# TYPICAL PEDESTRIAN ACCIDENT SCENARIOS FOR THE TESTING OF AUTONOMOUS EMERGENCY BRAKING SYSTEMS

**James Lenard**
**Russell Danton**
Loughborough University
UK
**Matthew Avery**
**Alix Weekes**
Thatcham
UK
**David Zuby**
Insurance Institute for Highway Safety
USA
**Matthias Kühn**
German Insurers Accident Research
Germany
Paper Number 11-0196

## ABSTRACT

The research objective of this work was to describe typical accident scenarios for pedestrian accidents.

The accident analysis forms a component of work by the AEB Test Group which aims to develop test procedures for assessing Autonomous Emergency Braking (AEB) systems. This technology is penetrating the vehicle market and is designed to offer protection against the occurrence and severity of collisions; however there is a need to evaluate the systems and their effectiveness since they are not yet subject to regulation or standardised assessment.

Case files for 175 pedestrians who were struck by the front of a passenger car were extracted from an in-depth accident database and reviewed in detail to establish the position and movement of road users before impact. A dataset of key parameters was formed from the detailed case reviews and subjected to a hierarchical cluster analysis to identify groups of similar accident scenarios. A second cluster analysis was performed on a dataset derived from the British national accident database for over 10,500 accidents where a pedestrian was struck by the front of a passenger car. This led to a second set of typical accident circumstances based on a comprehensive coverage of the accident population.

The national accident database for Great Britain, STATS 19, is compiled annually from police reports and effectively defines the national road accident population. In 2008 it registered over 28,000 pedestrian casualties from a total of around 230,000 road user casualties. The UK On-the-Spot (OTS) in-depth accident database was compiled by research teams at the scene of accidents in two regions of England from 2000 to 2010, including some non-injury accidents. Each team attended approximately 250 accidents per year, resulting in a total of over 4,700 accidents involving over 11,000 road users (including 288 pedestrians). This study was designed to collect a representative sample of accidents.

The cluster analyses show the association of accident circumstances such as speed limit, light conditions, weather, vehicle manoeuvre, pedestrian size, pedestrian movement, obstruction of line of sight, vehicle travel speed and change of speed to impact. The proportion of fatal, serious and slight casualties associated with these scenarios is quantified, showing for example that one scenario covered 12% of the population but 23% of fatal casualties.

Typical circumstances for pedestrian accidents in the dataset include (1) crossing from the kerb side without obstruction of the driver’s line of sight, (2) smaller pedestrians crossing from the kerb side with at least partial obstruction of the driver’s line of sight and (3) adult pedestrians crossing in inclement light and weather conditions. These scenarios were computed mathematically from large in-depth and national accident databases using cluster analysis and provide relevant information for the formulation of controlled tests of AEB systems.

## INTRODUCTION

Autonomous Emergency Braking is one of a number of modern safety technologies designed to prevent or mitigate the severity of vehicle impacts. There is scope for considerable variation among

AEB systems depending on the type of sensors fitted, the decision logic programmed into the control units, how and when the driver is alerted, how and when braking is activated and other factors. For this reason there is interest in developing and conducting physical tests to assess performance, compare systems and inform consumers.

The AEB Test Group is formed from insurance-based research centres around the world with a common interest in assessing AEB systems for their effectiveness in mitigating and preventing collisions. The members include Thatcham, the Insurance Institute for Highway Safety (IIHS), the German Insurers Accident Research (UDV) and Folksam. The Group is involved with assessing the effectiveness of these systems in real accidents, but since they show potential benefit of collision mitigation the group is authoring test procedures to assess the effectiveness of the systems. It is important to assess these systems since there is not yet any regulation or other consumer assessment that might influence the development of the AEB systems. The consumer rating of the systems will help to inform consumers of the most effective systems and help to drive design and development of systems that are best suited to addressing real-world collisions.

AEB systems can already work in collisions involving pedestrians and rear impacts and in the future will be able to address frontal, head-on collisions. However since the head-on systems are not yet widely fitted, this type of collision is not currently being considered by the AEB Test Group, but will be incorporated at a later date.

The setting of test conditions involves many considerations, one of which is the desirability of subjecting the vehicles to realistic accident conditions, i.e. circumstances that are encountered in real accidents, or at least to understand clearly how proposed test conditions relate to the circumstances of real accidents. The aim of this paper is to describe typical accident scenarios for pedestrian accidents based on empirical data. Some examples are given of possible test scenarios that could be based on this factual information.

## MATERIALS AND METHODS

### Source databases

Two major sources of information about accidents in Britain were used in this work: the national accident database STATS 19 and the in-depth On-the-Spot study (OTS). STATS 19 is compiled annually by the Department for Transport (DfT) based primarily on police reports and it effectively defines the road casualty population of Great Britain. OTS was a study run from 2000 to 2010 for the DfT and Highways Agency to collect in-depth information about a representative sample of road accidents based on approximately 500 at-scene investigations per year. Some key facts about these databases are presented in Table 1 and both are described more fully in the literature [1] [2] [3]. The analysis in this paper used STATS 19 for 2008 and OTS from 2000 to mid-2009, the latest versions available when work commenced.

A note on the relationship between the STATS 19 and OTS databases. While STATS 19 describes the whole reported road accident population for a year, the in-depth accident database OTS contains a sample of cases from two regions but over a greater period of time, 2000–2010. In addition, unlike STATS 19, OTS includes a proportion of non-injury accidents. There should consequently be some coverage of the same accidents, i.e. roughly one-third of the casualty accidents that occurred in the two OTS sample regions in 2008; however this overlap constitutes a distinct minority of both databases. Furthermore, as an in-depth database, OTS contains more information about accidents than STATS 19, especially quantitative information about velocity, location, injuries and causal factors based on accident investigation, reconstruction and follow-up data collection.

**Table 1.**
**Source databases STATS 19 and OTS**

| STATS 19 | OTS |
|---|---|
| *Period* | |
| 2008 | 2000–2010 |
| *Sample region* | |
| Great Britain | South Nottinghamshire<br>Thames Valley |
| *Purpose* | |
| National statistics | Detailed information to support casualty reduction programmes |
| *Source* | |
| Police reports | At-scene investigations by research teams at Loughborough University and TRL |
| *Inclusion criteria* | |
| Casualty on public road | Police attendance on rotating 8-hour shift |
| *Number of accidents* | |
| 170,591 | 4,744 |

The summary datasets prepared for the clusters analyses (described below) contain a selection of the most suitable fields available in each dataset. This resulted, for example, in taking vehicle speed from OTS but speed limit, the best available proxy, from STATS 19. The datasets derived from STATS 19 and OTS therefore contain only a partial overlap in (a) the variables used to describe the accidents and (b) the accidents covered. The two sources of information can accordingly be regarded as largely independent, the main link being that OTS was designed to be representative of the accident population as far as possible within the scope of the study.

The fields selected for detailed analysis were chosen in the context of their relevance to test conditions and the design of AEB technology. So for example an AEB system will be engineered to optimise its field of view, processing (recognition) speed and decision logic against the pre-impact location, speed, trajectory and size of pedestrians encountered in real accidents. Detrimental ambient light and weather conditions could diminish the effectiveness of certain sensors. It is highly relevant whether vehicles are typically turning or proceeding straight ahead in their approach to the point of impact and whether the line of sight from the vehicle to the struck pedestrian is fully or partially obscured by other vehicles or roadside objects in the seconds before impact. Information on the frequency and extent of braking before impact is relevant to the choice and effectiveness of a system that is fully automated or that reinforces avoidance actions initiated by the driver.

**Summary datasets**

The fields used for the STATS 19 and OTS cluster analyses are shown in Table 2 and Table 3. The fields derived from STATS 19 are mostly simplified versions of the originals obtained by aggregating and thereby reducing the number of categories, the exceptions being ‘pedestrian injury severity’ which is unchanged and ‘pedestrian age-sex’ which combines the original age and sex fields into a quasi-size category. The fields derived from OTS were recorded by an analyst after a review of full case materials.

The categorisation of each field as nominal, ordinal or scale is relevant to the operation of the cluster analysis algorithm. The basic concept is that scale variables are continuous parameters measured in units such as seconds or metres, ordinal variables provide categories with a natural order such as injury severity or speed limit, and nominal variables provide categories without a natural order such as vehicle type or precipitation.

**Table 2.**
**Variables in STATS 19 cluster analysis**

| | Field | Type |
|---|---|---|
| 1 | Pedestrian injury severity | Ordinal |
| 2 | Speed limit | Ordinal |
| 3 | Light conditions | Nominal |
| 4 | Precipitation | Nominal |
| 5 | Vehicle manoeuvre | Nominal |
| 6 | Pedestrian age-sex | Ordinal |
| 7 | Pedestrian movement | Nominal |
| 8 | Pedestrian masked by vehicle | Nominal |

As mentioned above, the choice of fields in the summary datasets was guided by their relevance for physical testing. While items such as light conditions, precipitation, vehicle speed and pedestrian crossing direction were included, other items such as the age and sex of the driver or the time of day of the accident were not, even though there could well be patterns in how these factors correlate in real accidents with other characteristics, for example it could be that female drivers experience a higher exposure to pedestrian accidents involving children in the morning and afternoon ‘school runs’. The underlying reasoning was that driver characteristics and time of day would not be reflected in the setup of physical tests of AEB performance.

**Table 3.**
**Variables in OTS cluster analysis**

| | Field | Type |
|---|---|---|
| 1 | Pedestrian injury severity | Ordinal |
| 2 | Light conditions | Nominal |
| 3 | Precipitation | Nominal |
| 4 | Vehicle manoeuvre | Nominal |
| 5 | Pedestrian age-sex | Ordinal |
| 6 | Pedestrian movement | Nominal |
| 7 | Pedestrian speed | Ordinal |
| 8 | Line of sight obscured (1 sec) | Nominal |
| 9 | Vehicle speed | Scale |
| 10 | Change of speed to impact | Scale |

**Cluster analysis**

The method employed in this analysis to move the from accident data to the formulation of accident scenarios was a data mining technique known as cluster analysis, in particular the hierarchical, ascending (agglomerative) variety. This works by progressively grouping together the most similar records of a dataset, where the notion of similarity is defined mathematically. As applied here, each record describes an accident and so the cluster analysis identifies groups of similar accidents. These groups or clusters have (by definition) common characteristics and can be interpreted as constituting accident scenarios. The foremost advantage of applying this method is that the results are objective and reproducible, with an additional benefit that the representativeness of the resultant accident scenarios is clearly defined.

The algorithm for computing the similarity or, on the analogy of points in space, 'distance' between clusters of accidents requires specification at three levels:

- at field level, the algorithm was set to compute a distance or (dis)similarity in the range 0–1 for any two values of a field with 0 signifying identity and 1 signifying maximum difference
- at record level, the distance between two accidents was defined as the sum of the distances between the fields—the *city block* or *Manhattan* distance
- at cluster level, the distance between two clusters was defined as the average of the distances between each pair of records in the groups—the *average linkage method*.

**Table 4.**
**Illustration of the assignment of numeric values for quantifying similarity**

| Field | Type | Numeric value | Field value |
|---|---|---|---|
| Vehicle manoeuvre | Nominal | 1 | Ahead |
| | | 2 | Turning |
| | | 3 | Other |
| Age-sex | Ordinal | 0.00 | 0–7 years |
| | | 0.33 | 8–15 years |
| | | 0.67 | Adult female |
| | | 1.00 | Adult male |
| Vehicle speed (km/h) | Scale | 0.0 | 40 (min.) |
| | | 0.2 | 50 |
| | | 0.8 | 80 |
| | | 1.0 | 90 (max.) |

For nominal fields, the distance or dissimilarity between two values is always either 0 or 1, depending whether the characteristic is the same or different for two accidents. Making reference to Table 4, if in two accidents the vehicles are both 'Turning', the distance is 0; if one is 'Going ahead' and the other 'Other', the distance is 1. For ordinal and scale values, the range is set to span 0–1 in equal increments for ordinal variables or continuously for scale variables. Accordingly the distance between an adult female and an adult male is 0.33 (1.00-0.67) and the distance between 50 and 80 km/h would be 0.6 (0.8-0.2) assuming minimum and maximum speeds in the dataset of 40 and 90 km/h respectively.

It remains to state briefly how the number of clusters for each analysis was determined. The hierarchical cluster analysis begins with one cluster for each record and iterates through a grouping procedure until it ends with one cluster for the whole dataset. No particular set of clusters is right or wrong: each is a valid representation of the data. The question is rather the usefulness of a set of clusters for a particular purpose. Clearly neither extreme—one for each record or one for the whole population—adds value. For the purpose of contributing to the design of testing procedures, it was considered relevant to have a relatively small number of clusters that covers much of the population. To this end supplementary programming code was written to assist in the identification of around six clusters to contain about 75–80% of the population, including the fatal and seriously injured sub-populations. In conjunction with further code to identify 'natural' gaps between the clusters, the final number of clusters for each accident type and source database was chosen manually after examination of the data.

The technical specifications of the algorithm underlying the cluster analysis were selected from a range of standard methods. Further details are available in the literature [4] [5]. This style of analysis has been applied to accident data before [6] but not, to the authors' knowledge, to STATS 19, OTS or another major British accident database. The details provided above are intended to suffice in principle for the clusters to be independently derived starting from the same datasets using any software. The order of cases in the input dataset should make no difference.

## RESULTS

### National accident database STATS 19

The casualty file for STATS 19 (2008) contains information on 230,905 road users, among whom were 28,482 pedestrians. There is provision to nominate a vehicle with which each pedestrian interacted. These constitute the pool of cases from which the pedestrian accidents were drawn. The primary criteria for the selection of pedestrian accidents from STATS 19 were:

- cars, including taxis and private hire cars, associated with a pedestrian casualty
- first point of impact of the front surface.

There were 13,257 vehicles that met these criteria (Table 5). A second filter was made (a) of vehicles that were parked or reversing and (b) of records with missing or unknown information in any field. This resulted in a drop in the number of cases from 13,257 to 10,574, the main contributor being unknown pedestrian movement (2,263).

**Table 5.**
**Vehicle type and first point of contact for pedestrian accidents (STATS 19)**

| | First point of contact | | |
|---|---|---|---|
| | Front | Other | Total |
| Car | 13257 | 9857 | 23114 |
| Other | 2833 | 2535 | 5368 |
| Total | 16090 | 12392 | 28482 |

Table 6 shows the distribution of maximum pedestrian injury severity for the cases with complete and partially incomplete information, providing a check on the number of fatally or seriously injured casualties excluded at this stage. The proportions are reasonably evenly distributed among the fatal, serious and slight categories and, as a practical matter, 13,257 was slightly beyond the technical capacity of the hardware and software to process in a standard manner. The cluster analysis of STATS 19 was therefore performed on the basis of 10,574 vehicles.

**Table 6.**
**Availability of information for pedestrian accidents of interest (STATS 19)**

| | Complete | Partially incomplete | Total |
|---|---|---|---|
| Fatal | 240 | 79 | 319 |
| Serious | 2463 | 559 | 3022 |
| Slight | 7871 | 2045 | 9916 |
| Total | 10574 | 2683 | 13257 |

The outcome of the cluster analysis is shown in Table 8 at a level where the accident population was partitioned into 23 groups. The characteristics of the largest six clusters which comprise 85% of the population are shown in detail. Cells printed in bold font indicate (a) that the distribution of numbers for the given field is significantly different from the distribution in the total population (chi-squared test to 99.5% significance) and (b) that the particular numbers highlighted are over-represented. To take an example, all cases in Cluster 1 occurred in daylight compared to a distribution of 67% daylight and 33% darkness in the overall population of 10,574. The probability that this would happen by chance is less than 0.5% and the value of 100% is over-represented.

The figures on cluster representativeness in Table 7 express the numbers for pedestrian injury severity in Table 8 as row percentages. This is useful in highlighting for example that Cluster 3, which comprises 12% of the overall population, contains 23% of the pedestrian fatalities. It can therefore be construed as a particularly dangerous scenario.

**Table 7.**
**Cluster representativeness by pedestrian injury severity (STATS 19): N=10,574 vehicles**

| | Cluster % | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7–23 | Total |
| Slight | 41 | 15 | 11 | 8 | 7 | 3 | 16 | 100 |
| Serious | 35 | 14 | 15 | 13 | 5 | 4 | 15 | 100 |
| Fatal | 24 | 4 | 23 | 19 | 3 | 14 | 14 | 100 |
| Total | 39 | 14 | 12 | 9 | 6 | 3 | 15 | 100 |

**Table 8.**
**Pedestrian accident clusters derived from national data (STATS 19): N=10,574 vehicles**

| | Cluster % | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7–23 | Total |
| Pedestrian injury severity | | | | | | | | |
| Slight | **78** | **77** | 68 | 63 | **81** | 60 | 76 | 74 |
| Serious | 21 | 22 | **28** | **32** | 17 | **30** | 22 | 23 |
| Fatal | 1 | 1 | **4** | **5** | 1 | **9** | 2 | 2 |
| Speed limit (mph) | | | | | | | | |
| 10-30 | 92 | **97** | 90 | 90 | **97** | 71 | 92 | 92 |
| 40-50 | 5 | 3 | **8** | 7 | 2 | **8** | 4 | 5 |
| 60-70 | 3 | 0 | 2 | 3 | 1 | **21** | 4 | 3 |
| Light conditions | | | | | | | | |
| Light | **100** | **100** | 0 | 0 | **98** | 0 | 46 | 67 |
| Dark | 0 | 0 | **100** | **100** | 2 | **100** | 54 | 33 |
| Precipitation | | | | | | | | |
| No | **96** | **100** | 71 | 73 | **100** | 79 | 42 | 83 |
| Yes | 4 | 0 | **29** | **27** | 0 | 21 | 58 | 17 |
| Vehicle manoeuvre | | | | | | | | |
| Ahead | **100** | **100** | **100** | **100** | 0 | **98** | 62 | 88 |
| Turning | 0 | 0 | 0 | 0 | **100** | 2 | 38 | 12 |
| Pedestrian age-sex | | | | | | | | |
| 0–7 yrs | **11** | **23** | 2 | 3 | 5 | 1 | 7 | 10 |
| 8–15 yrs | **34** | **42** | 18 | 16 | 11 | 9 | 26 | 28 |
| Female | 26 | 16 | 26 | 27 | **46** | 20 | 33 | 27 |
| Male | 28 | 18 | **53** | **55** | **38** | **71** | 34 | 35 |
| Pedestrian crossing from… | | | | | | | | |
| Left | **59** | **57** | **100** | 0 | **63** | 0 | 59 | 57 |
| Right | 33 | **40** | 0 | **100** | 31 | 0 | 37 | 36 |
| Other | 7 | 2 | 0 | 0 | 6 | **100** | 4 | 7 |
| Masked by vehicle | | | | | | | | |
| No | **100** | 0 | **100** | **100** | **100** | **100** | 54 | 79 |
| Yes | 0 | **100** | 0 | 0 | 0 | 0 | 46 | 21 |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

The highlighting of cells in bold font assists in four ways to interpret the clusters. Firstly, where all of the cases fall into a single category, the cluster can be thought of as "purely" something. For example in Cluster 1 all of the accidents occurred in daylight, all vehicles were going ahead and no pedestrians were masked by a vehicle. As a starting point in building up the concept of a scenario based on this cluster, these characteristics are unambiguous. Secondly, where a category or associated group of categories is over-represented and constitutes a majority of the cases, it also lends its character to the cluster. In Cluster 2 the vast majority of accidents occurred in a 10–30 mph speed zone where the pedestrian was either crossing from the left or from the right. Thirdly, where a category or associated group of categories is over-represented but constitutes a minority of the cases, this can be thought of as a tendency. In Cluster 6, serious and fatal casualties are significantly over-represented along with the higher speed limits 40–50 and 60–70 mph. Finally, where no cell is marked in bold, the column of numbers for a given characteristic is not significantly different from the overall population. This can be seen in the speed limit zones of Cluster 4.

Table 8 defines the accident scenarios precisely and succinctly and it would not necessarily be informative to re-express them in words. A few 'higher level' observations may however be of interest. The two largest clusters, 1 and 2, mostly amplify the dominant characteristics of the overall population (slight injury, 10-30 mph, daylight, fine, going ahead and pedestrian crossing) with two exceptions, (a) an over-representation of children and (b) in cluster 2, the pedestrian being masked (obscured) by a vehicle. Clusters 3 and 4, on the other hand, are weighted towards serious and fatal injury, occur in darkness with a tendency towards wet weather and adult males who are not masked, the really substantial difference between these two clusters being that the pedestrian was crossing from the left in one case and from the right in the other. Cluster 5 introduces a turning scenario at low speeds and low injury outcomes, mostly matching the dominant features of the overall population except for an over-representation of adults. Apart from the higher severity levels and speed zones in Cluster 6 already mentioned, it is worth noting that this group of accidents occurred in darkness with mostly men who were stationary in or moving along the carriageway, this being the meaning of the "Other" category. This is the only major cluster not dominated by pedestrian movement across the carriageway.

**In-depth accident database OTS**

The On-the-Spot study had compiled records on 7,665 vehicles at the commencement of work for this analysis, among which were 216 passenger cars that struck (219) pedestrians. After filtering out non-frontal impacts and cases where inadequate information from the scene of the accident was available to support a quantitative assessment of the movement of the pedestrian and striking vehicle before impact, 175 were subjected to a detailed case-by-case review.

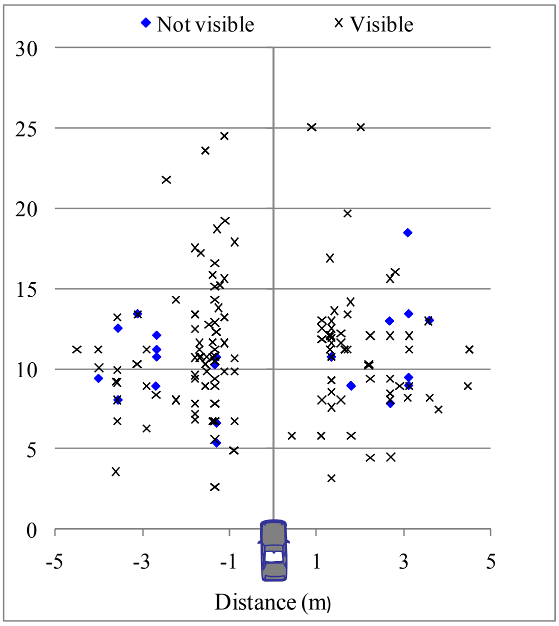


**Figure 1. Vehicle travel speed for clusters 1–6 (OTS): N=175 vehicles**

A focus of these reviews and reconstructions was the speed, direction of movement and distance apart of the road users and the presence or absence of a clear line of sight between the driver and pedestrian for up to five seconds before impact using established protocols where possible [7]. This is illustrated in Figure 1 which shows the location of the pedestrian relative to the striking vehicle one second before impact and whether there was a clear line of sight between the driver and pedestrian at this time. This parameter was not explicitly included in the cluster analysis because it is highly correlated with two items that were included: vehicle travel speed and change of speed to impact (braking). Including it would have provided double weight to essentially the same information.

The results of the cluster analysis of the OTS dataset are detailed at the level of 14 clusters. The largest six of these contain 79% of the population of the dataset (Table 9). The cells printed in bold font in Table 10 indicate (a) that the distribution of numbers for the given field is significantly different from the distribution in the whole population (chi-squared test to 95% significance) and (b) that the particular value highlighted is over-represented. This is similar to the treatment of STATS 19 above except that the statistical test is evaluated at 95% confidence instead of 99.5%. This level is better suited to the lower number of cases in OTS for providing an objective guide to differences between the clusters and the overall population.

**Table 9.**
**Cluster representativeness by pedestrian injury severity (OTS): N=175 vehicles**

| | Cluster % | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7–14 | Total |
| Slight | 29 | 20 | 12 | 9 | 8 | 0 | 23 | 100 |
| Serious | 30 | 9 | 15 | 24 | 4 | 0 | 19 | 100 |
| Fatal | 20 | 0 | 40 | 10 | 0 | 20 | 10 | 100 |
| Total | 29 | 15 | 14 | 14 | 6 | 1 | 21 | 100 |

**Table 10.**
**Pedestrian accident clusters derived from in-depth data (OTS): N=175 vehicles**

| | Cluster % | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7–14 | Total |
| Pedestrian injury severity | | | | | | | | |
| Slight/nil | 64 | 81 | 52 | 42 | 82 | 0 | 69 | 63 |
| Serious | 32 | 19 | **32** | **54** | 18 | 0 | 28 | 31 |
| Fatal | 4 | 0 | **16** | 4 | 0 | **100** | 3 | 6 |
| Light conditions | | | | | | | | |
| Light | **100** | **100** | 0 | 0 | **100** | 100 | 56 | 63 |
| Dark | 0 | 0 | **100** | **100** | 0 | 0 | 44 | 37 |
| Precipitation | | | | | | | | |
| No | 90 | 85 | **100** | 38 | 82 | 100 | 61 | 77 |
| Yes | 10 | 15 | 0 | **63** | 18 | 0 | 39 | 23 |
| Vehicle manoeuvre | | | | | | | | |
| Ahead | **100** | **100** | 72 | **100** | 55 | 100 | 69 | 87 |
| Turning | 0 | 0 | **28** | 0 | **45** | 0 | 31 | 13 |
| Pedestrian (age-sex) | | | | | | | | |
| 0–7 years | 8 | **22** | 4 | 0 | **55** | 50 | 14 | 13 |
| 8–15 years | 24 | **44** | 8 | 4 | **45** | 0 | 42 | 27 |
| Female | 36 | 11 | **24** | **38** | 0 | 0 | 14 | 23 |
| Male | 32 | 22 | **64** | **58** | 0 | 50 | 31 | 37 |
| Pedestrian movement from… | | | | | | | | |
| Left | 58 | **100** | **100** | 29 | 0 | 50 | 39 | 59 |
| Right | 34 | 0 | 0 | **58** | **100** | 50 | 58 | 37 |
| Other | 8 | 0 | 0 | **13** | 0 | 0 | 3 | 5 |
| Pedestrian speed | | | | | | | | |
| Walking | **100** | 0 | **96** | **100** | 0 | 0 | 42 | 65 |
| Running | 0 | **100** | 4 | 0 | **100** | 100 | 58 | 35 |
| Line of sight obstructed (1 sec) | | | | | | | | |
| No | 90 | 74 | **100** | **100** | 100 | 100 | 69 | 87 |
| Yes | 10 | **26** | 0 | 0 | 0 | 0 | 31 | 13 |
| Vehicle travel speed (km/h) | | | | | | | | |
| Mean | 43 | 35 | 48 | 51 | 37 | 87 | - | 44 |
| Change of speed to impact (km/h) | | | | | | | | |
| Mean | -7 | -6 | -6 | -7 | -11 | -7 | - | -7 |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

Cluster 1, the largest in the set comprising 29% of the population, has accidents in daylight involving vehicles going ahead and pedestrians walking (Table 10). Other majority characteristics are fine weather and an unobstructed line of sight one second before impact. The mean travel speed was 43 km/h with a reduction of 7 km/h before impact. The range of these last two parameters are shown in Figure 2 and Figure 3. Cluster 2, the second largest, has an over-representation of children running from the left with a tendency to be obscured. This compares interestingly with the corresponding STATS 19 cluster. There are also parallels with the STATS 19 results in clusters 3 and 4, with the tendencies towards serious injury outcomes, darkness, wet weather and adults. Cluster 5 is the closest that a major cluster approaches to a turning scenario, involving children running across from the right side; the mean travel speed is 37 km/h with 11 km/h reduction in speed before impact. This is consistent with the STATS 19 turning scenario which has speed limits and injury outcomes at the lower end of the range. Two of the ten fatalities are in cluster 6 which is too small to support any generalisations, but noteworthy for very high vehicle speeds.

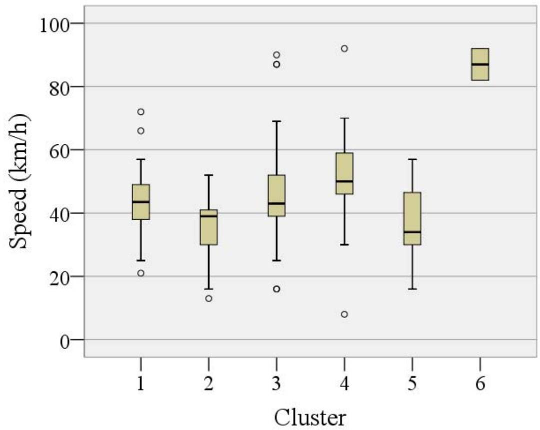


**Figure 2. Vehicle travel speed for clusters 1–6 (OTS): N=175 vehicles**

Figure 2 and Figure 3 show the median values, interquartile ranges (IQR) and outliers for vehicle travel speed and change of speed to impact using *Tukey's hinges* and outliers denoted as 'o' for 1.5–3 IQR and '*' for 3+ IQR [8].

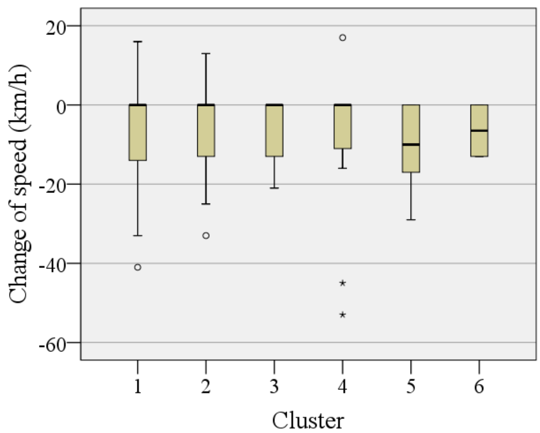


**Figure 3. Change of speed to impact for clusters 1–6 (OTS): N=175 vehicles**

## DISCUSSION

The decisive reason for using cluster analysis to identify groups and associated characteristics in the accident data is that the procedure is objective, reproducible and multivariate. It would not make sense to provide a subjective interpretation of the data to over-ride the key findings presented in Table 8 and Table 10. With this caveat, it is possible to discern some striking parallels between the two sets of clusters.

Firstly the set of characteristics of the largest clusters derived from STATS 19 and OTS mirror the most common features of the accident population, establishing a type of baseline scenario (Table 11).

**Table 11.**
**Baseline scenario**

| STATS 19 Cluster 1 | OTS Cluster 1 |
|---|---|
| ● 39% of population<br>● Daylight<br>● Fine<br>● Vehicle going ahead<br>● 10–30 mph limit<br>● Pedestrian crossing, especially from left<br>● Not masked<br>● Children over-represented minority | ● 29% of population<br>● Daylight<br>● Fine<br>● Vehicle going ahead<br>● Speed 43 km/h<br>● Pedestrian crossing, especially from left<br>● Walking<br>● Not obstructed |

The set of characteristics from the second largest clusters of each dataset differs from the first in having a smaller pedestrian who may be partially or fully obstructed from the line of sight of the driver and (in OTS) is moving faster than walking pace (Table 12).

**Table 12.**
**Smaller pedestrian with obstructed line of sight**

| STATS 19 Cluster 2 | OTS Cluster 2 |
|---|---|
| ● 14% of population<br>● Daylight<br>● Fine<br>● Vehicle going ahead<br>● 10–30 mph limit<br>● Children over-represented majority<br>● Pedestrians crossing, especially from left<br>● Masked by vehicle | ● 15% of population<br>● Daylight<br>● Fine<br>● Vehicle going ahead<br>● Speed 35 km/h<br>● Children over-represented majority<br>● Pedestrian crossing, especially from left<br>● Running<br>● Obstructed for over-represented minority |

The set of characteristics from the third and fourth largest clusters of each dataset involves darkness and potentially wet conditions, with a large pedestrian crossing at walking pace from either side of the carriageway without sight obstruction (Table 13).

**Table 13.**
**Larger pedestrian in darkness and some precipitation**

| STATS 19 Clusters 3–4 | OTS Cluster 3–4 |
|---|---|
| ● 21% of population (combined)<br>● Darkness<br>● Not fine over-represented minority<br>● Vehicle going ahead<br>● 10–30 mph limit<br>● Adult male over-represented majority<br>● Pedestrian crossing from both directions<br>● Not masked | ● 28% of population (combined)<br>● Darkness<br>● Fine (cluster 3) and not fine (cluster 4)<br>● Vehicle going ahead<br>● Speed 48–51 km/h<br>● Adults<br>● Pedestrian crossing from both directions<br>● Walking<br>● Not obstructed |

Having used cluster analysis on the accident data to define a set of pedestrian collision types that have similar features and represent over 75% of the selected cases, it is considered reasonable to use these as relevant scenarios for generation of an AEB testing protocol. In further on-going work, the UK data is being compared to other data sources from different countries to ensure that a global set of pedestrian collisions is represented and initial indications are of a high level of commonality. It is also necessary to ensure that testing procedures are feasible, repeatable and reproducible. The AEB Group is currently considering a number of provisional test conditions which are not, it must be stressed, a final precise list, but are subject to further discussion, definition and finalisation. These are:

- Pedestrian walks from near-side pavement into path of car
- Pedestrian walks from near-side pavement from behind an obstruction into path of car
- Pedestrian runs from far-side pavement into path of car
- Pedestrian walks along near side of carriageway ahead of car
- Pedestrian walks across junction from near-side pavement into path of car turning towards far side into junction.

This data analysis method has also been applied to car-to-car rear and head-on collision types [9]. For each of these collision types, including pedestrian collisions, the AEB Test Group is developing a test scenario along with procedures and targets that will evaluate the effectiveness of the AEB systems for preventing or mitigating these collisions [10].

A restriction on the scope of the results presented in this paper is that they are based on data from a single country. The frequency with which a certain event or combination of factors occurs is naturally dependent on the local road environment, vehicle fleet, driver characteristics and various social and legal factors. At a different level, the formation of clusters is determined in a substantial part by the fields on which accidents are compared. As mentioned above, fields relevant to physical testing were used in this work. If other factors were added or substituted for these, it would not be surprising to see this reflected in the constitution of the clusters. A further consideration relating to the effect of fields is that the number of fields that can be used meaningfully in a cluster analysis is limited by the number of cases. The risk of using too many fields is overfitting of the data with the consequential danger that at least some of the patterns observed would not be maintained with the addition of extra cases. With 175 cases for the OTS analysis and thousands of cases for the STATS 19 analysis, this is a relatively minor concern for the current work. Experience also indicates that the results obtained above are relatively insensitive to fine-tuning of the computational algorithm.

## CONCLUSION

The most common scenarios for pedestrian accidents identified in the STATS 19 and OTS databases are described in Table 8 and Table 10. These include a baseline scenario where a pedestrian steps out from the kerb without obstruction of the driver's line of sight; a similar second scenario where the pedestrian is smaller and at least partially obscured; and a third scenario in adverse meteorological conditions with adult pedestrians. The derivation of these situations from the accident data using cluster analysis is objective and mathematically reproducible, also providing a clear definition of the proportion of the accident population represented by the scenarios.

## REFERENCES

[1] Department for Transport. 2009. Reported road casualties Great Britain, Transport Statistics DfT, London. http://www.dft.gov.uk/pgr/statistics/datatablespublications/accidents/casualtiesgbar/rrcgb2008

[2] Department for Transport. 2004. Instructions for the completion of road accident reports. http://www.dft.gov.uk/collisionreporting/Stats/stats20.pdf

[3] Cuerden R., M. Pittman, E. Dodson and J. Hill. 2008. The UK On-the-Spot accident data collection study—Phase II report, Road Safety Research Report No. 73, Department for Transport, London. http://www.dft.gov.uk/pgr/roadsafety/research/rsrr/theme5/onthespotaccident.pdf

[4] Romesburg, H.C. 2004 Cluster analysis for researchers, Lulu Press, North Carolina.

[5] Martinez, W.L. and A.R. Martinez. 2005. Exploratory data analysis with MATLAB®, Chapman & Hall, London.

[6] For example: Skyving, M., H-Y. Berg and L. Laflamme. 2009. A pattern analysis of traffic crashes to older drivers, Accident Analysis and Prevention 41, pp. 253–8.

[7] Erbsmehl, C. 2009. Simulation of real crashes as a method for estimating the potential benefits of advanced safety technologies, ESV, 09-0162.

[8] SPSS Inc. 2009. PASW® statistics 18 command syntax reference, release 18.0.0, p. 683.

[9] Lenard, J. and R. Danton. 2010. Accident data study in support of development of autonomous emergency braking test procedures, project reference LUEL 5989/6175, Thatcham Research Centre, England.

[10] Further details at www.thatcham.org/AEB.

## ACKNOWLEDGEMENTS

The funding for this analysis of UK accident data was provided by Thatcham (Motor Insurance Repair Research Centre) and the Insurance Institute for Highway Safety.

The OTS project was funded by the Department for Transport (DfT) and the Highways Agency. This project was not possible without help and support from many individuals, especially the Chief Constables of Nottinghamshire and Thames Valley Police Forces and their officers. Permission from the DfT to use STATS 19 (2008) for the analysis of pedestrian, rear-end and head-on accidents is also gratefully acknowledged.