failure to obey licencing and registration regulations
10. Unsafe behaviour in or on vehicle
11. Failure to stop at or report an accident
12. Other (uninterpretable).

Table 8

Distribution of Offences by Year and Offence Category

Offence Category	Total Offences	Year 80	81	82	83	84	85	
1	52972	6829	6845	8378	9513	10599	9276	1532
2	6918	753	1226	1199	1061	1137	1208	334
3	620	119	101	99	83	92	100	26
4	4740	656	529	642	666	1036	1065	146
5	16093	2156	1879	2676	2540	3421	2942	479
6	309	69	76	42	43	41	36	2
7	14720	1985	2048	3174	2915	2757	1595	246
8	513	78	77	59	71	120	96	12
9	25101	3311	3403	4826	4851	4831	3529	350
10	699	71	86	177	118	120	97	30
11	612	84	124	130	105	95	69	5
12	14	2	9	1	1	1	0	0
Total	123311	16113	16403	21403	21967	24250	20013	3162

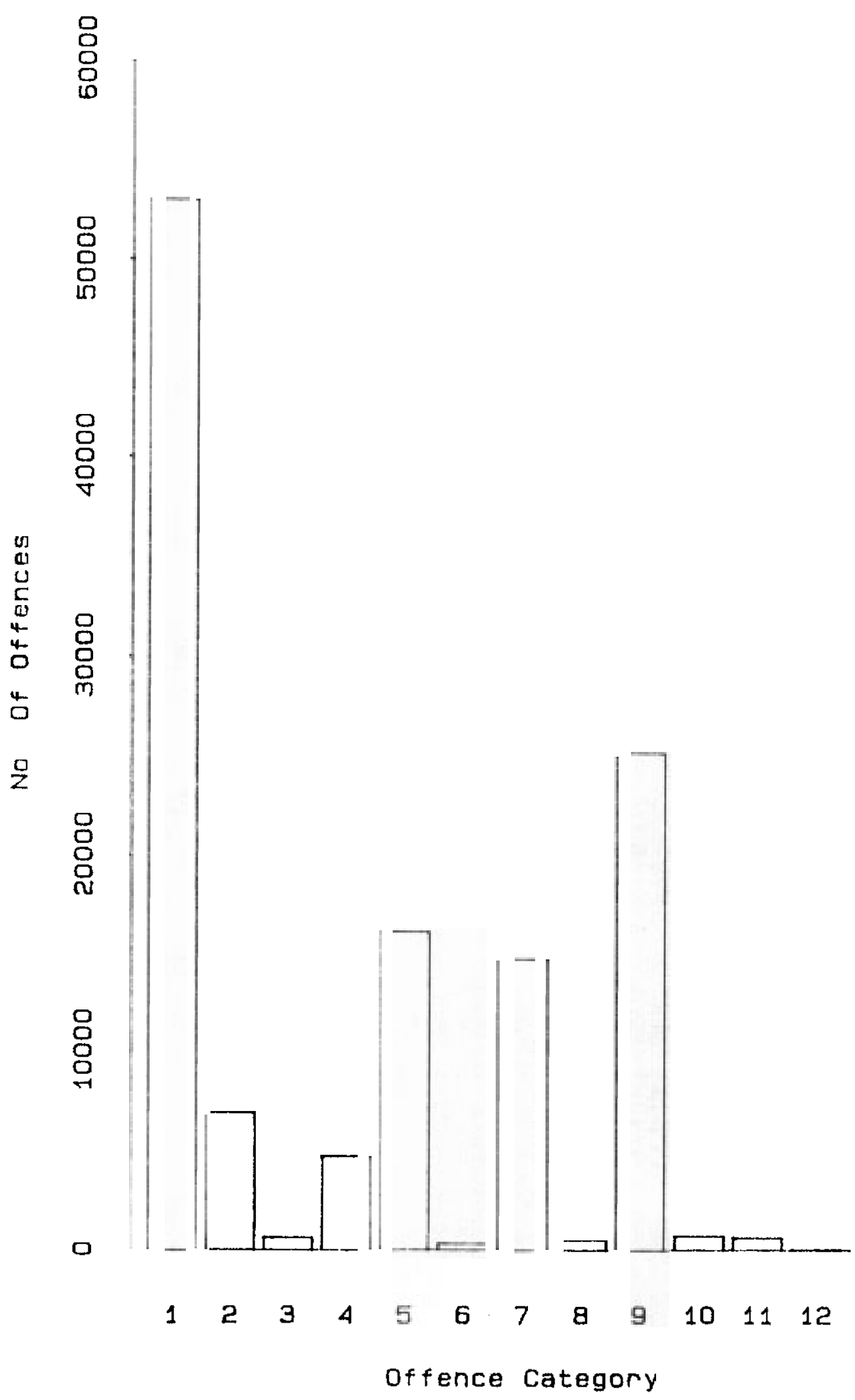


Figure 3: Number of offences in each category

We note the following with respect to Table 8

i. The reason for the decline in offences recorded in 1985 is unknown but could result from normal variations in ACT operational traffic procedures.

ii. For those categories of offences which have a substantial number of occurrences, there is stability in their relative contributions to the total from year to year. (Categories with small numbers of offences can be expected to show higher variability from year to year, and so it is not possible to discern trends in the amount of data available). The proportion of offences categorized as "speeding" (1) shows an increase in 1985; this may be due to the offences being mostly dealt with by infringement notice and hence appearing to be a higher proportion of offences dealt with at the time the data was provided. However, the proportion of offences in the "equipment" category (7) shows a sharp decline in 1985; it might have been expected to increase in proportion for the same reason as the speeding offences.

iii. Offences of driving at excessive speed account for the largest proportion of offences - 42.96% overall. The next largest category is that of failure to obey licencing and registration regulations, with 20.36% of all offences. Failure to obey lights, markings, signals and directions is a category accounting for 13.05% of all offences, and equipment offences comprise 11.93% all offences. The remaining categories make small contributions.

Table 9 shows the difference in offence distribution between those 32690 drivers who have recorded a single offence, and the 23477 drivers with multiple offences. (The multiple offenders have an average of 3.86 offences each).

Table 9

Distribution of Offences for Single and Multiple Offenders

Offence Category	Single Offenders	Multiple Offenders
1	18298	34674
2	2226	4692
3	90	530
4	1503	3237
5	6513	9580
6	46	263
7	2702	12018
8	14	499
9	1136	23965
10	135	564
11	19	593
12	8	6
TOTAL	32690	90621

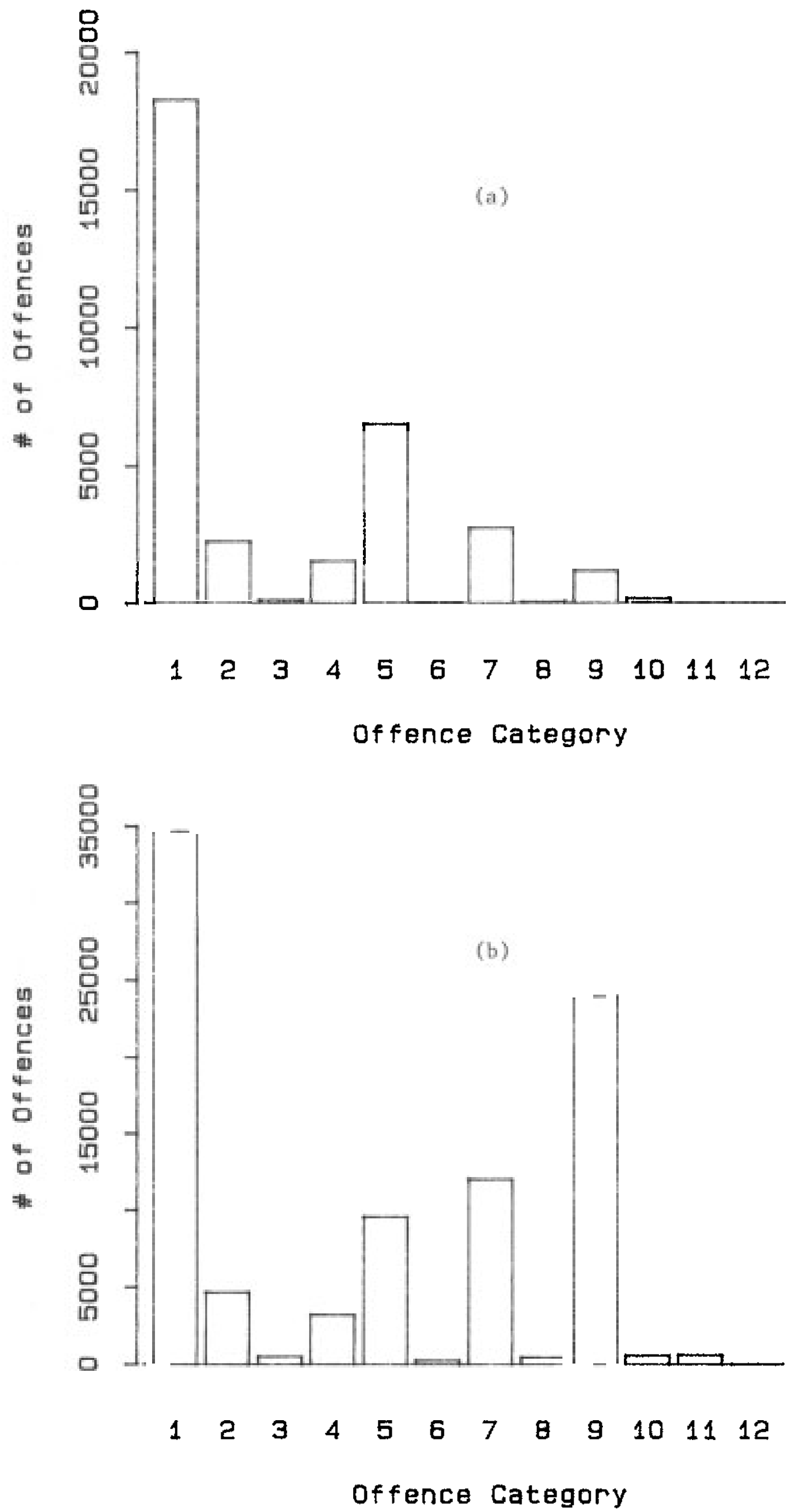


Figure 4 Offence distributions of drivers

(a) with a single offence recorded

(b) with more than one offence recorded

These distributions of offences are very different (see Figure 4). Note in particular that

i. drivers with a single offence have a much higher proportion of speeding offences than drivers with multiple offences (55.97% vs 38.26%);

ii. drivers with a single offence have a much higher proportion of offences in the category "failure to obey lights, markings, signals and directions" than drivers with multiple offences (19.92% vs 10.57%);

iii. drivers with multiple offences have a much higher proportion of offences in the category "failure to obey licencing and registration regulations" than do single offenders (26.45% vs 3.48%);

iv. drivers with multiple offences have a somewhat higher proportion of equipment offences than do single offenders (13.26% vs 8.27%).

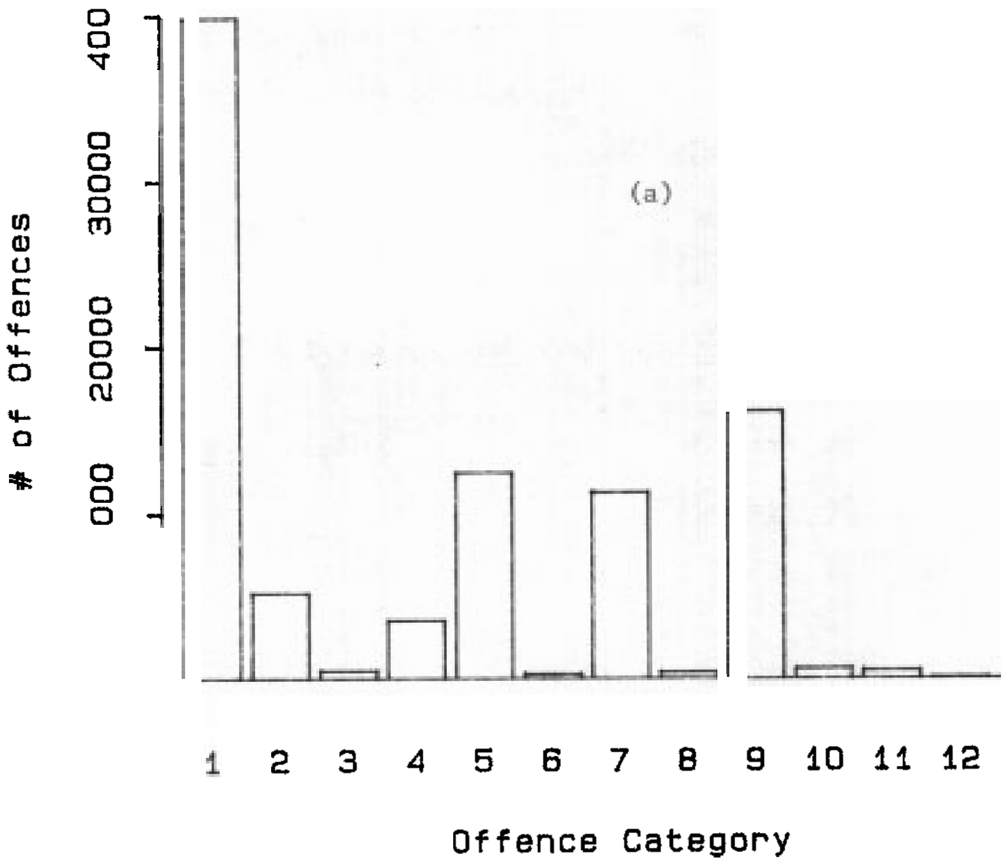
The above remarks tend to confirm a view that licencing, registration and equipment offences are on some occasions "add on" offences, ie offences detected at the same time as the driver is stopped for some offence related to driving behaviour. The relatively higher proportion of the offences of single offenders categorised as "speeding" and "failure to obey lights, markings, signals and directions" can be seen to show the deterrent effect of being apprehended for such offences.

Table 10 shows the offence distributions for drivers licenced in the Australian Capital Territory (ACT) and for those licenced in the State of New South Wales (NSW). The same results are shown graphically in Figure 5. (The "unknown" licence group accounted for only 9328 offences, and has not been included in this comparison). There is very little difference between the two distributions; the only difference worthy of note may be a marginally higher proportion of speeding offences amongst NSW-licenced drivers.

Table 10

Distribution of Offences for ACT and NSW Licenced Drivers

Offence Category	Drivers Licenced in ACT	NSW
1	39858	11203
2	5162	1339
3	459	112
4	3510	957
5	12507	2919
6	238	57
7	11302	2649
8	346	92
9	16136	3889
10	572	103
11	454	108
12	10	1
TOTAL	90554	23429



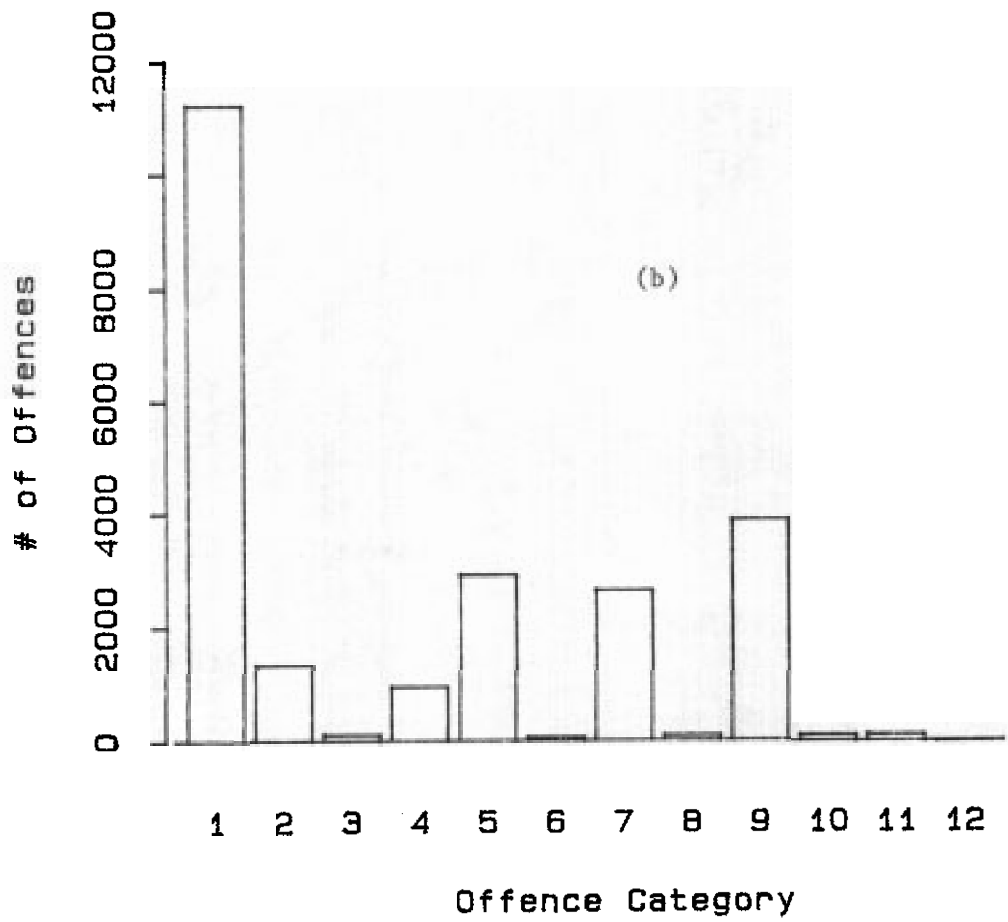


Figure 5 Offence distributions of
(a) ACT licenced drivers
(b) NSW licenced drivers

2. THE CRASH DATA

The crash data covers the years 1982, 1983, 1984 and 1985 (part only). It was supplied by the National Capital Development Commission and provides information on crashes and on the drivers involved in the crashes.

There were, in the records made available, 39076 crashes involving 73465 drivers. (Note that a few of the 73465 drivers will occur twice or more in the records - the number of individual drivers is 63473). We will present some tabulations of the data regarding crashes and drivers involved in crashes to illustrate the nature of the base data.

The distribution of the number of vehicles involved in each crash is as shown in Table 11.

Table 11

Distribution of number of vehicles in crash

Number of Vehicles in Crash	Number of Crashes
1	6100
2	31903
3	906
4	130
5	37

The great majority of reported crashes are two-vehicle crashes

The officer attending the reported crash characterises the crash according to its type. The distribution of crashes by type is shown in Table 12.

Table 12

Numbers of crashes for each crash type

Accident Type	Number of Crashes
Vehicle to Vehicle Crash	
Right turn into oncoming vehicle	1919
Right angle crash	5731
Acute angle - same direction	3766
Acute angle - opposite direction	687
Head-on crash	367
Rear - end crash	11498
Crash with parked vehicle	5154
Crash while one vehicle reversing	3256
Other	531
Single Vehicle Crash on Carriageway	
Struck pedestrian	382
Struck animal (not ridden)	593
Struck object	1546
Overturned	593
Fall from moving vehicle	214
Other	219
Single Vehicle Crash off Carriageway	
Struck pedestrian	40
Struck vehicle	0
Struck animal (not ridden)	8
Struck object	2220
Overturned	194
No object struck	48
Other	110

The types of vehicles involved in the reported crashes are shown in Table 13

Table 13

Vehicle types involved in crashes

Vehicle Type	Number of Vehicles
Car	31452
Taxi/Hire Car	414
Utility	1358
Panel Van	2308
Articulated Vehicle	121
Truck (Not articulated)	811
Bus	433
Bicycle	396
Motor Cycle/Scooter	1622
Other	45
Not Known	116

Note that the data record vehicles and vehicles towing trailers separately; we have amalgamated the figures so that "Cars" includes "Cars + trailers", etc. Some of the less-frequently occurring vehicle types are aggregated in "Other". The relatively low number of motor cycles and scooters involved in crashes is somewhat surprising (assuming that all the vehicles recorded as "bicycle" are in fact push-bikes). However, the ratio of motorcycles to cars registered in the ACT is also low - about one to twenty.

The type of road on which the crash occurred is distributed as in Table 14.

Table 14

Distribution of crashes by road character

Road Character	Number of Crashes
Normal road	31174
Bridge	213
Parking area	5471
Driveway or lane	859
Ramp	333
Private property	286
Construction site	53
Recreation area	56
Other off road	631

The distribution of crashes across the day of the week when the crash occurred is shown in Table 15.

Table 15

Distribution of crashes by day of week

	Number of Crashes
Monday	5477
Tuesday	5643
Wednesday	5646
Thursday	6420
Friday	7093
Saturday	5059
Sunday	3738

This distribution is consistent with that observed in other studies, where crash numbers increase through the working week and then are lower on the weekend days. Table 16 shows the time of day distribution of crashes, which again is in accord with the general distribution of traffic intensity.

The apparent disarity between Hours 1 and 24 probably arises from confusion as to the correct recording of time at the beginning and end of the 24-hour clock day.

Table 16

Distribution of crashes by time of day

Hour of Day (from midnight)	Number of Crashes
1	1097
2	461
3	254
4	213
5	105
6	198
7	1101
8	3577
9	2206
10	2080
11	2317
12	2715
13	2140
14	1929
15	3103
16	3651
17	4522
18	2040
19	1665
20	1193
21	867
22	670
23	940
24	25

The National Capital Development Commission assigns a severity index to crashes, according to whether the crash resulted in a fatality, a casualty or no injuries. There were 127 crashes involving one or more fatalities; there were 3640 crashes with casualties and 3481 with no casualties. The severity was not recorded for 468 crashes. There is also a damage rating provided by the attending police officer; the distribution of crashes according to damage level is shown in Table 17.

Table 17

Distribution of crashes by damage level

Damage Level	Number of Crashes
None	2552
Minor/Not towed away	29448
Minor/Towed away	2913
Major/Not towed away	513
Major/Towed away	3524
Unknown	126

We see that the great majority of crashes result in little or no damage as assessed on this scale.

Turning to the drivers involved in the crashes, we find the distribution of numbers of crashes per driver as in Table 18.

Table 18

Distribution of number of crashes per driver

Number of Crashes	Number of Drivers
1	55630
2	6145
3	1132
4	254
5	77
>5	35

Only 12% of drivers involved in crashes had more than one crash.

The sex distribution of the drivers in crashes is as shown in Table 19

Table 19

Sex distribution of drivers in crashes

Sex	Number of Drivers
Male	44754
Female	23877
Unknown	4834

The State of issue of licence of the drivers involved in the crashes is shown in Table 20.

Table 20

State of issue of licence of drivers involved in crashes

State of Issue	Number of Drivers
ACT	56184
NSW	8020
Victoria	820
Queensland	662
South Australia	378
Western Australia	236
Tasmania	122
Northern Territory	48
Papua New Guinea	0
Overseas	293
Diplomatic	7
Unlicenced	424
Other	1039
Unknown	5232

That only seven drivers were recorded as having diplomatic licences is surprising; we surmise that drivers with diplomatic licences also contribute to the "Other" group.

Information on the class of those licences is also available, but we will not tabulate it here. The great majority are standard licences to drive a car; there were only 1639 specific motor cycle licences represented and 2568 truck licences.

CRASHES AND OFFENCES - CROSS-TABULATION

1. NUMBERS OF CRASHES vs NUMBERS OF OFFENCES

The first step in examining relationships between crashes and offence histories was to cross-tabulate the numbers of crashes per driver by the numbers of offences per driver. From such a tabulation we can derive

- the proportion of drivers who are involved in a crash but who have not had any offences recorded; and
- some measure of the rate of crashes amongst drivers with various numbers of offences. We have chosen to use as a measure the number of crashes per 100 drivers.

We have considered three sets of data in these cross-tabulations -

- all the data described in the previous section of this Report;
- the subset of the data which relates only to ACT - licenced drivers; and
- a further subset of the data which relates only to ACT - licenced drivers and to offences recorded over the same time period as the crashes (ie 1982 - mid 1985).

In summary, we can identify these data sets as

Data Set I - all available offences and crashes

Data Set II - all available offences and crashes for ACT - licenced drivers

Data Set III - concurrent offences and crashes for ACT - licenced drivers

Table 21 shows the percentage of drivers involved in crashes who had no offence history within the relevant data set. These results can be compared with Stewart *et al* (1972) who found that in North Carolina 71% of crash drivers had no offence in a prior 2 year period.

Table 21

Percentages, for the three data sets, of crash drivers with no offence history

Data Set	Percentage of Crash Drivers with No Offence History
I	73.4
II	77.8
III	78.7

These percentages are high, particularly since there is at least an overlap between the offence and crash recording periods (in North Carolina the offence period was prior to the crash period). The relatively high percentage is possibly an indication of the "mobile" nature of the ACT driving community - the drivers who have crashes in the ACT may have an offence history recorded outside the ACT.

The rates per 100 drivers of crashes for various numbers of offences are shown in Table 22. Again, the rates are calculated for the data sets I, II and III defined above.

Table 22

Crash rates versus number of offences

Numbers of Offences	Crashes per 100 drivers		
	Data Set I	Data Set II	Data Set III
0	N/A	N/A	N/A
	24.8	22.4	23.8
2	31.6	27.5	29.3
3	37.2	32.6	34.9
4	42.7	37.4	40.5
5	46.5	40.2	43.0
6	52.3	44.8	48.4
7	51.3	44.4	47.7
8	59.2	50.8	52.6
>8	64.0	54.7	57.7

Note first that, since we have no information about drivers who have no offences and no crashes, we are unable to calculate a crash rate for zero offences. (To calculate this rate, we would require a study of the offence and crash histories of all drivers, rather than of drivers who appear on offence and/or crash records).

The above rates should be compared with those of Coppin et al (1965), shown in

Table 1. The rates quoted by Coppin et al are for a concurrent period (i.e., they are comparable to those for Data Set III), and with crash-related offences excluded. The rates we calculate are lower than those quoted by Coppin et al, particularly for higher numbers of offences. However, we have included all offences, whereas Coppin et al restrict attention to offences carrying points in the Californian Highway Code.

Peck et al (1982) calculated crash rates in a predictive fashion for Californian data, i.e. crash rate in 1972-74 for offence numbers in 1969-71. (See Table 3). Their rates are also higher than those we obtain when the number of offences is high. Campbell (1958) calculated crash rates for North Carolina drivers for numbers of "moving offences" - his rates (Table 4) are much higher than those we calculated (and higher than those calculated for California data).

Despite variations in the magnitudes calculated for crash rates for various numbers of offences, it never-the-less remains clear that the ACT data supports that of the other sources in pointing to an increase in crash rate with increasing numbers of offences.

2. CORRELATIONS OF NUMBERS OF CRASHES WITH NUMBERS OF OFFENCES

We calculated product moment correlations between the numbers of offences and the numbers of crashes per driver; the correlations were calculated for each offence category separately. The data set used for this exercise was Data Set III in the list above - ACT - licenced drivers and a concurrent period of crashes and offences. It was felt this would provide the "best" correlations if correlations were to be found. The results were as shown in Table 23.

Table 23

Correlations of number of crashes with number of offences

Offence Category	Correlation
1	-0.2539
2	-0.1215
3	-0.0081
4	-0.0013
5	-0.1574
6	-0.0060
7	-0.0996
8	-0.0257
9	-0.0958
10	-0.0296
11	-0.0047

("Other" category not considered).

These results are startling! All are negative, and all but the two smallest would be judged by the standard test of significance to be significantly different from zero. However, we believe that they demonstrate only that the product moment correlation is not an appropriate measure of association for such data.

Indeed, we believe that correlations calculated in such a manner will be negative, regardless of any true relationship.

Consider for, say, speeding offences the distribution for each number of speeding offences of numbers of crashes per driver. For each number of speeding offences, the distribution of numbers of crashes per driver will be approximately Poisson, that is, a distribution with a long tail. However, the number of samples observed will be large for small numbers of speeding offences but reducing rapidly as the number of speeding offences increases. Thus, even if the Poisson distribution were <u>identical</u> at each offence number, there will tend to be more high values of crash numbers at the lower speeding offence numbers because of the sample sizes. The correlation coefficient is sensitive to such values, and thus will be negative. The situation is illustrated for a small set of data in Figure 6.

We believe that this is what Lew *et al* (1966) were attempting to say. However, how they were able to obtain consistently positive correlation coefficients (see Table 2) is not clear to us.

The gross failure of assumptions described above leads us to discontinue any analysis by way of correlations; it would also cast doubt on any simple least squares regression analysis of such data. Regression analysis with a more appropriate model is discussed in a later Section.

```
crashes -
        - 2
        -
     4.5+
        -
        - 7    *    *    *
        -
        -
     3.0+ +    5    *    2    *    2                   *
        -
        -
        - +    +    +    4    3    4              *
        -
     1.5+
        -
        - +    +    +    +    2    3    2              *
        
        -
     0.0+ +    +    +    +    9    4    2    2    *    *      Speeding
          +---------+---------+---------+---------+---------+------ Offences
         0.0       2.0       4.0       6.0       8.0      10.0
```

Figure 6 Number of crashes for drivers having various numbers of speeding offences.

3. OFFENCE DISTRIBUTION COMPARISONS

As an indication of the relationship between offences and crashes, we can look at the offence distributions for various subsets of the data available. We will, initially, look at differences in offence distributions for several sub-divisions of the total data available (i.e., the data referred to earlier as Data Set I). Subsequently, we will carry out the same examinations for Data Set III to see what differences, if any, in interpretation occur.

By "offence distribution" in what follows, we refer to the total numbers of offences in each of the categories defined earlier (see the Section "Categorization of Offences") for all drivers in a specified group, e.g. drivers who have had a crash.

The first comparison we make is of offence distributions between drivers who have had a crash and drivers who have not had a crash.

We will not attempt, with regard to this comparison or any that follow, to carry out a statistical test of the null hypothesis that these two distributions are the same. The sample sizes are so large that even very minor differences will be seen as significant - i.e. they will be statistically significant but practically of little use. We think it more useful to comment on differences that appear to be relevant to the purpose of the study. The relevant distributions are shown in Table 24 and in Figure 7.

Table 24

Offence distributions of drivers who have had a crash and drivers who have not had a crash

Offence Category	Offence Numbers (and % of total)			
	Drivers who have had a crash		Drivers who have not had a crash	
1	16178	(39.6)	36794	(44.6)
2	2104	(5.2)	4814	(5.8)
3	293	(0.7)	327	(0.4)
4	2447	(6.0)	2293	(2.8)
5	5298	(13.0)	10795	(13.1)
6	129	(0.3)	183	(0.2)
7	5316	(13.0)	9404	(11.4)
8	155	(0.4)	358	(0.4)
9	8350	(20.5)	16751	(20.3)
10	249	(0.6)	450	(0.5)
11	305	(0.7)	307	(0.4)

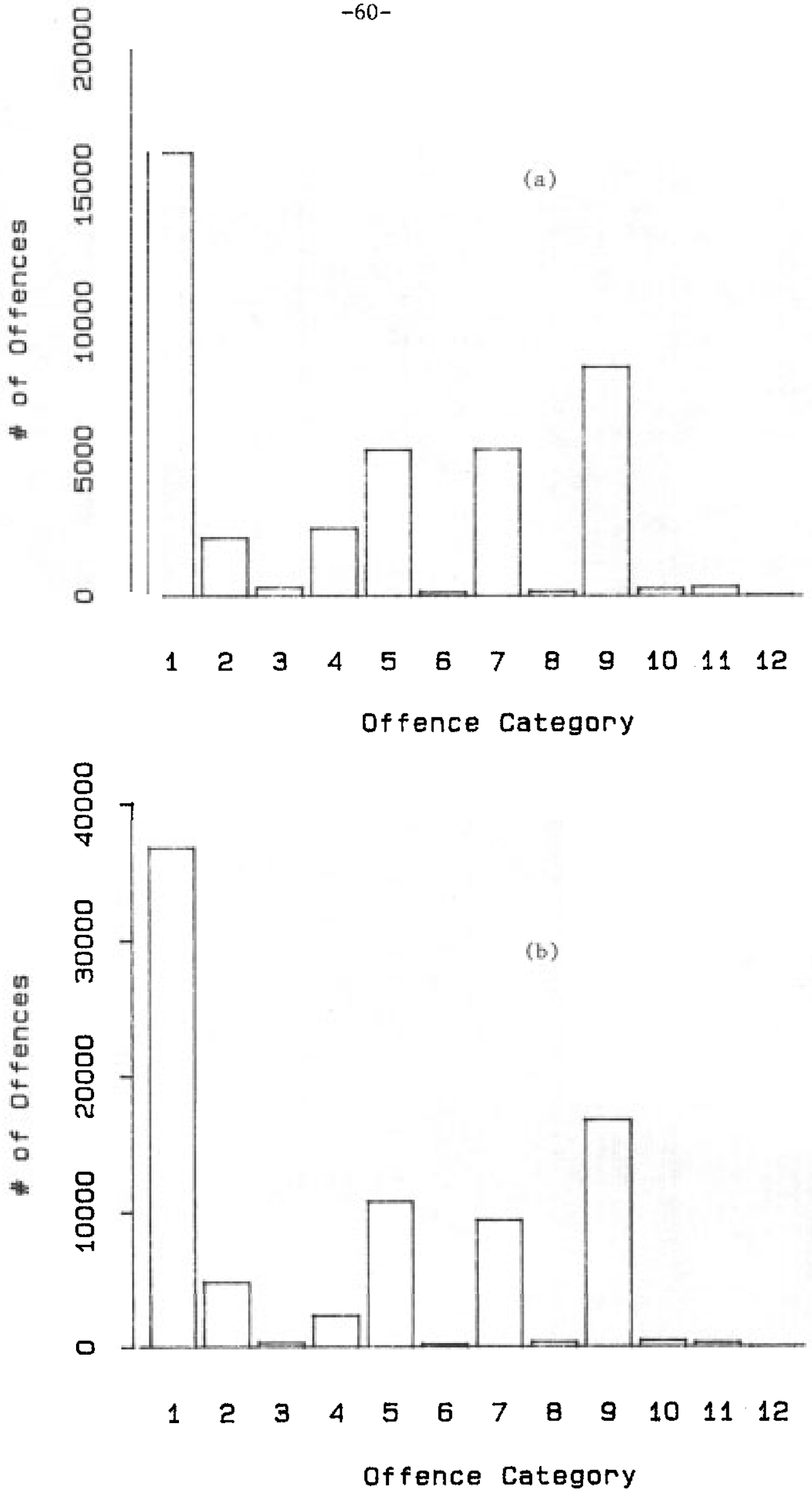


Figure 7 Offence distributions of drivers
(a) who have had a crash
(b) who have not had a crash

With reference to Table 24, we see that drivers who have not had a crash tend to have a higher proportion of speeding offences than those who have had a crash or crashes (44.6% to 39.6%). This is balanced by the drivers involved in crashes having a higher proportion of offences in Category 4 (driving / overtaking/passing/pulling out) - 6.0% to 2.8% for drivers not involved in crashes. The drivers involved in crashes also tend to have more offences in the "equipment" category - 13% to 11.4% for those not involved in crashes. This could reflect the suggestion that such offences are "added on" to offences of a more serious nature recorded at the time of the crash.

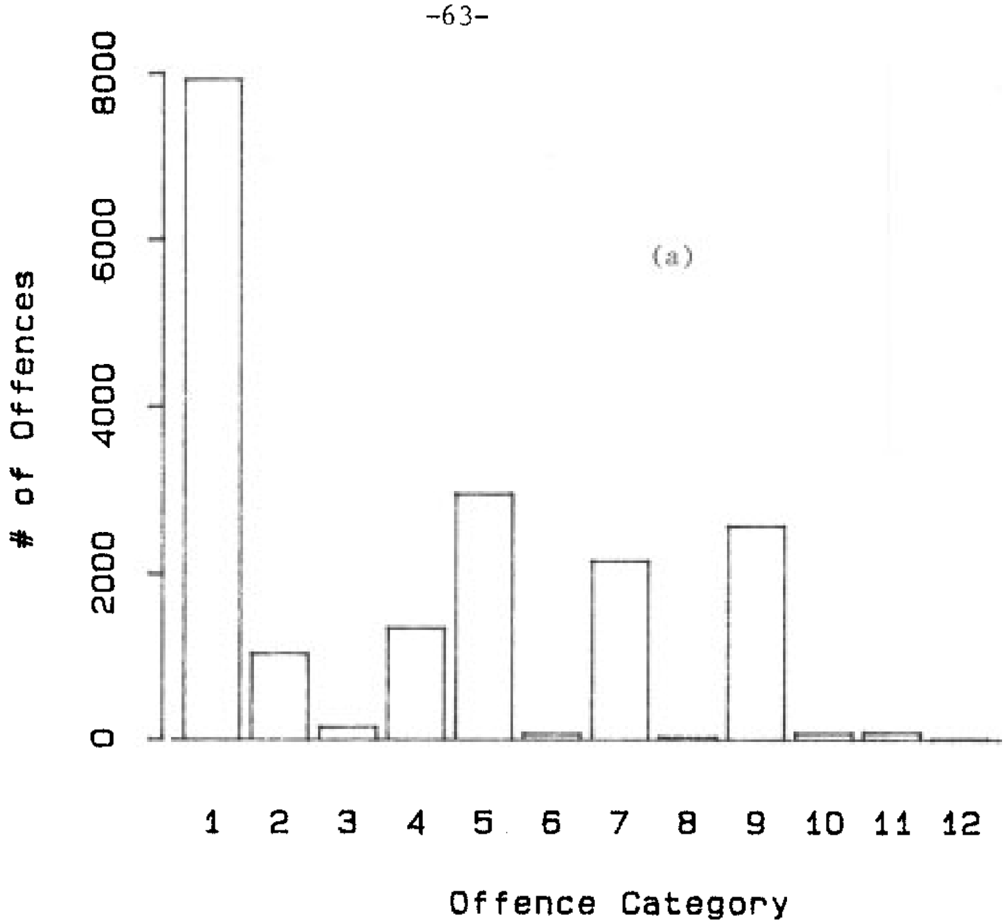
Another comparison of interest is that between drivers prior to their first crash and drivers subsequent to their first crash. The results are given in Table 25 and illustrated in Figure 8. (Note that offences recorded on the same date as a crash are treated as being prior to the crash.)

We observe that, again, speeding offences decrease as a proportion of the total after the first crash, as do offences of failure to obey lights, markings, signals and directions. There is, however, a marked increase in the proportion of offences relating to "equipment", and in offences of failure to obey licencing and registration regulations.

Table 25

Offence distributions of drivers prior to and after their first crash

Offence Category	Offence Numbers (and % of total)			
	Prior to 1st Crash		After 1st Crash	
1	7927	(43.1)	8235	(36.8)
2	1033	(5.6)	1071	(4.8)
3	144	(0.8)	149	(0.7)
4	1346	(7.3)	1097	(4.9)
5	2962	(16.1)	2326	(10.4)
6	69	(0.4)	56	(0.3)
7	2157	(11.7)	3152	(14.1)
8	26	(0.1)	129	(0.6)
9	2577	(14.0)	5765	(25.8)
10	76	(0.4)	173	(0.8)
11	81	(0.4)	223	(1.0)
12	2	-	0	-



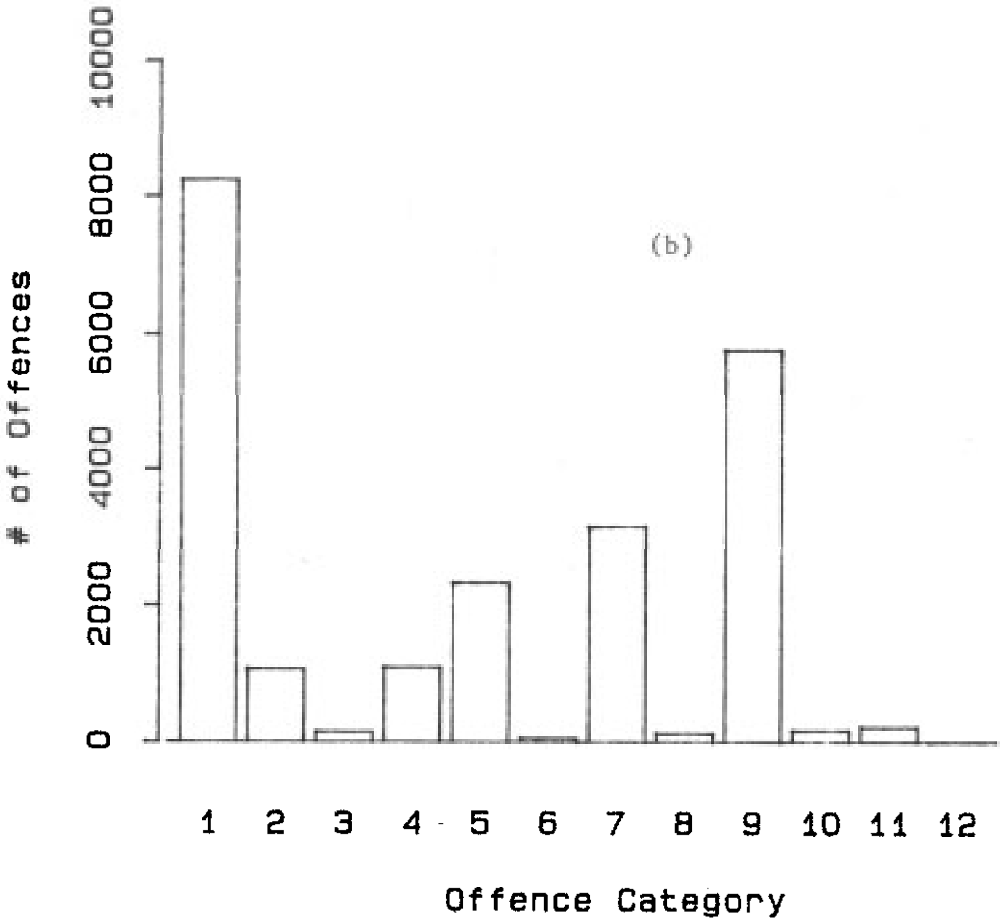


Figure 8 Offence distributions of drivers
(a) prior to their first crash
(b) after their first crash

We have also looked at the offence distribution of drivers after they have their only crash, i.e. we exclude the post - first crash histories of drivers who have multiple crashes. This distribution is essentially identical to that for drivers after their first crash (shown in Figure 8 (b)), i.e. post crash offence behaviour doesn't appear to be modified by subsequent crashes.

Table 26 presents figures for offence distributions of motor-cycle riders involved in crashes and the offence distribution for car drivers involved in crashes.

Table 26

Offence distributions of motor cycle riders and car drivers in crashes

Offence Category	Number of Offences (and % of total)			
	Motor-cycle Riders in Crashes		Car Drivers in Crashes	
1	156	(39.4)	13534	(39.8)
2	17	(4.3)	1730	(5.1)
3	2	(0.5)	246	(0.7)
4	29	(7.3)	2060	(6.1)
5	53	(13.4)	4360	(12.8)
6	1	(0.3)	102	(0.3)
7	48	(12.1)	4426	(13.0)
8	2	(0.5)	134	(0.4)
9	82	(20.7)	6958	(20.4)
10	2	(0.5)	201	(0.6)
11	4	(1.0)	250	(0.7)
12	0	-	1	-

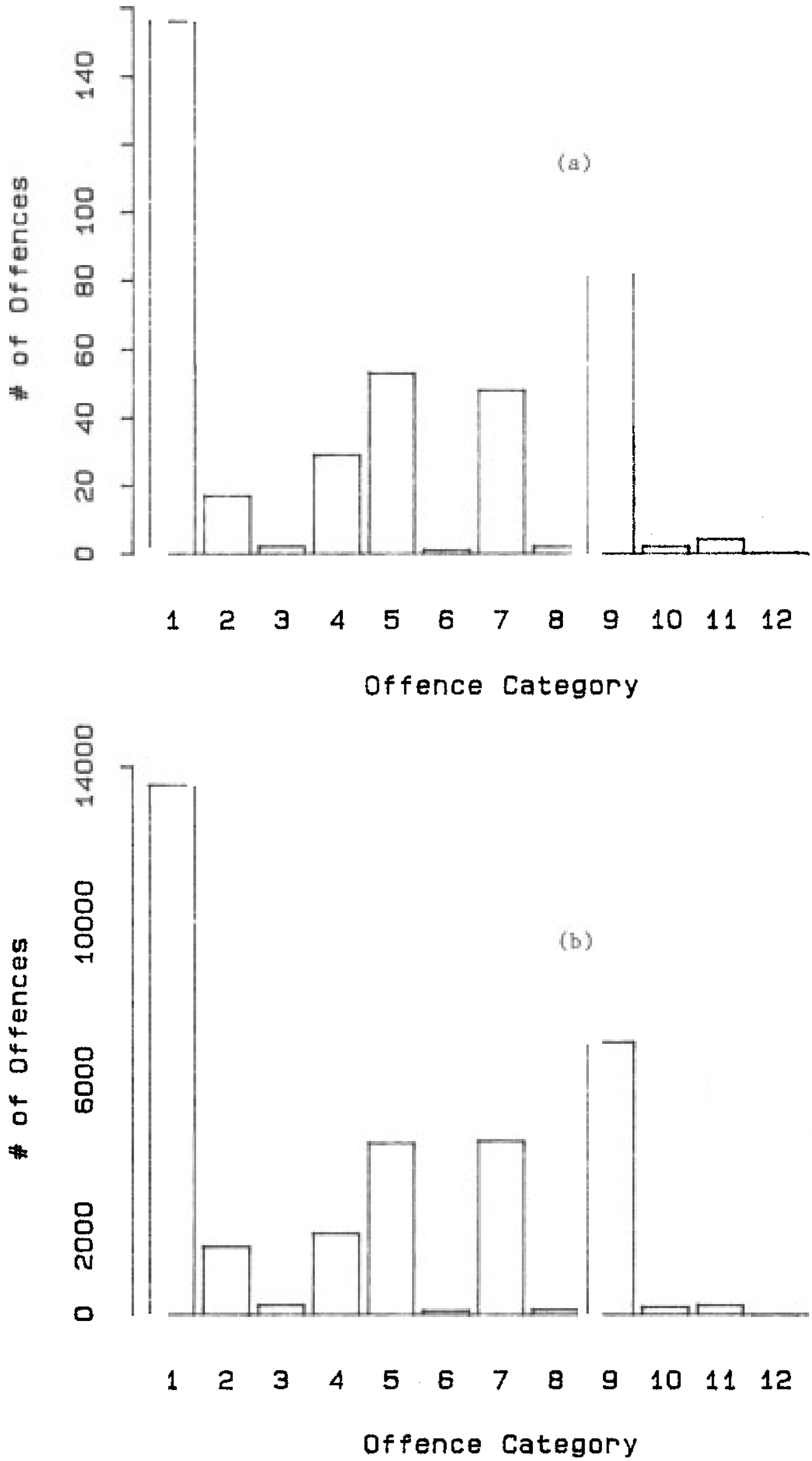


Figure 9 Offence distributions for
(a) motor cycle riders in crashes
(b) car drivers in crashes

The data are illustrated in Figure 9. Despite the sample size for motor cycle riders being low and hence the results more variable, there is no discernible difference between these distributions.

Some comparison is required between offence distributions for "young" and "old" drivers involved in crashes. Here definition is a problem on several fronts:-

- information is available only on the driver's date of birth, not on the number of years of driving experience (which may be more relevant);
- whatever cut-off point is used to divide "young" and "old", many drivers will pass from one category to the other over the period of the study; and
- the choice of a cut-off point is arbitrary.

We chose to use 25 years of age as a dividing age, and compared drivers who were under 25 at their first crash with drivers who were over 25 at their last crash (within the study period). The results are shown in Table 27 and Figure 10.

Little difference is discernible. Young drivers involved in crashes tend to have more offences of drunk driving and of failure to obey lights, markings, signals and directions, and fewer offences of driving, overtaking, passing and pulling out; the differences are not sufficiently large to be remarkable.

Table 27

Offence distributions for "young" and "old" drivers in crashes

Offence Category	Number of Offences (and % of total)	
	"Young" drivers in crashes	"Old" drivers in crashes
1	3577 (39.9)	2655 (39.7)
2	491 (5.5)	261 (3.9)
3	64 (0.7)	62 (0.9)
4	490 (5.5)	446 (6.7)
5	1204 (13.4)	792 (11.8)
6	30 (0.3)	18 (0.3)
7	1171 (13.0)	929 (13.9)
8	34 (0.4)	34 (0.5)
9	1799 (20.0)	1388 (20.8)
10	52 (0.6)	37 (0.6)
11	62 (0.7)	67 (1.0)
12	2 –	0 –

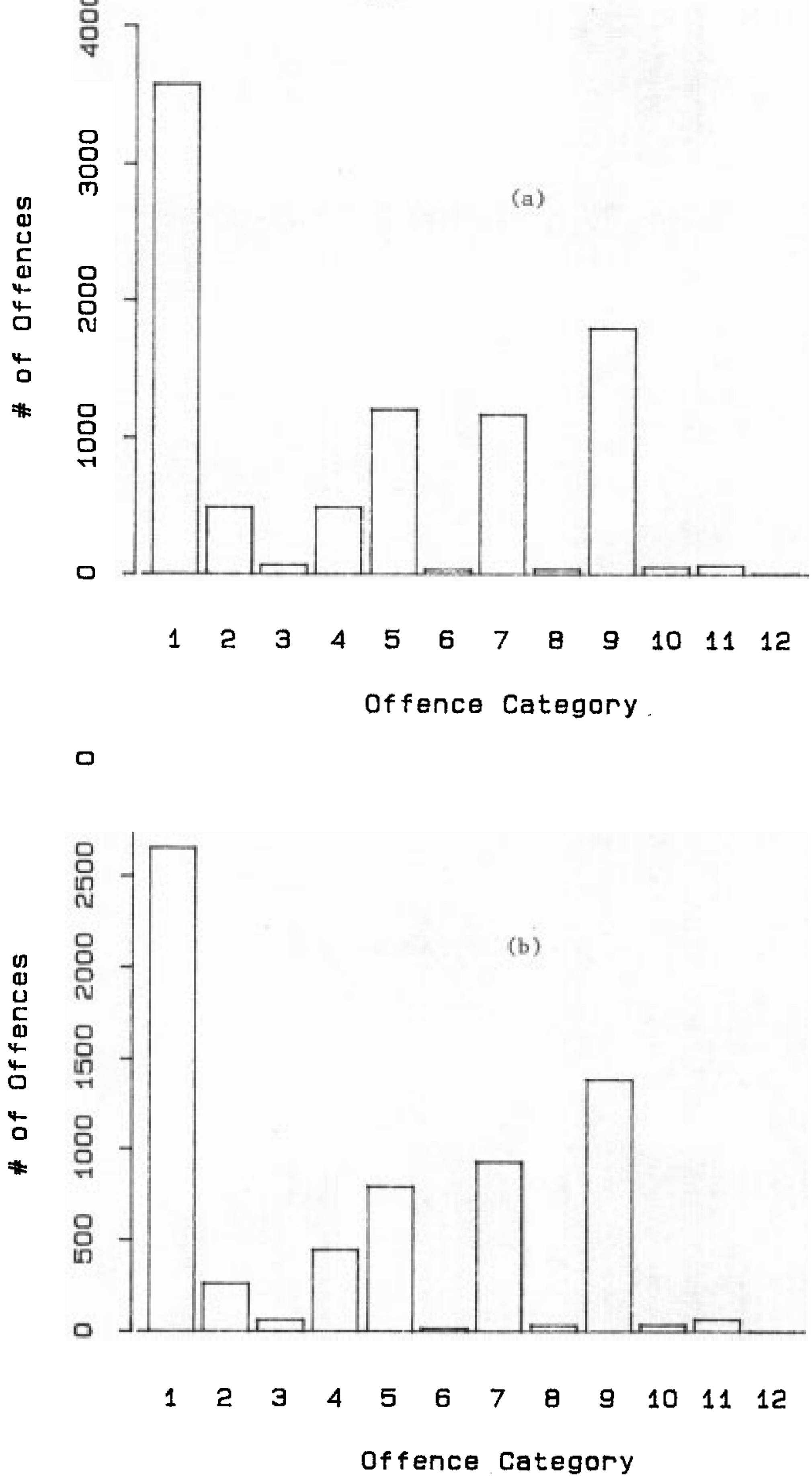


Figure 10 Offence distributions for

(a) "young" drivers in crashes

(b) "old" drivers in crashes

The final comparison we make is by severity of crash. Table 28 shows the offence distributions for drivers involved in non-casualty crashes and those involved in casualty crashes. (This is the National Capital Development Commission's rating of severity). **The figures are illustrated in Figure 11.**

These distributions are very similar, and do not suggest any difference in offence distributions for drivers in casualty crashes and drivers in non-casualty crashes.

Table 28

Offence distributions for drivers in non-casualty crashes and drivers in casualty crashes

Offence Category	Number of Offences (and % of total)	
	Drivers in Non-Casualty Crashes	Drivers in Casualty Crashes
1	13831 (39.7)	2233 (39.1)
2	1868 (5.4)	227 (4.0)
3	254 (0.7)	39 (0.7)
4	2100 (6.0)	319 (5.6)
5	4554 (13.1)	707 (12.4)
6	108 (0.3)	17 (0.3)
7	4486 (12.9)	782 (13.7)
8	129 (0.4)	24 (0.4)
9	6999 (20.1)	1284 (22.5)
10	214 (0.6)	35 (0.6)
11	259 (0.7)	43 (0.8)
12	2 –	0 –

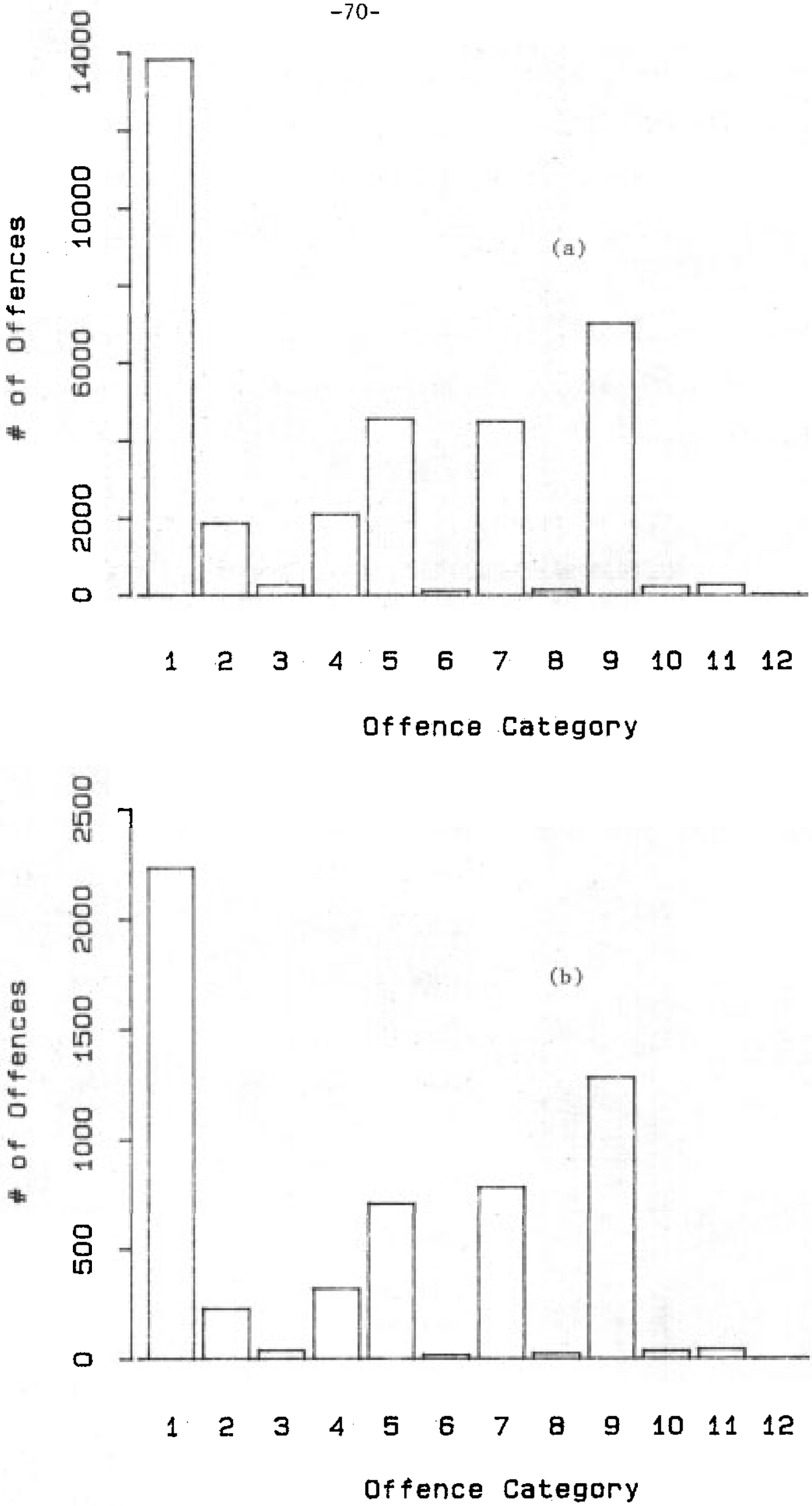


Figure 11 Offence distributions for

(a) drivers in non-casualty crashes

(b) drivers in casualty crashes

first crash and after the driver's first crash, the restriction of the data set leads to only minor random changes in the distributions. The remarks made about Table 25 therefore apply equally to the restricted data set.

The numbers of crashes experienced by motor-cycle riders is reduced, in the ACT-licenced, concurrent periods data set, to a small number, and hence the comparison between offence distributions for motor-cycle riders in crashes versus car drivers in crashes is even less reliable than in Table 26. However, there is again no discernible difference between the offence distributions.

The "young" versus "old" comparison shows, for the ACT-licenced, concurrent period data set, even less differences than were found in the total data set. (See Table 27 and the comments thereon). In fact, none of the differences remarked on in the comments on Table 27 can be seen in the restricted data set when viewed in this way. Similarly, a division of the restricted data set into drivers in non-casualty crashes and drivers in casualty crashes shows no difference in offence distributions.

We conclude, from the analyses described above, that no difference in interpretation of comparisons of offence distributions is necessitated when the data set is reduced from all that available to data on ACT- licenced drivers and offences and crashes in the same time period.

5. REMOVAL OF CRASH-RELATED OFFENCES

As noted in several of the references cited in the Review of Related Studies, it might be expected that offences recorded against drivers at a crash site have a distribution different from that of the general mass of offences. To examine this possibility, we isolated the offences recorded on the same date as the crash; we assume that these offences were recorded at the crash site. Considering all the available data (ie Data Set I), the distribution of these crash-related or coincident offences is as shown in Table 30 and Figure 12.

Clearly, the distribution of crash-related offences is very different from the general distribution of offences. It is dominated by offences relating to licencing and regulation - 61.9% of the total.

Next most prominent are the equipment offences (16.8%), offences relating to failure to obey lights, markings, signals and directions (7.8%) and speeding (5.9%). The preponderance of licencing and regulation offences amongst the crash-related offences is surprising; we envisaged that equipment offences would perhaps be as large a contributor to the crash-related offences. However, the offence categories represented strongly amongst the crash-related offences are those one would expect.

Table 30

Distribution of crash-related offences

Offence Category	Number of crash-related offences (and % of total)
1	1648 (5.9)
2	181 (0.7)
3	139 (0.5)
4	442 (1.6)
5	2161 (7.8)
6	188 (0.7)
7	4653 (16.8)
8	279 (1.0)
9	17204 (61.9)
10	349 (1.3)
11	530 (1.9)

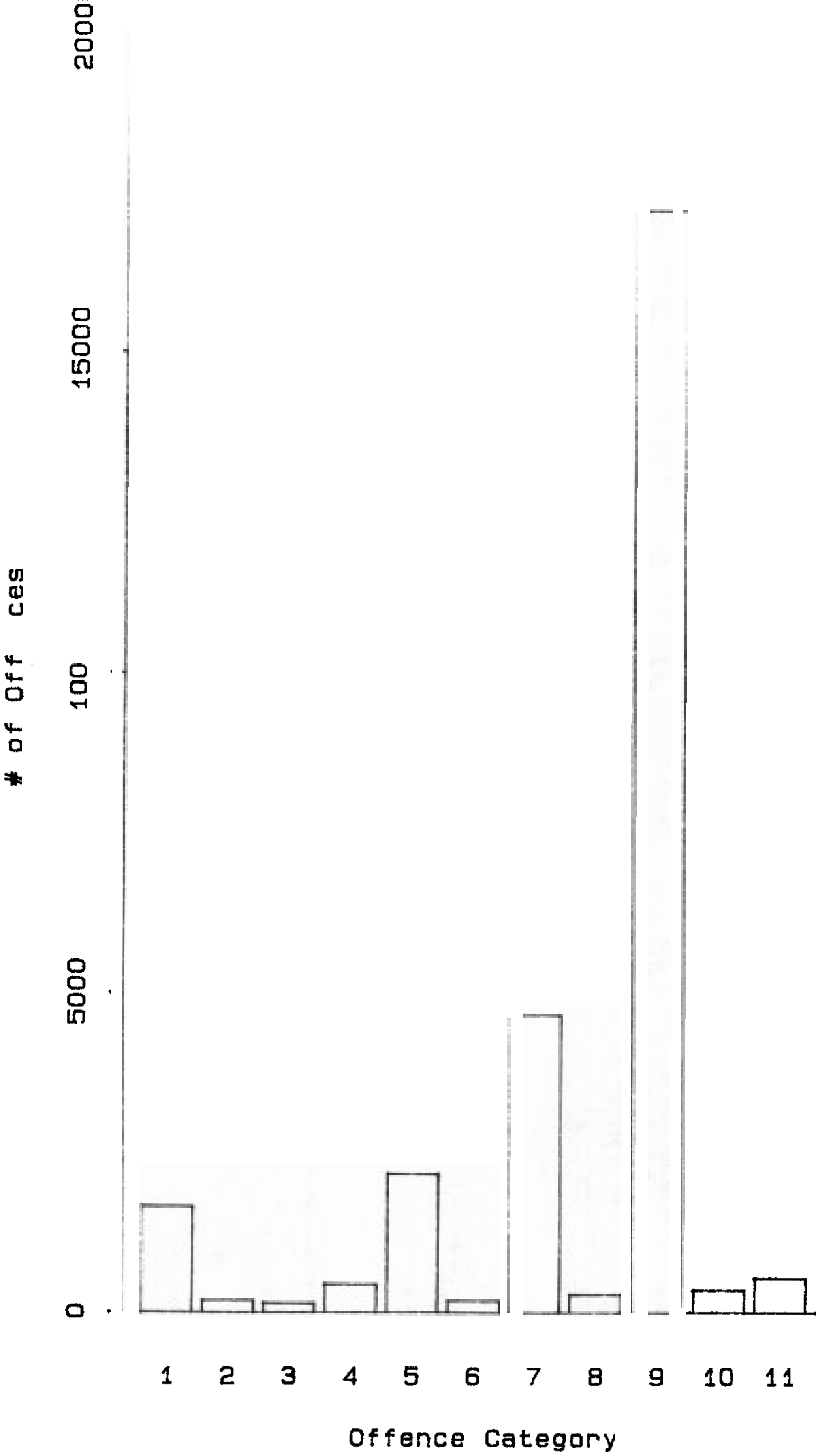


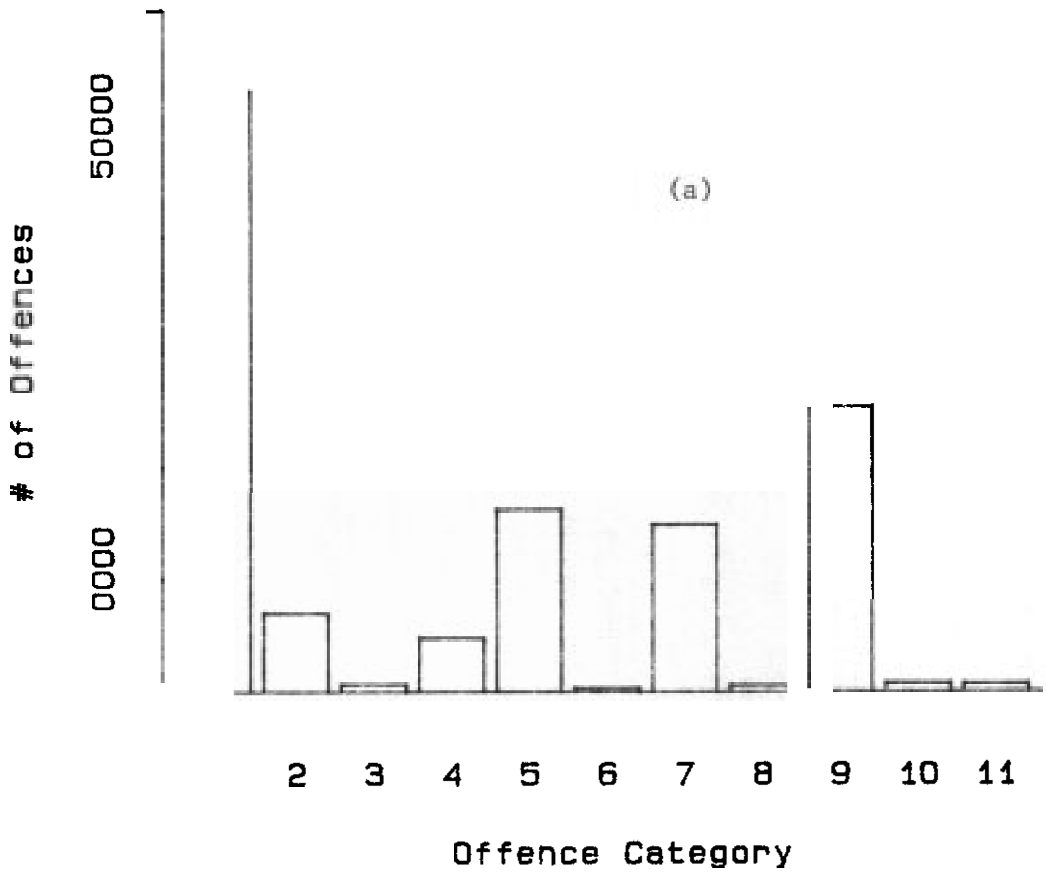
Figure 12 Distribution of crash-related offences

Table 31 and Figure 13 compare the offence distributions for the whole of Data Set I and for Data Set I with the crash-related offences deleted. We see that, as expected, the proportion of licencing and regulation offences is greatly reduced (from 20.4% to 8.4%). The major off-setting increase is in the proportion of speeding offences (43.0% to 53.7%), drunken driving (5.6% to 7.1%), failure to obey lights, markings, signals and directions (13.1% to 14.6%). There is a slight reduction in the proportion of equipment offences (from 11.9% to 10.5%).

Table 31

Distribution of offences with and without crash-related offences

Offence Category	Number of offences (and % of total) All data	Number of offences (and % of total) Crash-related offences removed
1	52972 (43.0)	51324 (53.7)
2	6918 (5.6)	6737 (7.1)
3	620 (0.5)	481 (0.5)
4	4740 (3.8)	4298 (4.5)
5	16093 (13.1)	13932 (14.6)
6	309 (0.3)	121 (0.1)
7	14720 (11.9)	10067 (10.5)
8	513 (0.4)	234 (0.2)
9	25101 (20.4)	7897 (8.3)
10	699 (0.6)	350 (0.4)
11	612 (0.5)	82 (0.1)



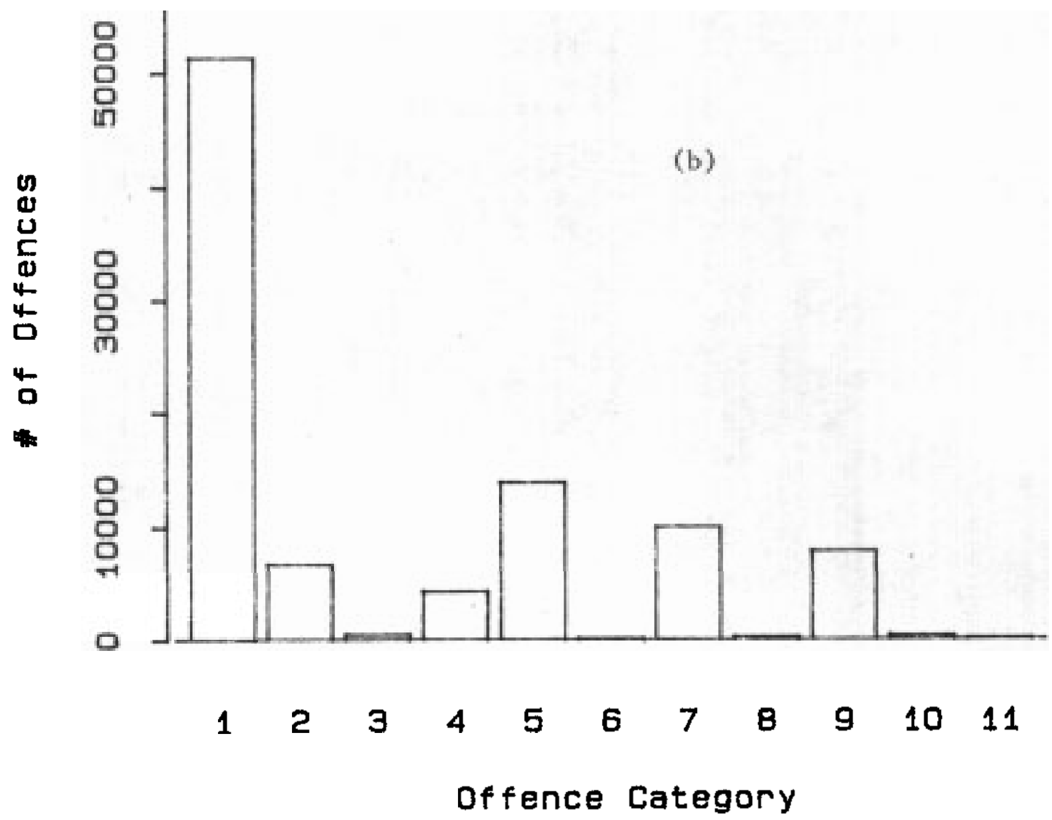


Figure 13 Offence distribution

(a) for all offences

(b) with crash-related offences removed

The differences brought about in offence distributions by the exclusion of crash-related offences make it essential to examine whether their exclusion would alter the interpretation of comparisons between offence distributions for various classes of drivers. In order to check the effect of deleting crash - related offences, we repeat the comparisons made in 3. above but for Data Set I with the crash related offences removed from the data set.

The comparison made in Table 24 is repeated in Table 32, where the distributions of offences for drivers who have had a crash, and drivers who have not had a crash, are compared in the absence of crash related offences.

Table 32

As for Table 24, but with crash-related offences removed

Offence Category	Offence Numbers (and % of total) Drivers who have had a crash	Drivers who have not had a crash
1	15888 (50.5)	35451 (55.3)
2	2047 (6.5)	4690 (7.3)
3	244 (0.8)	237 (0.4)
4	2269 (7.2)	2033 (3.2)
5	4579 (14.5)	9359 (14.6)
6	56 (0.2)	65 (0.1)
7	3562 (11.3)	6508 (10.2)
8	68 (0.2)	166 (0.3)
9	2584 (8.2)	5315 (8.3)
10	128 (0.4)	222 (0.3)
11	45 (0.1)	37 (0.1)

We see that the general shape of the distributions has been re-aligned as in Table 31, but that the interpretation of the comparison between the two classes of drivers is not altered (except that the difference in the equipment offence category is no longer evident).

Turning to the other comparisons made in 3. above, we found that the interpretations made in 3. are valid for the same comparisons made on data from which crash-related offences are deleted.

It would seem, then, that deleting of crash-related offences does not substantially alter any of the offence distribution comparisons we wish to make. The deletion does, however, alter the overall offence distribution and this must be taken into account in any attempt to predict crash propensity from offence histories.

6. SUMMARY OF CROSS-TABULATION OF OFFENCES AND CRASHES

We have examined the relationships between offences and crashes by various cross-tabulations, and by looking at correlations between numbers of crashes and numbers of offences. The use of correlations as a measure of association is, because of the form of the data, inappropriate and has not been pursued.

The cross-tabulations of the data have shown that

- there is an increase in the number of crashes per 100 drivers as the total number of offences increases;
- drivers who have not had a crash have a higher proportion of speeding offences and a lower proportion of offences related to driving/overtaking/passing/pulling out than have drivers who have had a crash;
- drivers after their first crash have a lower proportion of speeding offences and offences of failure to obey lights, markings, signals and directions, but an increase in equipment offences and licencing and regulation offences;
- there is no discernible difference between the offence distributions of motor-cycle riders in crashes and car-drivers in crashes (though the number of motor-cycle riders is small);
- the age of the driver in the crash has little effect on offence distribution;

- there is no difference in the offence distributions of drivers involved in non-casualty crashes in comparison to those involved in casualty crashes;
- restricting the data set to ACT-licenced drivers and to offences in a period concurrent with the crashes does not affect the conclusions;
- deleting crash-related offences radically changes the offence distribution itself, but not the conclusions drawn above.

LINEAR DISCRIMINANT ANALYSIS

Linear discriminant analysis is a multivariate statistical technique which aims to specify a linear combination of the variables measured on sample members which will best separate the sample into pre-specified groups. In our context here, we have eleven variables measured on each driver who appears on the records - the number of offences the drivers has had recorded against him or her in each of the eleven categories. (We disregard the "Other" offence category in this analysis). In addition, we can specify each driver on the records as belonging to one of two groups - one group of drivers who have had no crashes recorded, and the other group of drivers who have had one or more crashes recorded. Discriminant analysis provides a set of eleven weights to be applied to the driver's offence record to give a single indicator number, and a cut-off value such that if the driver's indicator number falls below the cut-off the driver is assigned to one group (and to the other group if the indicator number falls above the cut off). The theoretical details are given in most books on multivariate statistical analysis, e.g. that of Cooley and Lohnes (1971).

The following should be noted

i. The discriminant analysis will not correctly assign every driver as being one who has had a crash or one who has not had a crash. The proportion of correct assignments is an indication of how effectively the discriminant function (the linear combination) discriminates between drivers.

ii. The ultimate aim of discriminant analysis is to use a sample with known group assignment as a "training set" to develop a discriminant function which will assign future observational units whose status is unknown. In our context, we use drivers with known offence and crash records to develop a discriminant function; if it is effective we can use it to assign drivers, on the basis of their offence records, as likely to have a crash or not likely to have a crash.

We have used two sets of data in examining discriminant analysis as a tool for determining crash propensity.

i. The data for ACT-licence drivers and for offences and crashes in the period 1982 to mid-1985,but with the offences recorded on the same date as a crash deleted (the concurrent,non-coincidental data set);and

ii. A data set for ACT-licenced drivers consisting of offences recorded in 1980, 1981 and 1982 and crashes recorded in 1983, 1984 and part of 1985 (the predictive data set).

The predictive data set is similar to some used in overseas studies, and is designed to simulate a purely predictive use of the driver's offence record.

Note,however,that the data as described above will include many drivers who have no offence record,but who have had a crash. Indeed,the only way a driver with no offence record can appear is when a crash is recorded for them. The information about drivers with no offence records is severely biased,then - 100% of them are seen to have had a crash. This will influence the outcome of a discriminant analysis,and we have chosen to delete the drivers with no offence record from the data before carrying out the analysis.

Discriminant analysis was carried out using the GENSTAT statistical package. GENSTAT has a sequential read and calculate facility which enables large data sets to be analysed.

The results for the concurrent,non-coincident data set, with 36664 driver records, are shown in Table 33, and the accuracy of the prediction of crash status in Table 34. Note that the signs of the weights are not important - all the signs could be reversed, and the same discrimination achieved by looking above the cut-off rather than below for the members of a group. What is more important is the weight given to the more frequently occurring offence categories, and the accuracy of the prediction.

Looking first at the results in Table 34,showing the success with which the linear discriminant function assigns the members of the sample to either "crash" or "no crash" status,we see that the results are very disappointing. The discriminant function assigns 3562 drivers as not likely to have a crash when in fact they did,and a massive 21660 drivers as prone to have a crash when no crash was recorded against them. With the prediction being as poor as it is,there is little point in examining the weights used by the linear discriminant to arrive at its results.

Table 33

Weights calculated from linear discriminant analysis

Offence Category	Weights (Concurrent, non-coincident data set)
1	0.3539
2	0.5028
3	2.2812
4	2.5859
5	0.7793
6	2.7400
7	0.3565
8	-0.5742
9	0.2347
10	1.1583
11	4.2604

Cut-off value 0.9184

Table 34

Accuracy of prediction of crash status

		Actual Status (Concurrent, non-coincident data set)	
		No Crash	Crash
Predicted	No Crash	5968	3562
	Crash	21660	5474

We turn to the predictive data set, where offences in one time period are being used to predict crashes in a subsequent period. The results are shown in Tables 35 and 36.

The figures shown in Table 36 are just as disappointing as those shown earlier for the concurrent,non-coincidental data set. It would seem from these results that linear discriminant analysis does not provide a useful way of obtaining a relationship between crashes and offences that could be used to predict crash propensity from offence history. This inability may be due to the necessity to exclude those drivers who had a recorded crash but no recorded offence history - if more information had been available about the "no offence,no crash" drivers, better results may have been obtained.

Table 35

Weights calculated from linear discriminant analysis - predictive data set

Offence Category	Weights (Predictive data set)
1	-0.8755
2	0.0320
3	-2.2767
4	-1.4130
5	-0.8280
6	-5.2135
7	-1.0208
8	1.7209
9	-0.6221
10	-4.1208
11	-3.2926

Cut-off value -1.2309

Table 36

Accuracy of prediction of crash status - predictive data set

		Actual Status (Predictive Data Set)	
		No Crash	Crash
Predicted	No Crash	4701	1203
	Crash	13004	2094

FITTING A GENERALIZED LINEAR MODEL

As explained in the Appendix, modelling of the relationship between offence histories and crashes is most appropriately approached through a generalised linear model. Such an approach seeks to relate a parameter of the distribution of the number of crashes to the number of offences in each category. We have applied this to two data sets-

- the data for ACT-licenced drivers, with offences and crash recorded over a concurrent period and the crash related offences deleted (the concurrent, non-coincident data set), and

- the data for ACT-licenced drivers, with offences recorded for 1980-82 and crashes for 1983-mid 1985 (the predictive data set).

For each data set, we isolated the distinct offence histories, and for each different history counted the number of drivers with that history and the number of those who have recorded crashes. Note that a zero in each category, i.e. a history of no offences, is a history, and in fact the most commonly-occurring one. However, we must exclude it from our analysis because the null history occurs only for drivers who have had crashes. This is, again, a disadvantage of relying on data restricted to those who have had either on offence or a crash - we are deprived of information on drivers who have no offences and no crashes.

We fitted a generalised linear model which expressed the probability of crash involvement in terms of the numbers of offences in each of the eleven categories via the logistic model. (See the Appendix for details). The result of the fitting are shown in Table 37 for both data sets considered.

Table 37

Estimated coefficients for the Generalised Linear Model

Offence Category	Coefficients	
	Concurrent, non-coincident data set	Predictive data set
1	0.1492	0.1973
2	0.2200	-0.0491
3	0.9090	0.4986
4	1.0671	0.3245
5	0.3451	0.1867
6	1.0930	1.0620
7	0.1479	0.2236
8	-0.2820	-0.6760
9	0.0943	0.1345
10	0.5180	0.8950
11	1.6400	0.7370
Constant	-1.5366	-1.9685

Note that a constant term is included in the linear combination of numbers in offence categories. That is, although we have excluded offence histories with no offences from the data, the model enables us to estimate the probability of crash involvement for such drivers - 0.1770 for the concurrent, non-coincident data set and 0.1226 for the predictive data set.

However, before proceeding to further comment on the fitted model, it is appropriate to examine the adequacy of the fit. The deviances, and corresponding degrees of freedom, are

Concurrent, non-coincident data set:	1672 on 1128 degrees of freedom
Predictive data set :	600.3 on 518 degrees of freedom

Both of the deviances are very significantly large when compared with the appropriate X^2 distribution. Therefore the statistical evidence is that the generalised linear model using all the available offence category data does not provide a useful model for the probability of crash involvement. Further statistical inference based on the estimated weights, or on the estimated probabilities calculated from the weights, is therefore inappropriate.

One further analysis that was appropriate was to examine whether the fit was significantly degraded by aggregating the offences into group. To do this, we summed the numbers in offence categories 1-6 to form a count of "major" offences, and summed the number in offence categories 7-11 to form a count of "minor" offences. Using only those two variables in the generalised linear model, there was a statistically significant degradation in the adequacy of the fit of the model in both data sets. That is, there is statistical evidence that information about the individual offence categories leads to an improvement in the fit of the model (although not enough to make the fit adequate). This result is contrary to results in other studies, where very broadly aggregated offence numbers were found to be just as useful as detailed offence information in predicting crash numbers.

DISCUSSION

The study reported here has been based on two large and reasonably well maintained data sets -

- offences recorded against drivers in the ACT in the period 1980 - early 1986,and
- crashes recorded in the ACT in the period 1982 - mid-1985

We have linked these two data sets to achieve a data base of drivers who have had offences recorded against them in the ACT,been involved in crashes in the ACT or both. Unfortunately,we have no information on drivers who have had neither an offence nor a crash,and this lack of information appears to be a crucial factor in attempts to establish a relationship between offences and crashes.

In order to reduce the offence data to manageable dimensions,we have assigned the offences recorded against drivers into eleven categories. The categories were chosen by us to represent what we saw as different types of driver behaviour.

We summarised the offence data and the crash data in various ways;the following remarks could be made about the data.

- Offence distributions were consistent through the years for which data was available,with speeding accounting for 43% of all offences.

- Drivers against whom only a single offence was recorded had more speeding offences and offences of "failure to obey lights,signals,markings and directions" than did drivers against whom more than one offence had been recorded.

- drivers with multiple offences had more offences of "failure to obey licencing and registration regulations" and equipment offences than did drivers with a single offence.

In examining the relationship between crashes and offences,we used three basic data sets initially -

- all the data described above

- the data set restricted to ACT-licenced drivers only

- the data set restricted to ACT-licenced drivers and to offences recorded during the same time period as the crashes.

We found that about three-quarters of drivers involved in crashes had no previous offence history. However,we also found that crashes per 100 drivers increased steadily as the number of offences per driver increased. That is,the more offences a driver has,the more likely the driver is to be involved in a crash. One approach taken to quantifying such a relationship is to use the correlation coefficient and/or linear least squares regression analysis. Examination of the data led us to conclude that such an approach was inappropriate,and likely to give misleading results.

Comparisons were made between offence distributions for various sub-groupings of the total data set. Because of the very large sample sizes involved,any difference between the distributions would be seen to be statistically significant even if the difference was very minor;we have therefore not carried out such significance tests. When the comparisons were made,we observed the following.

- Drivers not involved in crashes had more speeding offences than did drivers who were involved in crashes. However,the drivers involved in crashes had a higher proportion of offences of "driving/overtaking/passing/pulling out" and more equipment offences.

- Drivers had fewer speeding offences and fewer offences of "failure to obey lights,markings,signals and directions" after their first crash. However, they had a higher proportion of equipment offences and offences related to licencing and registration.

- There was no difference between the offence distributions of motor-cycle riders involved in crashes and car drivers involved in crashes. The number of motor-cycle riders involved in crashes was small,and their offence distribution was correspondingly less reliable.

- There was no apparent difference between the offence distributions of young and old drivers involved in crashes.

- The severity of the crash did not seem to be a factor showing a difference between offence distributions.

The above observations held equally when the data set was restricted to ACT-licenced drivers,and when it was further restricted to a concurrent time period for offences and crashes.

When we examined crash-related offences,i.e. those offences recorded against a driver on the same date as a crash was recorded,we found that the distribution of the crash-related offences was substantially different from that of all offences. 61.9% of crash-related offences were of the category of licencing and registration offences. Removing the crash-related offences does change the offence distribution in the various groups considered,but does not alter the outcome of any of the comparisons discussed above.

A variety of other comparisons were of potential interest,and some (such as crashes on roadways versus crashes on other than roadways) were suggested to us. However,the comparisons necessary would have been unreliable because one of the groups would have had quite small numbers of drivers involved in such crashes.

We attempted,through two statistical techniques,to obtain some relationship between crash involvement and offence history that might be of value for prediction of likely future crash involvement. Neither linear discriminant analysis nor generalised linear models were able to establish a useful relationship. We believe that this failure is due,at least in part,to the available data covering only those who have had recorded against them an offence,a crash or both. No information is thus available about the number of drivers with no recorded offences or crashes. Because this information is not to hand, the large number of drivers who are involved in a crash without having any offence record must be excluded from the analysis in order to avoid bias in the results. The remaining data,when analysed,did not support any relationship.

REFERENCES

CAMPBELL, B.J. (1958) Driver Improvement: The Point System Institute of Government Universities of North Carolina, for Amercian Association of Motor Vehicles Administrators.

CHIPMAN, M.L. (1982) The role of exposure, experience and demerit point levels in the risk of collision Accident Analysis & Prevention 14(6) 475-483.

CHIPMAN, M.L. & MORGAN, P.P. (1976) The driver demerit point system in Ontario as a long-term predictor of collisions. Proceedings of the 20th Conference of the American Association for Automotive Medicine. 45-51

COOLEY, W.W., & LOHNES, P.P. (1971) Multivariate Data Analysis. New York, Wiley.

COPPIN, R.S., LEW, A & PECK, R.C. (1965) The 1964 California Driver Record Study - Part 4: The relationship between concurrent Accidents and Citations. California, Dept. of Motor Vehicles.

DOBSON, A.J. (1983) An Introduction to Statistical Modelling London, Chapman and Hall.

KWONG, K.W., KUAN, J. & PECK R.C (1976) Longitudinal study of California driver accident frequencies I: An explanatory multivariate analysis. California, Dept. of Motor Vehicles.

LEW, A., COPPIN, R.S. & PECK, R.C. (1966). The 1964 California driver record study - Part 7: the relationship between types of conviction and accidents. California, Dept. of Motor Vehicles.

McCONNELL, E.J. & HAGEN, R.E. (1980) Design and evaluation of a crash prediction strategy. California, Dept. of Motor Vehicles.

McCULLAGH, P. & NELDER, J.A. (1983) Generalised Linear Models. Landon, Chapman and Hall.

OFFICE OF ROAD SAFETY (1984) The Cost of Road Crashes. Pamphlet, 8 p

PECK, R.C. & COPPIN, R.S. (1967) The prediction of accident involvement using concurrent driver record data. Traffic Safety Research Review V.11, 34-41.

PECK, R.C. & KUAN, J (1982) A statistical model of individual risk prediction using driver record, territory and other biographical factors. California, Dept. of Motor Vehicles.

PECK, R.C., McBRIDGE, R.S. & COPPIN, R.S. (1971) The distribution and prediction of driver accident frequencies. Accident Analysis & Prevention 2(4), 243-279.

STEWART, J.R. & CAMPBELL, B.J. (1972) The statistical association between past and future accidents and violations. Chapel Hill, N.C. University of North Carolina Highway Safety Research Center.

APPENDIX - GENERALISED LINEAR MODELS

The linear models which are commonly used to model data express the mean of the distribution of observations at a particular set of observing conditions (measured by values of the independent or predictor variables) as a linear combination of the values of the independent or predictor variables. The values of the independent or predictor variables. The values of the weights in the linear combination are determined by minimising the squares of the differences between the observed values and their values as predicted by the model (i.e. fitted by "least squares"). In order to make statistical inferences about the weights, the fitted values, or the various other statistics calculated during the fitting of a linear model it is necessary to assume that the observations came from a normal (Gaussian) distribution.

When linear models have been used to examine offence-crash relationships, the independent variables have been taken to be the number of offences in various categories for the driver, and occasionally some demographic information about the driver. The observation to be modelled is either the number of crashes for that driver, or a 0/1 variable reflecting the driver's crash status. As discussed in the Review of Related Studies, the experience in such modelling has been disappointing. A measure of how successful the fitting of a linear model had been is given by the coefficient of determination (called R^2), which represents the proportion of the total variability in the observations which is accounted for by fitting the model. (Equivalently, R^2 is the square of the correlation between fitted and observed values). The studies referred to in the Review of Related Studies reported R^2 values of less than 0.1, which indicate a poor fit to the data. As discussed earlier, however, the sample sizes are so large that these R^2 values can (using tests based on the assumption of a normal distribution of the observations) be shown to be significantly different from zero. That is, there is statistical evidence of <u>some</u> relationships, even if the R^2 is small. We would feel that the relationship has little predictive value.

The assumptions usually made in order to fit linear models by least squares are not satisfied by data such as is involved in offence-crash studies. The observations most certainly do not come from a normal distribution, and so a linear model (and associated measures such as R^2) are not appropriate. A more relevant form of modelling is provided by generalised linear models - see, eg McCullagh & Nelder (1983) or Dobson (1983). Under a generalised linear model, flexibility is allowed in the choice of distribution of the observations - in particular discrete distributions such as the Poisson and binomial are catered for. Also, the model applies to the appropriate parameter of the distribution, and model forms are chosen to satisfy natural constraints on the parameter. For example, in a generalised linear model for binomially distribution observations, the binomial parameter p is fitted, and the model chosen constrained to give fitted values between 0 and 1. The model usually chosen for the binomial parameter p is the logistic, given by

$$p = \frac{\text{exp (linear combination of independent variables)}}{\text{1 + exp (linear combination if independent variables)}}$$

The model is fitted by maximum likelihood, i.e. the weights in the linear combination are chosen to maximise the likelihood of the results observed. This requires an iterative computational procedure, again available in the statistical package GENSTAT.

The adequacy of the fit of a generalised linear model can be judged from the maximum value achieved for the likelihood. If the model is an adequate fit to the data, -2 log (maximum likelihood) has a X^2 distribution with degrees of freedom given by the sample size less the number of parameters fitted. If the fit is judged adequate, we can calculate standard errors for the estimated weights.

An adequate model for the available data can be used to estimate the chosen parameter value for individuals in the sample used for fitting and, with appropriate caution about extrapolation, to estimate that parameter value for future individuals. In the context of offence/ crash relationships, an adequate model could be used to estimate the probability of crash involvement based on the driver's offence history.