

**TRL Limited**



**PUBLISHED PROJECT REPORT PPR056**

# **The innovatory analysis of road traffic accident data**

Version: 1

**by J P Hill (TRL Limited)**

**Prepared for: Project Record: 9/31/119 Statistical Analysis of Accident Data III**  
**Client: Road Safety Strategy Division, DfT**  
**(Dr S Marçal)**

Copyright TRL Limited July 2004

This report has been prepared for Road Safety Strategy Division, Department for Transport is unpublished and should not be referred to in any other document or publication without the permission of the DfT. The views expressed are those of the author and not necessarily those of the DfT

<b>Approvals</b>	
<b>Project Manager</b>	<input type="text"/>
<b>Quality Reviewed</b>	<input type="text"/>

This report has been produced by TRL Limited, under/as part of a Contract placed by the DfT. Any views expressed are not necessarily those of the DfT.

TRL is committed to optimising energy efficiency, reducing waste and promoting recycling and re-use. In support of these environmental goals, this report has been printed on recycled paper, comprising 100% post-consumer waste, manufactured using a TCF (totally chlorine free) process.

# CONTENTS

<b>Executive summary</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Cluster Analysis</b>	<b>1</b>
<b>3 Two-step analysis</b>	<b>2</b>
3.1 Description of technique	2
3.2 Discussion	3
3.3 Clusters identified	4
3.3.1 Cluster 1	4
3.3.2 Cluster 2	4
<b>4 CHAID Analysis</b>	<b>5</b>
4.1 Description of technique	5
4.2 Discussion	6
4.3 Gender	6
4.4 Clusters identified	7
4.4.1 Cluster 3	7
4.4.2 Cluster 4	8
<b>5 Cross-tabulation</b>	<b>8</b>
5.1 Description of technique	8
5.2 Discussion	8
5.3 Clusters identified	9
5.3.1 Cluster 5	9
5.3.2 Cluster 6	10
5.3.3 Cluster 7	10
<b>6 Sunlight</b>	<b>12</b>
6.1 Measuring the sun's position	12
6.2 Data preparation	13
6.3 Analysis	13
6.4 Manoeuvres	16
6.5 Road types and junctions	18
6.5.1 Single carriageway roads	18
6.5.2 Dual carriageway roads	19
6.5.3 Junctions	19
<b>7 Conclusions</b>	<b>19</b>
7.1 Cluster analysis	19
7.2 Solar position	20
<b>Acknowledgements</b>	<b>20</b>

<b>References</b>	<b>21</b>
<b>Appendix A. Two-step analysis</b>	<b>22</b>
<b>Appendix B. Two-step analysis output</b>	<b>23</b>
<b>Appendix C. CHAID analysis</b>	<b>26</b>
<b>Appendix D. CHAID analysis output</b>	<b>28</b>
<b>Appendix E. Cross-tabulation output</b>	<b>29</b>
<b>Appendix F. Solar Position</b>	<b>30</b>
F.1 Road Types	30
F.1.1 Single carriageway roads	30
F.1.2 Dual carriageway roads	32
F.2 Junctions	34

## Executive summary

The aim of this project was to assess the potential of innovative forms of statistical analysis to explore the STATS19 database of records of Road Traffic Accidents. The STATS19 database contains many different variables describing the circumstances surrounding every Personal Injury Accident reported by the police in Great Britain, and the vehicles and casualties involved in the accidents.

In the past these variables have tended to be analysed in relatively small groups, for example, breath test result together with age and gender. The first part of the report investigates a group of techniques called cluster analysis. Cluster analysis aims to identify homogeneous groups of accidents that may be of interest for further analysis. These groups will generally be smaller than those studied in the past, and may be defined by more variables. Because these subsets of accidents share a number of characteristics, it is likely that analysis of them will help to generate ideas for how the number of these accidents might be reduced. Ideally the technique will be objective rather than subjective, and identify groups of accidents that are reasonably consistent between years.

Three forms of cluster analysis were investigated:

- Two-step analysis, which groups records (in this case accidents or vehicles) that share some common characteristics.
- CHAID analysis, which progressively splits the entire set of records into smaller groups that have a similar target variable (in this case average accident severity).
- Cross-tabulation, which uses a table to compare the number of records that share a particular pair of characteristics to what would be expected if the characteristics were independent of each other.

The application of these techniques to STATS19 data from 1997 and 2002 has demonstrated that cluster analysis is a useful technique for studying accidents. Of the three types of cluster analysis studied, CHAID had the greatest potential. Where possible Exhaustive CHAID should be used as it is more thorough than CHAID. It is generally only necessary to run a CHAID analysis once on any set of accidents to obtain useful results. The results can be interpreted objectively and there is little scope for researchers to introduce bias. The clusters are produced by a transparent process that can be interpreted without any formal statistical training. There is also a clear measure of which clusters are of the greatest interest, in this case accident severity.

Cross-tabulation produced some interesting clusters. It is particularly useful if specialist software is not available and a quick analysis is desired. However, it can only be used for simple analyses of two variables. The clusters produced tended to include a relatively large number of vehicles.

Two-step analysis was of much more limited use. Although it did produce some interesting clusters it required a large amount of input from the researcher. This turned it into a far more subjective process than was desirable. The process by which the clusters had been produced was not transparent – it was very difficult to see how some of the clusters had been formed. It was also unclear which of the clusters were the most interesting. It is unlikely that there will be much scope for using two-step analysis in the future on Stats19 accident data.

The second part of the report investigates the effect that the position of the sun has on accidents. It investigates the relative propensity for drivers to be involved in an accident according to the position of the sun relative to the driver at the time of the accident. Strong evidence is found which suggests that the position of the sun has an effect on the occurrence of accidents. 6% more vehicles were involved in accidents when the sun was in the 45 degree segment of sky directly behind the driver (and above the horizon) than the 45 degree segment in front of the driver.

Vehicles that were ‘waiting to go ahead’ were 17% more likely to have the sun behind them than in front; vehicles that were ‘stopping’ were 33% more likely to have the sun behind them. This suggests that drivers may take longer to notice slowing vehicles when the sun is behind them.

The difference between the number of vehicles that had the sun behind them and the number that had the sun in front of them was much more pronounced on dual carriageway roads than single

carriageway roads. The sun appeared to have only a very small effect on accidents at junctions. On single-carriageway roads, the sun seemed to have a stronger effect on two-vehicle accidents than on single-vehicle accidents. This suggests that the effect may relate to the observation of other vehicles, perhaps including their signals and brake lights.

There is great scope for further research on this topic. Further analysis would look at the reasons why the position of the sun causes problems. Track trials could be used to test whether drivers were misled by 'false signals' when the lights on another vehicle catch the sunlight or are masked by it. When the reasons for the extra accident risk are better understood, consideration could be given to mitigation measures. Areas that could be considered would be the lighting on vehicles, such as the use of third brake lights, and mechanisms for reducing the glare in vehicle mirrors.

Analysis to determine at which elevations the sun has the greatest effect would also be conducted. The effect that the position of the sun has on pedestrians, particularly at signal controlled crossings, could also be studied.

## 1 Introduction

The aim of this project is to assess the potential of innovative forms of statistical analysis for exploring the STATS19 database of records of Road Traffic Accidents. This database contains many different variables describing the circumstances surrounding every Personal Injury Accident reported by the police in Great Britain, and the vehicles and casualties involved in those accidents. In the past, these variables have tended to be analysed in relatively small groups, for example, breath test result together with age and gender. Some variables have not been studied in great depth before. Analysis of the data is usually based on all accidents for a given period or a relatively large subset of them, such as all single vehicle accidents. This may result in several very different types of accident being examined at the same time.

Sections 2 to 5 of this report discuss ways of identifying homogeneous groups of accidents that may be of interest for further analysis. These groups will generally be smaller than those studied in the past, and may be defined by more variables. Because these subsets of accidents share a number of characteristics, it is likely that analysis of them will generate ideas for how the number of these accidents might be reduced.

Section 6 of the report looks at a very different topic - the effect that the position of the sun has on accidents. It investigates the relative propensity for drivers to be involved in an accident according to the elevation of the sun and its angle relative to the driver at the time of the accident.

## 2 Cluster Analysis

The first part of this report describes various forms of cluster analysis that were carried out on STATS19 data. Clusters are groups of accidents defined by a combination of variables that occur more frequently than would be expected by chance. Cluster analysis refers to a range of statistical techniques that can be used to identify these homogeneous subsets of data. By using these techniques to identify clusters it is hoped that some groups of accidents will be found that are of interest but have not been identified before. Cluster analysis removes some of the potential for researchers to introduce bias when selecting 'interesting' sets of accidents for analysis; in principle it is an objective process albeit data dependent. Accident groups identified may be ones that would not have been selected for study if cluster analysis were not used. This innovative approach should result in the exploration of aspects of accident data that would otherwise be overlooked.

The three forms of cluster analysis used in this report were two-step analysis, CHAID analysis, and cross-tabulation. Each analysis technique is discussed in the following sections.

Once a cluster had been identified further analysis was done to identify any other characteristics that accidents in the cluster might share and how the size or nature of the cluster had changed over time.

The cluster analysis was initially carried out on STATS19 data for 2002. Ideally, the clusters identified would be reasonably stable features of the accident data, rather than statistical artefacts that existed only in one year's data. Consequently, once the clusters had been identified in the 2002 data, the corresponding records were also extracted from the 1997 database and compared with the 2002 clusters.

When conducting cluster analysis it is important to be able to select the records and fields that are wanted for analysis simply and quickly. A Microsoft Access database application was designed that allowed users with no knowledge of Access to select one or more fields from the 1997 or 2002 STATS19 data and automatically import them into SPSS for further analysis. These fields could relate to a combination of accidents, vehicles and casualties. In addition a number of extra fields were added including markers for pedestrian accidents, vehicles that struck a pedestrian and accidents that involved a child.

Although the database did allow users to specify very complex combinations of variables it did not run as fast as was desired. The greater the number of variables selected the longer it took to extract the relevant data. This was due to the very large number of records, particularly when using both vehicle and casualty data. Therefore, it was necessary to be selective when choosing variables for analysis.

The database was designed so that when the data were imported into SPSS the appropriate value and variable labels were also imported. Without this feature a large amount of time would have been spent recoding the data in SPSS. SPSS syntax was used to automatically mark missing variables so they would be excluded from analysis where necessary.

### 3 Two-step analysis

#### 3.1 Description of technique

Two-step cluster analysis was carried out using the statistical software package SPSS. Further details of two-step analysis are given in Appendix A. Two-step analysis identifies groups of records with similar characteristics. In this case the records relate to accidents, vehicles or casualties. Two-step analysis can deal with both continuous variables, such as age, and categorical variables, such as weather. The user specifies the records and variables that should be considered for clustering; for example, the user might choose to analyse all single vehicle accidents and base the clustering on carriageway type, lighting, and junction detail.

In the first of the two steps SPSS takes each record in turn and decides whether it should be merged with a previously formed sub-cluster or, if it is sufficiently different from the existing sub-clusters, placed into a new sub-cluster. This process is carried out for each of the records. In the second step these sub-clusters are compared with each other and, where they have similar characteristics, are grouped appropriately. This two-step approach saves processing power compared to an analogous one-step approach, but this is at the expense of some accuracy.

Once the analysis has run various tables and charts are displayed. An example of the outputs is shown in Appendix B. It should be noted that the outputs shown are for a very simple analysis based on cars and taxis involved in accidents on non built-up roads. Only carriageway type and speed limit were used in the analysis; the analyses used for the report produced a far greater number of tables and charts. The tables show the number of cases in each cluster and the number of times that each variable occurs in each cluster. Pie charts and bar charts show the same information graphically. Before running the analysis the user can also choose to display a set of charts that shows the statistical contribution that each variable has when defining the clusters. Once these outputs have been studied the user can remove variables from the model that do not seem to have a significant effect on the clustering process and re-run the analysis. This process continues until a coherent set of clusters has been produced.

There are a number of settings in SPSS that can be adjusted to either reduce the time required to run the analysis or improve the coherence of the results. SPSS has a default limit on the number of sub-clusters that may be generated in the first step. Increasing this has the potential to improve accuracy significantly, particularly when the data being analysed are very diverse. However, this results in a large increase in the time required for the analysis to run.

SPSS uses an algorithm to automatically determine the final number of clusters that it should produce. It is possible to specify the number of clusters that should be produced before running the analysis. This saves processing power, as SPSS does not have to compare the merits of different numbers of clusters. It can also be used to force the generation of more clusters in a situation where SPSS has decided that there are only two. However, use of this facility is not advisable except in rare circumstances as it can lead to disparate data being inappropriately grouped.

### 3.2 Discussion

Two-step analysis was performed on a mix of accident and vehicle variables. From the outset it was decided to base most of the analyses on cars and taxis involved in accidents. Previous research has shown that different vehicle types tend to be involved in different types of accident. Including all of the 16 vehicle types recorded in STATS19 would have significantly increased the noise in the data. Some separate analyses were carried out on heavy goods vehicles and on two-wheeled motor vehicles (TWMVs). Pedestrian accidents were also analysed as a separate group; this required the generation of an extra field in STATS19 accident table to indicate whether or not the accident involved a pedestrian.

Most of the accident and vehicle variables in STATS19 were examined although it was not possible to analyse them simultaneously due to computational limitations. A few variables were not included in any of the analyses. These included the time and date, the local authority and police force, and the road number. Other variables were included at first but dropped after preliminary tests as they appeared to have little scope to influence clustering. This included variables such as 'Carriageway Hazards' which takes a value of 'none' in the vast majority of cases.

It was found from the start that two-step analysis did not generate many clusters and that the clusters that were produced were far more general than was desired. In some cases only two clusters were generated, one consisting of vehicles involved in accidents on built-up roads, the other consisting of vehicles in accidents on non built-up roads. To reduce the generality of the clusters four dichotomous variables were created to split the data into 16 clearly defined subsets. These specified whether the accident involved one or two vehicles, if the first road recorded was an 'A' road or a minor road, if the road was built-up or non built-up (BU/NBU), and whether or not the accident occurred at a junction. Accidents with three or more vehicles and accidents occurring on motorways were not included in this analysis.

Analysis was carried out individually on each of these 16 subsets. Subdividing the data forced SPSS to select variables other than BU/NBU and road class when generating the clusters. As a result the clusters were more specific and other variables such as manoeuvre, junction type and carriageway type were used to define the clusters produced. Splitting the data before analysis does in some ways go against the principle that the analyst should not have to make potentially bias-inducing decisions. However, it was clear that without this guidance two-step analysis would have been of little benefit.

Two-step analysis has a number of weaknesses which are summarised below and discussed in greater depth in Appendix A.

- The two-step method is highly data dependant. A small change in the input data can result in a very large change in the clusters that are produced.
- The result of the analysis depends on the order in which the records are analysed.
- It was often difficult to identify the decision process that had led to a cluster being created. Even the clusters that did seem to have a clear definition to them often had one or two extra records that had been allocated to them for no obvious reason.
- There is no single statistical measure of a cluster's importance with respect to the other clusters.

Although the two-step method did produce some interesting clusters, these tended to be larger and more general than was desirable. The clusters rarely made use of more than two or three of the variables. The process also involved a lot of human intervention when specifying the subsets of data and when interpreting the clusters that were found.

It is unlikely that two-step analysis would be of great practical benefit for accident analysis.

### 3.3 Clusters identified

A large number of clusters were produced using the two-step analysis. As explained above it was unclear what the distinguishing characteristics of many of the clusters were. Other clusters were too general to merit further analysis. Two clusters found by the two-step method were selected for further analysis. They were based on a suitably large number of variables and had not been studied in detail during predecessors of the current project.

#### 3.3.1 Cluster 1

- Accidents where a Goods Vehicle (>3.5 tonnes) jack-knifed and/or overturned.

**Table 3.1 Vehicles and casualties in Cluster 1**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	762	0.32%	1062	0.32%
2002	758	0.34%	1049	0.35%

This cluster was generated using two-step analysis on the set of Heavy Goods Vehicles (>7.5 tonnes) that were involved in accidents. However, prior to the quinquennial review of STATS19 in 1999 there was no distinction between goods vehicles of 3.5 to 7.5 tonnes and those over 7.5 tonnes. Therefore the cluster was adjusted to include all goods vehicles greater than 3.5 tonnes for both the 1997 and 2002 data.

The cluster was selected for further analysis because it was more coherent than many of the other clusters. It is possible that a similar set of vehicles might have been chosen for analysis if cluster analysis had not been used. However, it is likely that a set selected by a researcher would also have included HGVs that skidded (and did not overturn or jack-knife).

Only 11% of these accidents occurred at the weekend. This cluster had the highest proportion of accidents occurring between 8pm and 6am. Both these results are representative of the traffic patterns of goods vehicles. 37% of the accidents occurred when the road surface was not dry. 14% occurred in high winds, compared to between 2 and 4% of most other clusters. This was a very significant increase on 1997 when less than 10% of these accidents occurred in high winds. The vast majority of accidents occurred on roads with a 60 or 70 mph speed limit.

Very few (11%) drivers of goods vehicles were aged less than 25 or over 60. Only 2% were female, but this was a statistically significant increase on 1997 when 0.8% were female. The low number of younger and older drivers and women reflects the population of lorry drivers. Only 0.9% failed (or refused to provide) a breath test, less than half the proportion in 1997.

#### 3.3.2 Cluster 2

- Single vehicle accidents involving at least one pedestrian on a minor road (i.e. B, C or U) not at a crossing. The accident occurred in daylight hours and street lighting was present. The speed limit was 40mph or less.

**Table 3.2 Vehicles and casualties in Cluster 2**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	14465	6.02%	16165	4.93%
2002	12123	5.47%	13399	4.43%

Cluster 2 was found when conducting two-step analysis on single vehicle accidents that involved a pedestrian, on any road other than a motorway. This cluster was the largest of the nine found and included over one-third of the relevant accidents.

The cluster includes accidents where a pedestrian was struck as they attempted to cross the road and those where a vehicle lost control and struck a pedestrian on the footpath.

Only 1% percent of these accidents were fatal, perhaps surprisingly low for pedestrian accidents. 1.5% of the accidents occurred in a 40mph zone. Less than 1% of the drivers failed or refused to provide a breath test. Almost 15% of drivers hit and ran. This relatively high percentage will be due in part to the low level of damage to the vehicle involved.

## 4 CHAID Analysis

### 4.1 Description of technique

CHAID (Chi-squared Automatic Interaction Detector) analysis was carried out using SPSS's Answer Tree software. CHAID analysis uses a very different classification algorithm from two-step analysis. It aims to identify groups of records that are similar according to a specified target variable. Both the target variable and the predictor variables can be continuous (e.g. age), ordinal (e.g. speed limit) or categorical (e.g. weather). Starting with the complete dataset CHAID splits the data into two or more 'child' nodes using just one of the predictor variables. The predictor variable is chosen using chi-squared statistics to identify the optimal split. Each predictor goes through a process of merging response-similar categories so that the final set of categories will be optimally distinct on the response variable. Each of the resultant nodes is then analysed individually to determine whether it should be split and if so how. This process continues until no more splits can be made. This may happen because there are no more statistically significant splits to be made or because a stopping rule has been reached. Stopping rules are used to prevent the data being split into very small clusters that, despite being measurable different from neighbouring clusters, are too small to be of interest. They limit the minimum number of cases that can be in a child node and the minimum number of cases that can be in a parent node. Although Answer Tree has default values for the stopping rules these need to be adjusted according to the size of the dataset entered and how the analysis is being used. Nodes that cannot be split are called terminal nodes.

An alternative to CHAID in Answer Tree is Exhaustive CHAID. When trying to identify the optimal split for a node CHAID does not investigate every possibility; therefore, the true optimal split may not be identified. Exhaustive CHAID is more thorough when identifying the optimal split for a node. It does take longer to run than CHAID but with the data analysed for this project the difference was not substantial. Other than the possibility of extra time Exhaustive CHAID has no disadvantages when compared to CHAID. Any future analyses would almost certainly be done using Exhaustive CHAID.

## 4.2 Discussion

An example of a tree produced by CHAID is given in Appendix D. The example given is a trivial one based on only two predictor variables. Many of the trees produced for the proper analysis took up over 20 sheets of A4 paper, even when reduced to the smallest clearly readable font.

The accident clusters that were considered to be of greatest interest were those that had a relatively high severity. Therefore the target variable used for most of the CHAID analysis was accident severity. Because accident severity is highly dependent on vehicle type only car and taxis involved in accidents and their drivers were considered in the main analyses. It is important to note that the overall accident severity was used – not the severity of injury sustained by occupants of the car or taxi – and that other vehicle types may have been involved in the accident. In the first model run accident severity was defined as a nominal variable. In this case Answer Tree would have no concept that a fatal accident is less similar to a slight accident than a serious accident. The model was therefore re-run with accident severity specified as an ordinal variable i.e. a fatal accident is worse than a serious accident, which is worse than a slight accident. This made some minor changes to the tree but the overall structure looked fairly similar.

Further discussion on CHAID analysis is contained in Appendix C and is summarised below.

- In general the splitting of categories that had been merged at an earlier stage in the analysis should be allowed.
- The user can choose whether to specify a variable as nominal, ordinal or continuous - choosing carefully can improve the quality of results markedly.
- Exhaustive CHAID is more thorough than standard CHAID. As time penalties are relatively small it should be used where possible.
- Manually splitting a node can be a very useful technique. It can be used to ensure that non-ordered variables such as lighting are grouped appropriately. It can also be used to split nodes that would otherwise be terminal due to a stopping rule being reached
- Analysis of the 1997 data produced similar results to the analysis of the 2002 data. The first two levels of the tree were almost identical, but there were subtle differences further down the tree. This in part will be due to random variation in the data but also differences in STATS19 coding.

CHAID is highly intuitive and easy to interpret and the basis on which decisions have been made is very clear. The time taken for one run is typically around one minute and it is not necessary to make many adjustments to the analysis in order to get satisfactory results. Caution is needed when dealing with missing and unknown variables as these can get grouped inappropriately. However this can be dealt with by splitting a node manually or filtering the data in advance. Exhaustive CHAID is preferable to CHAID although there may be some situations in which it takes much longer to run.

It may be necessary to make some changes to the stopping rules. Where there is a very large amount of data the minimum cluster size needs to be increased to prevent over-splitting of the data. Similarly when the dataset is very small minimum cluster sizes should be reduced. It must be remembered that no split will be made if the statistical significance criteria are not met. In the models run for this report it was found that the minimum size stopping criteria were usually met before the statistical significance of the split became an issue. There is some subjectivity when selecting which clusters to study but this can be avoided by simply selecting the clusters with the highest proportion of fatal accidents.

## 4.3 Gender

A further set of CHAID analysis was run using male/female as the target variable. This would indicate if there are any accident types that might be regarded as particularly ‘masculine’ or ‘feminine’. Only car drivers who provided a negative breath test and whose age and gender were

known were included in the analysis. These conditions were used as there is a strong bias towards non-car drivers and drivers who provide a positive breath test being male. This would have made it less likely that previously unknown patterns would be found.

The most significant predictor for a driver's gender was their age. Almost 74% of drivers aged 19 or under were male. This proportion fell steadily to 63% for drivers aged 36 to 39 before rising again to 76% for drivers aged over 60. This result may be the result of young male drivers having a higher propensity to take risks than their female counterparts and there being more male driver over 60 than female drivers. Several clusters indicated that male drivers were more likely to have accidents on roads with higher speed limits and in the dark. This may simply reflect driving patterns or there may be some other reason for these differences.

#### 4.4 Clusters identified

CHAID analysis produced over 100 clusters. Most of these are of less interest as they are of a low or average severity, while those with high severity are of real interest. Two clusters of particular interest were identified using CHAID analysis and are examined in more detail below. Both clusters are considerably more sophisticated than those identified using Two-Step analysis. The cluster definitions are presented in the order in which the tree split, i.e. the most important variables appear first.

##### 4.4.1 Cluster 3

- Cars and taxis involved in accidents in the daylight with street lighting present. The first point of impact was the front and it was not at or near a junction. The speed limit was 30 mph or less and the driver was aged 50 or greater

**Table 4.1 Vehicles and casualties in Cluster 3**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	2981	1.24%	3976	1.21%
2002	3030	1.37%	4025	1.33%

This cluster was found using non-exhaustive CHAID on cars and taxis involved in accidents. The cluster may include a range of different accident types. Some accidents will have occurred when an older driver failed to respond to a vehicle ahead slowing down or stopped. Other accidents may include those where an older driver lost control of the vehicle and struck an object off the carriageway or collided head-on with another vehicle.

The severity of these accidents reduced slightly between 1997 and 2002. There was no change in distribution by either day of week or time of day. Accidents were slightly more likely to have occurred on dry roads in 2002 but there was no change in the distribution of weather conditions. There was no change in the age distribution of drivers (although it must be remembered that this cluster only includes drivers aged 50 or over). There was a slight increase in the proportion of drivers that were female. This change was seen in all the clusters except some of those that involved TWMVs. This may reflect increased mileage driven by women since 1997. In common with all other clusters there was an increase in the proportion of drivers who recorded a negative breath test. This was largely due to a decrease in the number of drivers who were not tested for some reason, presumably the result of a stricter testing policy by the police. There was a small increase in drivers who hit-and-ran.

#### 4.4.2 Cluster 4

- Motorcycle accidents occurring on 60 mph roads, not at a junction, where the first point of impact was the front and the engine capacity was greater than 125cc.

**Table 4.2 Vehicles and casualties in Cluster 4**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	1516	0.63%	2101	0.64%
2002	1386	0.63%	1883	0.62%

The second cluster was found using Exhaustive CHAID analysis on motorcycles. The split by engine size was selected manually as the stopping rules had already been reached because the parent node was too small. The tree generated for motorcycle accidents was shallower than that for cars mainly because there were fewer of them so the stopping rules were reached sooner. The 1997 STATS19 form did not make a distinction between different sizes of motorcycle (although it did distinguish between motor-scooters, mopeds and motorcycles). For the 1997 data 'motorcycle' was used in place of 'motorcycle (>125cc)'. This will mean that the results for the two years are not directly comparable.

As described in section 2 it was not possible to extract all the variables that might be of interest for analysis in one go. One variable that was not selected was manoeuvre. It is possible that had it been used it would have been one of the most important clustering variables.

This cluster had by far the highest accident severity of those analysed in this report. Over 13% of the accidents were fatal and 45% were serious. This is significantly higher than the cluster analysed for 1997 when 9% were fatal and 43% serious, although this difference is probably due to the inclusion of small motorcycles in the 1997 cluster. In 2002 44% of the accidents occurred at the weekend. As with cluster 6 it is likely that many of the motorcycle riders were leisure cyclists. Only 13% of accidents occurred when the road surface was damp and only 7% happened when the weather was not fine.

30% of the riders were aged 40-59 in 2002 compared to 18% in 1997. This may represent an increase in the number of older people buying powerful motorcycles but may be an artefact of the different vehicle types being compared in the two clusters. Only 3% of the riders were female.

## 5 Cross-tabulation

### 5.1 Description of technique

Cross-tabulation is perhaps the most basic method of performing a cluster analysis. Two variables are plotted against each other to show how many cases share those two specific characteristics. Chi-squared analysis can then be used to show whether particular pairings of variables occur more frequently than would be expected if they were independent. This can be done using standard spreadsheet software such as Excel.

### 5.2 Discussion

Cross-tabulation was used to identify whether particular pairings of vehicle manoeuvre occurred more frequently than might otherwise be expected. The STATS19 data were filtered for accidents involving two vehicles, both of which were cars, taxis or minibuses. The two vehicle manoeuvres

were tabulated with one vehicle manoeuvre appearing in the columns and the other in the rows. The lower value of vehicle manoeuvre (as listed in STATS19, variable 2.7) always appeared in the row heading. This meant that each manoeuvre pair appeared only once. For example the two accidents shown below are considered to be identical.

Vehicle 1 - Going Ahead Other                      Vehicle 2 - Stopping  
 Vehicle 1 - Stopping                                  Vehicle 2 - Going Ahead Other

The resultant table only contains data in the upper right hand half of the table (see Appendix E). The cells with the largest number of accidents in them are those of greatest interest. However, many of the largest clusters may be intuitively obvious. For example, the largest cluster involves two vehicles that were 'Going Ahead Other'. This cluster is too general to be of interest.

A slightly more sophisticated approach involves using chi-squared analysis to identify those pairs of manoeuvre that appear more often than would be expected if the two manoeuvres occurred independently. The expected number of accidents that would occur for each pair of manoeuvres is calculated. The difference between this and the observed value is squared and then divided by the expected value. This gives a chi-squared value for each cell. Cells with the largest chi-values are those which are considered to be most different from what might be expected and are of greatest interest.

A second cross-tabulation was performed on accidents involving one car, taxi or minibus, together with a two-wheeled motor vehicle. In this case, which of the vehicles is performing which manoeuvre is of relevance.

Cross-tabulation analysis is very quick to do and easy to understand. It can be done with the spreadsheet software that appears on any computer. The clusters produced are very simple and may relate to a large number of accidents. However, this analysis has been limited to two-dimensions so it is unlikely to produce particularly detailed clusters. It would be possible to use multi-level contingency table analysis to handle more than two variables; however, this is more complex and has not been investigated as yet.

### 5.3 Clusters identified

Three clusters were identified and examined more closely using cross-tabulation analysis.

#### 5.3.1 Cluster 5

- Accidents involving two cars (or taxis or minibuses); one going ahead on a left-hand bend (LHB), the other going ahead on a right-hand bend (RHB).

**Table 5.1 Vehicles and casualties in Cluster 5**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	3798	1.58%	7154	2.18%
2002	3355	1.51%	6209	2.05%

Most of these accidents involve two vehicles travelling in opposite directions along the same road. It is possible that in some cases the two vehicles do not impact because one has taken avoiding action.

There was little evidence of a change in the characteristics of these accidents between 1997 and 2002. However, there was an increase in the proportion of accidents that occurred on 40 or 50 mph roads;

this is probably a result of the tendency to reduce speed limits on rural roads. 58% of the accidents occurred on wet or damp roads, a much higher proportion than for any other cluster.

The two vehicles were analysed separately according to whether they were on the RHB or the LHB. and strong differences were found between the characteristics of the two vehicles. The vehicle that was travelling on the LHB was more likely to skid (38.5% compared to 22.3% for vehicles on the RHB in 2002). Their drivers tended to be several years younger and were slightly more likely to be male. They were also more likely to give a positive breath test and over twice as likely to hit-and-run. The characteristics exhibited by the LHB drivers are all associated with being at higher risk. This gives circumstantial evidence that the LHB drivers may be more likely to be at fault. It may be that drivers on the LHB are going wide and colliding with drivers on the RHB who are staying on the correct side of the road. This evidence suggests that at bends with high accident rates it may be worth targeting drivers about the risks on left hand bends, especially in wet conditions.

### 5.3.2 Cluster 6

- Accidents where a car (or taxi or minibus) is going ahead on a bend and a two wheeled motor vehicle is on the opposite bend.

**Table 5.2 Vehicles and casualties in Cluster 6**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	559	0.23%	725	0.22%
2002	593	0.27%	774	0.26%

This cluster is analogous to cluster 5 but involved a car (or taxi or minibus) and a TWMV. Of the 593 accidents in 2002, the motorcycle was on the LHB in 447 (75%). It is not known where these accidents occurred in relation to the bend. This would give greater understanding of which vehicle (if either) had crossed into the opposite lane and might be considered to be at fault.

Almost 8% of the accidents resulted in a fatality compared to 3% of accidents in cluster 5. Over 40% of the accidents occurred at the weekend suggesting that a high proportion of them involve leisure motorcyclists: only 11% occurred in darkness. Whereas the road surface was wet or damp for 25% of the cluster 6 accidents compared to 58% of the cluster 5 accidents. This may seem surprising given that the two clusters are identical, except for the vehicle types involved. This is likely to reflect the tendency for many motorcyclists to go out in better weather conditions and may also relate to differences in the handling of cars and TWMVs, i.e. if a TWMV brakes in a corner it is likely to go straight on instead of continuing round the bend.

The speed limit of the road on which the accident occurred was generally lower in 2002 than in 1997. As with cluster 5, this may be due to reduced speed limits.

94% of the riders of TWMV were male. A large proportion (59%) of the TWMV riders were not breath-tested, mainly because of their medical condition (they tended to suffer more serious injuries). The average age of the rider of the TWMV was 32 in 2002 compared to 29 in 1997. The average age of the car driver was 44. This was considerably older than those car drivers involved in cluster 5. It is not clear why this particular cluster should involve older drivers but it may be related to the higher speed of motorcycles meaning there is less time for an oncoming driver to react.

### 5.3.3 Cluster 7

- Accidents where a car (or taxi or minibus) turns right and a two wheeled motor vehicle is carrying out any manoeuvre.

**Table 5.3 Number of vehicles and casualties in Cluster 7**

Year	Accidents		Casualties	
	Number	% of total	Number	% of total
1997	5899	2.45%	6800	2.07%
2002	6979	3.15%	7841	2.59%

Cluster 7 includes accidents where the car was turning off the main road and those where the car was turning onto the main road. The cross-tabulation technique does not help to distinguish between these two types of accident. Although a cluster could have been formed by choosing a specific manoeuvre for the TWMV it was unclear how accurately STATS19 records distinguish between motorcyclists travelling in the stream of other vehicles and those that travel alongside and progressively overtake other vehicles.

In two-thirds of these accidents the manoeuvre of the TWMV was recorded as 'going ahead – other'. Another 15% of the TWMVs were overtaking a moving vehicle on its offside and 8% were overtaking a stationary vehicle on its offside.

Only 1% of the accidents in this cluster were fatal, a surprisingly low proportion for accidents that involved TWMVs. This cluster includes accidents that occurred on both built-up and non built-up roads. It is likely that the severity of accidents on non built-up roads is higher. The accidents were more likely to occur on a weekday than accidents in the other clusters that involved TWMVs. A relatively high proportion of the accidents (56%) occurred between 6am and 10 am or 4pm and 8pm. It is likely that for the majority of these riders a TWMV is a mode of transport for commuting rather than a leisure vehicle.

The proportion of the accidents that occurred in rain or when the road was damp increased significantly between 1997 and 2002. It is possible that TWMVs are being used more in wet weather conditions, whereas they used to be used only if the weather was good. Over 80% of the accidents occurred on roads with a speed limit of 30mph or less compared to 77% in 1997.

Over 36% of the car drivers were female, the highest proportion for any of the clusters. 93% of the TWMV riders were male, slightly less than for the other clusters that involved TWMVs. Only 1% of the TWMV riders hit-and-ran compared to 6% of the other drivers. This will be mainly due to the more serious injuries caused to the riders.

## 6 Sunlight

The second part of this report describes an innovative study into the effect of the position of the sun, relative to the driver, on the occurrence of accidents. It is based on the 2002 STATS19 data for the whole of Great Britain.

It was thought that the position of the sun relative to a driver might have an effect on their propensity to be involved in an accident. A driver heading into the sun might be dazzled by the sun making it harder for them to see the road and other vehicles. A driver changing lanes may fail to see another vehicle approaching as it is obscured by the sun. Earlier research has looked into the differences between accident rates in daylight and darkness and the effect that daylight saving time has on accidents; however, no research was identified that looked at how the position of the sun relative to the driver affected accidents.

This report describes an analysis of the effect that the sun's position has on the propensity for vehicles to be involved in accidents.

### 6.1 Measuring the sun's position

The sun's position is made up of two components – its elevation with respect to the notional horizon, and the angle between the sun and the north (the azimuth). These can be calculated using the grid reference, date and time of an accident, all of which appear in STATS19 data. Allowance must be made for the change between Greenwich Mean Time and British Summer Time.

The elevation of the sun is measured relative to the theoretical horizon. In the middle of summer the elevation of the sun varies between -16 degrees at night and 64 degrees during the day; in mid-winter the elevation varies between -64 degrees at night and 16 degrees during the day. The sun's elevation changes most rapidly around 6 a.m. and 6 p.m. (GMT) at all times of year. A small inaccuracy in the time of an accident around these times could lead to a substantial error in its calculated elevation. This could result in accidents where the sun was below the horizon being counted as though it was above the horizon and *vice versa*. The driver of a vehicle going over the crest of a hill will perceive the sun to be at different elevation from a driver on a level road.

The sun's elevation is related to many factors: it is determined by the time of day, date and location. The impact of the sun on accidents is also related to the weather; for example, in Britain it is more likely to be misty in the morning than in the afternoon, and in the summer it is more likely to rain in the afternoon than in the morning. It would require in-depth analysis to identify to what extent an increase in accidents around 6pm was due to low sun, fatigue or the levels of rush hour traffic. These complications mean that it would not be possible within the scope of this project to completely investigate the effect of the sun's elevation on the occurrence of accidents.

The azimuth of the sun depends almost entirely on the time of day (having adjusted for BST if necessary) and location, and is not substantially affected by the date. At noon (GMT) at Greenwich the azimuth is 180 degrees, i.e. the sun is due south: at 6 a.m. it is 90 degrees, due east; at 6 p.m. it is 270 degrees, due west; and at midnight it is 0 degrees, due north (but below the horizon). The azimuth changes at a similar rate throughout the day. An error in the time of the accident of 20 minutes would result in only a 5 degree error in the azimuth.

Analysing the effect of the azimuth of the sun relative to the driver is simpler. Although the azimuth of the sun at a given time and location is fixed, the angle relative to a driver depends on the direction in which they are travelling. At any given time a similar number of vehicles is likely to be travelling to and from each of the eight compass points. Therefore, if the angle of the sun does not have an effect on accidents, the angle of the sun relative to the accident involved vehicles should be uniformly distributed, between 0 and 360 degrees.

## 6.2 Data preparation

The position of the sun for each accident was calculated using a VBA macro (Pelletier, 2002). The azimuth and elevation of the sun were added to each STATS19 accident record for 2002. The elevation field was then added to each vehicle record without any need for adjustment. The azimuth was used in conjunction with the vehicle movement fields to calculate the direction of the sun relative to the driver of each vehicle. The direction of travel of each vehicle is only approximate as only eight directions are available on the STATS19 form. There is a strong bias to the four cardinal points when recording the to and from directions - over 58% of vehicle directions were recorded as travelling to or from the north, east, south or west. Two direction fields were generated for each vehicle, one relating to 'vehicle movement to' and the other to 'vehicle movement from'. For example, a vehicle travelling straight ahead with the sun directly in front would have 0 degrees recorded for both fields. A vehicle that started with the sun directly behind it and then turned left would have a relative direction of 180 degrees for the 'from' field and 270 degrees for the 'to' field (i.e. the sun was now on its left).

Since only eight values are available for the movement to and from fields it was necessary to group the vehicles by the relative position of the sun. Vehicles were put into eight groups, each representing a 45 degree segment. The first group included vehicles where the relative direction of the sun was between approximately -22.5 (equivalent to 337.5 degrees) and 22.5 degrees. This group included vehicles that were travelling almost directly into the sun and was labelled as the 0 degree group. The second group included those vehicles with the sun at between 22.5 and 67.5 degrees and was labelled as the 45 degree group. These had the sun ahead and to the right of them. The other groups were formed in a similar way.

## 6.3 Analysis

Analysis is based mainly on the 'vehicle movement to' field. This gives the direction in which the vehicle was travelling *after* the manoeuvre. This appears to be of greater interest than the 'vehicle movement from' field as it reflects where the driver was likely to be focussing most of his or her attention.

In each table the first row of data gives the total number of vehicles in the sample. The following rows give the ratio of vehicles observed in that 45 degree sector relative to the mean. A ratio greater than one means that more vehicles were observed in that sector than average.

Chi-squared analysis was used to compare the number of vehicles in each of the 45-degree sectors to the number that would be expected if the vehicles were evenly distributed with respect to the sun's position. A p-value of less than 0.05 is taken to mean that there is a statistically significant difference between the distribution of the recorded accidents and what would be expected if the sun had no effect. A value of less than 0.01 is regarded as providing strong statistical evidence and a value of less than 0.001 is considered to be very strong statistical evidence.

The back/front value is calculated by dividing the number of vehicles in the 180° group by the number in the 0° group. It provides a comparison of the effect of the sun when it is directly behind the vehicle and when it is directly in front of the vehicle.

The first analysis was based on vehicles involved in injury accidents during 2002 and is summarised in Table 6.1. It included all vehicle and road types. Vehicles whose manoeuvre was recorded as parked, reversing or missing were excluded. Vehicles whose movement to or from was recorded as '0' were also excluded as the relative position of the sun could not be calculated.

The analysis looked at the effect of the position of the sun at a range of different elevations. The first two columns in Table 6.1 compare the effect of the sun at negative and positive elevations; subsequent columns compare the effect of the sun at smaller ranges of positive elevations.

**Table 6.1 All vehicles, all road types**  
Excluding parked, reversing and missing manoeuvres

Sun angle (degrees)	All	Negative solar elevation	Positive solar elevation	Solar elevation 0-9°	Solar elevation 10-19°	Solar elevation 0-19°	Solar elevation 20-39°	Solar elevation >= 40°
Total	384582	122500	262082	37757	61338	99095	91144	71843
0	0.990	0.984	0.993	0.932	1.032	0.994	1.01	0.96
45	0.981	0.963	0.989	1.048	0.959	0.993	0.96	1.02
90	0.995	1.000	0.993	0.933	1.018	0.986	1.02	0.97
135	1.001	0.988	1.006	1.034	0.968	0.993	0.98	1.06
180	1.033	0.989	1.053	0.952	1.088	1.036	1.10	1.02
225	1.011	1.009	1.012	1.095	0.988	1.029	0.99	1.02
270	1.013	1.052	0.995	0.977	1.027	1.008	1.03	0.94
315	0.977	1.016	0.959	1.029	0.919	0.961	0.92	1.01
p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
back/front	1.043	1.005	1.061	1.022	1.054	1.042	1.083	1.057

The chi-squared analysis showed that there was very strong evidence of a difference in the number of vehicles involved in accidents depending on the location of the sun. However, it should be noted that due to the large number of vehicles in the analysis even a relatively small difference in percentage terms can still produce a highly significant result in a chi-squared test. Therefore it is necessary to look at the magnitude and pattern of the ratios when interpreting the results.

When the elevation of the sun was positive accidents were most likely to occur when the vehicle was being driven in the segment directly *away* from the sun rather than into the sun. Almost 2,000 more vehicles were involved in accidents while driving directly away from the sun than directly into the sun, a difference of almost 6%.

When the analysis was restricted to accidents where the sun's elevation was between 0 and 9 degrees there appeared to be a pattern where the number of vehicles per sector alternated between being above and below average. When the elevation was between 10 and 19 degrees this pattern was reversed. This is an artefact of the bias towards the four cardinal points when recording vehicle movements. When a specific range of solar elevations are considered the sun will tend to lie within a fairly narrow band of azimuths. Combined with the bias in recording vehicle movement this will cause a bias towards particular a range of relative azimuths. This effect will not be apparent when a wide range of elevations is considered or if sectors are grouped together as in Table 6.2.

A comparison of vehicles by whether the sun was in the 180 degree section in front of them or in the 180 degree section behind them and to the left or right of them produced some very interesting results.

**Table 6.2 All vehicles, all road types, grouped into 180° sections**  
Excluding parked, reversing and missing manoeuvres

Solar position	Negative elevation	Positive elevation	Elevation 0-9°	Elevation 10-19°	Elevation 0-19°	Elevation 20-39°	Elevation >= 40°
Total	122500	262082	37757	61338	99095	91144	71843
Front	1.000	0.982	0.990	0.984	0.986	0.977	0.984
Back	1.000	1.018	1.010	1.016	1.014	1.023	1.016
p-value	>0.05	<0.001	>0.05	<0.001	<0.001	<0.001	<0.001
Left	1.018	0.996	1.012	0.997	1.003	0.997	0.987
Right	0.982	1.004	0.988	1.003	0.997	1.003	1.013
p-value	<0.001	>0.05	<0.05	>0.05	>0.05	>0.05	<0.001

When the sun was below the horizon there was no evidence of any difference between the number of vehicles with the sun ahead and the number with the sun behind (which was expected as the sun is below the horizon). However, there was a significant difference (3.7%) between the number of vehicles that had the sun to the left and the number that had the sun to the right of them. This effect was diminished when only elevations of less than -20 degrees were considered. The sun may still have some effect when it is just below the horizon and there may also be some inaccuracies in the calculated elevation due to imprecision when recording the time of an accident.

During the daytime there was a significant difference (3.6%) between the number of vehicles with the sun in front of them and the number with the sun behind them. There was no statistically significant difference between the number with the sun to the left of them and the sun to the right of them. These findings were true for all positive solar elevations although there did seem to be a significant difference between the number of vehicles with the sun to the left and right of them when the sun was above 40 degrees.

An analysis was performed on cars and taxis whose manoeuvre was not recorded as parked, reversing or missing, and whose movement to and from fields were not zero. The sun may affect some vehicle types, such as lorries and vans, differently from cars. Their higher position on the road may mean that the sun is less likely to obscure their view of other vehicles in their wing mirrors. The absence of a rear-windscreen in many larger vehicles may reduce the danger of them being dazzled by the sun in their rear-view mirror.

**Table 6.3 Cars and taxis**  
Excluding parked, reversing and missing manoeuvres

Sun angle (degrees)	All	Positive solar elevation	Positive solar elevation, fine weather	Positive solar elevation, not fine weather
Total	294319	197397	154163	43234
0	0.990	0.993	0.999	0.972
45	0.979	0.988	0.985	0.997
90	0.998	0.996	0.990	1.014
135	1.001	1.007	1.005	1.012
180	1.031	1.051	1.055	1.039
225	1.009	1.012	1.017	0.995
270	1.015	0.994	0.991	1.008
315	0.977	0.958	0.957	0.963
p-value	<0.001	<0.001	<0.001	<0.01
back/front	1.041	1.059	1.056	1.070

The results were again highly significant. When only accidents where the solar elevation was positive were considered 6% more cars and taxis were involved in accidents when driving from the sun than into the sun.

It was thought that the effect of the sun might depend on the weather; in overcast and rainy conditions the sun is unlikely to affect a driver's vision. It would be desirable to compare accidents that occurred in sunny conditions with those occurring in other conditions. Unfortunately the weather field in STATS19 does not have a code for sunny. The standard code that is used for any weather condition that did not have an adverse effect on driving is 'fine without high winds'; this applied to over 75% of vehicles in 2002. Other weather categories include raining, snowing (both of which are divided into with or without high winds) and 'fog or mist – if hazard'. Many of the accidents coded as 'fine' will have occurred when the weather was somewhere between slightly cloudy and overcast. Conversely, during a light shower the sun may still be visible but the weather would still be recorded as 'raining'. A comparison of fine weather accidents with other accidents showed that there were statistically significant differences in their distributions. Accidents that occurred when the weather was not fine were more evenly distributed - more accidents were recorded with the sun to the side of the vehicle than in fine weather. However the position of the sun still seemed to have a substantial effect. In part this may be related to non-fine weather being associated with damp roads, which can cause problems with glare.

It is possible that if a code for sunny were available then a better comparison of different weather conditions could be made. It was decided that further analysis would be based on all weather conditions. This will ensure the results are representative of the weather encountered on Britain's roads. However, it is also likely to lead to a slight underestimation of the effect of the position of the sun on accidents.

#### 6.4 Manoeuvres

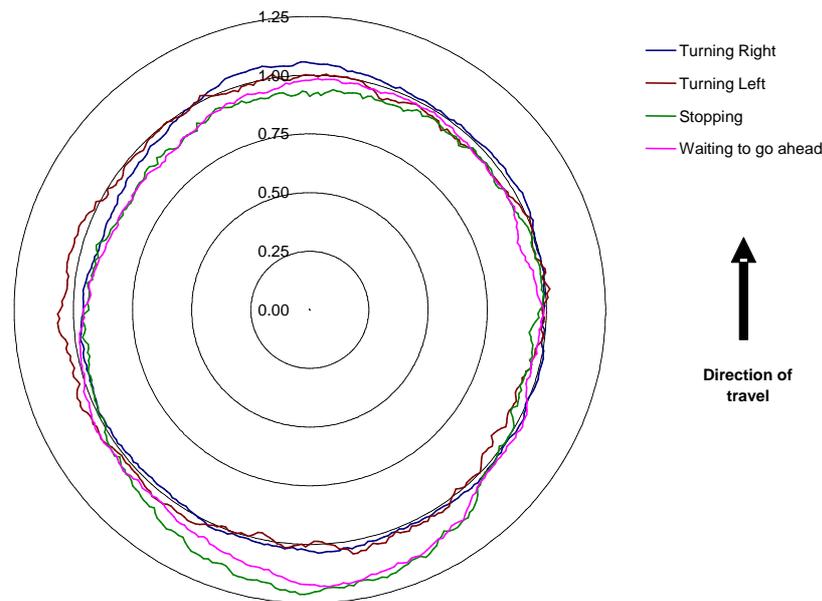
An analysis of vehicle manoeuvres was based on the position of the sun relative to the vehicle movement from field. This is more intuitive than using the vehicle movement to field as the resulting graph (see Figure 6.1) gives a snapshot view of the scene prior to the accident. There was statistically

significant evidence that the position of the sun relative to vehicles that had been going ahead – other was not evenly distributed. However, only 1% more vehicles had the sun directly in front of them than behind. A far more striking result was that of vehicles waiting to go ahead and those that were stopping. These were 17% and 33% more likely respectively to have the sun behind them than in front. This suggests that when the sun is behind the driver they fail to respond in time to vehicles slowing or stopped ahead, possibly because they failed to see the brake lights. It may be that when the sun is to the front of the vehicle drivers adapt and drive more cautiously. When the sun is behind the driver it may take them by surprise.

**Table 6.4 Cars and taxis, positive solar elevation**  
Excluding parked, reversing and missing manoeuvres

<b>Sun angle before manoeuvre (degrees)</b>	<b>All</b>	<b>Going ahead - other</b>	<b>Waiting to go ahead</b>	<b>Stopping</b>	<b>Turning right</b>	<b>Turning left</b>
Total	197397	97875	21127	10225	22598	6425
0	1.009	1.023	0.987	0.910	1.052	1.000
45	0.999	1.000	0.973	0.976	1.009	0.970
90	0.986	0.988	0.997	0.973	0.988	0.981
135	1.008	0.992	1.033	1.025	1.018	0.994
180	1.047	1.034	1.152	1.208	1.028	0.999
225	1.004	1.005	1.001	1.060	0.982	1.017
270	0.989	0.995	0.968	0.933	0.958	1.055
315	0.959	0.964	0.889	0.915	0.965	0.985
p-value	<0.001	<0.001	<0.001	<0.001	<0.01	>0.05
back/front	1.038	1.011	1.167	1.328	0.977	0.999

There was some evidence to suggest that the position of the sun had an effect of the occurrence of right-turn accidents. There was a larger than expected number of accidents when the sun was directly to the front to the turning vehicle, perhaps because of a failure to see oncoming vehicles. They were slightly more likely to occur with the sun on the right-hand side of the vehicle, although this particular distinction was not statistically significant ( $p=0.08$ ). A greater number of vehicles turning left were turning into the sun than away from the sun but, due to the relatively small numbers, there was no statistically significant difference. Vehicles turning left are generally faced with fewer potential conflicts making a temporary problem with visibility less of a hazard.



**Figure 6.1 Position of the sun relative to vehicle *before* manoeuvre, 45 degree moving average**

Note: A value >1.00 denotes more accidents than expected, and a value <1.00 denotes fewer.  
The manoeuvre 'going ahead – other' is not shown

## 6.5 Road types and junctions

Accidents were split into those that occurred on single-carriageway roads and those that occurred on dual carriageway roads. Accidents that occurred on a roundabout, one-way street or where the road type was unknown were not included in either category. The results are summarised below. More details can be found in Appendix F.

### 6.5.1 Single carriageway roads

Accident involved vehicles on single carriageway roads were 5% more likely to have been travelling away from the sun than into the sun. When only non-junction accidents were considered this figure rose to 8%. It is possible that drivers of vehicles at junctions have fewer difficulties with the position of the sun: if they have to give way they will have the opportunity to pause if their visibility is compromised. Vehicles continuing on the main road will have less time to adjust if they are suddenly dazzled by the sun.

There was no evidence that the position of the sun had a bearing on the occurrence of single-vehicle, non-junction accidents on single carriageway roads. However, there was strong evidence that two-vehicle, non-junction accidents were more likely to involve vehicles travelling away from the sun. This suggests that the main difficulty caused by the location of the sun is the observation of other vehicles, rather than the road.

The position of the sun did not have a significant effect on accidents occurring on non-built up roads (speed limit >40mph). However, there was strong evidence that the sun had an effect on accidents on built-up roads.

### 6.5.2 *Dual carriageway roads*

An analysis of vehicles travelling on dual carriageway roads showed that the position of the sun had a stronger effect on the occurrence of accidents than on single carriageway roads. 13% more vehicles were involved in accidents when heading away from the sun than into the sun. This contrasts with a 5% difference on single carriageway roads. There are several factors that might explain this:

- Many accidents on single carriageway roads involve two vehicles travelling in opposite directions. If one vehicle is heading into the sun, the other must be heading away from the sun. This will reduce the measurable effect of the sun's position. Conversely, accidents on dual carriageway roads will usually affect only one carriageway.
- On a dual carriageway road many accidents may involve one of the vehicles changing lane, into the path of another vehicle. When the sun is behind the driver it can be very difficult to see vehicles approaching from behind and judge their speed.
- It can also be difficult to make out the signals and brake lights of vehicles ahead as the sun may cause the lights to appear to be on when they are not and mask the lights when they are on.

The effect of the sun was even more pronounced when only non-junction accidents were considered. Over 20% more accidents occurred with the sun behind the vehicle than in front of the vehicle.

When only dual carriageways were considered it was found that the sun had a far greater effect on non-built up roads than on built up roads. This finding is the reverse of that found on single carriageway roads. However, the results for both road types were highly statistically significant. It is not obvious what the reason for these seemingly contradictory findings is.

### 6.5.3 *Junctions*

The sun had only a slightly significant effect on the occurrence of accidents at junctions despite a large sample size. The sun was behind the driver in only 3% more accidents than it was in front. It was thought that the position of the sun could pose particular problems at signalised junctions. The sun may make it difficult to see whether the signals are red or green; there may also be a problem with phantom signals caused by the sun reflecting off the lamps. However, no statistically significant difference was found ( $p=0.07$ ). A further analysis of accidents at signalised junctions when the sun was low in the sky (an elevation of between zero and 20 degrees) did not find a significant difference either.

## 7 **Conclusions**

### 7.1 **Cluster analysis**

The analyses reported above have demonstrated that cluster analysis is a useful technique for studying accidents. Of the three types studied, CHAID had the greatest potential as the results can be interpreted objectively and there is little scope for researchers to subconsciously introduce bias. The clusters are produced by a transparent process that can be interpreted without any formal statistical training. It is generally only necessary to run a CHAID analysis once on any set of accidents to obtain useful results. There is also a clear measure of which clusters are of the greatest interest, accident severity. Where possible Exhaustive CHAID should be used as it is more thorough than CHAID.

Cross-tabulation produced some interesting clusters. It is particularly useful if specialist software is not available and a quick analysis is required. However, it can only be used for simple analyses of two variables. The clusters produced tend to be relatively large which may mean the accidents in them do not have as much in common as those in clusters produced by CHAID analysis.

Two-step analysis is of much more limited use. Although it did produce some interesting clusters, it required a large amount of input from the analyst, which caused it to be far more subjective than was

desirable. The process by which the clusters had been produced was not transparent – it was very difficult to see how some of the clusters had been formed. It was also unclear which of the clusters were the most interesting. There is unlikely will be much scope for using two-step analysis in future accident research.

It is likely that further use of CHAID analysis on all Great Britain personal injury accident data would produce other interesting clusters that could be investigated in more detail. The technique will also be very useful when analysing smaller groups of accidents for specific projects. However, it should be used alongside traditional techniques, rather than replacing them.

## 7.2 Solar position

There appears to be very strong evidence that the position of the sun has an effect on the occurrence of accidents. 6% more accidents occurred when the sun was in the 45 degree segment of sky directly behind the driver (and above the horizon) than the 45 degree segment in front of the driver.

Vehicles that were ‘waiting to go ahead’ were 17% more likely to have the sun behind them than in front; vehicles that were ‘stopping’ were 33% more likely to have the sun behind them. This suggests that drivers may take longer to notice slowing vehicles when the sun is behind them.

On single-carriageway roads, not at a junction, 1% more single vehicle accidents had the sun behind the vehicle than in front. When two-vehicle accidents were considered the difference was 5%. The greater effect that the sun has on two-vehicle accidents suggests that the apparent effect may relate to the observation of other vehicles, perhaps including their signals and brake lights.

The effect of the sun’s position was greater on dual-carriageway roads and also away from junctions. On non-built up dual carriageway roads, away from junctions, the sun was 25% more likely to be behind a car or taxi driver than in front. The effect was less (only 5%) on built up dual carriageway roads. However, the converse was true on single carriageway roads. On non-built up roads only 2% more accidents occurred with the sun behind than in front: on built up roads the difference was almost 12%.

There is great scope for further research on this topic. A more sophisticated study would verify that the effect of the sun was not an artefact of the traffic flow in Great Britain at different times of day or the recording mechanism in STATS19.

Reasons for the differences between the effect of the sun on built-up and non built-up dual carriageway and single carriageway roads would also be investigated. The analysis of several years’ data would enable more thorough analysis of less common accident types (such as accidents at signalised junctions) to be investigated. Analysis to determine at which elevations the sun has the greatest effect would also be conducted.

The effect that the position of the sun has on pedestrians, particularly at signal controlled crossings, could also be studied.

Track trials could be used to determine some of the reasons that the sun causes problems. Tests could be carried out to assess whether drivers were misled by ‘false signals’ when the lights on another vehicle catch the sunlight or are masked by it.

When the reasons for the extra accident risk are better understood, consideration could be given to mitigation measures. Areas that could be considered would be the lighting on vehicles, such as the use of third brake lights, and mechanisms for reducing the glare in vehicle mirrors.

## Acknowledgements

The work described in this report was carried out in the Safety Group of TRL Limited. The author is grateful to Gavin Buckle for carrying out the analysis of the clusters and to Barry Sexton who carried out the quality review and auditing of this report. The author is also grateful to Greg Pelletier whose

VBA script for determining the elevation and azimuth of the sun at a given time and location was used.

## References

**Pelletier G** *Sunrise, sunset, solar position, and solar radiation calculators for Excel*  
<http://users.vei.net/pelican/sunrise.html>

## Appendix A. Two-step analysis

The two-step method is highly data dependant. A small change in the input data can result in a very large change in the clusters that are produced. This can be attributed largely to the varied nature of the data that are being analysed. A good example of this effect was the analysis of the 3,124 fatal accident records for 2002. These accidents were not divided into the 16 subsets as there are relatively few to start with. 13 accident variables were entered into the model and SPSS produced five accident clusters. One particular variable, 'pedestrian crossing – human control', did not seem to have any influence on the clustering of accidents. All but ten of the accident records were coded in the single category of 'no crossing facility or a physical crossing with no human control'; these ten accidents were distributed across four of the five clusters. As the variable was not significant it was removed from the model in line with standard statistical procedure. Instead of the clusters remaining fundamentally unchanged, as might be expected, the new model contained only two clusters. This instability in the model arises from the large number of variables and relatively small number of accidents analysed. Each accident can be placed into a cell that corresponds to a specific combination of the 13 variables analysed. Theoretically there are over 2.6 billion possible combinations of variable in total. It is unlikely that any of the 3,124 accidents will share all of its characteristics with another accident. This sparsity of data relative to the number of possible cells makes the model very sensitive to even the smallest change in data. Removing the 'pedestrian crossing - human control' variable subtly changed the shape of the data such that when the model was re-run only two clusters were produced. The model would be more stable if more data were analysed or fewer variables were used. However, since one of the key aims of this research was to identify previously unstudied clusters, reducing the number of variables used was not an ideal option.

Another weakness in the two-step method is that the analysis depends on the order in which the records are analysed. The first of the two steps involves comparing each record to those previously allocated to a sub-cluster. A change in the order in which records are presented, particularly at the start of the process, may result in different sub-clusters being produced. These may then be grouped into different clusters. This effect is likely to be more pronounced when SPSS has been restricted to a relatively small number of sub-clusters. In larger data sets the effect is unlikely to be very pronounced although the effect of changing the order of data was not tested for this report.

It was often difficult to identify the decision process that had led to a cluster being created. One cluster always consisted of all the records that didn't fit into other clusters. Even the clusters that did seem to have a clear definition to them often had one or two extra records that had been allocated to them for no obvious reason. The most important variable in a cluster was sometimes slightly counter-intuitive, although perfectly logical when considered in the context of the two-step algorithm. For example, instead of a cluster consisting of all of the non-junction accidents it may include all the accidents where the junction control is not known (either because there was no junction or the STATS19 form was not fully completed). Two-step analysis also has no way of identifying which variables one might reasonably expect to appear in the same cluster. Hence one cluster may contain only accidents that occurred either in the dark with street lighting or in the daylight where street lighting was not present. It is therefore necessary to check each cluster individually to ensure that it makes sense.

There is no single statistical measure of a cluster's importance with respect to the other clusters. The graphs produced by SPSS do give some indication of which clusters are more compact than others. However, it is not possible to tell without a close look at the data which of the clusters is the most important. This introduces subjectivity into the process, something that cluster analysis is intended to avoid.

## Appendix B. Two-step analysis output

### TwoStep Cluster

Cluster Distribution

	N	% of Combined	% of Total
Cluster 1	5779	17.7%	17.7%
Cluster 2	26798	82.3%	82.3%
Combined	32577	100.0%	100.0%
Total	32577		100.0%

### Cluster Profiles - Frequencies

Single or Dual

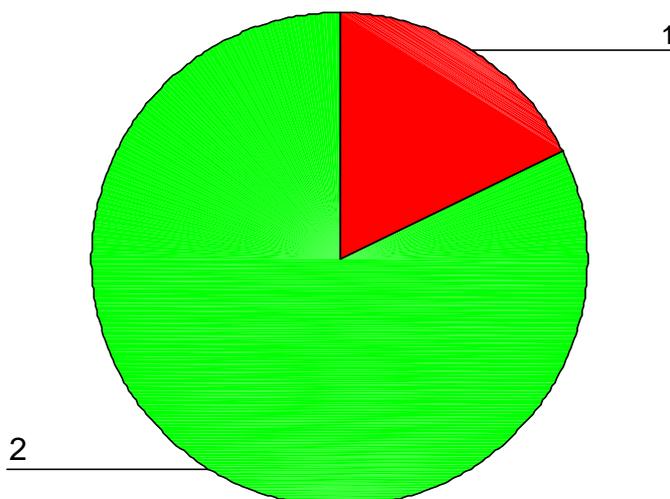
	Single		Dual	
	Frequency	Percent	Frequency	Percent
Cluster 1	2747	9.3%	3032	100.0%
Cluster 2	26798	90.7%	0	.0%
Combined	29545	100.0%	3032	100.0%

Speed Limit

	50		60		70	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Cluster 1	3688	100.0%	336	1.2%	1755	100.0%
Cluster 2	0	.0%	26798	98.8%	0	.0%
Combined	3688	100.0%	27134	100.0%	1755	100.0%

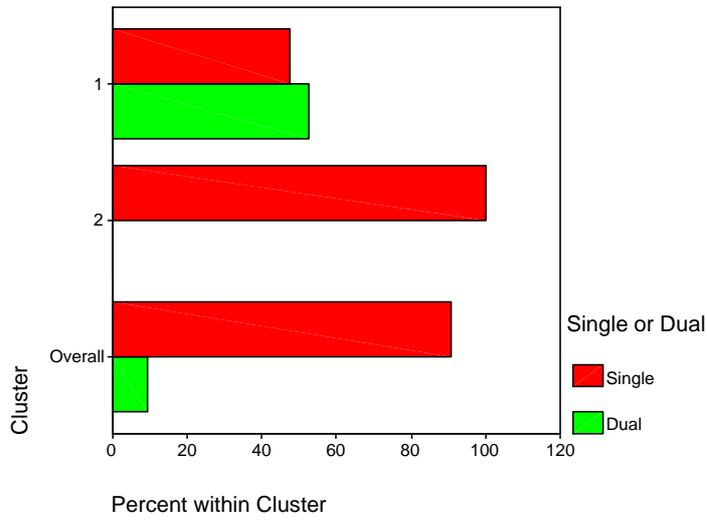
### Attribute Importance

Cluster Size

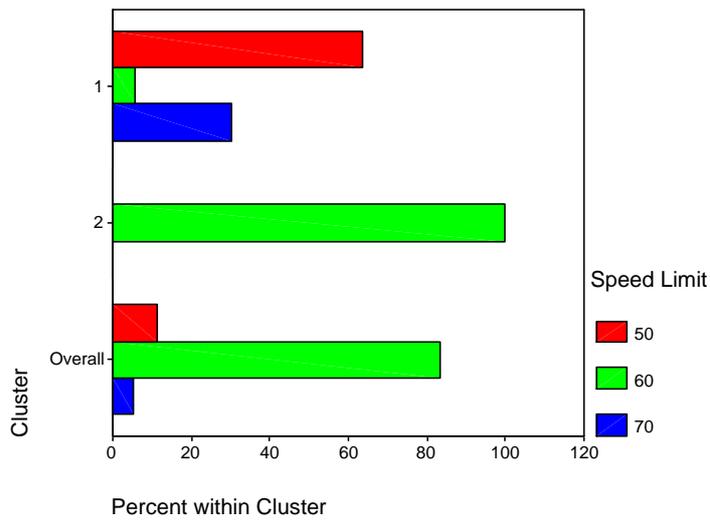


### Within Cluster Percentage

Within Cluster Percentage of Single or Dual



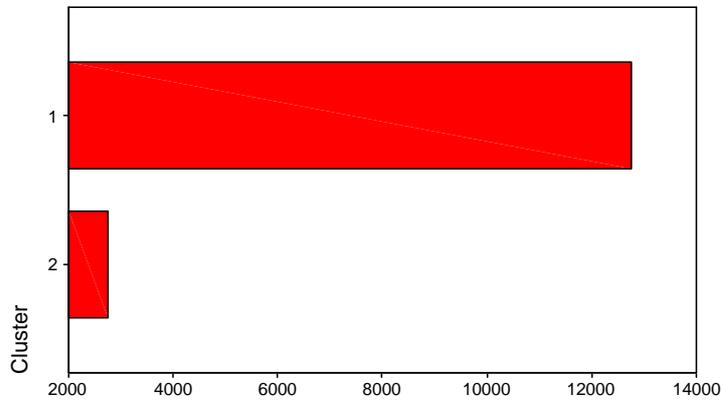
Within Cluster Percentage of Speed Limit



## Clusterwise Importance

Single or Dual

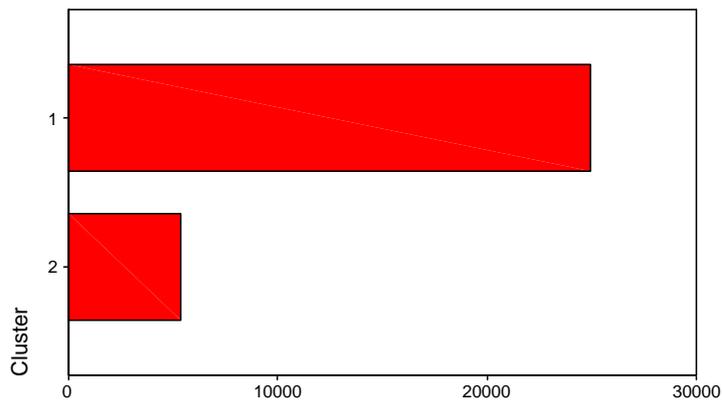
Bonferroni Adjustment Applied



Chi-Square

Speed Limit

Bonferroni Adjustment Applied



Chi-Square

## Appendix C. CHAID analysis

An option in CHAID is whether or not to allow the splitting of categories that had been merged at an earlier stage in the analysis. It is possible for two variables to be grouped together at one stage in the analysis only to be split later. This will most frequently happen with variables that can take many different values such as age. Although by default Answer Tree does not permit this there is no reason in this analysis why it should not be allowed. However, when the analysis was re-run with this option selected it was very similar to the original run.

Initially road class was included as a nominal variable. One particular node was split into four child nodes, namely Motorways & B roads, A(M) roads, A roads, and C & Unclassified roads. There is no reason why C & Unclassified roads should not be grouped together but it is clearly illogical for Motorways to be grouped with B roads, even if they do have similar accident severity ratios. The model was re-run with road class entered as an ordinal variable. The order used was that in STATS19 where Motorways are coded as '1', A(M) roads as '2' and so on down to Unclassified roads, which are coded as '6'. With an ordinal variable Answer Tree will only group adjacent categories preventing the illogical combination above.

This model was then re-run using Exhaustive CHAID. As described in the main text Exhaustive CHAID is more thorough when deciding how to split the tree. As a result it tends to produce a larger number of branches from each node. It also has the effect of making the tree shallower because the stopping rules are reached sooner. Since it was not substantially slower than CHAID all further runs were done using Exhaustive CHAID.

Answer Tree gives the option of manually selecting how a particular node should be split. As discussed above, road class was grouped illogically when it was entered as a categorical variable. In that case there was the option of specifying that road class was ordinal. In some situations the variable is categorical and definitely not ordinal, yet is subject to the possibility of illogical grouping. For example, Answer Tree might conclude that accidents occurring in daylight on lit roads were similar to those occurring in darkness on unlit roads. In this situation it is desirable to either choose another variable for splitting or to force the illogical combinations apart. For example one could instruct Answer Tree to split the accidents by daylight/darkness, or according to the presence or absence of lighting. Answer Tree allows this to be done very easily and will show what the second best split is if the first is not appropriate. Manually splitting a given node has no effect on how adjacent nodes or subsequent nodes are split.

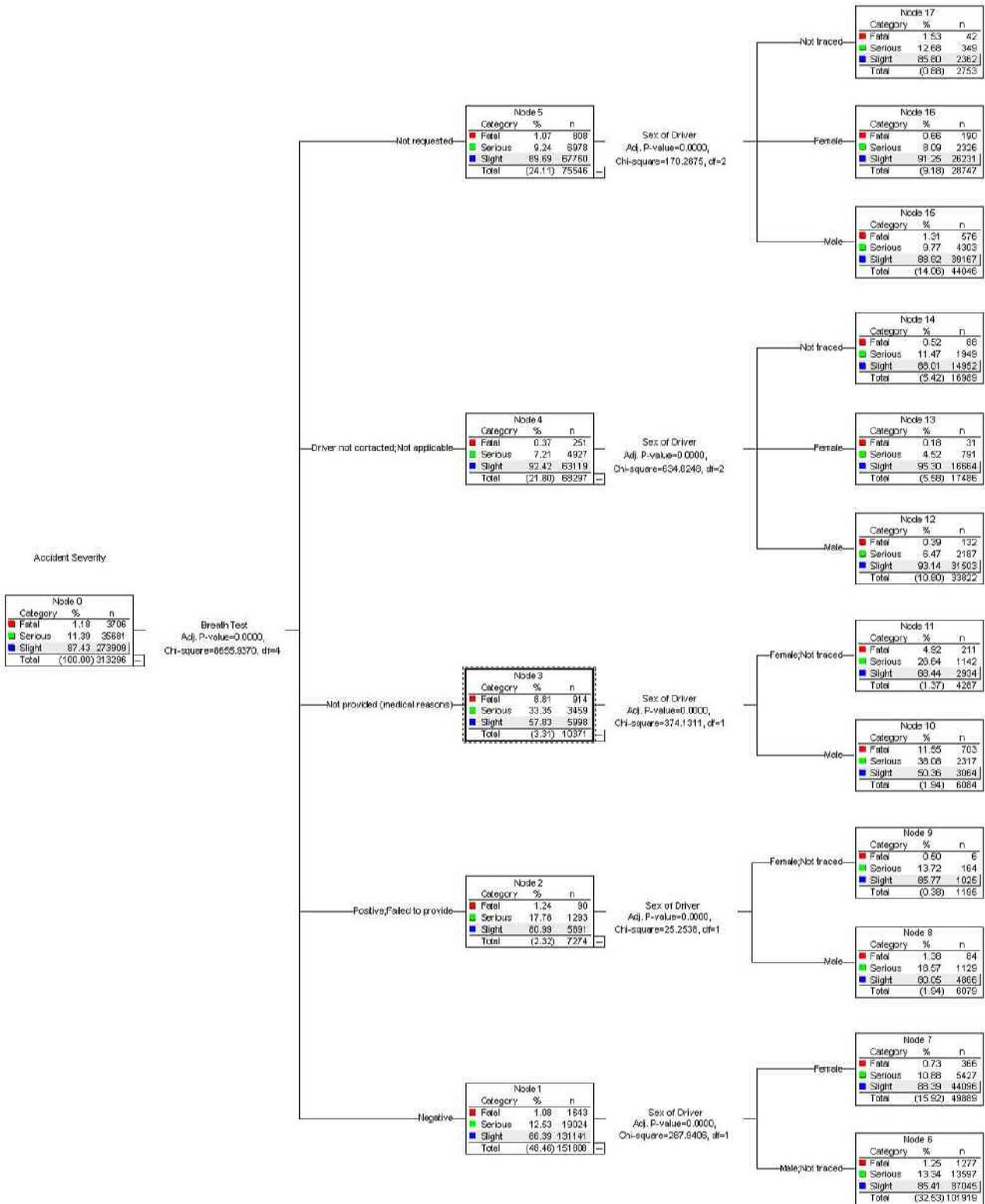
Another situation in which manual splitting is useful is when investigating how a terminal node might be split if the rules on parent and child sizes were relaxed. This facility was used when producing a cluster of motorcycle accidents. The minimum size for a parent node had been set to 2,000. One of the terminal nodes contained 1,623 motorcycle records. 12% of the records related to fatal accidents and a further 43% related to serious accidents. If the stopping rule had not been enacted the next split would have been between motorcycles with engine less than 125cc and those with larger engines. Of the 237 smaller-engined motorcycles 5% were involved in fatal accidents and 33% were involved in serious accidents; 13% of the 1,386 larger-engined bikes were involved in fatal accidents and 45% were involved in serious accidents. Since the accidents involving the larger bikes were of a much higher severity they are of greater interest. This split was still highly statistically significant so the node was split manually. The stopping rules are used largely as a matter of convenience, to prevent nodes being split into such small groups that they are not of interest. Forcing a split in this situation does not raise the same issues as splitting a node simply because you are not satisfied with how it has been done by default.

A desirable feature of any clustering technique is that two analyses of equivalent datasets from different years yield similar results. This would demonstrate that the clusters found were not short-term statistical quirks. To test this an analysis of cars and taxis involved in accidents was carried out on the STATS19 data from 1997. The analysis was an exact replication of the final analysis run for the 2002 data. The analyses produced very similar results. In both the 1997 and 2002 data the first variable to be used for splitting the tree was lighting; seven nodes, one for each category of lighting, were produced in both analyses. The second variable used for splitting the seven nodes was the same in the 1997 and the 2002 data for all but one of the nodes. That node included accidents occurring in the daylight where the street lighting was unknown. This node was split by speed limit in 1997 but by junction type in 2002. However, there are several reasons why some changes in the split should be expected. The number of

'daylight, street lighting unknown' accidents was much smaller in 2002 than in 1997. This suggests that the quality of information collected about this aspect of accidents has improved (ideally, 'street lighting unknown' wouldn't be used at all). Some junction categories were removed and there are also many more roads with 20 mph speed limits than there used to be. The combination of these factors will have led to the minor differences between the trees.

A comparison of the terminal nodes from 1997 shows that although many are similar to those found in 2002, few are identical. Subtle differences in the way in which junction types have been grouped are common. Where age, a continuous variable, was used it was almost inevitable that the precise age range used for grouping changed. As discussed earlier most of the terminal nodes were stopped because they were too small, not because no further significant splits could be made. As a comparison the 1997 data were re-analysed with a minimum parent node size of two and a minimum child node size of one. This analysis produced over 500 terminal nodes whereas the main analysis of 1997 data resulted in only 138 terminal nodes. Many of the terminal nodes that could not be split consisted only of slight accidents – thus there would be no way to split them further.

# Appendix D. CHAID analysis output



## Appendix E. Cross-tabulation output

Manoeuvre	Reversing	Parked	Waiting	Stopping	Starting	U-Turn	Turning left	Waiting to turn left	Turning right	Waiting to turn right	Changing lane to left	Changing lane to right	O/T offside	O/T stationary veh OS	O/T nearside	GA LHB	GA RHB	GA other	Total
Reversing	33	332	221	38	12	11	28	12	40	26	0	0	9	11	1	18	18	406	1216
Parked		22	14	49	67	20	71	1	83	2	24	9	13	333	34	153	199	2557	3651
Waiting			592	1527	243	12	120	38	159	42	57	56	20	74	37	167	93	9048	12285
Stopping				773	84	9	76	136	73	115	23	27	21	18	6	60	47	2669	4137
Starting					95	5	18	31	42	16	4	5	18	55	8	25	26	844	1192
U-Turn						3	10	2	14	2	0	5	70	38	13	25	14	963	1159
Turning left							223	153	226	106	12	13	56	32	37	46	90	2565	3559
Waiting to turn left								102	32	8	5	3	6	5	4	33	12	809	1019
Turning right									383	111	16	19	733	254	123	321	679	14986	17625
Waiting to turn right										46	8	5	56	54	36	110	74	2978	3367
Changing lane to left											23	40	9	4	19	16	29	523	663
Changing lane to right												41	111	18	6	16	24	798	1014
O/T offside													80	6	1	63	58	1031	1239
O/T stationary veh OS														32	3	21	30	312	398
O/T nearside															12	12	7	181	212
GA LHB																261	<b>3355</b>	695	4311
GA RHB																	230	636	866
GA other																		14998	14998
Total	33	354	827	2387	501	60	546	475	1052	474	172	223	1202	934	340	1347	4985	56999	72911

## Appendix F. Solar Position

### F.1 Road Types

Accidents were split into those that occurred on single-carriageway roads and those that occurred on dual carriageway roads. Accidents that occurred on a roundabout, one-way street or where the road type was unknown were not included in either category.

#### F.1.1 Single carriageway roads

**Table F.1 Cars and taxis, single carriageway road, positive solar elevation**  
Excluding parked, reversing and missing manoeuvres

Sun angle (degrees)	All	Non-junction	Non-junction, single vehicle	Non-junction, two vehicles
Total	144750	52229	13153	27311
0	0.981	0.968	0.997	0.993
45	0.996	0.999	1.041	0.995
90	1.007	1.009	0.982	1.006
135	1.011	1.023	1.005	1.029
180	1.033	1.046	1.008	1.042
225	1.011	1.024	0.990	1.014
270	0.996	0.994	0.993	0.998
315	0.964	0.937	0.985	0.924
p-value	<0.001	<0.001	>0.05	<0.001
back/front	1.052	1.081	1.011	1.049

Accident involved vehicles on single carriageway roads were 5% more likely to have been travelling away from the sun than into the sun. When only non-junction accidents were considered this figure rose to 8%. It is possible that drivers of vehicles at junctions have fewer difficulties with the position of the sun: if they have to give way they will have the opportunity to pause if their visibility is compromised. Vehicles continuing on the main road will have less time to adjust if they are suddenly dazzled by the sun.

There was no evidence that the position of the sun had a bearing on the occurrence of single-vehicle, non-junction accidents on single carriageway roads. However, there was strong evidence that two-vehicle, non-junction accidents were more likely to involve vehicles travelling away from the sun. This suggests that the main difficulty caused by the location of the sun is the observation of other vehicles, rather than the road. It was not possible within this project to identify whether the two vehicle accidents tended to involve vehicles travelling in the same or opposite directions. Further research on this group of accidents may yield interesting results.

**Table F.2 Cars and taxis, single carriageway, non-junction**  
Excluding parked, reversing and missing manoeuvres

<b>Sun angle (degrees)</b>	<b>All</b>	<b>NBU</b>	<b>BU</b>	<b>BU, vehicle hit pedestrian</b>
Total	52229	20774	31451	7213
0	0.968	0.971	0.965	1.009
45	0.999	1.016	0.988	1.010
90	1.009	1.017	1.004	0.997
135	1.023	1.018	1.027	1.005
180	1.046	0.994	1.079	1.012
225	1.024	1.025	1.024	0.983
270	0.994	0.997	0.992	0.989
315	0.937	0.962	0.920	0.995
p-value	<0.001	>0.05	<0.001	>0.05
back/front	1.081	1.023	1.118	1.002

The position of the sun did not have a significant effect on accidents occurring on non-built up roads (speed limit >40mph). However, there was strong evidence that the sun had an effect on accidents on built-up roads. It was thought that this may be due to the sun distracting the driver from the many moving hazards in built-up areas. When only vehicles that had a pedestrian casualty associated with them were considered it was found that the position of the sun had no effect. It is not clear why the sun should be so much more influential in built up areas than in non-built up areas

### F.1.2 Dual carriageway roads

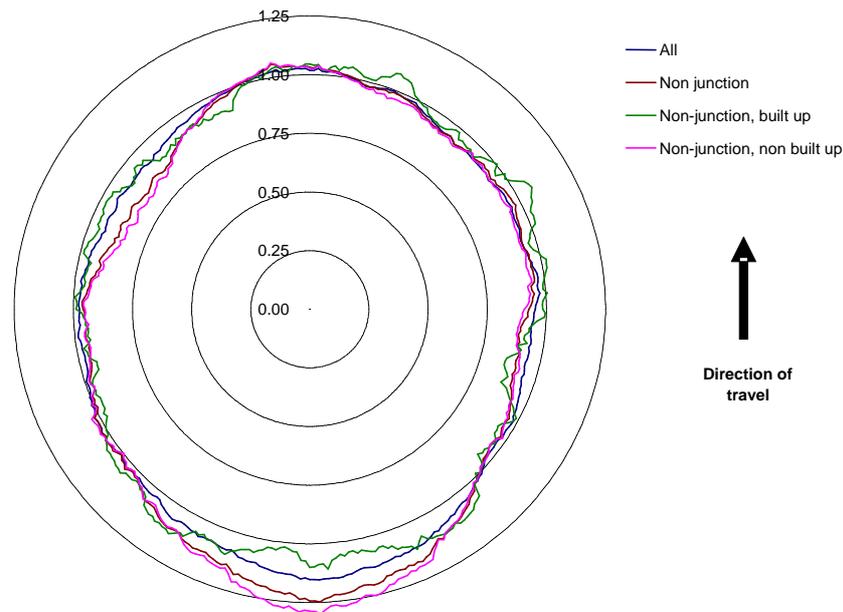
**Table F.3 Cars and taxis, dual carriageway**  
Excluding parked, reversing and missing manoeuvres

Sun angle (degrees)	All	Non-junction	Non-junction, NBU	Non-junction, BU	Vehicles changing lane to the right
Total	34642	18845	14361	4484	886
0	1.018	1.032	1.027	1.045	0.984
45	0.942	0.953	0.950	0.962	0.966
90	0.953	0.926	0.907	0.985	0.993
135	1.011	1.003	1.008	0.988	1.029
180	1.151	1.243	1.287	1.101	0.975
225	1.018	1.027	1.024	1.035	1.047
270	0.974	0.959	0.950	0.988	1.047
315	0.932	0.858	0.847	0.896	0.957
p-value	<0.001	<0.001	<0.001	<0.05	>0.05
Back/front	1.131	1.205	1.253	1.053	0.991

An analysis of vehicles travelling on dual carriageway roads showed that the position of the sun had a stronger effect on the occurrence of accidents than on single carriageway roads. 13% more vehicles were involved in accidents when heading away from the sun than into the sun. This contrasts with a 5% difference on single carriageway roads. There are several factors that might explain this:

- Many accidents on single carriageway roads involve two vehicles travelling in opposite directions. If one vehicle is heading into the sun, the other must be heading away from the sun. This will reduce the measurable effect of the sun's position. Conversely, accidents on dual carriageway roads will usually affect only one carriageway.
- On a dual carriageway road many accidents may involve one of the vehicles changing lane, into the path of another vehicle. When the sun is behind the driver it can be very difficult to see vehicles approaching from behind and judge their speed.
- It can also be difficult to make out the signals and brake lights of vehicles ahead as the sun may cause the lights to appear to be on when they are not and mask the lights when they are on.

The effect of the sun was even more pronounced when only non-junction accidents were considered. Over 20% more accidents occurred with the sun behind the vehicle than in front of the vehicle.



**Figure F.1 Position of the sun relative to car or taxi, positive solar elevation, dual carriageway, non-junction**

The data were split into built up and non-built up roads. When only dual carriageways were considered it was found that the sun had a far greater effect on non-built up roads than on built up roads. This finding is the reverse of that found on single carriageway roads. However, the results for both road types were highly statistically significant. It is not obvious what the reason for these seemingly contradictory findings is.

It was thought that vehicles changing lane to the right might have particular difficulty with the sun as they may be unable to see clearly in their driver's side wing mirror. However, there was no statistical evidence to substantiate this theory.

## F.2 Junctions

**Table F.4 Cars and taxis, junction accidents**  
Excluding parked, reversing and missing manoeuvres

Sun angle (degrees)	All	Signalised junctions	Signalised junctions, low sun
Total	124535	18878	7783
0	0.998	0.987	1.043
45	0.989	0.972	0.960
90	1.001	0.979	0.983
135	0.999	0.978	0.964
180	1.026	1.051	1.027
225	1.005	1.029	0.991
270	0.998	0.994	1.002
315	0.983	1.010	1.030
p-value	<0.05	>0.05	>0.05
back/front	1.028	1.065	0.984

The sun had only a slightly significant effect on the occurrence of accidents at junctions despite a large sample size. The sun was behind the driver in only 3% more accidents than it was in front. It was thought that the position of the sun could pose particular problems at signalised junctions. The sun may make it difficult to see whether the signals are red or green; there may also be a problem with phantom signals caused by the sun reflecting off the lamps. However, no statistically significant difference was found ( $p=0.07$ ). A further analysis of accidents at signalised junctions when the sun was low in the sky (an elevation of between zero and 20 degrees) did not find a significant difference either. However, a look at the last column of Table F.4 suggests that there are differences that may be significant if more data were analysed. In contrast to most of the other findings there was a greater problem when the sun was straight ahead, behind the signal heads. Some drivers, particularly if unfamiliar with the road may completely fail to see the presence of signals.