Accident Analysis and Prevention 37 (2005) 721–730

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

A statistical profile of road accidents during cross-flow turns

David D. Clarke[a,*], Richard Forsyth[b,1], Richard Wright[c,1]

[a] School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[b] Centre for English Language Teacher Education, University of Warwick, Coventry CV4 7AL, UK
[c] Unilever Research, Port Sunlight Laboratory, Bebington, Wirral L63 3JW, UK

Received 17 March 2005; received in revised form 17 March 2005; accepted 20 March 2005

Abstract

In-depth studies of behavioral factors in road accidents using conventional methods are often inconclusive and costly. In a series of studies exploring alternative approaches, 200 cross-flow junction road accidents were sampled from the files of Nottinghamshire Constabulary, England, coded for computer analysis using a specially devised 'Traffic Related Action Analysis Language', and then examined using different computational and statistical techniques. For comparison, the same analyses were also carried out on 100 descriptions of safe turns, and 100 descriptions of hypothetical accidents provided by experienced drivers. The present study used statistical methods to explore the database of cases. The youngest and oldest groups of drivers were found to be over-represented in the junction accidents, and were the least likely to stop before turning. The young drivers had particular problems turning onto major roads. Women were more likely than men to stop before turning; they tended to have their collisions with other women; and they were under-represented as drivers of the non-turning vehicle. In hypothetical accidents, informants tended to blame the younger driver, increasingly so for male informants as they got older. Female informants tended to blame male drivers.

Keywords: Accident causation; Road junction; Police records

1. Introduction

This paper is the fourth in series analyzing road junction accidents in the county of Nottinghamshire, England, using the same database of cases, but different methods of analysis. The first paper (Clarke et al., 1998a) used an Artificial Intelligence machine-learning technique to find rules distinguishing between accidents turning onto and off major roads; those of younger and older drivers; and those that resulted in injury or in damage only, for example. The second paper (Clarke et al., 1998b) used a different Artificial Intelligence technique to produce 'decision trees' which could distinguish injury from damage-only accidents, and those of Young Male drivers; and could pick out the properties of relatively more and less dangerous situations. The third paper (Clarke et al., 1999) used sequence analysis methods to examine the patterns of events preceding accidents of different types. Differences were found between real accidents, hypothetical ones and safe turns; between accidents turning onto and off a road with the right of way; between the accidents of younger and older drivers; between accidents on minor roads and major roads; and between the accident expectations (but not the real accidents) of Male and Female drivers. This fourth paper reports the statistical work done on the database of cases. The rationale for the project, and the methods for case selection and coding, are described in detail in the earlier papers, and briefly summarized here.

The purpose of this final paper in the series is (1) to 'complete the picture' by reporting the remaining findings from the large multi-method study; (2) to characterize the sample more fully on which the overall study was based; (3) to present the outcomes of conventional statistical analyses for comparison with the other techniques, derived from Artificial Intelligence and Ethology; and (4) to provide a summary evaluation of the different methods as they performed on these data.

* Corresponding author. Tel.: +44 115 951 5284; fax: +44 115 951 5324.
E-mail address: ddc@psychology.nottingham.ac.uk (D.D. Clarke).
[1] These authors were working at the University of Nottingham when this study was conducted.

0001-4575/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2005.03.013

Behavioral factors in road accidents are difficult to study by traditional research methods for a number of reasons. Accidents are relatively unpredictable and infrequent, so direct observation is often impossible. Statistical comparisons of accident rates for different kinds of driver and circumstance leave out 'the various stages in the causal sequence' (Wagenaar and Reason, 1990, p. 1373). Multidisciplinary Accident Investigation (MDAI) teams have a number of disadvantages. As (Grayson and Hakkert, 1987, p. 42) pointed out, "in spite of the tremendous amount of information collected in this type of study, the definitive conclusions reached on the crash process are very limited".

Simple self-report studies using interviews and questionnaires with accident-involved road users are limited by the inaccessibility of over-learned behaviors to verbal report; problems of post-hoc interference and forgetting; and deliberate concealment of critical information. Most in-depth accident studies use heterogeneous samples of accident types, which make it difficult to find meaningful general patterns, even though case-study approaches are known to work best with tightly defined samples of similar accidents (e.g. England, 1981; Midland, 1992). On-the-spot accident investigations also tend to be biased towards injury accidents and certain times of day.

Finally, most behavioral in-depth studies tend to sacrifice the 'glue' of a rich relational network in the data (Sheehy and Chapman, 1988) because they only look for complex patterns once the data have been aggregated over cases; rather than finding patterns in individual cases, and then aggregating those. This is a problem since most data coding methods can only preserve the elements of a case and not its structure, and the structural information, once lost, cannot be recovered by subsequent analysis. For instance, a given accident might involve three key events, A, B and C, with A causing B after a short delay, and causing C after a longer delay. It is important that the presence of the three events, their sequence, and their causal relations (which are not completely apparent from the sequence) should be recorded and compared with other cases in the search for common patterns. Most methods, however, would simply record the presence of A and/or B and/or C in each case, to produce frequency counts, from which neither sequence nor causal links could be established.

The present series of studies is intended to develop and demonstrate alternative approaches to accident causation research which could overcome many of these problems. Each study explores a different way of analyzing police case files describing right-turn junction accidents, sampled from the records of Nottinghamshire Constabulary (the regional police force for the county of Nottinghamshire, England) for the year of 1988. [Note that in Britain, drivers use the left side of the road, so a right turn involves crossing the oncoming stream of traffic when turning off a major road, and crossing one stream of traffic and merging with the second when turning onto a major road. In general, the implications and hazards of this maneuver are the same as a left turn in countries such as the USA and mainland Europe, where drivers use the right side of the road.]

Police files were chosen, partly because they were convenient and suitable for the purpose; partly because they cover all locations in the region, all seasons and all times of day; and partly in order to demonstrate that these methods can provide an effective way of using this large body of data, which have rarely been used in this way before. Accident research using summary statistics derived from police reports is nothing new, of course, but it is unusual to take account of all the information in complete accident files, such as witness statements, scale drawings of accident sites, and so on. Other instances of this kind of approach include work by Fell (1976), who advocated their use in the development of 'accident causal schemata'. Massie et al. (1993, p. 255) produced typologies for collision avoidance strategies using approximately 200 police files, concluding "The review of selected police accident reports is an essential element of the analytic process".

Following the recommendations of Grayson and Hakkert (op. cit.), a specific type of accident was targeted in order to cut down the diversity of the sample, and thus improve the chances of getting meaningful results which were consistent across cases. Right-turning accidents, either onto or off a larger road, were chosen because they met the following criteria:

(a) They occurred with sufficient frequency in Nottinghamshire to give a large enough pool of accidents to sample.
(b) There was a reasonably long chain of separable events leading up to the accident, which is a requirement for sequence analysis.
(c) This class of accident is of special interest to researchers in its own right.

Older drivers are over-represented in junction accidents (Moore et al., 1982; Viano et al., 1990; Verhaegen, 1995). So, with an ageing population which has greater susceptibility from side impacts (Viano et al., 1990), it seems that the human and financial cost of this type of accident will increase with time, just as the incidence of other types of accident is decreasing.

In addition to the accident case reports, a sample of experienced drivers were each asked to write an account of an imaginary right-turning accident. By comparing hypothetical with real accidents, we hoped to see how the accident-expectations of drivers differ from the real dangers, and thus where their precautions against accidents are likely to be misconceived.

2. Method

2.1. Real accidents

2.1.1. Case selection

Two hundred police case-files on right-turning accidents were randomly selected from the population of records of that type held at police headquarters for Nottinghamshire, England (Nottingham Constabulary) for 1988, to include 100 right turns off a main road and 100 right turns onto a main road. Police personnel helped us to locate the files for the correct time-period and maneuver, but the further selection of 200 cases for the study was made at random by the researchers. Of these, 16 failed one or more of the following exclusion criteria:

(a) the accident occurred at a roundabout (these were excluded to reduce the heterogeneity of the sample),
(b) the accident had obviously been mis-classified as a right turn, or
(c) there was insufficient data on the police report to allow the accident to be reliably coded. Naturally the police gather and record more information about the more serious accidents, but so few records had to be excluded for lack of detail that the representativeness of the sample was not seriously compromised. Any bias that does result from this factor, will only mean that our conclusions have more to say about serious than slight accidents, which seems a relatively acceptable emphasis.

As a further check on the representativeness of the sample of cases, several variables were spot-checked against their counterparts in the published figures for all road traffic collision casualties in Nottinghamshire for 1988, and all the variables were found to match barring a few percent discrepancy at most. The differences between our sample and the general figures were: the proportion occurring at weekends—a difference of 5%; the proportion in fine weather—a difference of 1%; the proportion occurring on dry roads—a difference of 1%; the proportion occurring in daylight—a difference of 3%; the proportions occurring in four seasons—differences of 1, 1, 0 and 0%; and the proportions occurring in four 6-h time-periods round the clock—differences of 0, 1, 5 and 4%.

This left 184 accident cases, 90 turning onto, and 94 turning off, a major road. This would seem like a very small sample, had the sole purpose of the project been to carry out a statistical survey. However, the aim was to carryout four rather different types of analysis on the same set of cases, some of which involved detailed computational analysis of case details which would not have been feasible with a larger sample. The statistical procedures we used were chosen to safeguard against misleading conclusions arising from the small sample. The very specific sampling frame also allowed us to examine a very homogeneous set of cases with similar accident mechanisms, which we felt to be essential if detailed conclusions were to be drawn validly about a class of accidents. The aim was to avoid the two common extremes of (a) summarizing a large number of accidents which do not share any detailed causal pattern; or else (b) going into so much detail on individual cases that none of them can be generalized beyond itself.

2.1.2. Coding

Each case file contained a wealth of detailed information about the accident, typically including the four page summary form completed by the attending police officers, witness statements, vehicle examiners' reports, breath test reports, scale drawings of the accident site, and photographs from numerous distances and positions. Some files also contained further information and correspondence concerning court proceedings and insurance claims. All these sources of information were considered when making the coded summaries for further analysis here. It is this wealth of initial information from different parties and different perspectives, some of it based on the work of independent experts such as vehicle examiners, which sets this study apart from the 'simple self-report' methods we criticize above. Nevertheless, it has to be admitted that the police are not a research organization, and that their information is gathered primarily in support of possible later prosecutions, in a context that is time and resource limited, so it is not without its limitations. To that extent, this study should be seen as one of the many 'converging operations' that are building us a picture of road accident causation, not the final definitive proof of any single point. It will be for future studies using similar or different methods to reinforce or challenge the findings, and to do so in ways that highlight the emphases that arise from this mode of data collection.

In the Artificial Intelligence parts of the study, the generalisability of the findings had been further safeguarded by special Machine Learning procedures reported in detail in Clarke et al. (1999). For this purpose, the results of each analysis were evaluated on a new set of cases (the 'test set'), which has been kept apart from the 'training set' in which the patterns were originally discovered. This provides an extra stringent test of the adequacy of the findings by reducing the danger of 'overfitting'—that is of describing anomalies in a particular set of cases which would not generalise to others. In experimental terms, this is like reporting the significance of a replication, rather than of the initial study.

'Static features' were recorded (such as weather and road conditions, time of day, carriageway type, and so on) together with the sequence of events making up the accident.

Each case was coded by one of two coders who had jointly devised and piloted the coding frame, and agreed the definitions of all the coding terms. A further check of inter-coder reliability was carried out by setting a rule-finder program ('Bionic Evolutionary Algorithm Generating Logical Expressions', BEAGLE) to try to predict which accidents had been coded by which coder. Where the rule-finder was successful, it not only indicated the presence of an inter-coder discrepancy, but also described its nature. On the first pass, some such rules were found, so all accidents were checked by both coders jointly and idiosyncratic codings removed. After this the rule-finder was used again, and it failed to find any effective rules to discriminate one coder's work from the other's (which is how it should be).

Missing information was handled by contacting the people involved in the accident, or the attending police officers, or else by visiting the accident site to gather further information. Photographs of the accident scene were used to check that the road layout had not been changed in the meanwhile. The research was carried out several years after the accidents had occurred, so there was no question of investigating crash scenes per se while relevant physical evidence was still present.

2.2. Hypothetical accidents

One hundred drivers, 52 males and 48 females, with a mean age of 40.7 years were paid a small honorarium to take part in the study. They were recruited through advertisements in the local press and shop windows, and by word of mouth.

Each subject was taken to a quiet room where they completed a questionnaire in which they were asked to imagine a right-turning accident based upon their own driving experience, write a free form description of how it occurred, and then fill in the background details. These details corresponded to the information contained in the police reports of real accidents. They also completed a second questionnaire relating to a ‘safe’ right turn—one which did not end in an accident—for another part of the project, not reported in detail here. The hypothetical accidents did not pose the same problems of missing information as the real accident data. Otherwise, however, they were checked in the same way as the real cases.

The remaining set of 184 coded real cases, together with the hundred hypothetical accidents (and a 100 ‘safe right turns’) then made up the ‘Nottingham Accident Database for Right Turns’ (NAD/RT).

3. Results

3.1. Basic statistics

Table 1 gives details of the main static features of the accidents in the NAD/RT Database. Note, the shorthand name used throughout for drivers turning onto and off the larger road is ‘Turners’, or ‘Turners Onto’ and ‘Turners Off’, respectively. The vehicle or pedestrian they hit is called the ‘Collider’.

Some of the accident features from this table are examined in more detail below. Others, such as the relation between severity and accident mechanism, emerge in detail from the analysis in other papers—in this case, Clarke et al. (1998b) in particular.

3.2. Age effects 1—overall involvement

There were significant differences between age groups. Young drivers (aged 16–24) accounted for more than their fair share of these accidents. Table 2 shows the ‘Turners’ of known age divided into three age groups, and the proportion in each group compared with expectations based on mileages driven, estimated from the National Travel Survey Report 1985–1986 (The Department of Transport, 1987).

Turners in the youngest age group are more than three times as numerous as mileage alone would suggest. Middle-aged drivers appear to be the safest group while older drivers are represented in proportion to miles driven. This replicates the findings of Rabbit (1991), among others.

3.3. Age effects 2—type of involvement

In this section, baseline figures will be drawn from part of Table 37, ‘Road Accidents: Great Britain 1988: The Casualty Report’ (The Department of Transport, 1989), dealing with involvement statistics for all motor vehicle drivers. For the sake of comparability, the accidents involving pedestrians and cyclists in our sample have been excluded from the analyses in this section.

The first question is whether drivers of different ages are particularly associated with one of the three main roles in right-turning accidents (Turning Off, Turning Onto, or Colliding). For this a Chi-square test was used as a measure of association. If the observed frequencies for the different age groups differed significantly from the expected values based on the national statistics for all accidents, it would indicate that certain age groups were over-represented in our sample relative to their involvement in other accident types.

First, the expected values were obtained by calculating the percentage involvement of each age group in road accidents of all types in 1988. These are shown in Table 3 (note that Table 3 is divided into slightly different age bands from Table 2 to correspond with the format of the national accident figures being used here as baseline data).

For each type of involvement in right-turn accidents, the age distribution is expected to follow the percentages shown in Table 3. For example, 31.5% of people involved in all accidents were under 25, therefore the number of under 25’s expected in this sample of 147 ‘Colliders’ is 0.315×147 or 46.4. Table 4 shows the observed and expected frequencies for each age group and each type of involvement.

Using a Chi-square test, it was found that the frequencies in different age groups for Turners Off and Colliders did not differ from the expected figures ($\chi^2 = 3.10$ and 1.70, respectively, d.f. = 3, $p > 0.25$ for both). However, the distribution for Turning Onto accidents did differ significantly from expectation ($\chi^2 = 10.79$, d.f. = 3, $p < 0.025$). This seems to be due to the over-representation of the youngest and oldest age groups in right-turning accidents compared with the figures for accidents of all types. [Note that Table 4 represents a slightly different subset of the data from Table 2, as it only draws on cases where the Off /Onto distinction had been recorded.]

A possible reason for this over-representation of the youngest and oldest age groups is suggested by Moore et al. (1982) analysis of disobedience of junction controls, which showed both young and old drivers to be coded as ‘disobey-

Table 1
Principal characteristic of accidents making up the 'NAD/RT' database

All cases	Onto	Off	Total
	90	94	184
Location			
Urban	79	84	163
Rural	11	10	21
Day			
Weekday (Monday to Friday)	70	71	141
Weekend (Saturday/Sunday)	20	23	43
Season			
Winter (December/January/February)	26	17	43
Spring (March/April/May)	21	23	44
Summer (June/July/August)	17	27	44
Autumn (September/October/November)	26	27	53
Time			
Small hours (00:00–05:59)	3	5	9
Morning (06:00–11:59)	17	28	45
p.m. (12:00–17:59)	41	44	83
Evening/night (18:00–23:59)	29	17	46
Speed limit			
Greater than 30 mph	14	17	31
30 mph or less	76	77	153
Weather			
Fine	71	80	151
Poor	19	13	32
Unknown	0	1	1
Surface			
Dry	49	66	115
Damp/frosty	40	28	68
Unknown	1	0	1
Lighting			
Daylight	55	65	120
Dark/streetlight	35	29	64
Turner's sex			
Male	58	73	131
Female	24	19	43
Unknown	8	2	10
Severity			
Injury	32	47	79
Non-injury	58	47	105
Turner's vehicle			
Cycle	3	2	5
Moped/motorcycle	2	7	9
Car	84	77	161
Larger vehicles	1	8	9
Mean age of turners (and standard deviation)	33.4 (16.3)	35.0 (16.1)	34.3 (16.2)

Table 2
Numbers of drivers involved in accidents while making right turns, as a function of age

Age group	Frequency	Percentage	Expected percentage
Under 25	66/170	39	12
25–59	83/170	49	77
60+	21/170	12	11

Expected frequencies were calculated from mileages given in The National Travel Survey Report 1985–1986. Note that sample size varies somewhat from table to table depending on the number of cases in which the police had recorded that variable.

ing sign/signal' more often than middle-age drivers. This suggests that these drivers may be failing to stop at junctions when perhaps the situation requires it. As very few drivers in our sample were coded by the police as 'disobeying sign/signal' (only nine, with three of these being either Turners or Colliders involved in Turning Off accidents) it is not possible to do a meaningful analysis on this basis. However, we have a more sensitive measure available. The sequential data in the NAD/RT database show which Turners Onto did and did not stop before making their right turn. Table 5 shows

Table 3
Road accident involvement as a function of age in 1988: all accident types

Age	No. involved	Percentage of total involvement
Under 25	119,443	31.5
25–34	99,333	26.2
35–54	116,950	30.9
55+	43,013	11.4
Total	378,739	100

The data here have been collapsed across age categories. This is in order to keep expected values in the following analyses above 5 so the assumptions of the Chi-squared test are not violated. The total excludes 25,832 unknown values which will be ignored here.

Table 4
Numbers of drivers involved in right-turn accidents in different ways, as a function of age

	Onto	Off	Collider
Under 25	33 (23.6)	28 (28.4)	47 (46.4)
25–34	12 (19.7)	24 (23.6)	44 (38.5)
35–54	17 (23.2)	23 (27.8)	39 (45.4)
55+	13 (8.5)	15 (10.2)	17 (16.7)
Total	75 (75)	90 (90)	147 (147)

Expected frequencies, in brackets, were calculated from national figures for all accident types, see Table 3.

how many times right-turning drivers in each age group were and were not coded as stopping before they turned (again collapsing across the age groups in Table 37 of 'Road Accidents in Great Britain . . .', The Department of Transport, 1989).

The proportions of drivers not stopping are higher in the youngest and oldest age groups than in the other two. If a Chi-square analysis is done on the drivers who do stop, again calculating the expected values from national statistics, a non-significant Chi-square of 4.00 is obtained (d.f. = 3, $p > 0.25$). So the youngest and oldest age groups are not statistically significantly over-represented in accidents of this type, provided they stop before turning. Hence their over-representation seems to arise because (some) members of these groups tend not to stop when the situation requires it. In addition, it may be that older drivers take too much time to cross in front of the oncoming traffic, or younger drivers are driving faster than the speed of the traffic stream. If the problem is failure to stop, this may be for very different reasons for the two groups. A previous study based on the same cases, using Artificial Intelligence machine-learning techniques to extract predictive rules from the data (Clarke et al., 1998b), had found a situation in which drivers in right-turn accidents

Table 5
Numbers of drivers stopping or not stopping before turning, as a function of age

Age	Did not stop	Stopped	Expected to stop
Under 25	13/33 [39%]	20	(16.7)
25–34	3/12 [25%]	9	(13.9)
35–54	1/17 [6%]	16	(16.4)
55+	5/13 [38%]	8	(6.0)

Expected frequencies were calculated from national figures for all accident types, see Table 3.

almost never stopped prior to the turn, even though the general tendency was for the majority of drivers to stop, in all age groups. The relevant rule was . . .

IF	It was Dark
AND	Turner was Male
AND	Season was NOT Winter (i.e. month was March to November)
AND	Turner was under 60 years old
THEN	Turner did NOT Stop (12/13 = 92% Non-stoppers)

This throws some more light, indirectly, on the behavior of younger Male drivers, since it picks out non-elderly Males under certain circumstances as being Non-stoppers.

By contrast the main group of 'stoppers' was picked out by the following rule:

IF	It was NOT Dark
AND	Turner was NOT driving a two-wheeler
THEN	Turner did Stop (39/51 = 77% Stoppers)

The AI program we used deals in expressions called 'rules', rather as other programs might use equations. The program creates a variety of rules and selects those that discriminate most effectively between the categories in question, such as drivers who stop at a junction and those who do not. There is no threshold number of instances, or of correct predictions, for something to become a rule, although we only report rules that achieve high levels of statistical significance.

It might seem strange that drivers are less likely to stop before turning in the dark than in daylight, but several possible reasons suggest themselves: firstly that drivers may expect less traffic in darkness, and the lower traffic densities may indeed reduce the need to stop. Secondly, they may rely on approaching headlights to signal the presence of another vehicle—which is less likely. Furthermore, Hancock et al. (1991) found, in a simulator experiment of drivers' gap acceptance under different vehicle approach velocities, that the type of approaching vehicle (i.e. truck/car/motorcycle) could affect drivers' decisions to turn. It could be that headlights alone provide much less information about the type of vehicle approaching during conditions of darkness, and thus the criterion for stopping is altered.

The association between certain groups of drivers and the decision not to stop may be partly mediated by other factors. Driver groups which tend not to stop may those which drive at less congested times of day, when there seems to be less need to stop. This might include older drivers, who do most of their driving in non-rush-hour times of the day, as well as the young drivers who are on the roads more at night. After all, there is seldom a legal requirement to stop before turning—most Onto turns in the UK are governed by 'Give Way' rather than 'Stop' signs; and Off turns depend on traffic conditions alone, with no specified stopping or yielding point.

The probable importance of stopping is further underscored by the finding in Clarke et al. (1999) that the typical sequences of events in drivers' reports of safe turns (and in their 'hypothetical' junction accidents) involved the turner stopping before turning, whereas the typical sequence for real accidents did not.

Table 6
Mean ages of drivers involved in accidents while making right turns, as a function of speed limit and turn type

Type	Speed Limit	
	Low	High
Onto	34.1	34.9
Off	32.9	46.6

Another possible reason for the over-representation of older drivers in right turns Onto major roads given by Moore et al. (1982), is that older drivers have relatively more accidents on faster rural roads than on slower urban roads. Moore et al. suggest that the slower information processing of older drivers becomes critical when turning Onto a fast major road.

"... on reaching a major road at a junction, particularly at a rural junction without signals, he has to accept the speed of the traffic on the major road and the rapidity with which the whole situation changes. Nothing the older driver can do can modify it and he has to do his best to cross or merge with the traffic, but his best might not be good enough." (p. 21)

Similarly Staplin (1995) found that older drivers had a relative insensitivity to vehicle approach speed when making cross-flow turns, and they tended to depend more on the unreliable cue of vehicle distance alone. Hancock et al. (1991) had found, surprisingly, that drivers in general were more reticent to turn across slower moving vehicles than faster moving vehicles at the same gap size, but this effect might be further moderated by driver age.

So even when older drivers do stop before turning, they may still be more prone to accidents due to misjudgments. If so, there should be an interaction between the speed limit of the road and type of turn, because the information processing account suggests the speed limit of the road will make less difference to the easier kind of turn (Off a major road). A two-way analysis of variance was carried out to test this: type of turn (Onto/Off) × speed limit (high/low—above or below 30 mph). The dependent variable was the age of the Turner. The means are shown in Table 6.

No significant effect of turn-type was found, but there was a significant main effect for speed limit ($F = 6.64$, d.f. = 1, 161, $p < 0.05$) and a significant turn-type × speed-limit interaction ($F = 4.13$, d.f. = 1, 161, $p < 0.05$). Surprisingly, this went against our expectations based on Moore et al. The mean age for drivers in Off accidents on fast roads is higher than in Onto accidents, which at first glance seems to indicate that older drivers have particular difficulties turning Off these roads, not Onto them as expected. However, Table 4 shows that the oldest group of drivers are somewhat over-represented as Turners in Onto accidents more than any other group of drivers, and more than they are in Off accidents, or as Colliders. So their worst problem as a group, like young drivers, is with turns Onto.

The age of collision-involved drivers may seem like a strange choice of dependent variable in this analysis, but it is used as an index of the relative accident liability of older and younger drivers. The greater the relative liability of older drivers, the higher the mean age of the drivers who crash. This may quite plausibly be influenced by the turn-type and speed limit, whereas the drivers' ages in themselves, of course, are not.

Tables 4 and 6 can be reconciled like this. Turns Onto a main road (whether it is fast or slow) are specially problematic for the youngest and oldest driver groups. Both are more numerous than expected (Table 4) and their effects on the mean age for that type of accident cancel out. Turns Off a low-speed road do not pose a special problem for young or old drivers, so they are relatively rare, and again their effects on mean age for the accident type cancel out. But Turns Off a fast road work differently. They are a problem for the oldest but not the youngest driver group, so the number of elderly Off accidents is higher than expected although the number of young Off accidents is not. Hence the mean age for this accident type is raised because there are few young drivers to cancel out the effect of the elderly ones. Put another way, the mean age data seem to show greater problems for older drivers with turns Off than with turns Onto, merely because the mean age for drivers in the Onto accidents is lowered by the young drivers involved, while for Off accidents it is not.

3.4. Sex differences

On the whole, these data provided little evidence that Male and Female drivers have different kinds of right-turning accidents at junctions. Other techniques of analysis which had been used on the same cases—rule-finding, sequential analysis and sequential pattern detection also failed to reveal consistent differences between Male and Female Turners, although a decision-tree analysis did find some rather complex sex differences, especially involving the distinctiveness of Young Male drivers (Clarke et al., 1998b). They were very common in Onto accidents with the following characteristics:

IF	Junction had only three arms
AND	Turner did NOT wait
AND	Turner did NOT slow down prior to maneuver
THEN	Turner was Young Male (89% Young Males)

They were absent from Onto accidents at junctions with four or more arms.

The picture was less clear for the Off accidents of Young Males. However, it seemed that an accident was very unlikely to involve a Young Male driver if the three following conditions were met: firstly, if the Turner got an invitation to proceed by another road user; secondly if the accident did NOT involve a failure to notice another road user by the Turner; and thirdly if the accident did not take place in the evening (6 p.m. to 11:59 p.m.). In fact, Young Males tended to have their right turning accidents on two-wheelers or on urban roads, though our data did not show how much of this preponderance was due to exposure.

Information from the UK Transport and Road Research Laboratory's 1990 seat-belt survey (Broughton, 1990) helped

Table 7
Proportions of Female drivers in right turn accidents on weekdays and at weekends, compared with exposure data from the Transport Research Laboratory (then TRRL) survey of seat belt use by car occupants (Broughton, 1990)

	Proportion of Female drivers	
	Turners in right-turn accidents (%)	SBS survey (%)
Weekdays (08:00–18:00 h)	32.56	31.54
Weekends (08:00–18:00 h)	25.93	25.04

to confirm the view that Males and Females have similar liabilities to right turning accidents. This was also conducted in Nottighamshire, and involved stopping vehicles randomly between the hours of 8 a.m. and 6 p.m. at selected sites—mostly traffic signals. The number of Male and Female drivers found by the seat belt survey matches remarkably well with our accident figures when day of the week and time of day are taken into account, as shown above in Table 7.

Hence neither sex appears especially prone to right-turning accidents. The close match between the gender ratio in our sample and that in a large independent survey further reinforces our view that our small random sample of accidents is adequately representative for its purpose.

However, using the Test for the Significance of Difference Between Two Proportions, we did find a lower proportion of Females among Colliders than among Turners ($z = 1.99$; $p < 0.05$) and a slight tendency for Female Turners to collide with other Females, as shown in Table 8.

The first point may arise because women drive more slowly (given that high speed accidents are predominantly male) and are thus less of a hazard to other road users. They also have different exposure patterns—less highway driving for example—which also reduces their liability to high speed collisions. The second point may arise because women drive relatively more during the day. Thus, women have their accidents when there is a higher proportion of Female drivers on the road. Conversely Males drive more during the hours of darkness when there are fewer women drivers to collide with. The National Travel Survey of 1985–1986, The Department of Transport, 1987, shows that women do a significantly higher proportion of their traveling between 07:00 and 17:59 than men.)

Table 8
Numbers of Male and Female ‘Turners’ and ‘Colliders’ in right-turn accidents

Turner	Collider		Total
	Female	Male	
Female	9	33	42
Male	14	103	117
Total	23	136	159

Excluding cases where the sex of Turner and Collider were not both known.

Table 9
Gender of the driver at fault in hypothetical accidents, as a function of the gender of the informant

	Sex of driver ‘at fault’		Total
	Male	Female	
Male informant	4	3	7
Female informant	14	3	17
Total	18	6	24

The data also showed that a higher proportion of Females than Males (79% versus 53%) stopped before an Onto accident, ($\chi^2 = 3.98$, d.f. = 1, $p < 0.05$).

3.5. Attribution of blame

The informants who gave us the 100 hypothetical accidents were also asked to attribute blame to the parties involved. They were given four choices: the Turner, Collider, Both or Neither. As a result it was possible to look at the ways in which the age and sex of the informant, and the ages and sexes of the (imaginary) accident participants, affected attribution of blame.

One point that emerged clearly was a tendency to blame the younger party. We defined ‘AGEDIFF’ as the difference between the Turner’s and Collider’s age, and ‘BLAMEDIF’ as +1 if the Turner was blamed and the Collider was not, −1 if the Collider was blamed and the Turner was not, and 0 otherwise. The correlation between AGEDIFF and BLAMEDIF was −0.40 ($p < 0.001$). This relationship held for both Male and Female informants but it was not constant across all age groups in our sample. Defining ‘GUILTAGE’ as the difference between the age of the ‘blamed’ party and the ‘innocent’ party (0 when neither or both parties are blamed), we found that older informants were more likely to blame the younger party. The correlation between age of informant and GUILTAGE was −0.275 ($p < 0.01$). This negative correlation was stronger when Male informants were considered alone ($r = -0.429$; $p < 0.001$) and absent for Females alone ($r = -0.068$; $p = 0.645$).

The effect was thus mainly produced by the older Males in our sample of informants, who seemed to ‘distance themselves’ from fault by imagining a situation in which a young driver or rider caused an accident to a more mature road user.

To see whether informants also differed in how they imagined the involvement of Males and Females in accidents, we selected the 24 hypothetical cases in which (a) the Turner and Collider were of different sexes, and (b) blame was unequally allocated. These are shown in Table 9 by sex of informant. This suggests that Female informants tend to envisage accidents which are caused by a Male driver.

Overall, it seems that the public is well aware of the danger posed by Young Male drivers. It may even be that they overestimate this danger.

4. Discussion

4.1. Main findings

Young drivers are over-represented as Turners in right-turn accidents, when compared with expectations based on mileage driven. Drivers over 60, are involved about as often as one would expect. Drivers between 25 and 59 are least involved, relative to mileage. More remarkably, the youngest and oldest groups of drivers also have more right turning accidents than one would expect from their involvement in accidents in general.

The youngest drivers tend to be involved as Turners Onto in particular, rather than Turners Off or Colliders, while the oldest group has problems with turns Off, as well as turns Onto. The mean age of Turners in Off accidents is higher than in Onto accidents, but this is because young drivers are relatively scarce in Off accidents, not because older drivers find turning Off more difficult than turning Onto. Older drivers may find turning Onto fast rural roads even more difficult because their information processing is slower. However, young drivers also tend to experience some problems with this maneuver (Table 4), reflecting its difficulty and the experience required to execute it safely. Hancock et al. (1991) commented that turn decisions appeared to result from a complex interplay of rate-of-change perceptual variables such as 'time to contact', which might be a particular problem for both these driver groups, although under apparently different circumstances.

The youngest and oldest drivers were also the least likely to stop before turning. Those who do stop have only about as many accidents as baseline figures would suggest, so the higher involvement of young and old drivers as Turners may be coming from the non-stoppers in particular.

Neither sex appears especially prone to right-turning accidents, although women are under-represented as Colliders relative to mileage (perhaps because they drive more slowly and are therefore less likely to be hit by a Turner coming out of a junction). They also tend to have their right-turn collisions with other women (almost certainly because of gender-specific exposure patterns). Women drivers in right-turn accidents are more likely to stop before turning than men.

In hypothetical accidents, informants tended to blame the younger driver, increasingly so for Male informants as they got older. Female informants tended to blame the Male driver, although this was not statistically significant in so few cases.

4.2. Methodology

This is the last paper in a series of four, based on the same set of cases, so it seems appropriate to overview the findings, and the relative merits of the methods used in the four sets of analyses. In general the results of the present statistical analysis complement those of the earlier papers (Clarke et al., 1998a,b, 1999). It was the statistical approach used here which showed most clearly the special problems associated with the youngest and oldest drivers in the sample, and with turns Onto major roads in particular. Its special advantage within this project lies in the way that other bodies of data can be used to estimate exposure rates and expected accident frequencies, so that significant departures from these can be detected.

However, it was the other techniques which gave the more detailed picture of the maneuvers and errors involved. The genetic algorithm used for the first paper, for example, showed that young Turners (under 25) are especially associated with right-turn accidents:

on minor roads or after dark,
riding two-wheelers on a dry road surface,

and that the key risk factors which make for a more serious accident are:

pulling over to an outside (passing) lane just prior to the turn,
colliding with a two-wheeler or pedestrian,
failing to notice another road user when turning Off,
failing to notice another road user in poor weather when turning Onto a larger road.

In that study, there was no estimate of exposure or expectation from other data sets, but the problem of 'over-fitting' (describing anomalies in a particular set of cases which would not generalize to others) was avoided by a form of internal cross-validation using 'training' and 'test' sub-sets of cases.

The decision-tree analyses used in the second paper found rather similar groups of features in the data, but the use of tree diagrams rather than rules to express them had some advantages. They showed that Young Males have their Onto accidents at simpler junctions than other drivers, and by not waiting or slowing before the turn. They have their Off accidents most typically on two-wheelers or on urban roads, and tend not to be invited to turn by another road-user, but probably fail to notice another road-user and have the accident in the evening. The finding that certain kinds of junction accidents typically involve two-wheelers is of course unrelated to findings that two-wheeler accidents typically involve certain problems such as visibility. In these analyses over-fitting was avoided by a process called 'tree pruning'. Excess (probably unreliable) detail was removed from the trees by balancing the information-theoretical cost of additional information, against the cost in classification errors of removing information.

The sequence analysis methods used in the third paper were particularly good for highlighting the differences between real and hypothetical accidents in the NAD/RT database. When constructing hypothetical accidents, informants tended to imagine right-turn accidents in which the Turner did everything correctly, just as in an idealized safe turn, until reaching the junction, and then went wrong by failing to observe a hazard. Hypothetical accidents also tended to be over-dramatic, with a large number of risk factors added together. The real accident sequences, by contrast, were shorter and different in character, with a crucial danger point occurring just after the Turner had stopped. Again, cross validation between training and test sub-sets of the data was used to reduce over-fitting.

In general, our experience with Nottinghamshire Constabulary accident records leads us to believe that police files, although unsuitable for some kinds of detailed moment-by-moment coding, do contain useful information that is not usually exploited for the purposes of road accident research. Such records give valuable behavioral information which is lost by the time the national statistics are compiled.

One part of this approach which we would firmly recommend to other researchers is the use of a relatively homogenous class of accidents, all involving the same maneuver or 'accident mechanism'. If we had taken a random sample of accidents, as some early 'in-depth' studies did, we would have found next to nothing. Moreover, our studies suggest that the appropriate 'grain size' (or fineness of detail) for defining the sample is surprisingly small. Similar accident mechanisms only seem to occur in very tightly specified groups of cases. 'Right-turn Accidents' is probably, on reflection, too broad a category. 'Right Turns Onto', and 'Right Turns Off' are better, but for some purposes it may be necessary to select even narrower groups. In some instances, a consistent picture only began to emerge when we looked at very specific subsets from the total database (such as drivers over 55 Turning Off a fast rural road).

We hope this work will help to open up new modes of road-accident research which provide a viable compromise between traditional in-depth studies and large scale, purely statistical methods. They have the potential of being cheaper than traditional in-depth studies, but at the same time, more informative about behavioral details than conventional statistical analyses of large aggregated data sets.

Acknowledgements

This study was sponsored by the United Kingdom Department for Transport and carried out under contract to the Transport Research Laboratory, Crowthorne, England. This paper is abridged with permission from the TRL final project report CR305 "The Analysis of Pre-Accident Sequences". We are most grateful to Nottinghamshire Constabulary for their patient assistance in locating suitable cases for analysis; to other members of the Accident Research Unit, Department of Psychology, University of Nottingham, for their helpful comments and suggestions; to Chris Ashton of the Nottinghamshire County Council Accident Investigation Unit for assistance with the selection of the sample; to Graham Grayson and Geoff Maycock of the Transport Research Laboratory, for their expert guidance and advice; and to Chrysanthi Lekka for assistance with revisions to an earlier draft.

References

Broughton, J. 1990. Restraint use by car occupants 1982–1989. Technical Report no. RR289. Transport and Road Research Laboratory, Crowthorne, UK.

Clarke, D.D., Forsyth, R.S., Wright, R.L., 1998a. Behavioural factors in accidents at road junctions: the use of a genetic algorithm to extract descriptive rules from police case files. Accid. Anal. Prev. 30 (2), 223–234.

Clarke, D.D., Forsyth, R.S., Wright, R.L., 1998b. Machine learning in road accident research: decision trees describing road-accidents during cross-flow turns. Ergonomics 41 (7), 1060–1079.

Clarke, D.D., Forsyth, R.S., Wright, R.L., 1999. Junction road accidents during cross-flow turns: a sequence analysis of police case files. Accid. Anal. Prev. 31 (1–2), 31–43.

England, L., 1981. The role of accident investigation in road safety. Ergonomics 24, 409–422.

Fell, J.C., 1976. A motor vehicle accident causal system: the human element. Hum. Factors 18, 85–94.

Grayson, G.B., Hakkert, A.S., 1987. Accident analysis and conflict behaviour. In: Rothengatter, J., de Bruin, R. (Eds.), Road Traffic Safety. Van Gorcum, The Netherlands.

Hancock, P.A., Caird, J.K., Shekhar, S., 1991. Factors influencing drivers' left turn decisions. In: Proceedings of the Human Factors Society 35th Annual General Meeting (sponsored by the Bay Area Chapter), San Francisco, CA.

Massie, D.L., Campbell, K.L., Blower, D.F., 1993. Development of a collision typology for evaluation of collision avoidance strategies. Accid. Anal. Prev. 25, 241–257.

Midland, K., 1992. In-depth accident investigation teams as a tool for traffic safety. Report 135/1992. Institute of Transport Economics [TØI], Oslo, Norway.

Moore, R.L., Sedgeley, I.P., Sabey, B.E., 1982. Ages of car drivers involved in accidents, with special reference to junctions. Supplementary Report 718. Transport and Road Research Laboratory, Crowthorne, Berkshire, UK.

Rabbit, P. M. A., 1991. Factors promoting accidents by elderly drivers and pedestrians. In: Grayson, G.B., Lester, J.F. (Eds.), Behavioural Research in Road Safety. Transport and Road Research Laboratory, Crowthorne, Berkshire, UK.

Sheehy, N., Chapman, A., 1988. Reconciling witness accounts of accidents. In: Rothengatter, J.A., de Bruin, R.A. (Eds.), Road User Behaviour: Theory and Practice. Van Gorcum, The Netherlands.

Staplin, L., 1995. Simulator and field measures of driver age differences in left-turn gap judgments. Transportation Res. Rec. 1485, 49–55.

The Department of Transport, 1987. National Travel Survey Report 1985–86. HMSO, London.

The Department of Transport, 1989. Road Accidents: Great Britain 1988: The Casualty Report. HMSO, London.

Verhaegen, P., 1995. Liability of older drivers in collisions. Ergonomics 38, 499–507.

Viano, D.C., Culver, C.C., Evans, L., Frick, M., 1990. Involvement of older drivers in multivehicle side-impact crashes. Accid. Anal. Prev. 22, 177–188.

Wagenaar, W.A., Reason, J.T., 1990. Types and tokens in road accident causation. Ergonomics 33, 1365–1375.