Federal Office of Road Safety

Grouping of Fatal Pedestrian Crashes

Authors

David Cairns
John Antill

ISBN 0 642 51269 8

ISSN 0810-770X

DEPARTMENT OF TRANSPORT
FEDERAL OFFICE OF ROAD SAFETY
DOCUMENT RETRIEVAL INFORMATION

Report No.	Date	Pages	ISBN	ISSN
CR 153	February 1996	72	642 - 51269 - 8	0810-770X

Title and subtitle

Grouping of Fatal Pedestrian Crashes

Authors

David Cairns, John Antill

Performing organisation

School of Behavioural Sciences
The Macquarie University

Sponsor

Federal Office of Road Safety
GPO Box 594
CANBERRA 2601

Available from	Price	Format
Federal Office of Road Safety GPO Box 594 CANBERRA 2601	No charge	Hard copy

Abstract

This research developed a typology of fatal pedestrian crashes using a multivariate statistical strategy which included homogeneity analysis and k-means clustering. Information on fatal pedestrian crashes was extracted from the 1990 version of the Federal Office of Road Safety's (FORS) "Fatality File" database. This database contained 378 variables recorded for each of 419 fatal pedestrian crashes. Using a hierarchical approach eight clusters were identified. The uniqueness of the solution lies in its hierarchical approach and the uncovering of important variables which could not have been chosen a priori. It is hoped that the results of this research will lead to strategies for the reduction of fatal pedestrian crashes

Keywords
Pedestrian crashes, homogeneity analysis, hierarchical approach.

Notes

(1) FORS research reports are disseminated in the interests of information exchange.
(2) The views expressed are those of the authors and do not necessarily represent those of the Commonwealth Government

TABLE OF CONTENTS

Abstract

This research developed a typology of fatal pedestrian crashes using a multivariate statistical strategy which included homogeneity analysis and k-means clustering. Information on fatal pedestrian crashes was extracted from the 1990 version of the Federal Office of Road Safety's (FORS) "Fatality File" database This database contained 378 variables recorded for each of 419 fatal pedestrian crashes. Using a hierarchical approach eight clusters were identified, being, (i) "Rural" (11.2%) - fatal crashes in rural areas; (ii) "School Student/Preschooler" (14.0%) - fatalities involving young students or children under school age; (iii) "8-18 Years of Age" (5.9%) - fatalities involving individuals aged 8 to 18 years; (iv) "Elderly Retired" (23 7%) - fatalities involving elderly retired pedestrians; (v) "Heavy Vehicle" (3.8%) - pedestrian fatalities involving heavy vehicles; (vi) "Adult Day" (11.5%) - day time fatalities involving adult pedestrians; (vii) "Adult Alcohol" (13.5%) - night time fatalities involving adults with alcohol; and (viii) "Adult Other" (16 5%) - other night time crashes involving adults. The uniqueness of the solution lies in its hierarchical approach and the identification of groups of pedestrian crashes which could not have been chosen a priori It is hoped that the results of this research will lead to strategies for the reduction of fatal pedestrian crashes.

Executive Summary

Introduction

In 1990, there were 419 pedestrian deaths in Australia. The aim of this research project was to develop a typology of these fatal pedestrian crashes which could lead to the development of specific counter measures to reduce the incidence of such crashes. This research differed from other published typologies of pedestrian crashes, identified in the research literature, in that it did not.-

- (i) include fatal with non-fatal pedestrian crashes,
- (ii) include fatal pedestrian crashes with all other fatal crashes,
- (iii) rely on the researcher's intuition to decide which variables were important,
- (iv) rely on the formation of clusters based on pre-defined grouping variables.

This research identified separate groups of fatal pedestrian crashes using an appropriate multivariate statistical strategy. This strategy was a particular strength of the research.

Methodology

Information on fatal pedestrian crashes was extracted from the 1990 version of the Federal Office of Road Safety's (FORS) "Fatality File" database. This database contained 378 variables recorded for each of the fatal pedestrian crashes.

A detailed, and logical seven-step multivariate strategy, was developed to deal with this complexity. As result of this strategy the database was eventually reduced to an information-rich set of 25 variables for each of 393 fatal pedestrian crashes. These 25 variables contained information concerning the vehicle, the pedestrian, the driver and the crash site. In addition to these 25 variables, on which the clustering solution was based, a further nine variables were used to help describe the obtained clusters. Thus, a total of 34 variables was used to describe the obtained clustering solution

Homogeneity analysis was chosen as the appropriate multivariate statistical method for clustering these data because of the categorical nature of most of the variables and the lack of a dependent/independent structure. Homogeneity analysis can be thought of as principal components analysis of nominal data.

Results and Conclusion

Homogeneity analysis was performed in a series of steps because the first homogeneity analysis was not able to separate all the crashes into a set of homogeneous groups Rather, each successive analysis separated one distinct group from the remaining crashes. Each distinctive, or "outlier group" was omitted before the next homogeneity analysis step proceeded. In this way, a "stepdown" or "zooming-in" approach was adopted to uncover the cluster structure. At the conclusion of this "zooming-in" approach eight clusters were identified.

These eight clusters were identified as:-

	Cluster Name	% of Fatalities	Description
1.	Rural	11.2%	Fatal crashes in rural areas
2.	School Student/ Preschooler	14.0%	Fatalities involving young students or children under school age
3.	8-18 Years of Age	5.9%	Fatalities involving individuals aged 8 to 18 years
4.	Elderly Retired	23.7%	Fatalities involving elderly retired people
5.	Heavy Vehicle	3.8%	Fatalities involving heavy vehicles
6.	Adult Day	11.5%	Day time adult fatalities
7.	Adult Alcohol	13.5%	Night time adult fatalities involving alcohol
8.	Adult Other	16 5%	Other night time adult fatalities

In interpreting the profiles it should be remembered that the procedure resulted in clusters within clusters. This overlap of clusters is outlined in the report. Additionally, an extensive profile of each cluster was developed based on the 34 important variables

This obtained cluster solution, performed on the 1990 Fatality file, was validated on the 1988 Fatality File database giving a remarkable confirmation of the 1990 "stepdown" clustering solution. The result of the validation with the 1988 file was a confidence in the solution gained with the 1990 pedestrian fatality file.

This research achieved its aim of developing a typology of pedestrian crashes. The uniqueness of the solution lies in its hierarchical approach and the uncovering of important variables which could not have been chosen a priori. In order, the variables of importance were region (urban/rural), age of pedestrian, type of vehicle, timing of crash (day/night), and the importance of alcohol. It is interesting that the characteristics of the crash site were not revealed as major classification variables.

It is hoped that the results of this research will lead to strategies for the reduction of fatal pedestrian crashes.

Grouping of Fatal Pedestrian Crashes

Chapter 1: Introduction

1.1. The rationale for clustering

The aim of this research project was to develop a typology of fatal pedestrian crashes which might suggest strategies for reducing these crashes. It was proposed that the identification of clusters of crashes with similar patterns would assist in the location of common causal variables and circumstances surrounding each group of crashes. This should then suggest different types of intervention procedures or counter-measures which may prevent fatal pedestrian crashes. A brief review of literature involving previous approaches to clustering pedestrian crashes is presented.

1.2. Approaches to clustering

1.2.1. Contrasting pedestrian crashes with other crashes

Fatal pedestrian crashes were contrasted with all other fatal crashes in a study conducted by Attewell and Dowse (1992). "Pedestrians accounted for 19% of all road fatalities and these crashes generally occurred in the afternoons and evenings on weekdays, and at night on weekends. A disproportionately high number occurred in Winter. Most pedestrians were killed in urban areas, away from intersections and while crossing the road where there were no marked crossings. Most (69%) of the pedestrians were considered to be responsible for the crashes. The pedestrians killed included children who did not look before crossing, young alcohol affected adults and the largest group (40%) comprised older persons who generally made misjudgments. Pedestrians, like cyclists, had a high incidence of death due to head injuries; 19% died instantly and half died in hospital. Many had lower extremity injuries. The drivers involved in these crashes tended to be younger, but not speeding or driving under the influence of alcohol." (p.1). Four subgroups of pedestrian crashes based on crash pattern were further investigated. They were: emerging (pedestrian came from in front of a stationary or parked vehicle and was hit from the right); near-side (pedestrian proceeded from the kerb, median or side of the road and was hit from the right); far-side (pedestrian was hit from the left); on carriageway (pedestrian was playing, working, lying, standing or walking with or against the traffic on the carriageway).

1.2.2. Intuitive/commonsense clustering

In Australia, there are three groups of pedestrian crashes which have been the focus of attention in advertising campaigns - children, intoxicated adults and the elderly The basis of this attention is their high representation among all fatal pedestrian crashes. Using an intuitive approach to clustering, Buhlman, Warren and Simpson (1983) divided 452 child pedestrian fatalities into three groups of victims, aged 1-4, 5-7, and 8-14 years.

Their results showed that: (i) Children aged 1-4 years were involved in daytime collisions on or near private driveways, (ii) Children aged 5-7 years were struck while crossing the road between parked cars immediately before or after school, (iii) Children aged 8-14 years were struck at night while walking along the road or crossing at an intersection.

1.2.3. Categories based on combinations of key factors

In reviewing pedestrian crashes, it is clear that there are major differences in the circumstances surrounding crashes and the people involved in them. For example, a crash may happen in the day or at night, at an intersection or midblock, or where there are traffic controls or not. The pedestrian may be young or old and both driver and pedestrian may be intoxicated or not Clusters can be formed on the basis of all combinations of these contrasts. For example, one cluster may be all crashes involving young intoxicated people which occur at night-time, at midblock where there are no controls.

In a study involving vehicle collisions, excluding pedestrians, 18 clusters were formed on the basis of the following factors. the involvement of a single or multiple vehicles; whether the crash occurred at an intersection or not; if so, whether there were traffic lights or only signs, whether the vehicles were travelling in the same or opposite direction or crossing paths, whether or not one vehicle was attempting to turn (Massie, Campbell & Blower, 1993).

1.2.4. Systematic review of crashes with confirmation of categories based on inter-rater agreement

The following study demonstrates the process of reviewing crashes, forming types on the basis of common events, and revising these types on the basis of other researchers reviewing further crashes. In this study of crash types, excluding pedestrians, reports of intersection crashes in one county were categorised by the similarity of the events they described (Retting, Williams, Preusser, & Weinstein, 1992). This process was repeated for non-intersection crashes. "A reviewer examined the police report narrative descriptions and diagrams in detail to identify precrash driver/vehicle behavior, develop a preliminary definition of crash type, and sort each crash correctly by type. Once the draft definitions were completed, a second person reviewed all the crash reports from four cities and independently assigned a crash type to each. Some revisions were then made to the crash type definitions. These revised (final) type definitions were used by a third person who independently read each crash report and assigned it to a crash type Discrepancies in crash type assignments among the three reviewers were reviewed and resolved." (p. 3). As a result of this process, thirteen distinct crash types plus an 'other' type for unusual, unclassifiable events were identified and defined.

1.2.5 Methods based on quantitative analysis

None of the four methods above involves any statistical analysis to form the typologies or clusters In contrast, there is a whole range of clustering techniques which vary in the way they measure similarity between pairs of items (crashes). On the basis of the precise measure or clustering algorithm a number of clusters can be formed.

In one study of driving under the influence (DUI) offenders, eight different clustering algorithms were used in order to assess the comparability of the clusters produced (Wells-Parker, Anderson, Pang & Timken, 1993) In another study, a form of clustering called homogeneity analysis was used to group 2689 pedestrian crashes, each measured on 29 variables, into four clusters (Gundy, 1990). These clusters were described as follows:

(i) This dry weather, daytime, type of accident tends to occur on straight road sections inside a built-up area. The driver, who is less likely to have been drinking, strikes a young, male child who suddenly crosses the street from the sidewalk, and possibly from behind an object. This less than lethal type of accident seems to represent the young child midblock dart-out.

(ii) This type of accident tends to occur during the winter months on weekend mornings under non-optimal weather or lighting conditions. The location tends to be an urban intersection or pedestrian crossing on a multi-lane road. The driver, who is turning or accelerating from a standstill then strikes the adult female pedestrian who is likely to be crossing the road on a pedestrian crossing. The driver is likely to be charged with failure to yield right-of-way or neglecting to obey a traffic light or sign.

(iii) This serious type of accident tends to occur on a rural, high-speed road. The mainly middle-aged pedestrian is more likely to be on the shoulder of the road or crossing in midblock, without the benefit of a pedestrian crossing, when he is struck

(iv) This weekend, night-time accident tends to occur inside built-up areas. The driver as well as the adult pedestrian may have been drinking. The driver strikes the pedestrian who is walking along or standing on the road (See Appendix 3, p.3).

1.3. Advantages/disadvantages of quantitative analyses

The dominant approach to clustering has been intuitive with few attempts being made to use multivariate statistical techniques. The main advantage of these latter methods is that new groupings may be found which may not be apparent to the eye or intuition. However, without some resort to intuition or commonsense there may be little relationship between the clusters produced via statistical methods and useful counter-measures.

1.4. Aim of this research project

As previously stated, the aim of this research was to develop a typology of fatal pedestrian crashes which might suggest strategies for reducing these crashes. With reference to the literature review, the approach of this research was quantitative, without reference to non-pedestrian crashes, and involving only fatal pedestrian crashes. The next chapter describes the clustering strategy developed for this research.

Chapter 2: The Clustering Strategy

The plan of the present chapter is to describe the clustering strategy and present its implementation to these data. Appropriate results are presented in their logical sequence

2.1. The clustering strategy

The developed strategy involved the following seven steps:-

(i) Understand the data file,
(ii) Understand how the data were collected,
(iii) Investigate the structure of the data,
(iv) Initial data analysis,
(v) Definitive data-analytic investigation,
(vi) Validation, and
(vii) Interpretation

The first four of these steps are presented in this chapter. The fifth, sixth and seventh steps are presented in the following two chapters. In this, and the following section, the first four steps of the strategy are described in general and as applied to these data.

2.1.1. Understand the data file - what was measured and why

The purpose of this first step was to understand the problem and clarify the objectives of the research It involved meetings with the Federal Office of Road Safety (FORS), to discuss the meaning of each variable, the meaning of special data codes, gain an understanding of prior knowledge, and consider how the results would be used. FORS also provided background reading to help in the process.

2.1 2. Understand how the data were collected

A full coding manual was supplied detailing the construction of the data file. The present analysis can be viewed as a secondary analysis of official statistics. The limitations of the data were also discussed.

2.1.3. Investigate the structure of the data

This third step involved reviewing the number of cases and number of variables and selecting the appropriate cases and variables for analysis. The selected variables were also grouped into meaningful categories for later analysis.

2.1.3 1 Select appropriate cases

The data were reviewed with respect to cases to determine which cases should be included in further analyses. This step was achieved in consultation with FORS. Originally the 1990 FORS "Fatality File" data file contained 438 cases.

It was decided to include only cases representing a fatal accident pedestrian death and to exclude cases which were coded as "suicide", "murder" or "runaway vehicle", as these cases represented deliberate injury. This reduced the file to 406 cases. A second decision was made to exclude pedestrian crashes where the pedestrian had been recorded as having suffered only "minor injury" This further reduced the number of cases to analyse to 393 cases. These 393 cases were where the pedestrian had severe injuries and died as a direct result of the accident. All further analyses were conducted on these remaining 393 cases.

2.1.3.2. Select appropriate variables

Because of the nature of the data file all variables at this stage were considered important for later investigation.

2.1.3.3. Group variables into meaningful "families"

The variables were grouped into four "families": vehicle level; crash level; driver level; and pedestrian level. The vehicle level family contained a set of variables pertaining to vehicles involved in the accident, such as type of vehicle, State of registration, extent of damage to vehicle etc. The crash level family contained a set of variables pertaining to general information concerning the accident, such as time of accident, date of accident, State in which accident occurred etc. The driver and pedestrian level families contained respectively, a set of variables pertaining to the people involved in the accident, such as age, sex, severity of injury etc.

These "families" were all considered to be of equal importance. The next step was then processed separately for each of these four "families" of variables.

2.1.4. Initial data analysis - carefully examine each variable

In this fourth step the importance and quality of each variable was assessed with the aim of reducing the dataset to a smaller number of information rich variables. As this stage was analysed separately for each of the four "families" of variables, it is described here in general and specifically for each of the "families" in the following section of the chapter.

2.1.4.1. Evaluate each variable with regard to its potential importance

This evaluation was conducted in consultation with FORS. The aim was to initially divide each "family" of variables into important and unimportant variables based on potential information contained in each variable. This division was not made with respect to statistical considerations of data quality. Rather, the important variables were those known to have some relationship to pedestrian crashes. Variables deemed to be unimportant were dropped from further consideration

2.1.4.2. Statistically investigate the important variables

Here, each variable was inspected to see if it was able to statistically contribute to possible major underlying patterns in the data. The step involved the use of descriptive statistics. The aim here was to divide the important variables into two groups: those to be investigated further for underlying patterns; and those to be put aside until after the pattern has emerged.

These put-aside variables are to be used later to help describe the emerged patterns Three statistical criteria were employed:

(i) Inspecting the variation in each variable. As a general rule variables with more than 70% in any one category were not statistically useful and were put aside

(ii) Inspecting the amount of missing data for each variable. Variables were put aside if they had more than 40% in the unknown or missing category.

(iii) Inspecting the overlap of information Where two or more variables measured the same information only one was selected for further analysis.

2.1.4.3. Modify (recode) the data where necessary

In this stage of the initial data analysis, selected variables were reviewed and recoded where necessary to reduce the number of categories with low frequencies. Without this step, these categories with low frequency would dominate the search for underlying patterns in the data by becoming peripheral points in the cluster analysis

2.1.4.4. Multivariate data-analytic variable reduction and search for pattern

Using the variables selected and recoded, the underlying pattern within the "family" was explored using Homogeneity Analysis (HOMALS). Homogeneity analysis is one of the names associated with correspondence analysis and multiple correspondence analysis (Greenacre, 1984; Gifi, 1990; SPSS, 1990). With homogeneity analysis, which can be thought of as a principal components analysis of nominal data, it is often possible to summarise the complex relationships between variables with a single two-dimensional plot (SPSS, 1990).

Homogeneity Analysis was particularly appropriate as most of the data were measured at the categorical level. Also, it should be noted that there was no structure in the data which separated variables into independent and dependent variables. Had there existed this structure, other techniques such as CHAID (chi squared automatic interaction detection) may have been more appropriate.

The HOMALS solution was able to indicate that some of the selected variables were not able to discriminate between groups of cases. These variables, which vary unsystematically across groups of cases, were put aside. Some of the variables put aside were later used in the description of the clusters formed by the definitive data-analytic investigation. The variables which did contribute to the HOMALS solution were retained for the definitive data-analytic investigation.

When all four "families" had reached this stage in the overall strategy the retained variables were collected together for the next step, which is presented in the following chapter. The next section describes the initial data analysis as applied to each of the four "families" of variables.

2.2. Initial data analysis of the four "families"

2.2.1. The vehicle level "family"

Initially there were 44 variables in the vehicle "family" to investigate. In terms of their theoretical or potential importance these variables were divided into 22 important and 22 unimportant variables. The unimportant variables were dropped from further consideration.

Next, each of the 22 important variables was inspected to see if it was able to statistically contribute to possible major underlying patterns in the data. The aim was to divide these 22 variables into two groups: those that were further analysed; and those that were put aside Using the three statistical criteria described above, seven variables had truncated variation, four variables had excessive missing data, and four variables were overlapping in information with other variables. As a result of the statistical criteria the 22 important variables were reduced to seven variables being:- distance from home, number of people in the vehicle, weight class of vehicle, speed limit/speeding, braked/swerved, point of primary impact and year of manufacture.

Next, the seven retained variables were reviewed and recoded where necessary. Each variable was inspected and recoded so as to eliminate categories with low frequencies. Without this step these categories with low frequency would dominate the search for major patterns in the data by becoming peripheral points in any spatial analysis. The variable, point of primary impact, was recoded in consultation with FORS.

Finally, using these seven variables, the pattern and structure of pedestrian deaths, at the vehicle level, was explored using Homogeneity Analysis (HOMALS). A number of homogeneity analyses were conducted. Initial analyses revealed that the variable braked/swerved did not help form homogeneous groups of pedestrian crashes Further, the unknown category of weight class of vehicle had low frequency and became a periphery point which distorted the initial HOMALS solutions. Accordingly, the variable braked/swerved was put aside and the unknown category of weight class of vehicle was recoded to missing.

2.2.2. The pedestrian level "family"

Initially there were 208 variables in the pedestrian "family" to investigate. In terms of their theoretical on potential importance these variables were divided into 16 important and 192 unimportant variables. The unimportant variables, which mainly detailed the injuries sustained to the pedestrian, were dropped from further consideration.

Next, each of the 16 important variables was inspected to see if it was able to statistically contribute to possible major underlying patterns in the data. The aim was to divide these 16 variables into two groups: those that were further analysed, and those that were put aside. Using the three statistical criteria, four variables had truncated variation, no variables had excessive missing data, and five variables were overlapping in information with other variables. As a result of the statistical evaluation of the variables, the 16 important variables were reduced to seven variables, being: pedestrian age, pedestrian sex, pedestrian employment status, pedestrian height, cause of death, ISS (Injury Severity Score), and pedestrian BAC.

Next, the seven retained variables were reviewed and recoded where necessary

Finally, using these seven variables, the pattern and structure of pedestrian deaths, at the pedestrian level, was explored using Homogeneity Analysis (HOMALS). Homogeneity analyses revealed that pedestrian sex and cause of death did not help form homogeneous groups of pedestrian crashes. Accordingly, these two variables were put aside.

2.2.3. The driver level "family"

Initially there were 19 variables in the driver "family" to investigate. In terms of their theoretical on potential importance these variables were divided into five important and 14 unimportant variables. The unimportant variables were dropped from further consideration.

Next, each of the five important variables was inspected to see if it was able to statistically contribute to possible major underlying patterns in the data. The aim was to divide these five variables into two groups: those that were further analysed; and those that were put aside. Using the three statistical criteria, one variable, driver BAC was put aside because of overlapping information with other variables. As a result of the statistical evaluation, four variables remained for further analysis, being: driver age, driver sex, seat belt wearing and driver BAC grouping.

Next, the four retained variables were reviewed and recoded where necessary Finally, using these four variables, the pattern and structure of pedestrian deaths was explored using Homogeneity Analysis (HOMALS). Homogeneity analyses revealed that driver sex did not help form homogeneous groups of pedestrian crashes. Accordingly, driver sex was put aside.

2.2.4. The crash level "family"

Initially there were 107 variables in the crash "family" to investigate. In terms of their theoretical or potential importance these variables were divided into 27 important and 80 unimportant variables. The unimportant variables were dropped from further consideration.

Next, each of the 27 important variables was inspected to see if it was able to statistically contribute to possible major underlying patterns in the data. The aim was to divide these 27 variables into two groups: those that were further analysed; and those that were put aside. Using the three statistical criteria described above, seven variables had truncated variation, two variables had excessive missing data, and one variable was overlapping in information with other variables. As a result of the statistical criteria the 27 important variables were reduced to 17 variables being:- day of week, landuse adjacent to the crash site, land classification, road type, road configuration, type of median, number of lanes, horizontal road alignment, speed limit at crash location, traffic controls, two variables concerning DCA event (Definition for Classifying Accidents), unit responsible for crash, major cause of accident, time of day/day of week, time, and day.

Next, the 17 retained variables were reviewed. It was decided further consultation was required with FORS to clarify four of the variables (traffic controls, two variables concerning DCA event, and major cause of accident). This process differed from the other families because of the complexity of these variables.

Finally, using 13 of the above 17 variables, the pattern and structure of pedestrian deaths, at the crash level, was explored using Homogeneity Analysis (HOMALS). Homogeneity analysis revealed that the variables speed limit at the crash location and unit responsible for crash did not help form homogeneous groups of pedestrian crashes. Accordingly, these two variables were put aside.

Regarding the four variables which required further clarification; traffic controls was put aside, one of the DCA variables was dropped while major cause of accident and the other DCA variable were retained.

2.3. Summary of the initial data analysis of the four "families"

At the beginning of the initial data analysis step the four "families" contained 378 variables. At the conclusion of this step:-

(i) 308 had been dropped from further consideration being considered unimportant,
(ii) 35 had been put aside due to failure to meet statistical criteria,
(iii) 8 had been put aside as a result of the homogeneity analysis, thus leaving
(iv) 27 deemed important in forming homogeneous groups.

The eight variables which had been put aside were; braked/swerved, pedestrian sex, cause of death, driver sex, traffic controls, one of the DCA variables, speed limit at crash location, and unit responsible for crash.

The 27 variables which went forward into the definitive data-analytic investigation were; distance from home, number of people in vehicle, weight class of vehicle, speed limit/speeding, point of primary impact, year of manufacture, driver age, seat belt, driver BAC, major cause of accident, pedestrian age, pedestrian employment status, pedestrian height, ISS score severity, pedestrian BAC, day of week, landuse adjacent to crash site, land classification, road type, road configuration, type of median, number of lanes, horizontal road alignment, DCA event, time of day/day of week, time, and day.

The final form of the 35 variables which were either put aside as a result of the homogeneity analysis or carried forward are shown in Table 3.1 of Chapter 3. For this table, only one of the DCA variables is included.

Chapter 3: The Obtained Clustering Solution

This chapter describes the fifth and sixth steps in the clustering strategy, namely, the definitive data-analytic investigation to obtain the clustering solution from the important variables, and then a validation of the obtained solution.

As a check in the data reduction process, the 27 important variables were reviewed with FORS to assure agreement, and to allow for final adjustments to the variables to be reviewed. As a result of this review, some variables were reinstated, some put aside, and some modified, resulting in 25 variables to be included in the definitive data-analytic investigation These 25 variables were·

(i) Vehicle level: year of manufacture, weight class of vehicle, distance from home, number of people in vehicle, speed limit/speeding, point of primary impact,
(ii) Pedestrian level· age, height, employment status, BAC, cause of death,
(iii) Driver level: age, seat belt, and
(iv) Crash level. unit responsible and major cause of accident combined together, day of week, landuse adjacent to crash site, land classification, road type, road configuration plus lanes. DCA event, time day/day of week, time, day, location, speed limit at crash location.

For the definitive data-analytic investigation, two data-analytic methods were employed, kmeans cluster analysis and "stepdown" homogeneity analysis. The following two sections reports on these differing approaches.

3.1. Kmeans cluster analysis

This procedure involved homogeneity analysis followed by kmeans cluster analysis. A similar approach has been suggested by Nakache (1981) Cluster analysis is the generic name for a wide variety of procedures which can be used for such analysis tasks as data reduction; identification of groups, the generation of classification schemes; and, the testing of hypotheses (Aldenderfer & Blashfield, 1984). The particular method employed in this study was kmeans cluster analysis (Everitt, 1974).

Together with the 393 pedestrian crashes, the 25 selected variables formed a 393 by 25 data matrix. The 25 variables consisted of 113 separate categorical levels. In this way, the data matrix can also be considered as a 393 by 113 data matrix.

Homogeneity analysis was conducted on this matrix obtaining a 10-dimensional solution. In this solution, each of the 393 crashes and 113 category levels of the 25 variables are located in the same 10-dimensional space. Each of these 506 points is therefore represented by a profile of 10 spatial coordinates. From this analysis these 506 10-dimensional coordinates were saved and then entered into the BMDP kmeans cluster analysis program.

Cluster solutions from 2 clusters up to 9 clusters were selected. In any cluster it was possible to have both crash points, representing individual crashes, and variable level points, representing category points of variables, clustered together. Having variable level points clustered with the crash points enabled the crash points clustered together to be interpreted by inspection of the variable level points.

From the eight cluster solution it was clear that there were distinct clusters corresponding to.-

rural crashes (40 coordinate points)
young student crashes (35 points)
intersection crashes (53 points)
crashes with unknown driver and vehicle characteristics (20 points)
driver at fault crashes (83 points)
heavy vehicle crashes (53 points)

However, the kmeans clustering procedure left many points, both crash points and variable level points, together in two large undifferentiated clusters containing 128 and 94 coordinate points respectively. This is not an infrequent problem with cluster analysis (Bergman, 1988). It was found that these points were clustered together, not because of their similarity with each other, but because of their large dissimilarity with the other identified clusters. It was therefore not possible to clearly identify the structure within these remaining two large clusters and hence the structure of the full data matrix. Because of this finding an alternative clustering procedure was adopted.

3.2. "Stepdown" homogeneity analysis of the 393 crashes

This step involved successive homogeneity analyses conducted in a "stepdown" or hierarchal fashion which zoomed in on the structure. This "zooming in" approach has been employed by other cluster analysts (Coolen & Hilkhuysen, 1992).

Firstly, when the 25 selected variables were combined in a full homogeneity analysis it was found that seven variables were not useful in forming homogeneous groups. These seven variables were year of manufacture, distance from home of driver, number of people in vehicle, point of primary impact, cause of death, seat belt and location. The remaining 18 variables were used in subsequent homogeneity analyses.

This "stepdown" procedure involved visually identifying a homogeneous cluster in the homogeneity analysis, removing those crashes identified by that cluster from the data matrix, and then repeating the homogeneity analysis on the remaining crashes. This procedure was repeated until all 393 crashes were allocated to specific clusters. In this way, the structure of the pedestrian crashes was successively zoomed in on in a "stepdown" fashion rather than clustering the matrix in a flat way. In other words, the full structure of the matrix was not evident until very distinct clusters were identified and removed.

These successive distinct clusters can be thought of as outliers or sample disjunctions (Gauch, 1982). Ordination techniques, such as homogeneity analysis, are a way of identifying these separate blocks of similar pedestrian crashes which are then omitted before further analysis (Gauch, 1980, 1982; Greenacre, 1984; Gifi, 1990).

The occurrence of distinct clusters of crashes with high internal similarity but from which all other crashes were dissimilar was evident in the successive homogeneity solutions. It was also evident from the obtained solutions how the structure of the crashes unfolded as distinct clusters were successively removed.

Eight clusters were identified using this "stepdown" procedure. These were:-

1.	Rural	44 crashes	(11.2%)
2.	School Student/Preschooler	55	(14 0%)
3.	8-18 Years of Age	23	(5.9%)
4.	Elderly Retired	93	(23.7%)
5.	Heavy Vehicle	15	(3 8%)
6.	Adult Day	45	(11.5%)
7.	Adult Alcohol	53	(13 5%)
8.	Adult Other	65	(16.5%)

Figure 3.1 shows the steps in the "stepdown" clustering analysis Table 3.1 presents the description of the population of pedestrian fatalities and all clusters of pedestrian fatalities Table 3.1 will also be referred to in Chapter 4.

Figure 3.1. Summary of Clustering Process

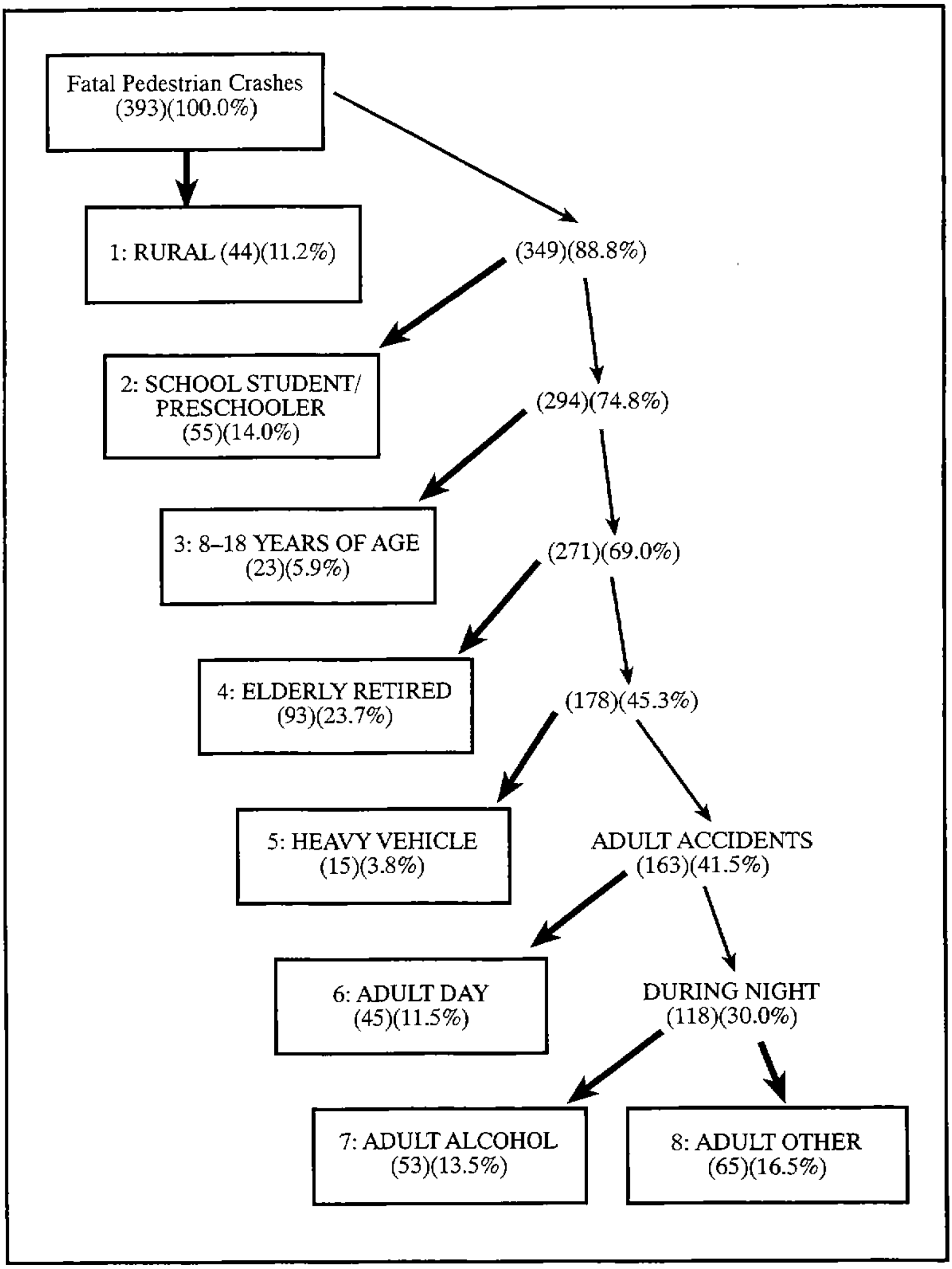


TABLE 3.1. Description of Population and All Clusters of Pedestrian Fatalities

Variable		Categories	Population n	Population %	1. Rural n	1. Rural %	2. School Student/ Preschooler n	2. School Student/ Preschooler %	3. 8 to 18 Years of Age n	3. 8 to 18 Years of Age %	4. Elderly Retired n	4. Elderly Retired %	5. Heavy Vehicle n	5. Heavy Vehicle %	6. Adult Day n	6. Adult Day %	7. Adult Alcohol n	7. Adult Alcohol %	8. Adult Other n	8. Adult Other %
(1)	Distance from home of driver	1-10 km	253	68	19	51	46	84	11	50	66	73	8	53	32	71	35	70	36	62
		11+km	119	32	18	49	9	16	11	50	24	27	7	47	13	29	15	30	22	38
(2)	No of people in vehicle	1	266	70	27	69	39	71	17	77	56	61	10	72	35	78	33	64	49	79
		2	81	21	8	21	9	16	4	18	28	30	2	14	9	20	12	23	9	15
		3+	34	9	4	10	7	13	1	5	8	9	2	14	1	2	7	14	4	7
(3)	Weight class of vehicle	Motor cycle	14	4	0	0	2	4	0	0	4	4	0	0	2	5	3	6	3	5
		Light car	63	17	7	18	10	18	3	14	18	20	0	0	9	21	5	10	11	18
		Medium car	94	25	8	21	14	26	7	32	28	31	0	0	7	17	15	29	15	25
		Heavy car	109	29	11	28	17	31	2	9	20	22	0	0	15	36	21	40	23	38
		Van/light truck	64	17	6	15	9	16	7	32	17	19	0	0	9	21	8	15	8	13
		Heavy vehicle	32	8	7	18	3	6	3	14	4	4	15	100	0	0	0	0	0	0
(4)	Speed limit/ speeding	<80 not speed	267	72	6	17	46	84	15	68	82	90	13	87	40	91	31	59	34	61
		<80 speed	39	10	0	0	5	9	3	14	7	8	1	7	2	5	13	25	8	14
		80+ not speed	62	17	29	80	4	7	3	14	2	2	1	7	2	5	8	15	13	23
		80+ speed	4	1	1	3	0	0	1	4	0	0	0	0	0	0	1	2	1	2
(5)	Braked/ swerved	Yes	195	54	18	47	30	55	8	40	55	66	10	71	18	42	30	64	26	46
		No	163	46	20	53	25	46	12	60	29	35	4	29	25	58	17	36	31	54
(6)	Point of primary impact	Left side	153	40	14	38	23	42	10	44	38	41	9	60	19	42	21	40	19	30
		Front	120	31	8	22	17	31	7	30	35	38	2	13	15	33	14	26	22	34
		Undercarriage	17	4	5	13	1	2	2	9	1	1	1	7	0	0	6	11	1	2
		Right side	95	25	10	27	14	25	4	17	19	20	3	20	11	25	12	23	22	34
(7)	Year of manufacture	Old (1962-84)	196	58	16	52	33	67	7	39	45	52	11	73	25	60	22	50	37	73
		New (1985-90)	142	42	15	48	16	33	11	61	42	48	4	27	17	41	23	50	14	28

Variable		Categories	Population n	Population %	1. Rural n	1. Rural %	2. School Student/ Preschooler n	2. School Student/ Preschooler %	3. 8 to 18 Years of Age n	3. 8 to 18 Years of Age %	4. Elderly Retired n	4. Elderly Retired %	5. Heavy Vehicle n	5. Heavy Vehicle %	6. Adult Day n	6. Adult Day %	7. Adult Alcohol n	7. Adult Alcohol %	8. Adult Other n	8. Adult Other %
(8)	Driver age	16-25 yrs	159	43	15	40	26	47	7	32	42	46	3	20	15	36	22	46	29	49
		26-92 yrs	212	57	23	61	29	53	15	68	50	54	12	80	27	64	26	54	30	51
(9)	Driver sex	Male	305	81	32	87	36	66	17	77	73	79	14	93	37	82	39	78	57	92
		Female	73	19	5	14	19	35	5	23	19	21	1	7	8	18	11	22	5	8
(10)	Seat belt	Worn	230	92	18	86	34	97	15	100	69	95	4	36	26	100	29	97	35	90
		Not worn	20	8	3	14	1	3	0	0	4	6	7	64	0	0	1	3	4	10
(11)	Blood alcohol concentration	Zero	273	89	24	80	40	91	12	75	74	97	13	100	36	95	37	84	37	79
		Above zero	35	11	6	20	4	9	4	25	2	3	0	0	2	5	7	16	10	21
(12)	Unit responsible for crash	Driver	50	13	5	11	3	6	6	26	14	16	1	7	8	18	3	6	10	16
		Driver and pedestrian	43	11	5	11	3	6	3	13	5	6	0	0	2	5	12	22	13	21
		Pedestrian	291	76	34	77	48	89	14	61	71	79	14	93	33	77	38	72	39	63
(13)	Major cause of accident	Alcohol/drugs	138	42	24	63	4	8	8	50	10	12	3	23	9	28	50	96	30	68
		Sensory impairment	6	2	0	0	0	0	0	0	6	8	0	0	0	0	0	0	0	0
		Vision obscured	9	3	0	0	0	0	0	0	5	6	0	0	4	13	0	0	0	0
		Not see other road user	49	15	5	13	13	26	1	6	17	21	2	15	6	19	0	0	5	11
		Excessive speed	9	3	0	0	2	4	0	0	5	6	0	0	0	0	1	2	1	2
		Ignore traffic controls	14	4	0	0	1	2	1	6	7	9	2	15	2	6	0	0	1	2
		Pedestrian step into path	101	31	9	24	30	60	6	38	31	38	6	46	11	34	1	2	7	16
(14)	Pedestrian age	1-18 yrs	90	23	12	27	55	100	23	100	0	0	0	0	0	0	0	0	0	0
		19-35 yrs	97	25	22	50	0	0	0	0	0	0	6	40	11	24	27	51	31	48
		36-65 yrs	104	26	7	16	0	0	0	0	0	0	8	53	32	71	26	49	31	48
		66+ yrs	102	26	3	7	0	0	0	0	93	100	1	7	2	4	0	0	3	4

Variable		Categories	Population		1. Rural		2. School Student/ Preschooler		3. 8 to 18 Years of Age		4. Elderly Retired		5. Heavy Vehicle		6. Adult Day		7. Adult Alcohol		8. Adult Other	
			n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
(15)	Pedestrian sex	Male	283	72	36	82	38	69	13	57	59	63	9	60	32	71	44	83	52	80
		Female	110	28	8	18	17	31	10	44	34	37	6	40	13	29	9	17	13	20
(16)	Pedestrian employment status	Pre-school/baby	20	6	2	5	18	33	0	0	0	0	0	0	0	0	0	0	0	0
		Manager/admin prof/para prof	18	5	0	0	0	0	0	0	0	0	1	7	6	15	6	13	5	10
		Trades persons	22	6	2	5	0	0	3	15	0	0	2	13	4	10	5	10	6	12
		Clerical/sales	12	3	0	0	0	0	2	10	0	0	0	0	2	5	2	4	6	12
		Plant/machine/ labour	47	13	9	24	0	0	2	10	0	0	1	7	13	33	13	27	9	18
		Housekeeping	13	4	1	3	0	0	0	0	0	0	1	7	5	13	3	6	3	6
		Unemployed	34	9	11	30	0	0	0	0	0	0	1	7	3	8	8	17	11	22
		Student	57	16	6	16	37	67	12	60	0	0	1	7	1	3	0	0	0	0
		Retired/ pensioner	135	38	6	16	0	0	1	5	93	100	8	53	6	15	11	23	10	20
(17)	Pedestrian height	80-145 cms	64	16	6	14	55	100	0	0	2	2	0	0	0	0	0	0	1	1
		149-169 cms	142	36	10	23	0	0	14	61	60	65	7	47	15	33	16	30	20	31
		170-195 cms	187	48	28	64	0	0	9	39	31	33	8	53	30	67	37	70	44	68
(18)	Cause of death	Head	53	14	8	19	12	22	2	9	9	10	0	0	9	21	7	13	6	10
		Body	50	13	10	23	8	15	4	17	12	13	0	0	2	5	5	10	9	14
		Multiple	279	73	25	58	34	63	17	74	69	77	14	100	32	74	40	77	48	76
(19)	ISS Score (Severity)	9-26	85	24	12	32	14	27	5	26	19	23	4	27	10	24	8	15	13	24
		27-35	89	25	7	18	13	25	2	11	26	31	2	13	11	26	12	23	16	29
		36-45	95	26	6	16	15	29	7	37	20	24	5	33	10	24	18	34	14	25
		48-75	89	25	13	34	10	19	5	26	19	23	4	27	11	26	15	28	12	22
(20)	Pedestrian BAC	0	143	55	7	25	13	100	9	60	55	85	9	82	30	81	0	0	20	53
		.001-.15	35	13	3	11	0	0	3	20	9	14	0	0	2	5	0	0	18	47
		>.15	82	32	18	64	0	0	3	20	1	1	2	18	5	14	53	100	0	0

Variable		Categories	Population		1. Rural		2. School Student/ Preschooler		3. 8 to 18 Years of Age		4. Elderly Retired		5. Heavy Vehicle		6. Adult Day		7 Adult Alcohol		8. Adult Other	
			n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
(21)	Day of week	Mon-Wed	152	39	18	41	24	44	6	26	43	46	10	67	28	62	8	15	15	23
		Thurs-Fri	121	31	11	25	15	27	12	52	31	33	3	20	10	22	16	30	23	35
		Sat	81	20	9	20	7	13	4	18	14	15	2	13	3	7	18	34	24	37
		Sun	39	10	6	14	9	16	1	4	5	6	0	0	4	9	11	21	3	5
(22)	Landuse adjacent to crash site	Residential	153	40	6	15	36	66	6	26	41	45	3	20	18	40	21	40	22	34
		Residential/Commercial	92	24	1	2	12	22	8	35	25	28	4	27	12	27	13	24	17	27
		Commercial/Industrial	77	20	1	2	3	6	3	13	21	23	8	53	13	29	12	23	16	25
		Urban/Parkland/Fwy	32	8	0	0	4	7	6	26	4	4	0	0	2	4	7	13	9	14
		Rural	33	8	33	81	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(23)	Land classification	Capital city/Metro-politan/major urban	283	72	0	0	37	67	19	83	81	87	15	100	36	80	43	81	52	80
		Towns (200-100,000)	66	17	0	0	18	33	4	17	12	13	0	0	9	20	10	19	13	20
		Rural	43	11	43	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(24)	Road type	Rural	44	11	44	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		Highways - urban	59	15	0	0	7	13	5	22	10	11	1	7	10	22	11	21	15	23
		Major arterial	137	35	0	0	18	33	9	39	41	44	9	60	19	42	21	40	20	31
		Other urban	153	39	0	0	30	54	9	39	42	45	5	33	16	36	21	40	30	46
(25)	Road configuration	2-way undivided	233	71	38	91	37	80	8	53	58	72	7	64	22	60	24	57	39	72
		Divided, dual carriageway, freeway	95	29	4	10	9	20	7	47	23	28	4	36	15	41	18	43	15	28
(26)	Type of median	None	62	20	8	20	16	36	5	36	9	11	2	22	6	16	4	10	12	24
		Low/narrow	65	20	3	7	5	11	4	29	18	23	3	33	11	28	13	31	8	16
		Not easily/intermittently surmountable	23	7	1	2	2	5	2	14	5	6	0	0	4	10	4	10	5	10
		Painted	169	53	29	71	21	48	3	21	48	60	4	45	18	46	21	50	25	50
(27)	Number of lanes	1 or 2	216	65	41	98	39	85	10	67	42	52	4	33	20	50	25	58	35	64
		3	57	17	1	2	3	6	3	20	16	20	5	42	11	28	9	21	9	16
		4 or 6	61	18	0	0	4	9	2	13	23	28	3	25	9	23	9	21	11	20

Variable		Categories	Population n	Population %	1. Rural n	1. Rural %	2. School Student/ Preschooler n	2. School Student/ Preschooler %	3. 8 to 18 Years of Age n	3. 8 to 18 Years of Age %	4. Elderly Retired n	4. Elderly Retired %	5. Heavy Vehicle n	5. Heavy Vehicle %	6. Adult Day n	6. Adult Day %	7. Adult Alcohol n	7. Adult Alcohol %	8. Adult Other n	8. Adult Other %
(28)	Horizontal road alignment	Mid-block straight	277	71	35	79	42	76	15	65	63	68	12	80	30	67	34	64	46	71
		Mid-block curved/ near curve	28	7	6	14	3	6	0	0	4	4	0	0	4	9	6	11	5	8
		Intersection	88	22	3	7	10	18	8	35	26	28	3	20	11	24	13	25	14	21
(29)	Speed limit at crash location	60	292	74	5	12	49	89	16	70	87	94	13	87	37	82	39	74	46	71
		70-90	57	15	8	19	3	6	3	13	6	7	2	13	8	18	12	23	15	23
		100-110	43	11	30	70	3	6	4	17	0	0	0	0	0	0	2	4	4	6
(30)	Traffic controls	Lights	43	11	0	0	4	7	5	22	8	9	3	20	9	20	7	13	7	11
		Non-electric eg stop	48	12	3	7	5	9	1	4	21	23	2	13	7	15	3	6	6	9
		None mid-block	264	67	38	86	40	73	14	61	51	55	9	60	26	58	39	74	47	72
		None intersection	38	10	3	7	6	11	3	13	13	14	1	7	3	7	4	7	5	8
(31)	DCA event	Near side/emerging	194	53	5	14	32	63	11	52	55	61	11	79	30	71	29	55	21	34
		Far side	82	22	6	17	17	33	1	5	29	32	3	21	6	14	8	15	12	20
		On carriageway	61	17	16	44	2	4	4	19	5	6	0	0	5	12	13	24	16	26
		With traffic	31	8	9	25	0	0	5	24	1	1	0	0	1	2	3	6	12	20
(32)	Time day/ day of week	Week day	155	39	11	25	31	56	12	52	52	56	11	73	38	84	0	0	0	0
		Weekend day	38	10	4	9	13	24	0	0	13	14	1	7	7	16	0	0	0	0
		Week night	78	20	12	27	5	9	3	13	17	18	1	7	0	0	18	34	22	34
		Weekend night	122	31	17	39	6	11	8	35	11	12	2	13	0	0	35	66	43	66
(33)	Time	Day	193	49	15	34	44	80	12	52	65	70	12	80	45	100	0	0	0	0
		Night	200	51	29	66	11	20	11	48	28	30	3	20	0	0	53	100	65	100
(34)	Day	Weekday	233	59	23	52	36	66	15	65	69	74	12	80	38	84	18	34	22	34
		Weekend	160	41	21	48	19	35	8	35	24	26	3	20	7	16	35	66	43	66

Notes 1. Unknowns and very small categories omitted.
2. Similar categories with small numbers combined.

3.3. Further interpretation of the eight cluster solution

The main interpretation of the eight cluster solution is contained in the next chapter. Because this stepdown procedure was quite novel, the resources on the INTERNET were used to solicit comments on the procedure. Cluster "experts" who subscribe to the CLASS-L list were asked their opinion on the procedure followed. Particularly, subscribers were asked to comment on the legitimacy of the approach and the problems which might be inherent. Several informative comments were received and are worthy of reporting.

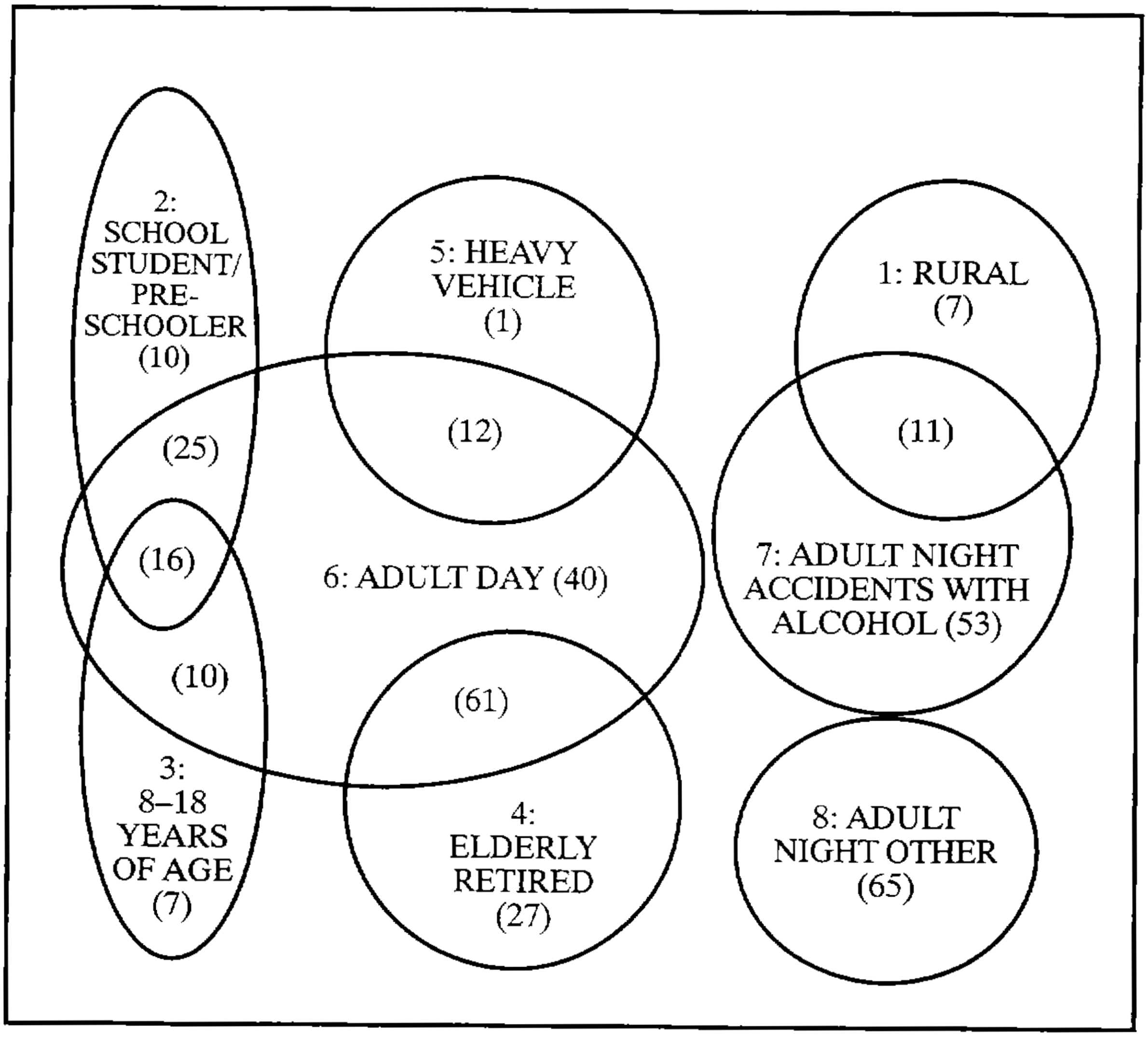
Firstly, the strategy was seen as legitimate but that it required careful interpretation. Professor James Rohlf (1995), commented that, "that type of procedure is often done. Sort of like using higher powers on a microscope. Main problem is that you are clustering by eye and hence others will not know what criteria you are using and it becomes very subjective."

Secondly, Guarino R. Colli (1995), commented that,

> With regards to your questions, I would like to point out the following:
>
> (1) The k-means procedure results do not seem surprising, for the procedure partitions the data into clusters which do not contain any other clusters. If the purpose of the analysis was to exhaust the data in order to find clusters within clusters, then a hierarchical clustering procedure should be used instead.
>
> (2) When you perform successive multiple correspondence analyses what you're doing is basically what a hierarchical clustering procedure will do, however it seems that cluster analysis is not appropriate for categorical data.
>
> (3) I would be cautious when interpreting the results produced by successive runs of multiple correspondence analysis, for the categories produced cannot be assigned the same ranks For example, if you take frogs, gorillas, chimps, and members of three human races, after the first run you are likely to get members of the frog category assigned to one cluster and all the others clumped into another. By removing members of the frog category you will be able to discern other clusters within the groups left. However, by the end, it will be misleading to assign members of each of the clusters obtained to the same rank. Hence, if the procedure you suggested is valid or not will depend on the question being asked at the beginning. If the goal is to identify distinct categories among the whole data set, I would stop after the first run. If the purpose is to reduce the data set to every possible recognizable category, then you will ultimately end up at the level of the individual, i.e each observation.

These are indeed important considerations. It is important when interpreting the obtained cluster solution generated by this "stepdown" procedure that the reader be aware of the overlap of clusters. The strategy has produced clusters with some overlap rather than distinct clusters. The interpretation in the next chapter should not interpret the assigning of the numbers 1 to 8 to the clusters as designating eight distinct clusters. The actual amount of overlap is schematically shown in the Figure 3.2 which displays the major areas of overlap between the eight clusters and represents 88% (345 cases) of the cases. Only overlaps which include at least 10 cases are included in this diagram. The remaining 48 cases (12%), represent other overlaps of minor importance.

Figure 3.2. Obtained Cluster Solution Showing Overlap



3.4. Validation of the obtained clustering solution

There are many methods recommended to determine the adequacy of clustering solutions (Blashfield et al, 1982). Among these are replication procedures, data alteration procedures, external criterion procedures and deletion procedures. The procedure employed here was based on the approach developed by Ehrenberg (1981)

Ehrenberg's (1981) approach is based on using prior knowledge to analyse new sets of data. Once the same results are found for different sets of data and different conditions of observations the results become practical and useful. In Ehrenberg's terminology the results become "lawlike", (1981, p. 65) Of course, if the same results are not found then a generalisable relationship was not found and the discrepancy would need to be interpreted. In terms of replicating cluster solutions, historical factors may often change clusters in a sample over time (Rapkin and Luke, 1993).

In this way, the 1988 pedestrian fatalities data file was analysed using the information gained by analysing the 1990 fatal pedestrian crashes. No attempt was made to separately cluster analyse the 1988 file from scratch. Rather, the validation process began at step five of the overall strategy. Homogeneity analysis was performed on the 1988 crash file using the variables established as important for the definitive data-analytic investigation of the 1990 file. For the 1990 file, this consisted of 18 variables. For the 1988 file, this consisted of 17 variables as the variable pedestrian height was not available on the 1988 file. As well, because of coding differences between 1988 and 1990 some of the variables were categorised slightly differently

However, even with these differences, there was a remarkable confirmation of the 1990 "stepdown" clustering solution. Using the 1988 file, the first cluster identified was a rural cluster, followed by a young student cluster. The third cluster was a pre-drinking student cluster followed by the elderly retired cluster. The remaining four clusters were also identified in space, although the heavy vehicle cluster was less distinct. Adding to this confirmation was the close similarity of the locations of the 1988 and 1990 clusters in the two-dimensional homogeneity spaces. The result of the investigation with the 1988 file was strong confidence in the solution gained with the 1990 pedestrian fatalities file. In Ehrenberg's (1981) terms, as the same results were found for the second set of data, the original results can be viewed as practical and useful.

Chapter 4: Interpretation of the Obtained Clustering Solution

In chapter three the data file was reduced through a "stepdown" procedure, to eight recognisable categories of fatal pedestrian crashes. The seventh step involves describing the obtained clustering solution.

In this chapter the definitions and defining characteristics of each of these eight clusters are separately presented. Following each cluster definition, a table is presented which compares the cluster to the population of pedestrian fatalities. In each table, variables which differ by over 10% between the cluster and the population are flagged with an upwards or downwards arrow signifying the direction of difference. This figure was arbitrarily chosen as a working rule-of-thumb to give consistency to the reporting. The results are also summarised in point form under the headings "environment", "driver", "vehicle", "timing", "pedestrian" and "crash" as appropriate

As decided with FORS, each cluster is compared to the population using the 34 important variables which had been used or put aside as a result of the separate homogeneity analyses on each family. The previous 35 variables which had been put aside as a result of not meeting the statistical criteria were not employed in these comparisons (see page 15).

The clusters are reported as though they are of equal rank. However, as has been described previously, there is overlap of clusters (see Figure 3.2) This overlap diagram needs to be kept in mind when reading this interpretation. For example, there is overlap between the Rural Cluster (Cluster 1) and the Adult Night Accidents with Alcohol Cluster (Cluster 7). Some of the Rural fatal pedestrian crashes share characteristics with Cluster 7 Had Cluster 7 been extracted before Cluster 1 then these shared crashes would have been allocated to Cluster 7. For now, however, the clusters are interpreted in the order of their emergence

4.1. Cluster 1 - Rural

4.1.1. Definition of cluster: Variable Road Type (Categories 1, 3, 5)

		Cluster 1		Total	
		n	%	n	%
1.	National Highway/Rural	8	18	8	2
2	National Highway within Urban Boundary			19	5
3.	State Highway/Rural	9	21	9	2
4	State Highway within Urban Boundary			40	10
5.	Other Rural Road	27	61	27	7
6.	Major Arterial Urban Road			137	35
7	Other Urban Road			153	39
		44		393	

For the purpose of presentation, categories 1, 3, 5 have been combined to form a Rural category and categories 2, 4 have been combined to form a Highways-urban category.

4.1.2. Characteristics of cluster

(i) Environment

- The landuse adjacent to the crash site was rural rather than residential, commercial or industrial.
- The land was rural rather than in a capital city, metropolitan area, major urban area or in a town.
- The road was 2-way undivided rather than divided, dual carriageway or freeway.
- The median was painted rather than low or narrow.
- The road had one or two lanes
- The speed limit was 100-110 rather than 60 kph.
- There were no traffic controls e.g., lights.
- The crash happened mid block rather than at an intersection.

(ii) Driver

- The driver was likely to be further rather than close to home.

(iii) Vehicle

- The vehicle was more likely to be classified as a heavy vehicle.

(iv) Timing

- The crash was more likely to have happened at night and less likely in the day during the week.

(v) Pedestrian

- The pedestrian was more likely to·
    - be male
    - be aged 19-35 years rather than older
    - be unemployed or employed as a plant/machine operator or labourer rather than be retired or a pensioner
    - be tall (170-190 cms) rather than average height (149-169 cms)
    - have a BAC of greater than .15 rather than zero

- The cause of death was more likely to result from body injuries rather than multiple injuries.

- The severity of the injury was less likely to be in the 36-45 (relatively high) range.

(vi) Crash

- The major cause of the crash was more likely to be alcohol or drugs.

- The crash was more likely to happen when the pedestrian was walking with the traffic or playing/ working/lying/standing on the road rather than being on the nearside when crossing the road or emerging from behind a car.

TABLE 4.1. Areas of Difference between the Rural Cluster and the Population of Pedestrian Fatalities

	Variable	Categories	Population		Rural Cluster		
			n	%	n	%	
(1)	Distance from home of driver	1-10 km	253	68	19	51	↓
		11+ km	119	32	18	49	↑
(3)	Weight class of vehicle	Motor cycle	14	4	0	0	
		Light car	63	17	7	18	
		Medium car	94	25	8	21	
		Heavy car	109	29	11	28	
		Van/light truck	64	17	6	15	
		Heavy vehicle	32	8	7	18	↑
(4)	Speed limit/speeding	<80 not speed	267	72	6	17	↓
		<80 speed	39	10	0	0	↓
		80+ not speed	62	17	29	80	↑
		80+ speed	4	1	1	3	
(13)	Major cause of accident	Alcohol/Drugs	138	42	24	63	↑
		Sensory impairment	6	2	0	0	
		Vision obscured	9	3	0	0	
		Not see other road user	49	15	5	13	
		Excessive speed	9	3	0	0	
		Ignore traffic controls	14	4	0	0	
		Pedestrian step into path	101	31	9	24	
(14)	Pedestrian age	1-18 yrs	90	23	12	27	
		19-35 yrs	97	25	22	50	↑
		36-65 yrs	104	26	7	16	↓
		66+ yrs	102	26	3	7	↓
(15)	Pedestrian sex	Male	283	72	36	82	↑
		Female	110	28	8	18	↓

Variable		Categories	Population		Rural Cluster		
			n	%	n	%	
(16)	Pedestrian employment status	Pre-school/baby	20	6	2	5	
		Manager/admin/prof/paraprof	18	5	0	0	
		Trades persons	22	6	2	5	
		Clerical/sales	12	3	0	0	
		Plant/machine/labour	47	13	9	24	↑
		Housekeeping	13	4	1	3	
		Unemployed	34	9	11	30	↑
		Student	57	16	6	16	
		Retired/pensioner	135	38	6	16	↓
(17)	Pedestrian height	80-145 cms	64	16	6	14	
		149-169 cms	142	36	10	23	↓
		170-195 cms	187	48	28	64	↑
(18)	Cause of death	Head	53	14	8	19	
		Body	50	13	10	23	↑
		Multiple	279	73	25	58	↓
(19)	ISS Score	9-26	85	24	12	32	
		27-35	89	25	7	18	
		36-45	95	26	6	16	↓
		48-75	89	25	13	34	
(20)	Pedestrian BAC	0	143	55	7	25	↓
		.001-.15	35	13	3	11	
		>.15	82	32	18	64	↑
(22)	Landuse adjacent to crash site	Residential	153	40	6	15	↓
		Residential/Commercial	92	24	1	2	↓
		Commercial/Industrial	77	20	1	2	↓
		Urban parkland/Freeway	32	8	0	0	
		Rural	33	8	33	81	↑
(23)	Land Classification	Capital city/Metro. Maj. Urban	283	72	0	0	↓
		Towns (200-100.000)	66	17	0	0	↓
		Rural	43	11	43	100	↑

	Variable	Categories	Population n	Population %	Rural Cluster n	Rural Cluster %	
(24)	Road type (DEFINITION)	Rural	44	11	44	100	↑
		Highways - urban	59	15	0	0	↓
		Major arterial	137	35	0	0	↓
		Other urban	153	39	0	0	↓
(25)	Road configuration	2-way undivided	233	71	38	91	↑
		Divided, dual carriage, freeway	95	29	4	10	↓
(26)	Type of median	None	62	20	8	20	
		Low/narrow	65	20	3	7	↓
		Not easily/intermittently surmountable	23	7	1	2	
		Painted	169	53	29	71	↑
(27)	Number of lanes	1 or 2	216	65	41	98	↑
		3	57	17	1	2	↓
		4 or 6	61	18	0	0	↓
(28)	Horizontal road alignment	Mid block straight	277	71	35	79	
		Mid block curved/near curve	28	7	6	14	
		Intersection	88	22	3	7	↓
(29)	Speed limit at crash location	60	292	74	5	12	↓
		70-90	57	15	8	19	
		100-110	43	11	30	70	↑
(30)	Traffic controls	Lights	43	11	0	0	↓
		Non-electric e.g. stop	48	12	3	7	
		None mid-block	264	67	38	86	↑
		None intersection	38	10	3	7	
(31)	DCA event	Near side/emerging	194	53	5	14	↓
		Far side	82	22	6	17	
		On carriageway	61	17	16	44	↑
		With traffic	31	8	9	25	↑

	Variable	Categories	Population		Rural Cluster		
			n	%	n	%	
(32)	Time of day/day of week	Week day	155	39	11	25	↓
		Weekend day	38	10	4	9	
		Week night	78	20	12	27	
		Weekend night	122	31	17	39	
(33)	Time	Day	193	49	15	34	↓
		Night	200	51	29	66	↑

4.2. Cluster 2 - School Student/Preschooler

4.2.1. Definition of cluster: Variable C24a Height of Person Killed (Category 1)

		Cluster 2 n	Cluster 2 %	Total n	Total %
1.	80 - 145 cms	55	100	64	16
2.	146 - 169 cms			142	36
3.	170 - 179 cms			147	38
4	180 - 195 cms			40	10
		55		393	

For the purpose of presentation, categories 3 and 4 have been combined

and Variable C7 Employment Status/Occupation - Pedestrian (Categories 0, 14)

		Cluster 2 n	Cluster 2 %	Total n	Total %
0	NA/Preschool or Baby	18	33	20	5
1.	Manager/Administrator			6	2
2.	Professional			9	2
3.	Military Service Personnel			5	1
4	Para-professional			3	1
5.	Tradespersons			22	6
6	Clerical			7	2
7.	Sales & Personal Service Worker			5	1
8.	Plans & Machine Operator			13	3
9	Labourers & Related Workers			34	9
10.	Keeping House			13	3
11.	Employed: Other Occupation			1	0
12.	Employed: Unspecified			0	0
13.	Unemployed			34	9
14	At School, University, College	37	67	57	15
15.	Retired/Pensioner			135	34
16.	Other			2	1
99.	Unknown			27	7
		55		393	

For the purpose of presentation, categories 3, 11, 16, 99 have been removed. Categories 1, 2, 4 have been combined to form a Manager/Administrator/Professional/Para-professional category; categories 6, 7 have been combined to form a Clerical/Sales category and categories 8, 9 have been combined to form a Plant/Machine/Labour category.

4.2.2. Characteristics of cluster

(i) Pedestrian

- The pedestrian was more likely to:
  - be aged up to 18 years rather than older
  - have a BAC reading of zero.

- The cause of death was less likely to be from multiple injuries.

(ii) Driver

- The driver was more likely to:
  - be closer rather than further from home
  - be female.

(iii) Timing

- The crash was more likely to occur during the day rather than at night.

(iv) Environment

- The landuse adjacent to the crash site was more likely to be residential rather than commercial or industrial.

- The land was more likely to be in towns (200 - 100,000) rather than in a rural area.

- The road type was relatively minor urban rather than rural.

- There was less likely to be a median strip.

- There were more likely to be 1 or 2 lanes rather than 3.

- The speed limit was more likely to be 60 kph.

(v) Crash

- The responsibility for the crash was more likely to lie with the pedestrian.

- The cause was more likely to be due to the pedestrian stepping into the path of the car or someone not seeing the other road user rather than to alcohol or drugs.

- The pedestrian was more likely to be crossing the road (near side, far side or emerging) rather than being on the road (playing, working, lying or standing).

TABLE 4.2. **Areas of Difference between the School Student/Preschooler Cluster and the Population of Pedestrian Fatalities**

	Variable	Categories	Population		School Student/ Preschooler Cluster		
			n	%	n	%	
(1)	Distance from home of driver	1-10 km	253	68	46	84	↑
		11+ km	119	32	9	16	↓
(4)	Speed limit/speeding	<80 not speed	267	72	46	84	↑
		<80 speed	39	10	5	9	
		80+ not speed	62	17	4	7	↓
		80+ speed	4	1	0	0	
(9)	Driver sex	Male	305	81	36	66	↓
		Female	73	19	19	35	↑
(12)	Unit responsible for crash	Driver	50	13	3	6	
		Driver and pedestrian	43	11	3	6	
		Pedestrian	291	76	48	89	↑
(13)	Major cause of accident	Alcohol/Drugs	138	42	4	8	↓
		Sensory impairment	6	2	0	0	
		Vision obscured	9	3	0	0	
		Not see other road user	49	15	13	26	↑
		Excessive speed	9	3	2	4	
		Ignore traffic controls	14	4	1	2	
		Pedestrian step into path	101	31	30	60	↑
(14)	Pedestrian age	1-18 yrs	90	23	55	100	↑
		19-35 yrs	97	25	0	0	↓
		36-65 yrs	104	26	0	0	↓
		66+ yrs	102	26	0	0	↓

Variable		Categories	Population		School Student/ Preschooler Cluster		
			n	%	n	%	
(16)	Pedestrian employment status (PART DEFINITION)	Pre-school/baby	20	6	18	33	↑
		Manager/admin/prof/paraprof	18	5	0	0	
		Trades persons	22	6	0	0	
		Clerical/sales	12	3	0	0	
		Plant/machine/labour	47	13	0	0	↓
		Housekeeping	13	4	0	0	
		Unemployed	34	9	0	0	
		Student	57	16	37	67	↑
		Retired/pensioner	135	38	0	0	↓
(17)	Pedestrian height (PART DEFINITION)	80-145 cms	64	16	55	100	↑
		149-169 cms	142	36	0	0	↓
		170-195 cms	187	48	0	0	↓
(18)	Cause of death	Head	53	14	12	22	
		Body	50	13	8	15	
		Multiple	279	73	34	63	↓
(20)	Pedestrian BAC	0	143	55	13	100	↑
		001-.15	35	13	0	0	↓
		>.15	82	32	0	0	↓
(22)	Landuse adjacent to crash site	Residential	153	40	36	66	↑
		Residential/Commercial	92	24	12	22	
		Commercial/Industrial	77	20	3	6	↓
		Urban parkland/Freeway	32	8	4	7	
		Rural	33	8	0	0	
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	37	67	
		Towns (200-100,000)	66	17	18	33	↑
		Rural	43	11	0	0	↓

	Variable	Categories	Population		School Student/ Preschooler Cluster		
			n	%	n	%	
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	7	13	
		Major arterial	137	35	18	33	
		Other urban	153	39	30	54	↑
(26)	Type of median	None	62	20	16	36	↑
		Low/narrow	65	20	5	11	
		Not easily/intermittently surmountable	23	7	2	5	
		Painted	169	53	21	48	
(27)	Number of lanes	1 or 2	216	65	39	85	↑
		3	57	17	3	6	↓
		4 or 6	61	18	4	9	
(29)	Speed limit at crash location	60	292	74	49	89	↑
		70-90	57	15	3	6	
		100-110	43	11	3	6	
(31)	DCA event	Near side/emerging	194	53	32	63	↑
		Far side	82	22	17	33	↑
		On carriageway	61	17	2	4	↓
		With traffic	31	8	0	0	
(32)	Time of day/day of week	Week day	155	39	32	56	↑
		Weekend day	38	10	13	24	↑
		Week night	78	20	5	9	↓
		Weekend night	122	31	6	11	↓
(33)	Time	Day	193	49	44	80	↑
		Night	200	51	11	20	↓

4.3. Cluster 3 - 8-18 Years of Age

4.3.1. Definition. Variable C5 - Age-pedestrian (Category 2)

		Cluster 3		Total	
		n	%	n	%
1.	1 - 7 years			43	11
2.	8 - 18 years	23	100	47	12
3	19 - 25 years			46	12
4.	26 - 35 years			51	13
5.	36 - 50 years			52	13
6	57 - 65 years			52	13
7.	66 - 75 years			41	10
8.	76 - 98 years			61	16
		23		393	

For the purpose of presentation, categories 1 & 2, categories 3 & 4, categories 5 & 6, categories 7 & 8 have been combined.

4.3 2 Characteristics of cluster

(i) Pedestrian

- The pedestrian was more likely to:
  - be female
  - be a student rather than retired or a pensioner
  - be of medium height (149-169 cms) rather than short (80-145 cms).
- The pedestrian was less likely to have a BAC reading >.15.
- The severity of the injury was more likely to be in the 36-45 (relatively high) range and less likely to be in the 27-35 (relatively low) range.

(ii) Driver

- The driver was more likely to:
  - be aged 26-92 years rather than younger
  - have a BAC reading above zero
  - be further rather than closer to home.
- The driver was less likely to have braked or swerved

(iii) Vehicle

- The vehicle was more likely to:
  - be a van or a light truck rather than a heavy car
  - be manufactured more recently (1985-90)

(iv) Timing

- The crash was more likely to have occurred.
  - on a weekday rather than a weekend day
  - on Thursday or Friday rather than Monday-Wednesday.

(v) Environment

- The landuse adjacent to the crash site was more likely to be residential/commercial or urban parkland/freeway rather than residential.

- The land was more likely to be in a capital city, metropolitan or major urban area rather than a rural area.

- The road type was less likely to be rural.

- The road was more likely to be divided, dual carriageway or freeway and less likely to be 2-way undivided.

- There was more likely to be no median and less likely to be a painted median

- The crash was more likely to occur at an intersection.

- There were more likely to be traffic lights.

(vi) Crash

- The driver was more likely to be responsible for the crash with the pedestrian less likely.

- The pedestrian was more likely to be walking with the traffic and less likely to be crossing it and on the far side of the road.

TABLE 4.3. Areas of Difference between the 8-18 Years of Age Cluster and the Population of Pedestrian Fatalities

	Variable	Categories	Population		8-18 Years of Age Cluster		
			n	%	n	%	
(1)	Distance from home of driver	1-10 km	253	68	11	50	↓
		11+ km	119	32	11	50	↑
(3)	Weight class of vehicle	Motor cycle	14	4	0	0	
		Light car	63	17	3	14	
		Medium car	94	25	7	32	
		Heavy car	109	29	2	9	↓
		Van/light truck	64	17	7	32	↑
		Heavy vehicle	32	8	3	14	
(5)	Braked/swerved	Yes	195	54	8	40	↓
		No	163	46	12	60	↑
(7)	Year of manufacture	Old (1962-84)	196	58	7	39	↓
		New (1985-90)	142	42	11	61	↑
(8)	Driver age	16-25 yrs	159	43	7	32	↓
		26-92 yrs	212	57	15	68	↑
(11)	Blood alcohol concentration	Zero	273	89	12	75	↓
		Above zero	35	11	4	25	↑
(12)	Unit responsible for crash	Driver	50	13	6	26	↑
		Driver and pedestrian	43	11	3	13	
		Pedestrian	291	76	14	61	↓
(14)	Pedestrian age	1-18 yrs	90	23	23	100	↑
	(DEFINITION)	19-35 yrs	97	25	0	0	↓
		36-65 yrs	104	26	0	0	↓
		66+ yrs	102	26	0	0	↓

Variable		Categories	Population n	Population %	8-18 Years of Age Cluster n	8-18 Years of Age Cluster %	
(15)	Pedestrian sex	Male	283	72	13	57	↓
		Female	110	28	10	44	↑
(16)	Pedestrian employment status	Pre-school/baby	20	6	0	0	
		Manager/admin/prof/paraprof	18	5	0	0	
		Trades persons	22	6	3	15	
		Clerical/sales	12	3	2	10	
		Plant/machine/labour	47	13	2	10	
		Housekeeping	13	4	0	0	
		Unemployed	34	9	0	0	
		Student	57	16	12	60	↑
		Retired/pensioner	135	38	1	5	↓
(17)	Pedestrian height	80-145 cms	64	16	0	0	↓
		149-169 cms	142	36	14	61	↑
		170-195 cms	187	48	9	39	
(19)	ISS Score (Severity)	9-26	85	24	5	26	
		27-35	89	25	2	11	↓
		36-45	95	26	7	37	↑
		48-75	89	25	5	26	
(20)	Pedestrian BAC	0	143	55	9	60	
		.001-.15	35	13	3	20	
		>.15	82	32	3	20	↓
(21)	Day of week	Mon-Wed	152	39	6	26	↓
		Thurs-Fri	121	31	12	52	↑
		Sat	81	20	4	18	
		Sun	39	10	1	4	
(22)	Landuse adjacent to crash site	Residential	153	40	6	26	↓
		Residential/Commercial	92	24	8	35	↑
		Commercial/Industrial	77	20	3	13	
		Urban parkland/Freeway	32	8	6	26	↑
		Rural	33	8	0	0	

	Variable	Categories	Population		8-18 Years of Age Cluster		
			n	%	n	%	
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	19	83	↑
		Towns (200-100,000)	66	17	4	17	
		Rural	43	11	0	0	↓
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	5	22	
		Major arterial	137	35	9	39	
		Other urban	153	39	9	39	
(25)	Road configuration	2-way undivided	233	71	8	53	↓
		Divided, dual carriageway, fwy	95	29	7	47	↑
(26)	Type of median	None	62	20	5	36	↑
		Low/narrow	65	20	4	29	
		Not easily/intermittently surmountable	23	7	2	14	
		Painted	169	53	3	21	↓
(28)	Horizontal road alignment	Mid-block straight	277	71	15	65	
		Mid-block curved/near curve	28	7	0	0	
		Intersection	88	22	8	35	↑
(30)	Traffic controls	Lights	43	11	5	22	↑
		Non-electric e.g.stop	48	12	1	4	
		None mid-block	264	67	14	61	
		None intersection	38	10	3	13	
(31)	DCA event	Near side/emerging	194	53	11	52	
		Far side	82	22	1	5	↓
		On carriageway	61	17	4	19	
		With traffic	31	8	5	24	↑
(32)	Time of day/day of week	Week day	155	39	12	52	↑
		Weekend day	38	10	0	0	↓
		Week night	78	20	3	13	
		Weekend night	122	31	8	35	

4.4. Cluster 4 - Elderly Retired

4.4.1. Definition of cluster: Variable C5 Age-pedestrian (categories 7, 8)

		Cluster 4 n	Cluster 4 %	Total n	Total %
1.	1 - 7 years			43	11
2.	8 - 18 years			47	12
3.	19 - 25 years			46	12
4.	26 - 35 years			51	13
5.	36 - 50 years			52	13
6.	51 - 65 years			52	13
7.	66 - 75 years	36	39	41	10
8.	76 - 98 years	57	61	61	16
		93		393	

For the purpose of presentation, categories 1 & 2, categories 3 & 4, categories 5 & 6, categories 7 & 8 have been combined.

and: Variable C5 Employment Status/Occupation - Pedestrian (Category 15)

		Cluster 4 n	Cluster 4 %	Total n	Total %
0	N/A Preschool or Baby			20	5
1.	Manager/Administrator			6	2
2.	Professional			9	2
3.	Military Service Personnel			5	1
4.	Para-professional			3	1
5.	Tradespersons			22	6
6.	Clerical			7	2
7.	Sales & Personal Service Worker			5	1
8	Plant & Machine Operator			13	3
9.	Labourers & Related Workers			34	9
10.	Keeping House			13	3
11.	Employed: Other Occupation			1	0
12.	Employed: Unspecified			0	0
13.	Unemployed			34	9
14.	At School/University/College			57	15
15.	Retired/Pensioner	93	100	135	34
16.	Other			2	1
99.	Unknown			27	7
		93		393	

For the purpose of presentation, categories 3, 11,16, 99 have been removed. Categories 1, 2, 4 have been combined to form a Manager/Administrator/Professional/Para-professional category; categories 6, 7 have been combined to form a Clerical/Sales category and categories 8, 9 have been combined to form a Plant/Machine/Labour category.

4.4 2. Characteristics of cluster

(i) Pedestrian

- The pedestrian was more likely to be average height (149-169 cms) rather than taller or shorter.
- The pedestrian was less likely to have a BAC reading >.15 and more likely to have one of zero.

(ii) Driver

- The driver was more likely to have braked or swerved.

(iii) Timing

- The crash was more likely to have occurred during the week rather than the weekend and during the day rather than at night

(iv) Environment

- The crash occurred in a capital city/metropolitan or major urban rather than rural area.
- The road type was less likely to be rural.
- The number of lanes was more likely to be 4 or 6 and less likely to be 1 or 2.
- The speed limit was more likely to be 60 rather than 100-110 kph.
- There were more likely to be non-electric controls such as stop signs/pedestrian crossings with the accident less likely to occur mid block where there are no controls

(v) Crash

- The crash was less likely to be caused by drugs or alcohol.
- The pedestrian was more likely to be on the far side of the road crossing over it rather than on it (playing, working, lying or standing).

TABLE 4.4. Areas of Difference between the Elderly Retired Cluster and the Population of Pedestrian Fatalities

	Variable	Categories	Population		Elderly/Retired Cluster		
			n	%	n	%	
(4)	Speed limit/speeding	<80 not speed	267	72	82	90	↑
		<80 speed	39	10	7	8	
		80+ not speed	62	17	2	2	↓
		80+ speed	4	1	0	0	
(5)	Braked/swerved	Yes	195	54	55	66	↑
		No	163	46	29	35	↓
(13)	Major cause of accident	Alcohol/Drugs	138	42	10	12	↓
		Sensory impairment	6	2	6	8	
		Vision obscured	9	3	5	6	
		Not see other road user	49	15	17	21	
		Excessive speed	9	3	5	6	
		Ignore traffic controls	14	4	7	9	
		Pedestrian step into path	101	31	31	38	
(14)	Pedestrian age (PART DEFINITION)	1-18 yrs	90	23	0	0	↓
		19-35 yrs	97	25	0	0	↓
		36-65 yrs	104	26	0	0	↓
		66+ yrs	102	26	93	100	↑
(16)	Pedestrian employment status (PART DEFINITION)	Pre-school/baby	20	6	0	0	
		Manager/admin/prof/paraprof	18	5	0	0	
		Trades persons	22	6	0	0	
		Clerical/sales	12	3	0	0	
		Plant/machine/labour	47	13	0	0	↓
		Housekeeping	13	4	0	0	
		Unemployed	34	9	0	0	
		Student	57	16	0	0	↓
		Retired/pensioner	135	38	93	100	↑

Variable		Categories	Population		Elderly/Retired Cluster		
			n	%	n	%	
(17)	Pedestrian height	80-145 cms	64	16	2	2	↓
		149-169 cms	142	36	60	65	↑
		170-195 cms	187	48	31	33	↓
(20)	Pedestrian BAC	0	143	55	55	85	↑
		.001-.15	35	13	9	14	
		>.15	82	32	1	1	↓
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	81	87	↑
		Towns (200-100,000)	66	17	12	13	
		Rural	43	11	0	0	↓
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	10	11	
		Major arterial	137	35	41	44	
		Other urban	153	39	42	45	
(27)	Number of lanes	1 or 2	216	65	42	52	↓
		3	57	17	16	20	
		4 or 6	61	18	23	28	↑
(29)	Speed limit at crash location	60	292	74	87	94	↑
		70-90	57	15	6	7	
		100-110	43	11	0	0	↓
(30)	Traffic controls	Lights	43	11	8	9	
		Non electric e.g.stop	48	12	21	23	↑
		None mid-block	264	67	51	55	↓
		None intersection	38	10	13	14	
(31)	DCA event	Near side/emerging	194	53	55	61	
		Far side	82	22	29	32	↑
		On carriageway	61	17	5	6	↓
		With traffic	31	8	1	1	

	Variable	Categories	Population		Elderly/Retired Cluster		
			n	%	n	%	
(32)	Time of day/day of week	Week day	155	39	52	56	↑
		Weekend day	38	10	13	14	
		Week night	78	20	17	18	
		Weekend night	122	31	11	12	↓
(33)	Time	Day	193	49	65	70	↑
		Night	200	51	28	30	↓
(34)	Day	Weekday	233	59	69	74	↑
		Weekend	160	41	24	26	↓

4.5. Cluster 5 - Heavy Vehicle

4.5.1. Definition of cluster: Variable Weight class of vehicle (Category 7)

		Cluster 5		Total	
		n	%	n	%
1	Bike			1	0
2.	Motor cycle			14	4
3.	Light car			63	17
4.	Medium car			94	25
5.	Heavy car			109	29
6.	Van/Light truck			64	17
7.	Heavy vehicle	15	100	32	8
		15		377	

For the purpose of presentation, category 1 has been removed.

4.5.2. Characteristics of cluster

(i) Vehicle

- The vehicle was likely to have been manufactured earlier (1962-84) rather than later (1985-90)

(ii) Driver

- The driver was more likely to·
  - be further (11+ kms) rather than close to home
  - have braked or swerved
  - be aged 26 years or older
  - be male
  - not be wearing a seatbelt
  - have zero BAC.

(iii) Environment

- The landuse adjacent to the crash site was more likely to be commercial/industrial and less likely to be residential.

- The crash was more likely to have occurred in a capital city, metropolitan or major urban area rather than in towns or rural areas

- The road type was more likely to be major arterial rather than rural.

- The median was more likely to be low or narrow.

- The number of lanes was more likely to be three rather than one or two.

- The speed limit was more likely to be 60 rather than 100-110 kph

(iv) Timing

- The crash was more likely to have occurred in the day time during the week rather than at the weekend.

- It was more likely to have occurred on Monday to Wednesday rather than Thursday or Friday.

(v) Pedestrian

- The pedestrian was more likely to
  - be aged 19-65 rather than 1-18 or over 66 years
  - be female
  - be retired or a pensioner
  - be average height (149-169 cms) rather than shorter
  - have zero BAC.

- The cause of death was more likely to be from multiple injuries rather than head or body injuries only.

- The severity of the injury was less likely to be in the 27-35 (relatively low) range.

(vi) Crash

- The point of primary impact was more likely to be on the left hand side of the vehicle and less likely to be the front.

- The responsibility for the crash was more likely to rest with the pedestrian

- The major cause of the accident was more likely to be the pedestrian ignoring traffic controls and stepping into the path of the vehicle and less likely to be due to drugs or alcohol

- The crash was more likely to occur when the pedestrian was crossing the road and was on the near side or emerging from behind a vehicle It was less likely to occur when the pedestrian was on the road (working/playing/lying/standing).

TABLE 4.5. Areas of Difference between the Heavy Vehicle Cluster and the Population of Pedestrian Fatalities

	Variable	Categories	Population n	Population %	Heavy Vehicle Cluster n	Heavy Vehicle Cluster %	
(1)	Distance from home of driver	1-10 km	253	68	8	53	↓
		11+ km	119	32	7	47	↑
(3)	Weight class of vehicle	Motor cycle	14	4	0	0	
	(DEFINITION)	*Light car*	63	17	0	0	↓
		Medium car	94	25	0	0	↓
		Heavy car	109	29	0	0	↓
		Van/light truck	64	17	0	0	↓
		Heavy vehicle	32	8	15	100	↑
(4)	Speed limit/speeding	<80 not speed	267	72	13	87	↑
		<80 speed	39	10	1	7	
		80+ not speed	62	17	1	7	↓
		80+ speed	4	1	0	0	
(5)	Braked/swerved	Yes	195	54	10	71	↑
		No	163	46	4	29	↓
(6)	Point of primary impact	Left side	153	40	9	60	↑
		Front	120	31	2	13	↓
		Undercarriage	17	4	1	7	
		Right side	95	25	3	20	
(7)	Year of manufacture	Old (1962-84)	196	58	11	73	↑
		New (1985-90)	142	42	4	27	↓
(8)	Driver age	16-25 yrs	159	43	3	20	↓
		26-92 yrs	212	57	12	80	↑
(9)	Driver sex	Male	305	81	14	93	↑
		Female	73	19	1	7	↓

Variable	Categories	Population		Heavy Vehicle Cluster		
		n	%	n	%	
(10) Seat belt	Worn	230	92	4	36	↓
	Not worn	20	8	7	64	↑
(11) Blood alcohol concentration	Zero	273	89	13	100	↑
	Above zero	35	11	0	0	↓
(12) Unit responsible for crash	Driver	50	13	1	7	
	Driver and pedestrian	43	11	0	0	↓
	Pedestrian	291	76	14	93	↑
(13) Major cause of accident	Alcohol/Drugs	138	42	3	23	↓
	Sensory impairment	6	2	0	0	
	Vision obscured	9	3	0	0	
	Not see other road user	49	15	2	15	
	Excessive speed	9	3	0	0	
	Ignore traffic controls	14	4	2	15	↑
	Pedestrian step into path	101	31	6	46	↑
(14) Pedestrian age	1-18 yrs	90	23	0	0	↓
	19-35 yrs	97	25	6	40	↑
	36-65 yrs	104	26	8	53	↑
	66+ yrs	102	26	1	7	↓
(15) Pedestrian sex	Male	283	72	9	60	↓
	Female	110	28	6	40	↑
(16) Pedestrian employment status	Pre-school/baby	20	6	0	0	
	Manager/admin/prof/paraprof	18	5	1	7	
	Trades persons	22	6	2	13	
	Clerical/sales	12	3	0	0	
	Plant/machine/labour	47	13	1	7	
	Housekeeping	13	4	1	7	
	Unemployed	34	9	1	7	
	Student	57	16	1	7	
	Retired/pensioner	135	38	8	53	↑

	Variable	Categories	Population		Heavy Vehicle Cluster		
			n	%	n	%	
(17)	Pedestrian height	80-145 cms	64	16	0	0	↓
		149-169 cms	142	36	7	47	↑
		170-195 cms	187	48	8	53	
(18)	Cause of death	Head	53	14	0	0	↓
		Body	50	13	0	0	↓
		Multiple	279	73	14	100	↑
(19)	ISS Score (Severity)	9-26	85	24	4	27	
		27-35	89	25	2	13	↓
		36-45	95	26	5	33	
		48-75	89	25	4	27	
(20)	Pedestrian BAC	0	143	55	9	82	↑
		.001- 15	35	13	0	0	↓
		>.15	82	32	2	18	↓
(21)	Day of week	Mon-Wed	152	39	10	67	↑
		Thurs-Fri	121	31	3	20	↓
		Sat	81	20	2	13	
		Sun	39	10	0	0	↓
(22)	Landuse adjacent to crash site	Residential	153	40	3	20	↓
		Residential/Commercial	92	24	4	27	
		Commercial/Industrial	77	20	8	53	↑
		Urban parkland/Freeway	32	8	0	0	
		Rural	33	8	0	0	
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	15	100	↑
		Towns (200-100,000)	66	17	0	0	↓
		Rural	43	11	0	0	↓
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	1	7	
		Major arterial	137	35	9	60	↑
		Other urban	153	39	5	33	

Variable		Categories	Population n	Population %	Heavy Vehicle Cluster n	Heavy Vehicle Cluster %	
(26)	Type of median	None	62	20	2	22	
		Low/narrow	65	20	3	33	↑
		Not easily/intermittently surmountable	23	7	0	0	
		Painted	169	53	4	45	
(27)	Number of lanes	1 or 2	216	65	4	33	↓
		3	57	17	5	42	↑
		4 or 6	61	18	3	25	
(29)	Speed limit at crash location	60	292	74	13	87	↑
		70-90	57	15	2	13	
		100-110	43	11	0	0	↓
(31)	DCA event	Near side/emerging	194	53	11	79	↑
		Far side	82	22	3	21	
		On carriageway	61	17	0	0	↓
		With traffic	31	8	0	0	
(32)	Time of day/day of week	Week day	155	39	11	73	↑
		Weekend day	38	10	1	7	
		Week night	78	20	1	7	↓
		Weekend night	122	31	2	13	↓
(33)	Time	Day	193	49	12	80	↑
		Night	200	51	3	20	↓
(34)	Day	Weekday	233	59	12	80	↑
		Weekend	160	41	3	20	↓

4.6. Cluster 6 - Adult Day

4.6.1. Definition of cluster: Variable Time of day/week (Categories 1, 2,)

		Cluster 6		Total	
		n	%	n	%
1.	Week day	38	84	155	39
2.	Weekend day	7	16	38	10
3.	Week night			78	20
4.	Weekend night			122	31
		45		393	

4.6 2. Characteristics of cluster

(i) Timing

- The crash was more likely to occur on a weekday rather than at the weekend.
- It was more likely to occur from Monday to Wednesday.

(ii) Environment

- The crash was less likely to be in a rural area or on a rural-type road.
- The road was more likely to be divided, dual-carriageway or freeway and less likely to be 2-way undivided.
- The number of lanes was more likely to be 3 and less likely to be 1 or 2.
- The speed limit was less likely to be 100-110 kph.

(iii) Driver

- The driver was less likely to have braked or swerved.

(iv) Pedestrian

- The pedestrian was more likely to:
  - be aged 36-65 years rather than 1-18 or over 66 years
  - be a plant/machine operator or labourer or alternatively a manager, administrator or professional
    rather than a student or be retired or a pensioner
  - be above average height (170-195 cms) rather than small (80-145 cms)
  - have a BAC of zero rather than above .15.

(v) Crash

- The crash was more likely to be due to vision being obscured and less likely to be due to alcohol or drugs.

- It was more likely to occur on the near side when the pedestrian was crossing the road or emerging from behind a parked vehicle.

TABLE 4.6. Areas of Difference between the Adult Day Cluster and the Population of Pedestrian Fatalities

	Variable	Categories	Population n	Population %	Adult-Day Cluster n	Adult-Day Cluster %	
(4)	Speed limit/speeding	<80 not speed	267	72	40	91	↑
		<80 speed	39	10	2	5	
		80+ not speed	62	17	2	5	↓
		80+ speed	4	1	0	0	
(5)	Braked/swerved	Yes	195	54	18	42	↓
		No	163	46	25	58	↑
(13)	Major cause of accident	Alcohol/drugs	138	42	9	28	↓
		Sensory impairment	6	2	0	0	
		Vision obscured	9	3	4	13	↑
		Not see other road user	49	15	6	19	
		Excessive speed	9	3	0	0	
		Ignore traffic controls	14	4	2	6	
		Pedestrian step into path	101	31	11	34	
(14)	Pedestrian age	1-18 yrs	90	23	0	0	↓
		19-35 yrs	97	25	11	24	
		36-65 yrs	104	26	32	71	↑
		66+ yrs	102	26	2	4	↓
(16)	Pedestrian employment status	Pre-school/baby	20	6	0	0	
		Manager/admin/prof/paraprof	18	5	6	15	↑
		Trades persons	22	6	4	10	
		Clerical/sales	12	3	2	5	
		Plant/machine/labour	47	13	13	33	↑
		Housekeeping	13	4	5	13	
		Unemployed	34	9	3	8	
		Student	57	16	1	3	↓
		Retired/pensioner	135	38	6	15	↓

Variable		Categories	Population		Adult-Day Cluster		
			n	%	n	%	
(17)	Pedestrian height	80-145 cms	64	16	0	0	↓
		149-169 cms	142	36	15	33	
		170-195 cms	187	48	30	67	↑
(20)	Pedestrian BAC	0	143	55	30	81	↑
		.001- 15	35	13	2	5	
		>.15	82	32	5	14	↓
(21)	Day of week	Mon-Wed	152	39	28	62	↑
		Thurs-Fri	121	31	10	22	
		Sat	81	20	3	7	↓
		Sun	39	10	4	9	
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	36	80	
		Towns (200-100,000)	66	17	9	20	
		Rural	43	11	0	0	↓
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	10	22	
		Major arterial	137	35	19	42	
		Other urban	153	39	16	36	
(25)	Road configuration	2-way undivided	233	71	22	60	↓
		Divided, dual carriageway, fwy	95	29	15	41	↑
(27)	Number of lanes	1 or 2	216	65	20	50	↓
		3	57	17	11	28	↑
		4 or 6	61	18	9	23	
(29)	Speed limit at crash location	60	292	74	37	82	
		70-90	57	15	8	18	
		100-110	43	11	0	0	↓

	Variable	Categories	Population n	Population %	Adult-Day Cluster n	Adult-Day Cluster %	
(31)	DCA event	Near side/emerging	194	53	30	71	↑
		Far side	82	22	6	14	
		On carriageway	61	17	5	12	
		With traffic	31	8	1	2	
(32)	Time of day/day of week	Week day	155	39	38	84	↑
		Weekend day	38	10	7	16	
		Week night	78	20	0	0	↓
		Weekend night	122	31	0	0	↓
(33)	Time (DEFINITION)	Day	193	49	45	100	↑
		Night	200	51	0	0	↓
(34)	Day	Weekday	233	59	38	84	↑
		Weekend	160	41	7	16	↓

4.7. Cluster 7 - Adult Alcohol

4.7.1 Definition of cluster: Variable C9a Blood Alcohol Content-Pedestrian (Category 4)

		Cluster 7		Total	
		n	%	n	%
1.	zero			143	36
2.	.001 - .10			25	6
3	.101 - 150			10	3
4.	>.15	53	100	82	21
5.	not tested			99	25
6	unknown			34	9
		53		393	

For the purpose of presentation, categories 2 and 3 have been combined, and categories 5 and 6 have been removed.

4.7.2. Characteristics of cluster

(i) Pedestrian

- The pedestrian was more likely to
  - be aged 19-65 years rather than 1-18 or above 66 years
  - be male
  - be a plant/machine operator or labourer rather than a student or retired/pensioner
  - be tall (170-195 cms) rather than short (80-145 cms).

(ii) Driver

- The driver was more likely to
  - be speeding
  - have braked or swerved

(iii) Vehicle

- The vehicle was more likely to be a heavy car

(iv) Timing

- The crash was more likely to have occurred:
  - at the weekend rather than during the week
  - at night rather than in the day.

(v) Environment

- The road and its environment was less likely to be rural.
- The road was more likely to be divided, dual carriageway or freeway rather than 2-way undivided
- The median was more likely to be low or narrow with there being less likelihood of being no median.

(vi) Crash

- Responsibility for the crash was more likely to be due to the driver and pedestrian combined.
- The cause of the crash was more likely to be alcohol and drugs and less likely to be not seeing the other road user or the pedestrian stepping into the path of the car.

TABLE 4.7. Areas of Difference between the Adult Alcohol and the Population of Pedestrian Fatalities

Variable		Categories	Population		Adult-Alcohol Cluster		
			n	%	n	%	
(3)	Weight class of vehicle	Motor cycle	14	4	3	6	
		Light car	63	17	5	10	
		Medium car	94	25	15	29	
		Heavy car	109	29	21	40	↑
		Van/light truck	64	17	8	15	
		Heavy vehicle	32	8	0	0	
(4)	Speed limit/speeding	<80 not speed	267	72	31	59	↓
		<80 speed	39	10	13	25	↑
		80+ not speed	62	17	8	15	
		80+ speed	4	1	1	2	
(5)	Braked/swerved	Yes	195	54	30	64	↑
		No	163	46	17	36	↓
(12)	Unit responsible for crash	Driver	50	13	3	6	
		Driver and pedestrian	43	11	12	22	↑
		Pedestrian	291	76	38	72	
(13)	Major cause of accident	Alcohol/Drugs	138	42	50	96	↑
		Sensory impairment	6	2	0	0	
		Vision obscured	9	3	0	0	
		Not see other road user	49	15	0	0	↓
		Excessive speed	9	3	1	2	
		Ignore traffic controls	14	4	0	0	
		Pedestrian step into path	101	31	1	2	↓
(14)	Pedestrian age	1-18 yrs	90	23	0	0	↓
		19-35 yrs	97	25	27	51	↑
		36-65 yrs	104	26	26	49	↑
		66+ yrs	102	26	0	0	↓

Variable		Categories	Population		Adult-Alcohol Cluster		
			n	%	n	%	
(15)	Pedestrian sex	Male	283	72	44	83	↑
		Female	110	28	9	17	↓
(16)	Pedestrian employment status	Pre-school/baby	20	6	0	0	
		Manager/admin/prof/paraprof	18	5	6	13	
		Trades persons	22	6	5	10	
		Clerical/sales	12	3	2	4	
		Plant/machine/labour	47	13	13	27	↑
		Housekeeping	13	4	3	6	
		Unemployed	34	9	8	17	
		Student	57	16	0	0	↓
		Retired/pensioner	135	38	11	23	↓
(17)	Pedestrian height	80-145 cms	64	16	0	0	↓
		149-169 cms	142	36	16	30	
		170-195 cms	187	48	37	70	↑
(20)	Pedestrian BAC (DEFINITION)	0	143	55	0	0	↓
		.001- 15	35	13	0	0	↓
		>.15	82	32	53	100	↑
(21)	Day of week	Mon-Wed	152	39	8	15	↓
		Thurs-Fri	121	31	16	30	
		Sat	81	20	18	34	↑
		Sun	39	10	11	21	↑
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	43	81	
		Towns (200-100,000)	66	17	10	19	
		Rural	43	11	0	0	↓
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	11	21	
		Major arterial	137	35	21	40	
		Other urban	153	39	21	40	

	Variable	Categories	Population		Adult-Alcohol Cluster		
			n	%	n	%	
(25)	Road configuration	2-way undivided	233	71	24	57	↓
		Divided, dual carriage, freeway	95	29	18	43	↑
(26)	Type of median	None	62	20	4	10	↓
		Low/narrow	65	20	13	31	↑
		Not easily/intermittently surmountable	23	7	4	10	
		Painted	169	53	21	50	
(32)	Time of day/day of week	Week day	155	39	0	0	↓
		Weekend day	38	10	0	0	↓
		Week night	78	20	18	34	↑
		Weekend night	122	31	35	66	↑
(33)	Time	Day	193	49	0	0	↓
		Night	200	51	53	100	↑
(34)	Day	Weekday	233	59	18	34	↓
		Weekend	160	41	35	66	↑

4.8. Cluster 8 - Adult Other

4 8.1. Definition: The cases remaining after the removal of the other clusters

4.8.2. Characteristics of cluster

(i) Pedestrian

- The pedestrian was more likely to:
  - be aged 19-65 years rather than 1-18 or over 66 years
  - be unemployed rather than a student or be retired or a pensioner
  - be taller than average (170-195 cms) rather than shorter (80-145 cms)
  - have a BAC reading of .001-.15 but less likely to be above .15

(ii) Driver

- The driver was more likely to.
  - be a male
  - have a BAC reading above zero
  - be speeding in a <80 kph speed zone

(iii) Vehicle

- The vehicle was more likely to be old i.e., manufactured prior to 1985.

(iv) Timing

- The crash was more likely to have occurred
  - at night rather than during the day
  - on the weekend (particularly Saturday) rather than during the week (particularly not Monday to Wednesday).

(v) Environment

- The crash was less likely to have occurred in a rural area.

(vi) Crash

- The point of primary impact was less likely to have been on the left hand side of the vehicle

- The driver and pedestrian together were more likely to be responsible for the crash with the pedestrian alone being less likely to be responsible.

- The major cause was more likely to be drugs and alcohol and less likely to be the pedestrian stepping into the path of the car.

- The pedestrian was more likely to be walking with the traffic than crossing the road on the near side or emerging from behind a parked car.

TABLE 4.8. Areas of Difference between the Adult Other Cluster and the Population of Pedestrian Fatalities

	Variable	Categories	Population		Adult-Other Cluster		
			n	%	n	%	
(4)	Speed limit/speeding	<80 not speed	267	72	34	61	↓
		<80 speed	39	10	8	14	
		80+ not speed	62	17	13	23	
		80+ speed	4	1	1	2	
(6)	Point of primary impact	Left side	153	40	19	30	↓
		Front	120	31	22	34	
		Undercarriage	17	4	1	2	
		Right side	95	25	22	34	
(7)	Year of manufacture	Old (1962-84)	196	58	37	73	↑
		New (1985-90)	142	42	14	28	↓
(9)	Driver sex	Male	305	81	57	92	↑
		Female	73	19	5	8	↓
(11)	Blood alcohol concentration	Zero	273	89	37	79	↓
		Above zero	35	11	10	21	↑
(12)	Unit responsible for crash	Driver	50	13	10	16	
		Driver and pedestrian	43	11	13	21	↑
		Pedestrian	291	76	39	63	↓
(13)	Major cause of accident	Alcohol/drugs	138	42	30	68	↑
		Sensory impairment	6	2	0	0	
		Vision obscured	9	3	0	0	
		Not see other road user	49	15	5	11	
		Excessive speed	9	3	1	2	
		Ignore traffic controls	14	4	1	2	
		Pedestrian step into path	101	31	7	16	↓

Variable		Categories	Population n	Population %	Adult-Other Cluster n	Adult-Other Cluster %	
(14)	Pedestrian age	1-18 yrs	90	23	0	0	↓
		19-35 yrs	97	25	31	48	↑
		36-65 yrs	104	26	31	48	↑
		66+ yrs	102	26	3	4	↓
(16)	Pedestrian employment status	Pre-school/baby	20	6	0	0	
		Manager/admin/prof/paraprof	18	5	5	10	
		Trades persons	22	6	6	12	
		Clerical/sales	12	3	6	12	
		Plant/machine/labour	47	13	9	18	
		Housekeeping	13	4	3	6	
		Unemployed	34	9	11	22	↑
		Student	57	16	0	0	↓
		Retired/pensioner	135	38	10	20	↓
(17)	Pedestrian height	80-145 cms	64	16	1	1	↓
		149-169 cms	142	36	20	31	
		170-195 cms	187	48	44	68	↑
(20)	Pedestrian BAC	0	143	55	20	53	
		.001-.15	35	13	18	47	↑
		>.15	82	32	0	0	↓
(21)	Day of week	Mon-Wed	152	39	15	23	↓
		Thurs-Fri	121	31	23	35	
		Sat	81	20	24	37	↑
		Sun	39	10	3	5	
(23)	Land Classification	Capital city/Metropolitan Major Urban	283	72	52	80	
		Towns (200-100,000)	66	17	13	20	
		Rural	43	11	0	0	↓
(24)	Road type	Rural	44	11	0	0	↓
		Highways - urban	59	15	15	23	
		Major arterial	137	35	20	31	
		Other urban	153	39	30	46	

Variable		Categories	Population		Adult-Other Cluster		
			n	%	n	%	
(31)	DCA event	Near side/emerging	194	53	21	34	↓
		Far side	82	22	12	20	
		On carriageway	61	17	16	26	
		With traffic	31	8	12	20	↑
(32)	Time of day/day of week	Week day	155	39	0	0	↓
		Weekend day	38	10	0	0	↓
		Week night	78	20	22	34	↑
		Weekend night	122	31	43	66	↑
(33)	Time	Day	193	49	0	0	↓
		Night	200	51	65	100	↑
(34)	Day	Weekday	233	59	22	34	↓
		Weekend	160	41	43	66	↑

Chapter 5: Discussion

This research project achieved its aim to develop a typology of fatal pedestrian crashes Homogeneity analysis (HOMALS) was chosen for this purpose. While the variables all measured something about each fatal pedestrian crash, the level of measurement of each variable ranged from binary through categorical to ordinal. This was a difficulty for genuine cluster analysis (Kmeans) but not for homogeneity analysis, where all variables were treated at the binary or categorical level. The HOMALS solution displayed each pedestrian crash as part of a "cloud" of crashes in multidimensional space. By inspection, definite clusters were found in this "cloud". While not being the same as a genuine cluster analysis, the result is probably quite similar (Van de Geer, 1993). The structure that was uncovered contained eight clusters which are summarised below

The first cluster to emerge was the Rural cluster. The main characteristics of this cluster, in addition to being rural, were that the crash happened mid block; the driver was likely to be further from home, driving a heavy vehicle at night; and the pedestrian was likely to be a middle aged male with some alcohol in his blood walking with the traffic or being near the road side.

Once these rural crashes were removed, the second cluster to emerge was the School Student/Preschooler cluster of pedestrians under 18 years old, including babies and pre-school children. The main characteristics of this cluster, in addition to being non-rural, were that the pedestrian was free of alcohol yet responsible for the crash by stepping into the path of the driver while crossing the road; the driver was more likely to be close to home and female, driving during the day and perhaps not seeing the pedestrian; and the crash was more likely to occur in a residential area on a minor road

The third cluster to emerge was the 8-18 Years of Age cluster of urban, young pedestrians all between the ages of 8 and 18 years, without the presence of blood alcohol. These pedestrians were more likely to be female students of medium height at an intersection controlled by traffic lights; the driver, driving on a weekday further from home, was more likely to be middle aged and responsible for the crash with some blood alcohol and driving a van or light truck; and the crash was more likely to occur in a metropolitan urban residential/commercial or urban parkland/freeway area.

The fourth cluster to emerge contained the urban, elderly pedestrian, generally older than 66 years who was retired or a pensioner and most likely free of alcohol. This cluster was labelled the Elderly Retired. For this cluster, the driver was more likely to have braked or swerved and the crash was more likely to have occurred during the week in daylight hours on a wider road controlled by non-electric controls with the pedestrian on the far side of the road from where they crossed.

The fifth cluster to emerge, called the Heavy Vehicle cluster, was comprised of male drivers, away from home, driving a heavy vehicle in a commercial/industrial area during the day. The crash, which generally involved braking or swerving was more likely to involve an adult victim, but not elderly, retired or pensioner female with no blood alcohol but who was likely to be at fault for the crash by ignoring traffic controls while crossing the road.

The remaining three clusters all contained adult, non elderly, urban pedestrians. The first split for these three clusters separated day crashes (called Adult Day), from night crashes, while the second split separated the night crashes into those with alcohol (Adult Alcohol), from the remaining crashes (Adult Other).

Cluster 6, the Adult Day cluster, was characterised by a week day crash where the driver was more likely to have had their vision obscured and the pedestrian hit on the nearside of the road For cluster 7, the adult night crashes where alcohol was involved (Adult Alcohol cluster), the pedestrian was more likely to be male and the driver more likely to be speeding during the weekend The last cluster, cluster 8, the Adult Other cluster, contained other adult, night crashes Here the driver was more likely to be speeding in an older car, unemployed and under the influence of some alcohol or drugs. The pedestrian, also with some blood alcohol was more likely to be walking on the weekend with the traffic rather than crossing the road

There were a number of strengths associated with this analysis. Firstly, it was performed following a logical and precise clustering strategy as outlined in Chapter 2. This clustering strategy involved seven defined steps

A second strength was the uncovering of structure by successive removal of "outlier" clusters Homogeneity analysis tends to be sensitive to these outlier cases (Van de Geer, 1993), which then cause the remaining cases to be merged. This successive procedure clarified the hierarchical importance of variables which could not have been chosen a priori. In order, the variables of importance were region (rural/urban), age of pedestrian, heavy vehicle, timing of crash (day/night) and the importance of alcohol. Also, this successive procedure indicated that the clusters emerged like a classification tree rather than by a factorial combination of variables. For instance, the first cluster separated rural crashes from urban crashes and then successive clusters looked for structure within the urban crashes and so on.

None of the previous research, reported in Chapter 1, selected this combination of important variables a priori or allowed for a hierarchical structure of the selected variables. As a result, the clustering strategy followed in this research adds a new method of investigating road safety data. By comparison, Attewell and Dowse (1992) concentrated on point of primary impact and major cause of crash for their classification; Buhlman, Warren and Simpson (1983) classified crashes by age of pedestrian; and Massie, Campbell and Blower (1993) formed 18 clusters by the factorial combination of variables including how many vehicles were involved, whether the accident occurred at an intersection or not, and whether the intersection was controlled by traffic lights or not among others.

Clearly, for these researchers, the number of clusters reported depended on the number of a priori variables chosen for use. Interestingly, Grundy (1990) also used homogeneity analysis in an investigation of 2689 pedestrian crashes which resulted in four clusters. The crashes were not limited to fatal crashes, as they were in the current analysis, and their strategy does not seem to have been hierarchical. Their solution has a number of similarities with the current solution but their low number of obtained clusters, four, suggests that some clusters were merged.

The present analysis has a number of limitations. Firstly, the cases were limited to fatal pedestrian crashes This limited the analysis in that there was one set of interrelated variables without a natural dependent/independent structure Had the cases not been limited to fatal crashes, then severity of crash could have been treated as the dependent variable to predict. This approach was seen with some of the previous research. This prediction of severity could proceed using a classification tree approach such as an automatic interaction detection program. Secondly, the analysis was limited to homogeneity analysis which treats the variables as one set of categorical variables. Other approaches, such as non-linear principal components analysis, could allow a mixture of measurement levels with some variables treated as ordinal or numerical thereby possibly extracting more explained variance in the data. Also, other approaches, such as non-linear canonical correlation, could allow the variables to be organised into logical sets for analysis as well as allowing a mixed measurement level. These logical sets could be composed of driver variables, pedestrian variables, crash variables and vehicle variables. Important variables within logical sets and between logical sets could then be investigated. Of course, each of these limitations can be seen as opportunities for further research

It is hoped that a full range of methodological approaches will be employed in the future to extract the maximum information from the set of pedestrian crash variables, correlated in unknown ways, with the aim of developing a typology of pedestrian crashes Achievement of this will help suggest strategies for reducing the incidence of these crashes. It was not the aim of this research to outline such policy measures but it is hoped that these results will help lead to successful counter measures being instigated. These counter measures could be developed to take advantage of the hierarchical nature of the clusters and also to take advantage of the known overlap between the clusters.

References

Aldenderfer, M.S , & Blashfield, R K (1984). *Cluster analysis.* Sage University Paper series on Quantitative Applications in the Social Sciences, 07-044. Beverly Hills and London: Sage Pubns.

Attewell, R.G , & Dowse, M J (1992) *Fatal crash types; Analysis of 1988 fatality file.* Canberra. Federal Office of Road Safety.

Bergman, L.R. (1988) You can't classify all of the people all of the time. *Multivariate Behavioral Research, 23,* 425-441.

Blashfield, R.K., Aldenderfer, M.S., & Morey, L.C. (1982) Validating a cluster analytic solution. In H.C. Hudson and Associates. (Eds.). *Classifying social data,* (pp. 167-176) San Francisco: Jossey-Bass.

Buhlman, M A., Warren, R.A., & Simpson, H.M. (1983). Profile of fatal collisions involving child pedestrians. *Voice of the Pedestrian, 14,* 53-56.

Colli, G.R. (1994) Personal communication via INTERNET

Coolen, H., & Hilkhuysen, G. (1992) The structure of do it yourself activities; an application of nonmetric principal components analysis *OTB Working Paper, No. 92-01.* Delft: Delft University Press.

Ehrenberg, A.S.C (1981). *Data reduction.* Chichester. John Wiley and Sons

Everitt, B. (1974) *Cluster analysis.* London: Heinemann Educational Books.

Gauch, H.G., Jr. (1980). Rapid initial clustering of large data sets. *Vegetatio, 42,* 103-111.

Gauch, H.G., Jr. (1982). *Multivariate analysis in community ecology* Cambridge: Cambridge University Press.

Gifi, A. (1990) *Nonlinear multivariate analysis.* Chichester: Wiley.

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis.* London. Academic Press.

Gundy, C M (1990). *Accident typology.* Leidschendam: SWOV Institute for Road Safety Research, The Netherlands.

Massie, D.L , Campbell, K.L., & Blower D.F. (1993). Development of a collision typology for evaluation of collision avoidance strategies. *Accident Analysis and Prevention, 25,* 241-253.

Nakache, J.P (1981). Some methods of discriminant analysis on binary variables. In J F Bithell and R. Coppi. (Eds.). *Perspectives in medical statistics: Proceedings of the European Symposium on medical statistics, Rome 1980,* (pp. 133 - 154). London: Academic Press

Rapkin, B.D., and Luke, D.A. (1993). Cluster analysis in community research: Epistemology and practice. *American Journal of Community Psychology, 21(2),* 247-277

Retting, R., Williams, A., Preusser, D., & Weinstein, H. (1992) *Classifying urban crashes for countermeasure development.* Arlington, Va· Insurance Institute for Highway Safety.

Rohlf, J. (1994). Personal communication via INTERNET.

SPSS. (1990) *SPSS categories.* Chicago: SPSS Inc

Van de Geer, J.P. (1993). *Multivariate analysis of categorical data: Applications.* Newbury Park: Sage.

Wells-Parker, E., Anderson, B., Pang, M., & Timken, D (1993). An examination of cluster-based classification schemes for DUI offenders. *Journal of Studies on Alcohol, 54,* 209-218.