ROAD RESEARCH LABORATORY

Ministry of Transport

RRL REPORT NO. 6

ALCOHOL AND ROAD ACCIDENTS

A discussion of the Grand Rapids study

by

R. E. Allsop

HARMONDSWORTH

ROAD RESEARCH LABORATORY

1966

CONTENTS

Ownership of the Transport Research Laboratory was transferred from the Department of Transport to a subsidiary of the Transport Research Foundation on 1st April 1996.

ALCOHOL AND ROAD ACCIDENTS

A discussion of the Grand Rapids study

ABSTRACT

In an Indiana University study led by Professor R. F. Borkenstein, drivers involved in accidents in Grand Rapids, Michigan from July, 1962 to June, 1963 were compared with a control group of drivers. Observations were analysed with respect to nine variables, one of which was the blood alcohol level. Accident risk was found to vary significantly with each variable; in particular it was significantly higher for drivers with blood alcohol levels of 80 mg/100 ml and above than for those with blood alcohol levels lower than 10 mg/100 ml. Information was also obtained about many aspects of drivers' consumption of alcohol.

The Report comprises a summary of the Borkenstein report and discussion of some of the methods and results. Data concerning drivers with blood alcohol levels of 10-49 mg/100 ml suggest, generally at a low level of statistical significance, that accident risk rises steadily with alcohol consumption, but at the same time shows more certainly that more frequent drinkers have significantly lower accident risk than non- or infrequent drinkers. If the risk of accident involvement due to alcohol intake is ignored below 80 mg/100 ml, as a basis for estimating the minimum number of accidents due to alcohol, it is calculated on the basis of the Grand Rapids study that the number of drivers involved in accidents would have been reduced by six per cent. Further analysis of the data is suggested.

1. INTRODUCTION

In a study led by Professor R. F. Borkenstein, the Department of Police Administration of Indiana University examined the effect of a number of factors on drivers' involvement in road accidents in the city of Grand Rapids, Michigan, in 1962-63. The methods and results are described in an extensive preliminary report.[1]

The Report provides a summary of Borkenstein's report, followed by a discussion of some of the methods and results, based on analysis carried out at the Road Research Laboratory.

2. SUMMARY OF THE REPORT ON THE GRAND RAPIDS STUDY

2.1 Background

The report briefly reviews earlier investigations of the effects of alcohol, including its effects on driving skill and accident involvement, using both laboratory and survey techniques. These investigations revealed the need for an extensive survey to determine the significance of alcohol in relation to all kinds of road accidents.

A year's accidents were studied in the city of Grand Rapids, Michigan, which was chosen because it has a balanced and steady population (201 000), is distinct from any larger conurbation, and is large enough to have several thousand accidents annually. The city police and administration were ready to co-operate in the study and made detailed records of past accidents available.

2.2 Design of the survey

Two groups of drivers were selected and compared; the accident group consisted of drivers involved in accidents during the study year, and the control group was selected from the city's traffic.

Information about each driver was obtained by roadside interview and breath sampling. Interviews were conducted informally, but were based on carefully designed questionnaires which were subjected to thorough field trials in Bloomington, Indiana, and final tests in Grand Rapids. The control group questionnaire covered time, place, road and traffic conditions, social and personal information, purpose of journey, and drinking habits. For drivers involved in accidents, a shorter questionnaire was used because some of the information could be obtained from police accident records; the survey interview usually followed police investigations at the scene of the accident. Drivers were assured in writing that all information given would be treated as confidential, and were told that a refusal to co-operate was preferred to misleading answers. Rapid questioning made systematic lying difficult. Because some of the questions were personal, interviewers were trained to adopt a positive, sincere, adaptable attitude, avoiding any suggestion of moral judgment of drivers' answers.

The interviewer was free to decide when to ask for the breath sample, which the driver gave by breathing out as deeply as possible through a drinking straw into a plastic bag. A Mylar-polyethylene bag was used, with a capacity considerably greater than the volume of air required to operate the breathalyser. Breath samples were re-heated to mouth temperature before being passed through the instrument, which was designed to give blood alcohol levels correct to 1 mg/100 ml on the basis of the relation:

$$\frac{\text{Weight of alcohol per unit volume of blood}}{\text{Weight of alcohol per unit volume of breath}} = 2100$$

The interview and breath sample together provided sufficient information for the two groups of drivers to be compared with respect to nine major variables, namely:

Blood alcohol level
Age
Estimated annual mileage
Education level
Race or nationality
Marital status
Occupation status
Reported drinking frequency
Sex

The survey was preceded and accompanied by extensive publicity in Grand Rapids, informing and reassuring the public, and appealing for their co-operation in the interests of road safety.

2.3 Selection of the accident and control groups

Each road accident occurring in Grand Rapids must be reported to the police at once; usually the police go straight to the scene of the accident to make a routine investigation, but if the accident is only slight, the drivers may instead be allowed to make a full report later at a police station. Such accidents are known as late-reported. The accident group consisted of all drivers involved in other than late-reported accidents in Grand Rapids between 1st July, 1962 and 30th June, 1963. Pedestrians involved in accidents were not included in the study.

In selecting the control group the aim was to obtain a group similar in size to the accident group, and consisting of drivers whose exposure to accident risk was similar to that of drivers in the accident group. To achieve this the road system of the city was divided into small sections called blocks (a block usually comprised an intersection, or a length of road joining two intersections). A punched card was prepared for each of the 27000 accidents occurring in Grand Rapids between 1st May, 1959 and 30th April, 1962, recording the time, date and day of week that the accident occurred, and the block in which it occurred. The control group was selected by stopping drivers on the road at times and places determined by drawing accident cards from the pack of 27000. Cards corresponding to accidents involving buses and emergency vehicles, which would be difficult to stop were removed first.

Two thousand cards were drawn at random, together with a reserve of 200, the use of which is described in 2.4. Each card determined a sample site (the block in which the accident had occurred) and a time of sampling (the hour in which the accident had occurred, on the same day of the week, and as nearly as possible the same date in the survey year). Within this block and hour, four vehicles would be stopped at random and their drivers would form the control group, giving a total of 8000 drivers. The distribution of control driver selection in time and space would therefore be similar to that of accident risk, measured by the occurrence of accidents in the three-year period May, 1959 to April, 1962.

2.4 Collection of data

The survey team could visit only two control sites in any one hour, but there were 68 hours in which the schedule determined by the 2000 accident cards required more than two sites to be visited. The third and subsequent cards for such hours were replaced from the reserve. About ten other sites were replaced from the reserve because traffic conditions made it dangerous to attempt to stop vehicles.

Occasionally sampling at a control site was prevented because no police officer was available to stop the traffic, or because weather was too severe; in such cases, sampling was postponed by one week exactly, to the hour.

If one of the drivers stopped was willing to co-operate, but had urgent business, the interviewer took the breath sample and arranged to complete the interview later. A few such interviews were not completed, and other control drivers (less than one per cent of the total) were lost because less than four vehicles appeared at the site during the hour of sampling. Co-operation was given by more than 97 per cent of the drivers in the control group.

During the survey year, 9353 drivers were involved in accidents (other than late-reported ones). The first intention was that most interviews of accident-involved drivers would be carried out by police officers, but more pressing duties often prevented officers from completing the interviews; moreover, drivers were understandably more ready to co-operate when interviewed by research workers. It was therefore decided that the research team should cover as many accidents as possible, but even then 2764 accident-involved drivers were not interviewed, and are described as <u>missed</u>. The age and sex distributions of the missed drivers were obtained from police accident reports, and these distributions did not differ significantly (at the five per cent level) from those of drivers who were interviewed. In analyses by variables other than age and sex, missed drivers have to be omitted from the accident group. Co-operation was given by more than 95 per cent of those who were interviewed in the accident group.

The design of the control group was examined by comparing the distributions of the scheduled control sites by hour of day, by day of week, and by month with the corresponding distributions of accidents occurring in the survey year. The effectiveness of the design in practice was examined by comparing the corresponding distributions of completed observations in the accident and control groups. The differences between the distributions were considered small enough for the data to be accepted without attempting to correct for them.

2.5 Single-factor analysis

For each of the nine major variables separately, the distributions of the accident and control groups were compared using the χ^2 test. In every case the difference between the two distributions was significant at the five per cent level, showing that each of the nine major variables is a statistically significant discriminator of accident risk.

Over-representation of a particular class of drivers in the accident group, compared with the control group, indicates that drivers in this class are subject to a higher than average accident risk. As a conservative measure of such under- or over-representation the <u>accident involvement index</u> was defined as follows:

Let a_i = number of accident drivers in class i

c_i = number of control drivers in class i

$$A = \sum_i a_i \qquad C = \sum_i c_i$$

If the two groups were similarly distributed with respect to the variable of classification, the expected number e_i of accident drivers in class i would be

$$e_i = \frac{A}{A + C}(a_i + c_i)$$

Then the accident involvement index for class i was defined as

$$I_i = \frac{100\,(a_i - e_i)}{e_i}$$

i.e. the number of accident drivers in excess (or otherwise) of the expected number, expressed as a percentage of the expected number. The index and its interpretation are discussed in 3. 2.

The nine pairs of distributions with values of the accident involvement index are given in Tables 19 and 21-28 of Borkenstein's report, which provide the basis for Tables 1-9 of this Report. Calculation of the involvement index requires grouping of the higher alcohol levels, but complete distributions of the accident and control groups with respect to blood alcohol level in intervals of 10 mg/100 ml are given in Table 17 of the Borkenstein report and Table 10 of this Report. The lowest alcohol level class includes all drivers whose blood alcohol level, indicated by the breathalyser, was less than 10 mg/100 ml. Drivers with alcohol levels 0-39 mg/100 ml are under-represented in the accident group, but over-representation begins in the 40-49 mg/100 ml class, and increases rapidly as the high alcohol levels are reached. By comparing each alcohol level class in turn with the 0-9 mg/100 ml class, the over-representation of the accident group was found to be significant in the 80-89 mg/100 ml and higher classes.

With respect to the other variables, the classes over-represented in the accident group were generally those that had been expected; the young and the aged, those whose annual mileage is low, and those with little education or low occupation status. However, the distributions by reported drinking frequency (Table 8) show that frequent and infrequent drinkers are respectively under and over-represented in the accident group. This result is discussed in 3. 5.

Although the analysis in Borkenstein's report is concerned mainly with the effect of alcohol, the data provide equally valuable information about the effect of other variables on accident involvement.

2. 6 Two-factor analysis

The difference between the accident risks for drivers found in different alcohol level classes may be partly attributable to other variables, because the groups of drivers in different alcohol level intervals will not be matched with respect to other variables

TABLE 1

Distribution of accident and control groups by blood alcohol level

Blood alcohol level (mg/100 ml)	Number of drivers			Accident involvement index	Relative accident rate	95 per cent confidence limits for relative accident rate	
	Accident group	Control group	Total				
0-9	4992	6756	11 748	–3.62	1.00	—	—
10-19	188	276	464	–8.10	0.92	.76	1.11
20-29	95	134	229	–5.91	0.96	.74	1.25
30-39	57	96	153	–15.50	0.80	.58	1.12
40-49	66	83	149	0.46	1.08	.78	1.49
50-59	50	56	106	6.98	1.21	.83	1.77
60-69	46	44	90	15.92	1.41	.94	2.14
70-79	36	32	68	20.07	1.52	.95	2.44
80-89	39	28	67	32.01	1.88	1.16	3.05
90-99	39	27	66	34.02	1.95	1.20	3.18
100-119	92	21	113	84.65	5.93	3.70	9.50
120-139	73	20	93	78.03	4.94	3.02	8.07
140-159	54	7	61	100.78	10.44	4.84	22.51
160 and over	158	10	168	113.30	21.38	11.39	40.14
Total	5985	7590	13 575				

TABLE 2

Distribution of accident and control groups by age

Age	Number of drivers			Accident involvement index
	Accident group	Control group	Total	
15	18	17	35	— 7.97
16	221	82	303	30.55
17	343	105	448	37.04
18-19	900	485	1385	16.32
20-24	1612	950	2562	12.63
25-34	1910	1608	3518	— 2.82
35-44	1708	1682	3390	— 9.81
45-54	1192	1317	2509	—14.96
55-64	855	767	1622	— 5.64
65-69	246	216	462	— 4.69
70-74	179	76	255	25.65
75 or older	125	49	174	28.59
Total	9309	7354	16 663	

TABLE 3

Distribution of accident and control groups by estimated annual mileage

Estimated annual mileage	Number of drivers			Accident involvement index
	Accident group	Control group	Total	
Up to 1000	189	303	492	45.09
1001-5000	609	1178	1787	28.72
5001-15 000	1145	3589	4734	— 8.65
More than 15 000	685	2228	2913	—11.18
Total	2628	7298	9926	

TABLE 4

Distribution of accident and control groups by education level

Years of education	Number of drivers: Accident group	Control group	Total	Accident involvement index
Less than 8	363	303	666	18.32
8	670	756	1426	2.00
9-11	1688	1654	3342	9.65
12	2011	2440	4451	— 1.92
13-15	1002	1256	2258	— 3.66
16	334	563	897	—19.17
More than 16	169	331	500	—26.62
Total	6237	7303	13540	

TABLE 5

Distribution of accident and control groups by race or nationality

Race or nationality class	Number of drivers: Accident group	Control group	Total	Accident involvement index
White	6220	7272	13492	— 1.50
Non-white	633	517	1150	17.60
Total	6853	7789	14642	

TABLE 6

Distribution of accident and control groups by marital status

Marital status	Number of drivers			Accident involvement index
	Accident group	Control group	Total	
Single	1868	1592	3460	17.05
Married	3898	5198	9096	— 7.09
Widowed	204	218	422	4.81
Separated	108	115	223	5.00
Divorced	234	250	484	4.82
Total	6312	7373	13685	

TABLE 7

Distribution of accident and control groups by occupation status

Occupation status	Number of drivers			Accident involvement index
	Accident group	Control group	Total	
Lower	2625	2380	5005	14.43
Skilled	1354	1719	3073	— 3.87
Lower white collar	1366	1646	3012	— 1.05
Upper white collar	844	1569	2413	—23.69
Total	6189	7314	13503	

TABLE 8

Distribution of accident and control groups by reported drinking frequency

Reported drinking frequency	Number of drivers			Accident involvement index
	Accident group	Control group	Total	
Less than yearly	1815	1874	3689	6.43
Yearly	983	1078	2061	3.17
Monthly	1107	1101	2208	8.45
Weekly	1417	1665	3082	– 0.55
Three times weekly	565	815	1380	–11.44
Daily	379	755	1134	–27.71
Total	6266	7288	13554	

TABLE 9

Distribution of accident and control groups by sex

Sex	Number of drivers			Accident involvement index
	Accident group	Control group	Total	
Male	7235	6213	13448	– 0.94
Female	2087	1630	3717	3.39
Total	9322	7843	17165	

TABLE 10

Complete distribution of accident and control groups by blood alcohol level

Blood alcohol level (mg/100 ml)	Number of drivers		
	Accident group	Control group	Total
0-9	4992	6756	11748
10-19	188	276	464
20-29	95	134	229
30-39	57	96	153
40-49	66	83	149
50-59	50	56	106
60-69	46	44	90
70-79	36	32	68
80-89	39	28	67
90-99	39	27	66
100-109	54	14	68
110-119	38	7	45
120-129	43	16	59
130-139	30	4	34
140-149	21	3	24
150-159	33	4	37
160-169	27	2	29
170-179	30	1	31
180-189	21	2	23
190-199	14	1	15
200-209	14	2	16
210-219	12	0	12
220-229	6	0	6
230-239	9	1	10
240-249	6	0	6
250-259	3	1	4
260-269	3	0	3
270-279	3	0	3
280-289	2	0	2
290-299	0	0	0
300-309	1	0	1
310-319	1	0	1
320-329	0	0	0
330-339	0	0	0
340-349	1	0	1
350-359	1	0	1
360-369	2	0	2
370-379	2	0	2
Total	5985	7590	13575

affecting accident risk. The ideal would be to classify by all nine variables simultaneously, and consider the variation of accident risk with alcohol level, keeping the other eight variables fixed. Even in such a large survey as this there are far too few observations to permit such detailed analysis, but Borkenstein did find it possible to study the variation with alcohol level when any one of the other variables was kept fixed. For this purpose, alcohol levels were combined into five classes, as in Table 11.

TABLE 11

Alcohol level classes used in two-factor analyses

Class number (j)	Blood alcohol levels (mg/100 ml)
0	0-9
1	10-49
2	50-79
3	80-109
4	110 and over

Suppose that the variable to be kept fixed is v. To avoid small class frequencies v classes were combined where possible. As the aim was to study the distributions of the accident and control groups with respect to alcohol level within each v class, Borkenstein used the following test to decide whether it was appropriate to combine two v classes. In the first v class,

let a'_j = number of accident drivers
let c'_j = number of control drivers
} in alcohol level class j $(j = 0, 1, \ldots 4)$

and let a''_j, c''_j be the corresponding numbers of drivers in the second v class. The distributions

$$(a'_0, a'_1, a'_2, a'_3, a'_4) \text{ and } (a''_0, a''_1, a''_2, a''_3, a''_4)$$

were compared using the χ^2 test, and distributions

$(c'_0, c'_1, c'_2, c'_3, c'_4)$ and $(c''_0, c''_1, c''_2, c''_3, c''_4)$ were also compared in this way.

The two v classes were combined only if each χ^2 value was not significant at the five per cent level.

Let p_v = number of v classes remaining after appropriate combinations had been made; then the basis of the two-factor analysis by alcohol level and variable v is a $p_v \times 5$ table of the numbers of accident and control drivers.

Let a_{ij} = number of accident drivers $\}$ in v class i and alcohol level class j
c_{ij} = number of control drivers $\}$ $(i = 1, 2, \ldots, p_v; j = 0, 1, \ldots, 4)$

A list of variables v, corresponding values of p_v and relevant table numbers in Borkenstein's report and this Report are given in Table 12.

TABLE 12

Key to two-factor analyses

Second Variable v	Number of v classes (p_v)	Table number in Borken-stein's report	Table number in this Report
Age	4	30-33	13
Estimated annual mileage	4	36	14
Education level	6	37	15
Race or nationality	2	38	16
Marital status	5	39	17
Occupation status	4	40	18
Reported drinking frequency	5	41	19
Sex	2	42	20
	32		

Within each v class, an accident involvement index was calculated just as for the the single-factor analysis, and to test the significance of the variation in accident risk with alcohol level within v classes the distributions

$(a_{k0}, a_{k1}, a_{k2}, a_{k3}, a_{k4})$ and $(c_{k0}, c_{k1}, c_{k2}, c_{k3}, c_{k4})$

for each k in turn were compared using the χ^2 test. In every case the value of χ^2 was significant at the five per cent level, indicating that within every v class, alcohol level is a significant discriminator of accident risk.

To determine, within the kth v class, the alcohol level classes in which the representation of the accident group is significantly different from that in the 0-9 mg/100 ml class, the χ^2 test was applied to the 2 × 2 matrices

$$\begin{pmatrix} a_{k0} & a_{kj} \\ c_{k0} & c_{kj} \end{pmatrix} \quad (j = 1, 2, 3, 4)$$

The cases in which the value of χ^2 was significant at the five per cent level are indicated in Tables 13-20 of this Report by marking the corresponding a_{kj} with an asterisk. Not all of these significant differences are mentioned in the discussion on pp. 141-162 of Borkenstein's report. Over-representation of the accident group is always significant at alcohol levels 110 mg/100 ml and above, and often in the 80-109 mg/100 ml and 50-79 mg/100 ml classes. However, under-representation of the accident group in the 10-49 mg/100 ml class is significant in only two of the 17 cases in which it occurs.

Finally, to test whether other variables are still significant discriminators of accident risk at high alcohol levels, let

$$A_{kh} = \sum_{j=h}^{j=4} a_{kj} \quad C_{kh} = \sum_{j=h}^{j=4} c_{kj} \quad \begin{cases} h = 0, 1, 2 \ldots 4 \\ k = 1, 2, \ldots p_v \end{cases}$$

A_{kh} and C_{kh} were calculated and for each h in turn the distributions

$$(A_{1h}, A_{2h}, \ldots, A_{p_v h}) \text{ and } (C_{1h}, C_{2h}, \ldots C_{p_v h})$$

were tested for proportionality using the χ^2 test. A value of χ^2 not significant at the five per cent level indicated that in the alcohol level class h and above, any variation of accident risk between v classes could have arisen by chance. At least one non-significant value of χ^2 was obtained with each variable except occupation status. The alcohol level classes for which non-significant values were obtained are marked with double asterisks in Tables 13-20 of this Report. This analysis reveals that at high alcohol levels, accident experience does not vary significantly with any of the other major variables except occupation status and sex, but care is needed in interpreting this lack of significance (see 3. 6). The variation with sex is the only one that is significant at high alcohol levels but not at low alcohol levels; women had a significantly higher accident rate than men at alcohol levels 80 mg/100 ml and above. Contrary to page 161 of Borkenstein's report, the variation of accident risk with drinking frequency is not significant at alcohol levels above 110 mg/100 ml.

2. 7 Relation between alcohol level and the probability of causing an accident

No attempt was made by judging evidence about any particular accident to identify the driver causing it. However, an estimated distribution of accident-causing drivers by alcohol level was obtained using the following somewhat arbitrary process.

It was assumed that all single-vehicle accidents were caused by the driver involved. The distribution of these drivers (622 in number) by alcohol level is known.

It was also assumed that of the 5366 drivers involved in multiple-vehicle accidents, exactly half were accident-causing drivers and the other half were innocently involved, and were typical of the drivers of vehicles passing the scenes of the accidents at the times at which the accidents occurred.

Then the process of selection of the innocent drivers would have been similar to that of the control drivers, so the distribution of the 2683 innocent drivers by alcohol level was assumed to be proportional to that of the control group. Since the distribution of all 5366 drivers involved in multiple-vehicle accidents is known, that of the 2683 drivers assumed to have caused such accidents may be obtained by subtraction.

Addition of the known distribution of 622 drivers assumed to have caused single-vehicle accidents gives the distribution of all 3305 drivers assumed to have caused accidents. This distribution is given in Table 43 of Borkenstein's report and since it has already been shown that overall accident rate rises with alcohol level we are bound to find that the risk of causing an accident rises even faster as high alcohol levels are reached than does the risk of being involved in one. However, in view of the assumptions on which the calculation is based, the actual values in the distribution must be treated with considerable caution.

2.8 Alcohol and accident characteristics

The alcohol level distribution of drivers in single-vehicle accidents was compared with that of drivers in multiple-vehicle accidents, showing that at high alcohol levels a significantly higher proportion of drivers was involved in single-vehicle accidents. Accidents were classified according to estimated value of property damage, and severity of injury caused; both were found to be significantly greater in accidents involving drivers with high alcohol levels (Tables 44-46 of Borkenstein's report).

2.9 Characteristics of drinking drivers

To find out what type of driver drives after drinking, the joint distributions of the control group with respect to alcohol level and each of the other eight major variables were examined. High alcohol levels were over-represented in the middle age groups, the 5001-15 000 mile/year class, the middle education levels, the non-white class, the separated and divorced classes, the lowest occupation status classes and the higher drinking frequency classes. Almost all drivers with high alcohol levels were male (Tables 47-54 of Borkenstein's report).

To show which drinking habits are associated with driving after drinking, the control group was analysed jointly with respect to alcohol level and certain aspects of the use of alcohol. High alcohol levels were over-represented among those who drink at home and in public houses, drink in the evening and at the weekend, drink alone or with business contacts, prefer to drink beer, sometimes but not always have drink available in their homes, take many drinks at a time, think they can take many drinks and still drive safely, often drive after drinking, have had an alcohol problem, most frequently 'get high', or suffer from blackouts or hangovers (Tables 56, 58, 60, and 62-73 of Borkenstein's report).

To show which drinking habits are associated with higher than average accident risk, the accident and control groups were compared with respect to time and place of drinking, drinking companions, and type of drink preferred. Each comparison revealed statistically significant differences between the two groups; high accident risk was associated with drinking in the morning, at parties and with casual acquaintances. Low accident risk was associated with drinking in public houses, before dinner and with business contacts (Tables 55, 57, 59 and 61 of Borkenstein's report).

TABLE 13

Distribution by alcohol level within age groups

Age		Blood alcohol level (mg/100 ml) 0-9	10-49	50-79**	80-109	110 and over**	All alcohol levels
16-17	a †	352	18				370
	c	181	3				184
	I	–1.12	128.33	‡			—
	r	1	3.09				—
	s	2.34	7.23				2.42
18-24	a	1426	108	32	26*	47*	1639
	c	1276	93	25	5	3	1402
	I	–2.08	–0.31	4.16	55.61	74.40	—
	r	1	1.04	1.15	4.65	14.02	—
	s	1.35	1.40	1.54	6.26	18.87	1.41
25-34 and 55-64	a	1400	126	42	58*	130*	1756
	c	2034	181	43	27	14	2299
	I	–5.86	–5.22	14.09	57.56	108.47	—
	r	1	1.01	1.42	3.12	13.49	—
	s	0.83	0.84	1.18	2.59	11.18	0.92
35-54	a	1483	138	49*	44*	134*	1848
	c	2519	247	53	31	25	2875
	I	–5.29	–8.39	22.77	49.93	115.38	—
	r	1	0.95	1.57	2.41	9.10	—
	s	0.71	0.67	1.11	1.71	6.46	0.77
Totals	Σa	4661	390	123	128	311	5613
	Σc	6010	524	121	63	42	6760

* } The meaning of the asterisks is explained in 2.6
** }

† a = number of drivers in the accident group
c = number of drivers in the control group
I = accident involvement index
r = relative accident rate
s = accident rate

‡ The presence of an empty cell in any row of Tables 13-20 indicates that the last completed cell in the same row may include a few drivers at higher alcohol levels

TABLE 14

Distribution by alcohol level within estimated annual mileage classes

Estimated annual mileage		Blood alcohol level (mg/100 ml) 0-9	10-49	50-79	80-109	110 and over**	All alcohol levels
Up to 1000	a†	160	6	16*			182
	c	276	12	4			292
	I	−4.43	−13.19	108.35	†		—
	r	1	0.86	6.90			—
	s	1.62	1.40	11.19			1.74
1001-5000	a	498	36	13	34*		581
	c	1035	70	21	9		1135
	I	−4.05	0.30	12.91	133.53		—
	r	1	1.07	1.29	7.85		—
	s	1.35	1.44	1.73	10.56		1.43
5001-15000	a	905	84	21	31*	66*	1107
	c	3091	274	57	28	29	3479
	I	−6.18	−2.80	11.53	117.66	187.81	—
	r	1	1.05	1.26	3.78	7.77	—
	s	0.82	0.86	1.03	3.10	6.36	0.89
more than 15000	a	550	46	24*	9	29*	658
	c	1878	201	47	28	9	2163
	I	−2.88	−20.16	44.92	4.27	227.16	—
	r	1	0.78	1.74	1.10	11.00	—
	s	0.82	0.64	1.43	0.90	9.01	0.85
Totals	Σa	2113	188	58	74	95	2528
	Σc	6280	561	125	65	38	7069

† See footnote to Table 13

TABLE 15

Distribution by alcohol level within education level classes

Education level		Blood alcohol level (mg/100 ml) 0-9	10-49	50-79	80-109**	110 and over	All alcohol levels
Less than 8 years	a†	238	26	65*			329
	c	257	23	10	†		290
	I	–9.54	–0.17	63.06			—
	r	1	1.22	7.02			—
	s	1.14	1.39	8.00			1.40
8 years	a	508	43	18	52*		621
	c	625	61	15	16		717
	I	–3.40	–10.92	17.52	64.76		—
	r	1	0.87	1.48	4.00		—
	s	1.00	0.87	1.48	4.00		1.07
9-11 years	a	1283	122	37	97*		1539
	c	1418	129	35	38		1620
	I	–2.50	–0.23	5.48	47.49		—
	r	1	1.05	1.17	2.82		—
	s	1.11	1.16	1.30	3.14		1.17
12 years	a	1573	129	38	110*		1850
	c	2105	178	40	25		2348
	I	–2.95	–4.65	10.55	84.90		—
	r	1	0.97	1.27	5.89		—
	s	0.92	0.89	1.17	5.42		0.97
13-15 years	a	825	52*	12	45*		934
	c	1095	99	20	13		1227
	I	–0.58	–20.32	–13.24	79.51		—
	r	1	0.70	0.80	4.60		—
	s	0.93	0.65	0.74	4.26		0.94
16 years or more	a	415	27	10	15*		467
	c	789	56	12	6		863
	I	–1.84	–7.36	29.45	103.43		—
	r	1	0.92	1.58	4.75		—
	s	0.65	0.59	1.03	3.08		0.67
Totals	Σa	4842	399	180	319		5740
	Σc	6289	546	132	98		7065

† See footnote to Table 13

TABLE 16

Distribution by alcohol level within race or nationality classes

Race class		Blood alcohol level (mg/100 ml)					All alcohol levels
		0-9	10-49	50-79**	80-109**	110 and over**	
White	a†	4618	350	104	100*	236*	5408
	c	5919	508	108	48	33	6616
	I	—2.56	—9.30	9.07	50.23	95.06	—
	r	1	0.88	1.23	2.67	9.17	—
	s	0.93	0.82	1.14	2.48	8.50	0.97
Non-white	a	344	56	26	31*	84*	541
	c	370	40	18	16	10	454
	I	—11.39	7.29	8.68	21.31	64.35	—
	r	1	1.51	1.55	2.08	9.03	—
	s	1.10	1.66	1.72	2.30	9.98	1.42
Total	Σa	4962	406	130	131	320	5949
	Σc	6289	548	126	64	43	7070

† See footnote to Table 13

TABLE 17

Distribution by alcohol level within marital status classes

Marital status		Blood alcohol level (mg/100 ml) 0-9	10-49	50-79**	80-109**	110 and over	All alcohol levels
Single	a†	1511	111	34	70*		1726
	c	1394	105	26	13		1538
	I	–1.64	–2.82	7.16	59.49	†	—
	r	1	0.98	1.21	4.97		—
	s	1.33	1.30	1.61	6.62		1.38
Married	a	2973	251	80*	236*		3540
	c	4448	380	87	68		4983
	I	–3.55	–4.23	15.34	86.91		—
	r	1	0.99	1.38	5.19		—
	s	0.82	0.81	1.13	4.26		0.87
Widowed	a	170	5*	14*			189
	c	185	20	5			210
	I	1.10	–57.78	55.56			—
	r	1	0.27	3.05			—
	s	1.13	0.31	3.44			1.11
Separated	a	57	7	33*			97
	c	81	17	11			109
	I	–12.28	–38.06	59.28			—
	r	1	0.59	4.26			—
	s	0.86	0.51	3.69			1.09
Divorced	a	142	24	44*			210
	c	192	26	21			239
	I	–9.10	2.63	44.73			—
	r	1	1.25	2.83			—
	s	0.91	1.13	2.57			1.08
Totals	Σa	4853	398	205	306		5762
	Σc	6300	548	150	81		7079

† See footnote to Table 13

TABLE 18

Distribution by alcohol level within occupation status classes

Occupation status		Blood alcohol level (mg/100 ml) 0-9	10-49	50-79	80-109	110 and over	All alcohol levels
Lower	a†	1896	189	72*	67*	157*	2381
	c	2006	187	51	34	20	2298
	I	−4.51	−1.22	15.03	30.36	74.31	—
	r	1	1.07	1.49	2.08	8.31	—
	s	1.17	1.26	1.75	2.45	9.76	1.29
Skilled	a	1064	87	27	29*	41*	1248
	c	1464	142	32	13	14	1665
	I	−1.76	−11.32	6.82	61.17	74.00	—
	r	1	0.84	1.16	3.07	4.03	—
	s	0.90	0.76	1.05	2.77	3.64	0.93
Lower white-collar	a	1117	72	16	21*	42*	1268
	c	1439	111	22	9	4	1585
	I	−1.67	−11.48	−5.26	57.50	105.40	—
	r	1	0.84	0.94	3.01	13.53	—
	s	0.96	0.81	0.90	2.90	13.05	0.99
Upper white-collar	a	704	45	13	5	16*	783
	c	1374	106	20	6	5	1511
	I	−0.74	−12.69	15.41	33.17	123.22	—
	r	1	0.83	1.27	1.63	6.25	—
	s	0.64	0.53	0.81	1.04	3.98	0.64
Totals	Σa	4781	393	128	122	256	5680
	Σc	6283	546	125	62	43	7059

† See footnote to Table 13

TABLE 19

Distribution by alcohol level within reported drinking frequency classes

Reported drinking frequency		Blood alcohol level (mg/100 ml)					All alcohol level
		0-9	10-49	50-79	80-109	110 and over**	
Yearly or less	a†	2485	94	34*			2613
	c	2778	88	5			2871
	I	−0.91	8.40	82.97	†		—
	r	1	1.19	7.60			—
	s	1.09	1.30	8.30			1.11
Monthly	a	934	55	49*			1038
	c	1010	56	10			1076
	I	−2.15	0.91	69.14			—
	r	1	1.06	5.30			—
	s	1.13	1.20	5.98			1.18
Weekly	a	959	122	50*	51*	97*	1279
	c	1374	160	39	17	12	1602
	I	−7.41	−2.55	26.55	68.94	100.46	—
	r	1	1.09	1.84	4.30	11.58	—
	s	0.85	0.93	1.57	3.66	9.87	0.97
Three times weekly	a	340	64	29	25*	52*	510
	c	616	103	32	17	13	781
	I	−9.97	−2.99	20.34	50.68	102.51	—
	r	1	1.13	1.64	2.66	7.25	—
	s	0.67	0.76	1.11	1.80	4.88	0.80
Daily	a	160	63*	22	27*	67*	339
	c	505	136	43	27	16	727
	I	−24.34	−0.45	6.43	57.23	153.84	—
	r	1	1.46	1.61	3.16	13.22	—
	s	0.39	0.57	0.62	1.22	5.11	0.57
Totals	Σa	4878	398	184	103	216	5779
	Σc	6283	543	129	61	41	7057

† See footnote to Table 13

TABLE 20

Distribution by alcohol level within sexes

Sex		Blood alcohol level (mg/100 ml) 0-9**	10-49**	50-79**	80-109	110 and over	All alcohol levels
Male	a†	3785	349	121*	120*	282*	4657
	c	4889	478	114	59	43	5583
	I	—4.05	—7.21	13.22	47.41	90.79	—
	r	1	0.94	1.37	2.63	8.47	—
	s	0.92	0.87	1.26	2.41	7.78	0.99
Female	a	1204	58	11	53*		1326
	c	1422	76	11	5		1514
	I	—1.80	—7.30	7.09	95.71	†	—
	r	1	0.90	1.18	12.52		—
	s	1.00	0.91	1.19	12.57		1.04
Totals	Σa	4989	407	132	173	282	5983
	Σc	6311	554	125	64	43	7097

† See footnote to Table 13

2.10 Conclusions reached by the Grand Rapids investigators

On the basis of the single-factor analysis the report concludes (p. 136) 'The nine variables of classification are all statistically significant discriminators of accident experience.' Following the two-factor analysis it concludes (p. 162). 'In every case, the higher alcohol levels are associated with more frequent accident experience...... This association is so strong that any other explanation of the frequency of the accident experience of drivers in the higher alcohol ranges is substantially ruled out.'

Concerning the accident experience of drivers at blood alcohol levels 10-49 mg/100 ml the report concludes (p. 163) 'Either in point of fact drivers perform better at this alcohol level than they do on the average in the complete absence of alcohol, or drivers who drink only enough to attain the 0.01-0.049 per cent [i.e. 10-49 mg] alcohol level are for some unconsidered reason better than average drivers.' This conclusion is discussed in 3.4, and it is not found to be entirely supported by the data.

On the estimate of the effect of alcohol on the risk of causing an accident the report concludes (p. 167) 'The results indicate that the moderate use of alcohol is not inconsistent with traffic safety..... At higher alcohol levels the picture quickly changes. Even before alcohol levels are reached which are legally recognized as a public hazard in driving [in the U.S.A.] the probability of causing an accident has multiplied several fold.'

The studies of the type of accident in which drinking drivers are involved, and the drivers' drinking habits, lead to the following conclusions (pp. 177 and 197). 'A successful programme to reduce the amount of driving by drivers with relatively high blood alcohol levels would not only reduce the number of accidents but also the average cost and severity of accidents', and 'Many people do not know or do not admit the effect of alcohol on their driving behaviour'.

3. DISCUSSION

3.1 Design and method of survey

The aim in selecting the control sample was to obtain a group of drivers matching the accident group with respect to exposure to accident risk, to the extent that this varies in time and from one stretch of road to another. Accident records for a recent three year period were used to determine the control sample sites. In the Toronto study of 1951-52[2] in which the aim was the same, the control group was obtained by going to the scene of each accident from which a driver was included in the accident group and stopping drivers passing as soon as possible after the accident. It is not clear how far the two groups of drivers in the Bratislava study of 1956-57[3] were matched.

There are two important differences between the Toronto and Grand Rapids studies; in Toronto only a selection of the accidents occurring between 6.30 p.m. and 10.30 p.m. were studied, but in Grand Rapids all accidents occurring during the survey year were included. Accident and control drivers in Toronto were matched by type and age of vehicle, but control drivers in Grand Rapids were selected at random subject to the time and place schedule, so that if the accident and control groups in the Grand Rapids study differ significantly with respect to any variable except time and place of selection, there is evidence of an association between that variable and accident experience. It should be remembered, however, that two distinct groups of drivers are being compared; no attempt was made to examine the behaviour of the same drivers with and without alcohol.

There appear to be several possible sources of small errors in control sample selection. Each of the 2000 accident record cards selects four drivers for the control group, whatever the number of drivers involved in the corresponding accident. Drivers who are relatively highly exposed to the risk of single-vehicle accidents will therefore be over-represented in the control group; this would cause error if the time and space distribution of single-vehicle accidents differed from that of multiple-vehicle accidents. Examination of accident records would show whether these distributions differ widely, and any error could be avoided by making the number of drivers selected at a control site proportional to the number of drivers involved in the corresponding accident. In a study of fatal accidents in three English police districts Jeffcoate[4] found that the proportion of single-vehicle accidents was higher in the night hours from 10 p.m. to 4 a.m. than during the rest of the day. If we assume that high alcohol levels are more common in the night hours, it follows that in England such alcohol levels will be more common among drivers who are relatively highly exposed to the risk of single-vehicle accidents than among drivers generally. Borkenstein's comparison of alcohol level distributions of drivers involved in single-driver and two-driver accidents (see Table 21) suggests that the

TABLE 21

Relative accident rates for single-driver and two-driver accidents

	Blood alcohol level (mg/100 ml)					Total
	0-9	10-49	50-79	80-109	110 and over	
Control drivers	6756	589	132	69	44	7590
Drivers involved in two-driver and pedestrian accidents*	4699	378	111	105	211	5504
Relative rate for two-driver and pedestrian accidents†	1	0. 92	1. 21	2. 19	6. 90	—
Drivers involved in single-driver accidents	260	29	20	24	110	443
Relative rate v_i for single-driver accidents (see 3. 8)	1	1. 28	3. 94	9. 04	64. 96	—

* The distribution of these drivers is similar to that of drivers involved in two-driver accidents only

† an approximation to the relative rate u_i (see 3. 8)

same may be true in Grand Rapids. If so, the above error should take the form of a slight over-representation of high alcohol levels in the control group, and consequent slight underestimation of the increase in accident risk as the alcohol level rises.

Record cards corresponding to late-reported accidents were used to select control sites, but drivers involved in late-reported accidents were not included in the accident group. Since about one-third of the accidents in Grand Rapids are late-reported, error could arise if the time and space distribution of late-reported accidents differs from that of accidents investigated on the spot, as seems quite likely because late-reported accidents are usually slight. This difficulty could have been avoided by removing record cards corresponding to late-reported accidents from the pack of 27 000 before drawing the 2000 cards used to select the control sites, thus using the same type of accident to select both the accident group and the control sites.

Difficulties in keeping to the control site schedule combine to reduce the number of control drivers selected at times when accidents are frequent. Such difficulties were dealt with in two ways; replacement sites were chosen from the reserve, or sampling was postponed by exactly one week. The latter procedure seems preferable since it preserves randomness with respect to time of day, day of week, and place; this randomness is lost when sites are replaced from the reserve, and nothing seems to be gained in return.

The report mentions that a new motorway had been opened shortly before the survey began; control sample selection was based mainly on the accident pattern before the motorway was opened, whereas the accident group was selected by accidents occurring when the motorway was open (excluding accidents occurring on the motorway). The matching of the two groups will have been affected to the extent that the accident pattern on the city's streets has been altered by the opening of the motorway.

In general, however, the design of the survey appears to us to be satisfactory, and correction of the above errors should not substantially affect the conclusions.

3.2 Measures of differences in accident risk

The sign of the accident involvement index described in 2.5 is interpreted as follows:

$I_i > 0$ more } accident drivers in class i than if the distribution of the accident
$I_i < 0$ less } group were proportional to that of the control group.

However, the magnitude of I_i is difficult to interpret; there seems to be a need for a numerical measure of the accident risk of one class of drivers relative to that of another class. With the notation given in 2.5, suppose that the variable of classification is blood alcohol level, and that $i = 0$ refers to the 0-9 mg/100 ml alcohol level interval.

Let the relative accident rate $r_i = \dfrac{a_i c_0}{c_i a_0}$

i.e. the ratio of the number of accident drivers to the number of control drivers in class i divided by the corresponding ratio for the 0-9 mg/100 ml class.

r_i provides a measure of the accident risk in the alcohol level i compared with that in the 0-9 mg/100 ml interval. Values of r_i are given in Table 1 (the relative accident rate is referred to briefly in Borkenstein's report, in which it is called the alcohol level factor).

With the notation given in 2.5, the relative accident rate can be related to the accident involvement index:

$$I_i = \frac{100\,(a_i - e_i)}{e_i}$$

$$= 100\left\{\frac{(A + C)\,a_i - A(a_i + c_i)}{A(a_i + c_i)}\right\}$$

$$= 100\left\{\frac{C}{A} - \frac{(A + C)c_0}{A(c_0 + a_0 r_i)}\right\}$$

I_i is thus a monotonic increasing function of r_i

r_i can take any positive rational value: as $r_i \to \infty$, $I_i \to \frac{100C}{A}$

as $r_i \to 0$, $I_i \to -100$

A graph of I against r when $A = 5985$, $C = 7590$, $a_0 = 4992$, $c_0 = 6756$ (the values in Table 1) is given in Figure 1.

The statistical significance of the difference between relative accident rates r_i and r_j for two different classes may be tested by applying the χ^2 test of proportionality to the 2×2 matrix

$$\begin{pmatrix} a_i & c_i \\ a_j & c_j \end{pmatrix}$$

With the notation given in 2.6, the relative accident rate for the alcohol level class j within v class i is

$$r_{ij} = \frac{a_{ij}c_{io}}{c_{ij}a_{io}}$$

Values of r_{ij} are given in Tables 13-20 of this Report. Relative accident rates may only be compared within any one v class; to permit direct comparison between any two cells in the same two-factor analysis table, let the accident rate for v class i and alcohol level class j be

$$s_{ij} = \frac{a_{ij}C_v}{c_{ij}A_v}$$

where $A_v = \sum_{i=1}^{p_v} \sum_{j=0}^{4} a_{ij}$ and $C_v = \sum_{i=1}^{p_v} \sum_{j=0}^{4} c_{ij}$

(Because of incomplete observations, A_v and C_v are not the same for all variables v) s_{ij} is a measure of the accident risk in v class i and alcohol level class j compared with the average risk for all drivers observed.

3.3 Confidence limits for the relative accident rate

The relative accident rates given in Tables 1, and 13-20 are calculated from numbers of drivers in the accident and control groups, and are therefore subject to sampling error; the rate r_i in the ith alcohol level interval was obtained from the 2 × 2 matrix

$$\begin{pmatrix} a_i & c_i \\ a_0 & c_0 \end{pmatrix} \qquad (1)$$

Any 2 × 2 matrix with the same row- and column-sums is of the form

$$\begin{pmatrix} a_i + x & c_i - x \\ a_0 - x & c_0 + x \end{pmatrix} \qquad (2)$$

$$\text{Let } \rho_i = \frac{(a_i + x)\,(c_0 + x)}{(c_i - x)\,(a_0 - x)} \qquad (3)$$

Then there will be just one value of x (not in general a whole number) such that ρ_i represents the true accident risk of those who drive in Grand Rapids with alcohol levels in the ith interval, compared with the risk of those who drive with alcohol levels less than 10 mg/100 ml, allowing for differences between times of driving, distance driven, and routes taken. This particular value of x is determined as a function of a_i, c_i, a_0 and c_0 by equation (3), and with this value of x, let

$$\alpha\,(a_i, c_i, a_0, c_0) = x^2 \left(\frac{1}{a_i + x} + \frac{1}{c_i - x} + \frac{1}{a_0 - x} + \frac{1}{c_0 + x} \right) \qquad (4)$$

In a long series of experiments each yielding a 2 × 2 matrix as in (1) with the same set of row- and column-sums, α should be distributed approximately as χ^2 with one degree of freedom. This fact may be used to obtain from the survey data 95 per cent confidence limits for the relative accident rate in the ith alcohol level interval, as follows. If α is given the five per cent value of χ^2 (1 d.f.), and a_i, c_i, a_0, and c_0 are given the values obtained in the survey, equation (4) becomes an equation in x having just two roots such that all four entries in the 2 × 2 matrix in (2) are positive. These two values of x, when substituted in equation (3) give two values ρ_i' and ρ_i'' such that if the true relative risk did not lie between ρ_i' and ρ_i'', there would be a less than five per cent probability of chance variation producing a difference between r_i and the true relative risk at least as large as that obtained in the survey. ρ_i' and ρ_i'' are thus 95 per cent confidence limits for the accident risk in the ith alcohol level interval compared with that in the 0-9 mg/100 ml interval, and their values are included in Table 1.

3.4 Accident experience at blood alcohol levels 10-49 mg/100 ml

The data in Borkenstein's report warrant careful examination in order to assess whether or not they provide evidence that drivers perform better at alcohol levels 10-49 mg/100 ml than at alcohol levels less than 10 mg/100 ml.

When the accident and control groups are distributed with respect to alcohol level only, the relative accident rate is less than unity for alcohol levels 10-39 mg/100 ml; Table 22 is based on the first four rows of Table 1.

TABLE 22

Distributions by alcohol level in the range 0-39 mg/100 ml

Blood alcohol level (mg/100 ml)	Numbers of drivers		Relative accident rate	χ^2 (1 degree of freedom)*
	Accident	Control		
0-9	4992	6756	1	—
10-19	188	276	0.92	0.74
20-29	95	134	0.96	0.10
30-39	57	96	0.80	1.70

* These values of χ^2 are obtained by comparing each alcohol level interval in turn with the 0-9 mg/100 ml interval. The five per cent value of χ^2 with one degree of freedom is 3.84

When the accident and control frequencies in Table 22 are tested for proportionality, the value of χ^2 is 2.44 with three degrees of freedom. The corresponding five per cent significance level is 7.82.

Hence any difference revealed by the single-factor analysis between the accident experience of drivers with alcohol levels in the 10-39 mg/100 ml interval and that of drivers with alcohol levels less than 10 mg/100 ml is too small to be statistically significant at the five per cent level in a survey of this size.

The two-factor analyses of the accident and control groups with respect to alcohol level and each of the other variables (see Table 13-20) give distributions with respect to alcohol level within a total of 32 classes, and for each class a 2 × 2 matrix of the form

$$\begin{pmatrix} a & a_0 \\ c & c_0 \end{pmatrix}$$

may be extracted, where a and c are the numbers of accident and control drivers in the class and having alcohol levels in the interval 10-49 mg/100 ml, a_0 and c_0 are the corresponding numbers of drivers in the 0-9 mg/100 ml interval.

Let $\Delta = ac_0 - a_0c$

$s = a + c + a_0 + c_0$

$$y = \Delta\sqrt{\frac{1}{s}\left(\frac{1}{as - \Delta} + \frac{1}{cs + \Delta} + \frac{1}{a_0s + \Delta} + \frac{1}{c_0s - \Delta}\right)}$$

According to the χ^2 test of proportionality, on the null hypothesis that accident risks in the alcohol level intervals 0-9 and 10-49 mg/100 ml are equal, y^2 should be distributed approximately like χ^2 with one degree of freedom, so that y should be approximately a standard normal variate.

Let w be the area under the normal curve to the left of y.

Then on the above null hypothesis, w should be approximately uniformly distributed in the interval (0, 1). A bias towards or away from zero would indicate respectively lower or higher accident risk in the 10-49 mg/100 ml interval than in the 0-9 mg/100 ml interval.

Values of w for the 32 classes, together with corresponding relative accident rates for the 10-49 mg/100 ml alcohol level interval are given in Table 23 and the distribution for the 32 values of w is shown in Figure 2. The deviation of the observed distribution from uniformity was tested for statistical significance by means of the statistic U_N^2 developed by Watson.[5] If N points are distributed in the interval $0 \leqslant x \leqslant 1$ with distribution function $F_0(x)$, and the hypothetical distribution is F(x), then

$$U_N^2 = N \int_0^1 \{F_0(x) - F(x) - \int_0^1 (F_0(y) - F(y))\, dy\}^2\, dx$$

In the present case $F(x) = x$ and $N = 32$; it may be shown that if the 32 values of w are denoted by w_i $(1 \leqslant i \leqslant 32)$, and their arithmetic mean by $\bar{w}$, then

$$U_{32}^2 = \sum_{i=1}^{32} \left(w_i - \frac{2i-1}{64}\right)^2 - 32\,(\bar{w} - 1/2)^2 + 1/384$$

This gives the value 0.034, which is not significant according to the table of percentage points for U_N^2 give by Stephens.[6]

The conclusion is that the two-factor analyses do not reveal a significant difference at the five per cent level between the accident experience of drivers in the 0-9 and 10-49 mg/100 ml alcohol level intervals.

3. 5 Two-factor interaction: the danger of comparing ill-matched groups

It was mentioned in 2. 6 that a difference between the accident risks for drivers found in different alcohol level intervals can arise from the fact that the groups of drivers in the different intervals are not matched with respect to other variables.

TABLE 23

Relative accident rates at blood alcohol levels 10-49 mg/100 ml compared with 0-9 mg/100 ml when drivers are classified by one other variable

Second variable of classification		Relative accident rate	Value of w
Age	16-17	3.09	0.97
	18-24	1.04	0.60
	25-34 & 55-64	1.01	0.54
	35-54	0.95	0.32
Estimated annual mileage	1000 & under	0.86	0.38
	1001-5000	1.07	0.63
	5001-15 000	1.05	0.64
	over 15 000	0.78	0.07
Education level	less than 8 years	1.22	0.75
	8 years	0.87	0.25
	9-11 years	1.05	0.63
	12 years	0.97	0.40
	13-15 years	0.70	0.02
	16 years or more	0.92	0.36
Race or nationality	White	0.88	0.04
	non-white	1.51	0.97
Marital status	single	0.98	0.43
	married	0.99	0.44
	widowed	0.27	0.003
	separated	0.59	0.13
	divorced	1.25	0.77
Occupation status	lower	1.07	0.73
	skilled	0.84	0.12
	lower white-collar	0.84	0.13
	upper white-collar	0.83	0.15
Reported drinking frequency	yearly or less	1.19	0.88
	monthly	1.06	0.62
	weekly	1.09	0.75
	3 times weekly	1.13	0.75
	daily	1.46	0.98
Sex	male	0.94	0.21
	female	0.90	0.28

The danger of reaching false conclusions when comparing two ill-matched groups of individuals is clearly explained in a non-mathematical way by Bradford Hill[7], and a further apt illustration is provided by Table 24, which is an extract from Table 19.

TABLE 24

Comparison of the 0-9 and 10-49 mg/100 ml alcohol level intervals within reported drinking frequency classes

Reported drinking frequency	Blood alcohol levels (mg/100 ml)						
	0-9			10-49			
	Number of drivers		Accident rate	Number of drivers		Accident rate	Relative accident rate
	Accident	Control		Accident	Control		
Yearly or less	2485	2778	1. 09	94	88	1. 30	1. 19
Monthly	934	1010	1. 13	55	56	1. 20	1. 06
Weekly	959	1374	0. 85	122	160	0. 93	1. 09
3 times weekly	340	616	0. 67	64	103	0. 76	1. 13
Daily	160	505	0. 39	63	136	0. 57	1. 46
Total	4878	6283	0. 95	398	543	0. 90	0. 94

The last column shows that within each drinking frequency class the accident rate appeared higher in the higher alcohol level interval, whereas when all drinking frequency classes are combined, the reverse result occurs. The paradox arises from the fact that the proportion of infrequent drinkers is much higher in the 0-9 mg/100 ml interval than in the 10-49 mg/100 ml interval, and in both intervals the accident rate is significantly higher at the five per cent level for infrequent drinkers than for frequent drinkers. It should be noted that it is only for the daily drinkers that the difference between the accident rates in the two alcohol level intervals in Table 24 is by itself statistically significant at the five per cent level but the last column suggests that although greater drinking frequency is associated with a low accident rate, for each drinking frequency class the lowest accident rate may well occur in the 0-9 mg/100 ml alcohol level interval, i.e. even the frequent drinker should keep his drinking and driving separate in order to minimise his accident risk.

Borkenstein's report does not give details of two-factor analyses by drinking frequency and the other seven variables, but it does state (p. 135) that age-groups having the best and worst accident experience are respectively over- and under-represented among the frequent drinkers compared with the infrequent drinkers, and that the same is true when age is replaced by any other major variable except alcohol level. It thus appears that part of the difference between the accident risks of frequent and infrequent drinkers may be attributable to the effect of other variables.

The way in which the paradox in Table 24 arises is examined algebraically in the Appendix.

3.6 Effect of alcohol on drivers of different ages

In Table 13, which gives the distributions of accident and control drivers with respect to alcohol level within various age groups it can be seen that the relative accident rate rises faster as alcohol level rises above 80 mg/100 ml for drivers aged 25 to 34 and 55 to 64 than for those aged 35 to 54, and a little faster still for those aged 18 to 24. These differences are statistically significant at the five per cent level and it appears that in Grand Rapids the increase in accident risk resulting from high alcohol levels is about half as great again for the young and elderly drivers as for the middle-aged drivers.

This result contradicts the conclusion reached by the Grand Rapids investigators that accident risk does not vary significantly with age at high alcohol levels (see the last paragraph of 2.6). It should be noted that such a lack of significance may arise simply because the numbers of drivers found at the high alcohol levels are small, even when the variation of risk with the second variable is numerically greater at high alcohol levels than at low alcohol levels where it is statistically significant (as it is when the second variable is age). In such a case, it is hardly appropriate to interpret the lack of significance as indicating that there is no real variation of risk with the second variable.

3.7 Estimation of the proportion of drivers whose involvement in accidents in Grand Rapids was attributable to alcohol

The single-factor analysis with respect to alcohol level can be used to estimate the number and alcohol level distribution of drivers in the accident group whose involvement in accidents could be regarded as attributable to alcohol.

In the ith alcohol level there were, in the notation of 2.5 and 3.2, a_i accident-involved drivers and c_i control drivers. If the accident risk for drivers in the ith interval had been the same as for drivers observed to have alcohol levels less than 10 mg/100 ml, the expected ratio of the number of accident drivers to the number of control drivers would be $\frac{a_0}{c_0}$. Thus the expected number of accident drivers would be $\frac{c_i a_0}{c_0} = \frac{a_i}{r_i}$; $a_i - \frac{a_i}{r_i}$ therefore provides an estimate of the numbers of drivers in the ith interval whose involvement in accidents was attributable to alcohol. This calculation is made for each alcohol level in Table 25 (a).

The assumption implicit in the above calculation is that drivers observed to have alcohol levels above 10 mg/100 ml would, if they kept below 10 mg/100 ml, experience an accident rate similar to that of drivers observed to have alcohol level below 10 mg/100 ml. Because drivers found at different alcohol levels differed with respect to other variables, this assumption is no more than a useful approximation; it should be noted for instance, that some negative entries occur at medium alcohol levels indicating an increase of accident risk if drivers with these concentrations had their alcohol content reduced to 0-9 mg/100 ml (see 3.4 and 3.5). The general conclusion of Table 25 (a) is confirmed by the following calculation.

TABLE 25 (a)

An estimation of the proportion and alcohol level distribution of drivers whose involvement in accidents in Grand Rapids was attributable to alcohol

Blood alcohol level (mg/100 ml)	Relative accident rate (r_i)	Number of drivers in accident group (a_i)	Drivers whose involvement in accidents was attributable to alcohol: Estimated number $a_i - \frac{a_i}{r_i}$	Drivers whose involvement in accidents was attributable to alcohol: Percentage of accident group $\frac{100}{5985}\left(a_i - \frac{a_i}{r_i}\right)$
0-9	1.00	4992	0	0
10-19	0.92	188	—16.3	—0.27
20-29	0.96	95	—4.0	—0.07
30-39	0.80	57	—14.3	—0.23
40-49	1.08	66	4.9	0.08
50-59	1.21	50	8.7	0.15
60-69	1.41	46	13.4	0.22
70-79	1.52	36	12.3	0.20
80-89	1.88	39	18.3	0.30
90-99	1.95	39	19.0	0.32
100-119	5.93	92	76.5	1.27
120-139	4.94	73	58.2	0.97
140-159	10.44	54	48.8	0.82
160 and over	21.38	158	150.6	2.52
Totals	—	5985	376.1	6.28

For alcohol levels below 80 mg/100 ml, the difference between the alcohol level distributions of the accident and control groups is not statistically significant at the five per cent level, and therefore it is not very certain what effect a reduction of these alcohol levels to below 10 mg/100 ml would have on accident involvement. In Table 25 (b) therefore, in order to obtain a minimum estimate of the proportion of drivers whose involvement in accidents was attributable to alcohol, the levels up to 80 mg/100 ml are combined in the first row of the table; thus assuming for this one calculation that levels up to 80 mg/100 ml would not affect accident involvement, the ratio of the numbers of accident and control drivers which would be expected in the higher alcohol level intervals if no accidents were attributable to alcohol is taken to be the ratio observed in the 0-79 mg/100 ml interval. The calculation proceeds as in the ninth and subsequent rows of Table 25 (a) except that r_i is replaced by

$r_i' = \dfrac{a_i c_0'}{c_i a_0'}$, where a′ and c_0' are the numbers of accident and control drivers with alcohol levels less than 80 mg/100 ml. It so happens that in this survey r_i and r_i' do not differ until the fourth significant figure.

TABLE 25(b)

Estimation of the proportion of drivers whose involvement in accidents in Grand Rapids was attributable to alcohol, ignoring any such drivers who may have had alcohol levels of less than 80 mg/100 ml

Blood alcohol level (mg/100 ml)	Relative rate r_i'	Number of drivers in accident group	Drivers whose involvement in accidents was attributable to alcohol	
			Estimated number	Percentage of accident group
0-79	1. 00	5530	0. 0	0. 00
80-89	1. 88	39	18. 3	0. 31
90-99	1. 95	39	19. 0	0. 32
100-119	5. 92	92	76. 5	1. 28
120-139	4. 94	73	58. 2	0. 97
140-159	10. 43	54	48. 8	0. 82
160 and over	21. 36	158	150. 6	2. 52
Totals	—	5985	371. 4	6. 21

The conclusion from these two estimates is that if the same journeys had been made, but all drivers had kept their alcohol levels below 80 mg/100 ml, the number involved in accidents would have been reduced by about six per cent. It should be noted that this estimate is based upon comparison of the accident and control groups as a whole; no attempt was made to decide by assessing evidence which particular accidents were attributable to alcohol. The estimate is likely to be a conservative one because it takes no account of drivers with low alcohol levels who may have been innocently involved in accidents with drunken drivers. The estimation in Table 25 (b) could be repeated with a different base-level in place of 80 mg/100 ml, and the result of taking higher levels is shown in Fig. 3; for the reasons discussed above, the effect of taking base-levels below 80 mg/100 ml is uncertain.

The above calculations are based on single-factor analysis only, using two-factor analyses and making the above calculations for each level of the second variable, eight further pairs of estimates are obtained, each of which allows for the effect of one other variable. These estimates, given in Table 26, show comparatively little variation, and further support the above conclusion.

TABLE 26

Estimates, bases on two-factor analyses, of the percentage of the accident group whose involvement in accidents was attributable to alcohol

Variable the effect of which is allowed for	Estimated percentage of accident group	
	By method used in Table 25(a)	By method used in Table 25(b)
Age	7.3	6.9
Estimated annual mileage	6.4	6.4
Education level	5.1	5.2
Race	6.2	6.1
Marital status	5.5	5.8
Occupation status	5.3	5.5
Reported drinking frequency	7.5	6.7
Sex	6.4	6.8

3.8 Estimation of the proportion of accidents attributable to alcohol

Either of the methods of the preceding section can be extended to estimate the proportion of accidents which may in the same way be regarded as attributable to alcohol. Only single-driver and two-driver accidents are considered, because it appears that few accidents in Grand Rapids involve more than two drivers. Suppose that drivers are distributed over $n + 1$ alcohol level intervals denoted by suffixes 0, 1, 2. . . . n, the suffix 0 denoting the interval the lower limit of which is zero; this will be called the zero interval.

Let p_{ii} = proportion of accidents which involved two drivers, both having alcohol levels in the ith interval

$2p_{ij}$ = $2p_{ji}$ = proportion of accidents which involved two drivers, having alcohol levels one in the ith and the other in the jth interval

q_i = proportion of accidents which involved one driver having an alcohol level in the ith interval

$$p = \sum_{i=o}^{n} \sum_{j=o}^{n} p_{ij}$$

$$q = \sum_{i=o}^{n} q_i$$

Note: The data suggest that with the alcohol level intervals in Tables 25(a) and 1, $p_{oo} \sim 1, p_{io} \sim 10^{-2}$ $(i \neq o)$, and $p_{ij} \sim 10^{-4}$ $i, j \neq o$
Also, since few accidents involve more than two drivers, $p + q \doteqdot 1$

Separate relative accident rates can be defined for involvement in accidents of particular kinds. In the ith alcohol level interval, let

u_i = relative rate of involvement in two-driver accidents in which the other driver had an alcohol level in the zero interval

v_i = relative rate of involvement in single driver accidents

Then $u_0 = v_0 = 1$

Suppose there were originally N accidents; Nq of these would be single-driver and Np would be two-driver. Now consider new conditions in which the same journeys are made, but all drivers have alcohol levels in the zero interval. The expected rate of involvement in single driver accidents for drivers originally in the ith alcohol level interval will have been divided by v_i, so that instead of the original Nq_i single-driver accidents involving such drivers, $\frac{Nq_i}{v_i}$ would now be expected. The expected total number of single-driver accidents in the new conditions would therefore be

$$N \sum_{i=o}^{n} \frac{q_i}{v_i} = Nq^1 \qquad \text{(say)}$$

For two-driver accidents involving one driver in the ith alcohol level interval and one in the zero interval, the only change between the original and the new conditions is in the alcohol level of the former driver, whose expected rate of involvement in accidents of this kind will have been divided by u_i; there were originally $N(p_{io} + p_{oi})$ of these accidents, so that $\frac{N}{u_i}(p_{io} + p_{oi})$ would be expected in the new conditions. Since $u_0 = 1$, the expected number may be rewritten $N\left(\frac{p_{io}}{u_i u_o} + \frac{p_{io}}{u_o u_i}\right)$. It remains to consider two-driver accidents involving drivers in the ith and jth alcohol level intervals; for such accidents, the alcohol levels, of both drivers will be different in the new conditions, and their respective expected rates of involvement in accidents with drivers whose alcohol levels remain unchanged will have been divided by u_i and u_j. It therefore seems reasonable to expect that the original number $N(p_{ij} + p_{ji})$

of accidents involving one driver in the ith and one in the jth alcohol level interval will be divided by $u_i u_j$*, so that the expected number in the new conditions will be $N\left(\frac{p_{ij}}{u_i u_j} + \frac{p_{ji}}{u_j u_i}\right)$.

The expected total numbers of two-driver accidents in the new conditions would therefore be

$$N \sum_{i=0}^{n} \sum_{j=0}^{n} \frac{p_{ij}}{u_i u_j} = Np^1 \quad \text{(say)}$$

The expected reduction in accidents would therefore be

$$N[(p - p^1) + (q - q^1)]$$

and the percentage reduction in accidents would be

$$R_a = \frac{100[(p - p^1) + (q - q^1)]}{(p + q)}$$

The percentage reduction in the number of drivers involved would be

$$R_d = \frac{100[2(p - p^1) + (q - q^1)]}{2p + q}$$

It appears from Table 21 that $u_i < v_i$ from which it follows that $\frac{p - p^1}{p} < \frac{q - q^1}{q}$

Hence $R_d < R_a$

However $\frac{R_a}{R_d} = \left(\frac{(p - p^1) + (q - q^1)}{2(p - p^1) + (q - q^1)}\right)\left(\frac{2p + q}{p + q}\right) < 2$

Hence $1 < \frac{R_a}{R_d} < 2$

The data published in Borkenstein's report are insufficient for the above calculation to be made. The report does, however, give sufficient information to enable the relative accident rates v_i to be calculated for wide alcohol level intervals together with approximations to the relative rates u_i. The results are given in Table 21 (p. 25).

---

* The assumption here is that the effects of the changes in the two classes of drivers on the number of accidents involving one driver from each class are independent. This does not mean that the alcohol levels of two drivers involved in an accident are independent; any correlation between these levels (which might arise, for example, if drivers with high alcohol levels tended to be concentrated at particular times and places) will be reflected in the values of p_{ij}

It has not been found possible to estimate from Borkenstein's published data the proportion of casualties attributable to alcohol.

3.9 An estimate of the effect of a change in the proportion of drivers drinking

It will be convenient to describe in the form

$$(P_0, P_1, P_2, P_3, P_4)$$

the percentage distribution of a group of drivers over the five alcohol level intervals 0-9 mg, 10-49 mg, 50-79 mg, 80-109 mg, and 110 mg and above, where P_0 is the percentage of drivers in the 0-9 mg interval, etc.

The percentage distribution of Borkenstein's control group was

(89.04, 7.76, 1.74, 0.91, 0.58)

and that of his accident group was

(83.42, 6.78, 2.21, 2.21, 5.40)

Suppose that the proportion of drivers on the roads of Grand Rapids in each alcohol level interval except the first is multiplied by x. The expected distribution of the control group would then be

$$(100 - 10.99x, 7.76x, 1.74x, 0.91x, 0.58x)$$

Assuming that the accident rate within each alcohol level interval remained unchanged, the expected distribution of the accident group, expressed in percentages of the original accident group, would be

$$\left(\frac{83.42\,(100 - 10.99x)}{89.04}, 6.78x, 2.21x, 2.21x, 5.40x\right)$$

The sum of these percentages is 100 + p, where p is the expected percentage increase in accidents resulting from the change in the alcohol level distribution of drivers on the road. This gives the equation

$$p = 6.3\,(x - 1)$$

Note: When $x = 0, p = -6.3$, which agrees with the estimate in Table 25(a), because $x = 0$ corresponds to all drivers having alcohol levels below 10 mg/100 ml

3.10 Effect of other variables upon accident risk

The discussion in this Report has been concerned with the effect of alcohol upon accident risk, but the data in the report provide similar information about the effect of other variables; further work of this kind has not yet been done here but the observations below are relevant.

The relative accident rate defined in 3. 2 depends upon the choice of a base class to which to relate the accident rates in other classes. When considering the variation of accident risk with alcohol level, it is natural to choose the lower alcohol level interval as the base class, but with other variables the choice of a base class will be more arbitrary, and the relative accident rate may therefore be less useful.

The dangers of comparing ill-matched groups, discussed in 3. 5, is, however, very relevant to the examination of the effect of other variables upon accident risk; because the variation with other variables is much smaller than with alcohol level, the danger of attributing an observed variation to the wrong variable is correspondingly greater.

3. 11 Suggestions for further analysis of the data

There appear to be several ways in which the wide range of analyses in the report could be usefully complemented.

It would be interesting to have two-factor analyses by drinking frequency and each of the other major variables, so that the association between accident risk and drinking frequency could be examined more closely. It should also be possible to obtain from the basic survey data values of the p_{ij} of 3. 8 corresponding, for example, to the five wide alcohol level intervals used in the two-factor analyses; the estimation proposed in 3. 8 could then be carried out.

Even the abundant data collected in Grand Rapids are only partly sufficient for two-factor analysis and certainly insufficient for detailed three-factor analysis, but some such analyses would be most useful in studying accident experience in the alcohol level interval 10-49 mg/100 ml, where there are relatively many drivers, even if the numbers of drivers at higher alcohol levels were too small to support useful conclusions. The variables most likely to lead to useful three-factor analyses are those that separate the accident and control groups meaningfully into a small number of classes of similar size. Age, education level, and occupation status satisfy this condition well, and estimated annual mileage and reported drinking frequency satisfy it fairly well, but race, material status and sex give classes varying widely in size.

4. CONCLUSIONS

The Grand Rapids study shows positively that:

(i) at blood alcohol levels above 80 mg/100 ml, the risk of being involved in an accident is higher than at alcohol levels below 10 mg/100 ml, and the risk increases more and more rapidly as the highest alcohol levels are reached.

(ii) the increase in accident risk resulting from high alcohol levels is greater for young and elderly drivers than for middle-aged drivers.

(iii) assuming drivers correctly stated how frequently they drank alcohol, those who drank at least once a week had a lower accident risk, when their alcohol levels were low, than those who rarely or never drank alcohol.

The data are insufficient to show positively whether the accident risk at alcohol levels of between 10 and 50 mg/100 ml differs from that at alcohol levels below 10 mg/100 ml.

5. ACKNOWLEDGEMENT

The work in this Report is published with the agreement of Professor R. F. Borkenstein.

6. REFERENCES

1. BORKENSTEIN, R. F., R. F. CROWTHER, R. P. SCHUMATE, W. B. ZIEL, and R. ZYLMAN. The role of the drinking driver in traffic accidents. 1964 (Indiana University, Department of Police Administration).

2. LUCAS, G. H. W., W. KALOW, J. D. McCOLL, B. A. GRIFFITH, and H. W. SMITH. Quantitative studies of the relationship between alcohol levels and motor vehicle accidents. Proceedings of the Second International Conference on Alcohol and Road Traffic. Toronto, 1955 (International Conference on Alcohol and Road Traffic), pp. 139-42.

3. VAMOSI, M. Determination of the amount of alcohol in the blood of motorists. Traff. Saf., 1960, **57** (3), Res. Rev. 4 (3), 8-11.

4. JEFFCOATE, G. O. An examination of reports of fatal road accidents in three police districts from the point of view of the effect of alcohol. Brit. J. Addict., 1958, **54** (2), 81-103.

5. WATSON, G. S. Goodness of fit tests on the circle: I, Biometrika, 1961, **48** (1 & 2) 109-14.

6. STEPHENS, M. A. The distribution of the goodness of fit statistic U_N^2: II Biometrika, 1964, **51** (3 & 4) 393-97.

7. BRADFORD HILL, A. Principles of medical statistics. London, 1939 (The Lancet Ltd.) pp. 125-8.

7. APPENDIX

Algebraic examination of the paradox in Table 24

Consider the simplest case, in which the accident and control groups are analysed with respect to two variables R and K and there are two R-classes and two K-classes (rows and columns respectively in the tables of results). The accident and control group frequencies form 2 × 2 matrices

$$A = (a_{ij}) \text{ and } C = (c_{ij})$$

and the accident rates s_{ij} (as defined in 3. 2) form a 2 × 2 matrix

$$S = (s_{ij}) \text{ where } s_{ij} \propto \frac{a_{ij}}{c_{ij}}$$

Suppose that (i) within each R-class the first K-class has the lower accident rate i.e.

$$s_{i1} < s_{i2} \qquad (i = 1, 2)$$

so that $\quad a_{i1}c_{i2} < a_{i2}c_{i1} \qquad (i = 1, 2) \qquad (1)$

(ii) When the two R-classes are combined, the second K-class has a lower accident rate, i. e.

$$\frac{a_{11} + a_{21}}{c_{11} + c_{21}} > \frac{a_{12} + a_{22}}{c_{12} + c_{22}}$$

This condition can be rewritten

$$(a_{11}c_{22} - a_{22}c_{11}) + (a_{21}c_{12} - a_{12}c_{21}) > (a_{12}c_{11} - a_{11}c_{12}) + (a_{22}c_{21} - a_{21}c_{22}) \qquad (2)$$

It follows from (1) that both brackets on the right hand side of (2) are positive and that the brackets on the left have opposite signs. For (2) to hold it is necessary that the positive term on the left should be dominant.

Example

$$A = \begin{pmatrix} 110 & 12 \\ 9 & 100 \end{pmatrix} \quad C = \begin{pmatrix} 100 & 10 \\ 10 & 100 \end{pmatrix} \quad S = \begin{pmatrix} 1.05 & 1.14 \\ 0.86 & 0.95 \end{pmatrix}$$

R-classes combined $(119 \;\; 112) \qquad (110 \;\; 110) \qquad (1.03 \;\; 0.97)$

In this example the diagonal elements dominate the frequency matrices, i.e. the first R-class is over-represented in the first K-class and under-represented in the second class. If corresponding columns of A and C are multiplied by the same constant, (1) and (2) will still hold.

It is interesting to note that (1) and (2) can hold simultaneously even in certain cases where the effects of variables R and K on accidents risk are independent. If α and β are positive numbers representing the effects of R and K, and if

$$C = c\begin{pmatrix} 1 & y \\ x & z \end{pmatrix}$$

then $A = a\begin{pmatrix} 1 & \beta y \\ \alpha x & \alpha\beta z \end{pmatrix}$

In this case (1) becomes $\beta > 1$
and (2) becomes

$$z(1 - \alpha\beta) + xy(\alpha - \beta) > y(\beta - 1) + xz\alpha(\beta - 1)$$

In considering whether (3) can hold there are three cases:

(i) $\frac{1}{\beta} < \alpha < \beta$; both terms on the left are negative so that (3) cannot hold.

(ii) $\alpha > \beta$; then for sufficiently large x

$$xy(\alpha - \beta) > y(\beta - 1)$$

and, although $(1 - \alpha\beta) < 0$ and $\beta - 1 > 0$, (3) will hold for sufficiently small z

(iii) $\alpha < \frac{1}{\beta}$; then for sufficiently small x

$$z(1 - \alpha\beta) > xz\alpha(\beta - 1)$$

and, although $\alpha - \beta < 0$ and $\beta - 1 > 0$, (3) will hold for sufficiently small y.

Examples

(i) $\alpha > \beta$; $(\alpha = 3, \beta = 2, x = 10, y = 1, z = 0.2)$

$$A = \begin{pmatrix} 100 & 200 \\ 3000 & 120 \end{pmatrix} \qquad C = \begin{pmatrix} 100 & 100 \\ 1000 & 20 \end{pmatrix} \qquad S = \begin{pmatrix} 0.36 & 0.71 \\ 1.07 & 2.14 \end{pmatrix}$$

R-classes combined (3100 320) (1100 120) (1.01 0.95)

(ii) $\alpha < \frac{1}{\beta}$; $(\alpha = 0.25, \beta = 2, x = 1, y = 0.05, z = 1)$

$$A = \begin{pmatrix} 100 & 10 \\ 25 & 50 \end{pmatrix} \qquad C = \begin{pmatrix} 100 & 5 \\ 100 & 100 \end{pmatrix} \qquad S = \begin{pmatrix} 1.65 & 3.30 \\ 0.41 & 0.82 \end{pmatrix}$$

R-classes combined (125 60) (200 105) (1.03 0.94)

It is thus possible for two variables to affect the accident rate quite independently, while the composition of the accident and control groups is such that one variable causes single-factor analysis to give a completely false picture of the effect of the other variable on accident risk. Tests of statistical significance will not detect an error of this kind, so that caution is needed in interpreting even statistically significant differences between distributions of the accident and control groups.

8. LIST OF SYMBOLS

(Omitting some symbols used only in the sections where they are defined)

a_0, c_0	Numbers of accident and control drivers with blood alcohol levels of less than 10 mg/100 ml
a_i, c_i	Numbers of accident and control drivers in class i
$A = \sum_i a_i$	
$C = \sum_i c_i$	
a_{ij}, c_{ij}	Numbers of accident and control drivers in v-class i and alcohol level interval j
I_i	Accident involvement index for class i
j	Number assigned to an alcohol level class in two-factor analysis
n + 1	Number of alcohol level intervals in estimation of number of accidents attributable to alcohol
p_v	Number of v-classes remaining after appropriate combinations have been made
p_{ii}	Proportion of accidents which involved two drivers both having alcohol levels in the ith interval
p_{ij}	Half of the proportion of accidents which involved two drivers having alcohol levels one in the ith and one in the jth interval
$p = \sum_{i=o}^{n} \sum_{j=o}^{n} p_{ij}$	
q_i	Proportion of accidents which involved one driver having an alcohol level in the ith interval
$q = \sum_{i=o}^{n} q_i$	
r_i	Relative accident rate for ith alcohol level interval
r_{ij}	Relative accident rate for alcohol level class j within v-class i
R_a	Percentage reduction in number of accidents
R_d	Percentage reduction in number of drivers involved
ρ_i', ρ_i''	95 per cent confidence limits for r_i
s_{ij}	Accident rate for alcohol level class j within v-class i
u_i	Relative rate of involvement for drivers in the ith alcohol level inter-

val in two-driver accidents in which the other driver had an alcohol level in the zero interval

v_i Relative rate of involvement in single-driver accidents for drivers with alcohol levels in the ith interval

v Variable other than alcohol level in two-factor analysis

w Statistic used in comparing accident risk at alcohol levels 0-9 and 10-49 mg/100 ml

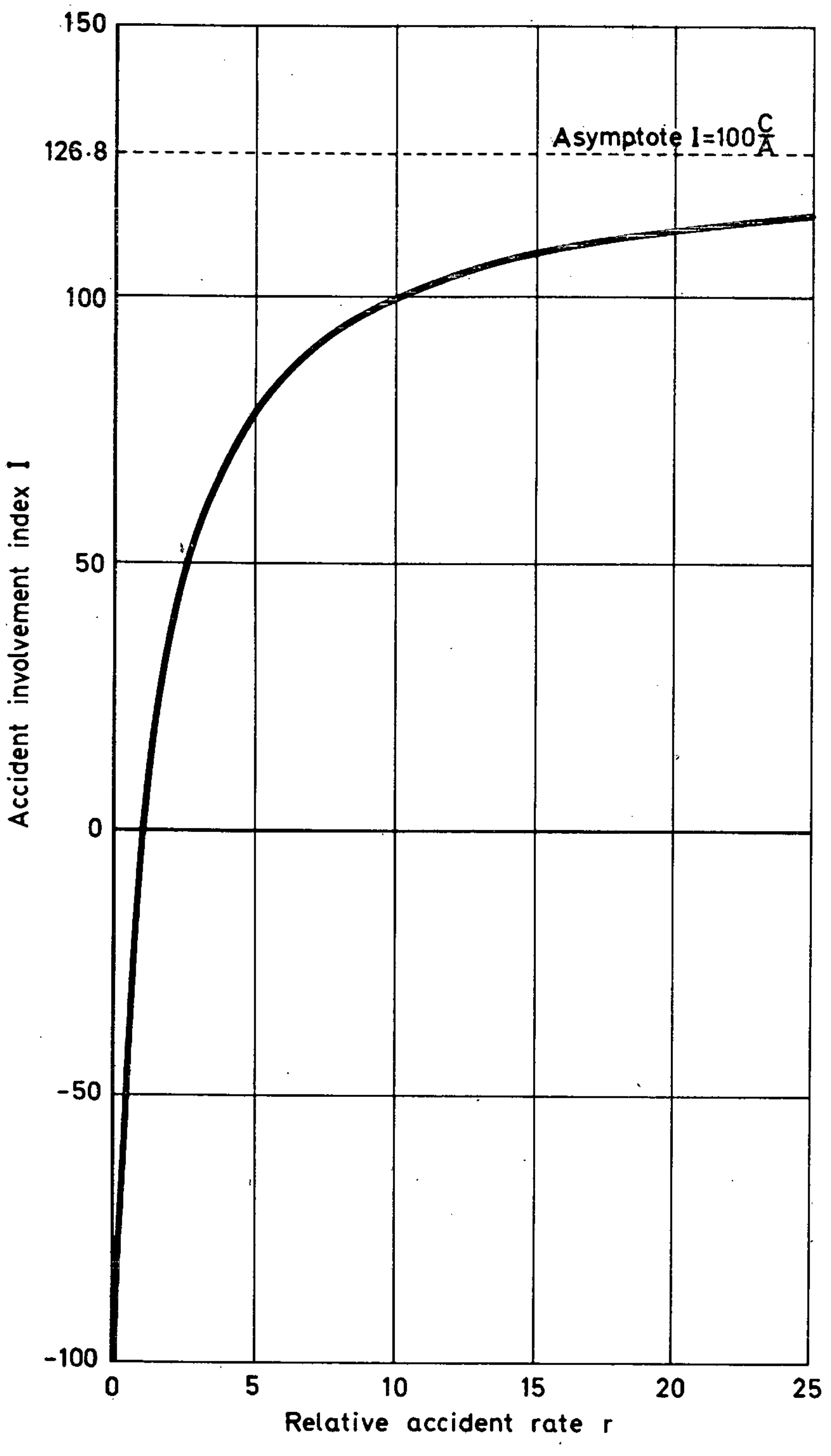

Fig. 1. RELATION BETWEEN ACCIDENT INVOLVEMENT INDEX AND RELATIVE ACCIDENT RATE WHEN THE ACCIDENT GROUP CONTAINS A=5965 DRIVERS OF WHICH 4992 HAVE ALCOHOL LEVELS BELOW 10 MG/100 ML AND THE CONTROL GROUP CONTAINS C = 7590 DRIVERS OF WHICH 6756 HAVE ALCOHOL LEVELS BELOW 10 MG/100 ML

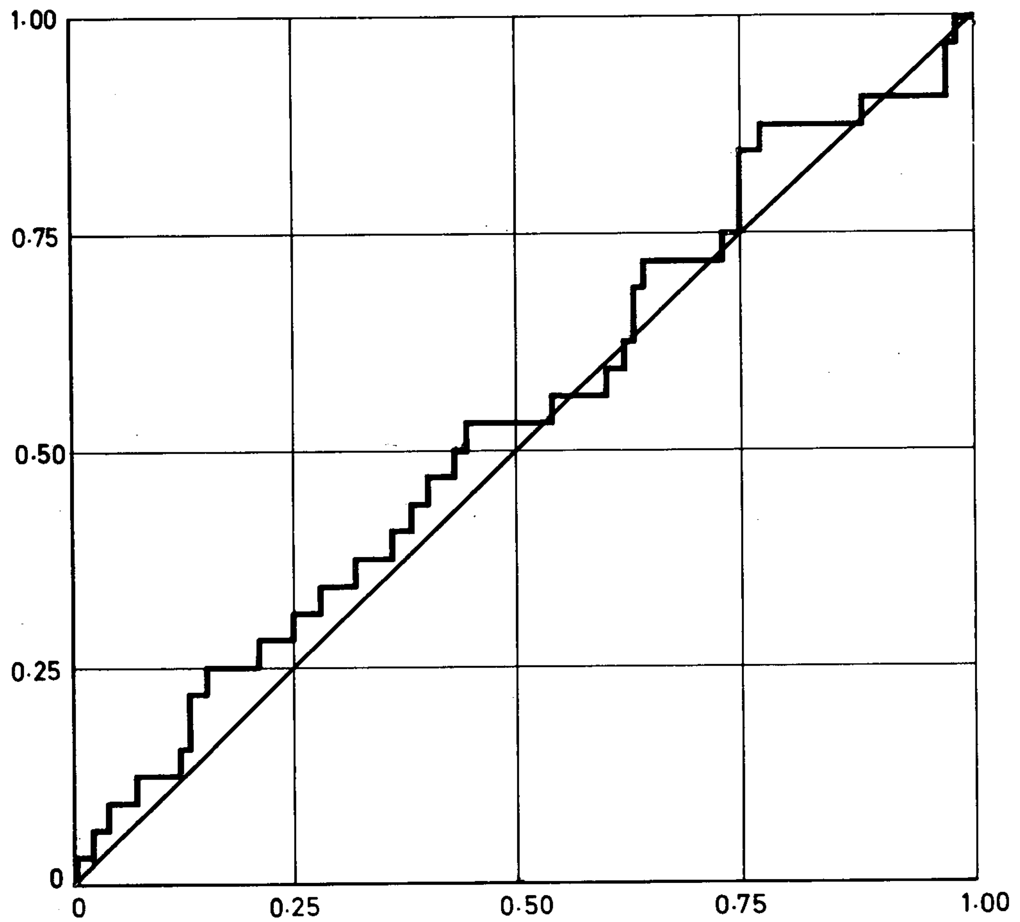

Fig. 2. DISTRIBUTION FUNCTION FOR THE 32 VALUES OF STATISTIC w OBTAINED FROM TWO-FACTOR ANALYSES

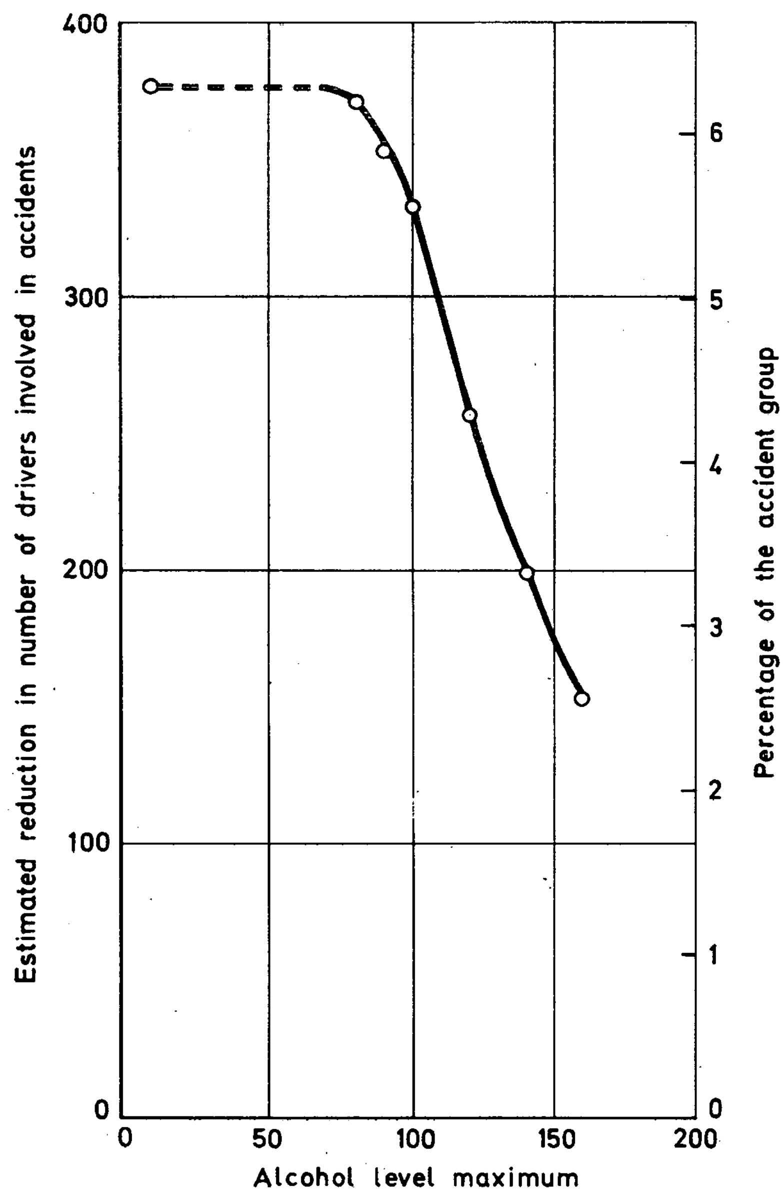

Fig. 3. ESTIMATED REDUCTION IN ACCIDENT INVOLVEMENT IF ALL DRIVERS KEPT THEIR ALCOHOL LEVELS BELOW A GIVEN MAXIMUM

Printed at the Road Research Laboratory, Harmondsworth