Transport Research Laboratory

Creating a greater understanding of pedestrians' injuries received in road traffic accidents using STATS19 and HES data

by R Cookson, D Richards, W Chislett, and R Cuerden.

PPR424

11109475

PUBLISHED PROJECT REPORT

Transport Research Laboratory

PUBLISHED PROJECT REPORT PPR424

Creating a greater understanding of pedestrians' injuries received in road traffic accidents using STATS19 and HES data

by R Cookson, D Richards, W Chislett, and R Cuerden. (TRL)

Prepared for: Project Record:	**11109475**
	Creating a greater understanding of the injuries which vulnerable road users receive in road traffic accidents using STATS19 and HES data
Client:	**TRF, Scientific Executive**
	(Rod Kimber)

This Published Report has been prepared for TRF. Published Project Reports are written primarily for the Client rather than for a general audience and are published with the Client's approval.

The views expressed are those of the author(s) and not necessarily those of TRF.

	Name	Date Approved
Project Manager	Rebecca Cookson	08/12/2008
Technical Referee	Richard Cuerden	08/12/2008

Contents Amendment Record

This report has been issued and amended as follows

Version	Date	Description	Editor	Technical Referee

Contents

List of Figures

List of Tables

Executive summary

Every year in the UK thousands of people are killed and tens of thousands seriously injured in traffic accidents. From these accidents an understanding of how injuries occur in accidents is sought in order to implement ideas to mitigate them, and a major part of gaining this understanding is looking at national level statistics. There are a number of accident research studies in Great Britain, on both a macro and a micro scale. This project explored the possibilities of using a large medical database for accident research purposes. The report documents the possible methodology for analysis of medical databases for accident research and shows some of the ways in which the data could be used.

TRL successfully collaborated with the South East Public Health Observatory (SEPHO) who provided Hospital Episode Statistics (HES) data for pedestrian casualties admitted to hospitals in England from April 1998 to March 2007.

The HES database, along with the Police STATS19 database, was used to investigate the demographics of pedestrian accidents. This showed how the two databases can be used in their current form to determine casualty groups which may be of interest.

Apart from being a large group of the road casualties in Great Britain, there are a number of reasons for investigating pedestrians in HES and STATS19. Arguably, compared to car occupants, the injury epidemiology and characteristics of pedestrians are less well understood, because of the lack of large pedestrian-focused accident studies. Also, new European Union pedestrian regulations have meant that the design of cars has changed, so it is important to find any corresponding change in pedestrian casualties.

Because it is a medical database, the strength of the HES database is the amount of detail recorded on the injuries received, and the operations undergone by each casualty. These were investigated for the pedestrians in the HES dataset. This shows that the most frequently recorded injuries for pedestrians are head and leg injuries, which agrees with previous studies on smaller samples of pedestrians. The most frequent operations were the reduction of bone fractures.

The large number of casualties recorded in the HES database (every patient admitted to hospital in England) means that the injuries due to road traffic accidents can be investigated on a scale not previously available in this country. This enables changes over time to be investigated; for example, fractures of the tibia and fibula are seen to have declined over time. This points towards an improvement in car design having a positive impact on the injuries received by pedestrians in impacts.

1 Introduction

Every year in the UK thousands of people are killed and tens of thousands are seriously injured in traffic accidents. As well as the personal tragedy of these events, road traffic accidents have economic implications. An understanding of how injuries occur in accidents is sought in order to implement ideas to mitigate them, and a major part of gaining this understanding is looking at national level statistics. The importance of these statistics is such that they can affect Government and local authority initiatives, policy, spending, and legislation, and even vehicle manufacturing decisions. Hence the need for them to be accurate and reliable is apparent.

There are a number of accident research studies in Great Britain, on both a macro and a micro scale. This project explores how another source of data can be used for accident research: the Hospital Episode Statistics, which are recorded by hospitals in England.

A traditional source of traffic accident data recorded by the Police (STATS19) and the data recorded by the hospitals (Hospital Episode Statistics) will be investigated. While the purpose of STATS19 is to record traffic accidents, recording details of traffic casualties is only a small part of the Hospital Episode Statistics (HES). As such, only limited accident analysis has been performed using HES in the past. This project provides the opportunity to display how HES can be used in accident investigation, and how it relates to STATS19. This report concentrates on the characteristics of the pedestrian casualties in the two datasets.

1.1 Overview of pedestrian accidents in Great Britain

The number of killed and seriously injured pedestrians per year has been decreasing from 1996 to 2006 (see Figure 1-1); however the decline has been less from 2004 to 2006. This trend is the same for pedal cyclist and car occupant casualties. Motorcyclist casualties increased between 1996 and 2003, decreasing back to the 1994-98 average by 2006.

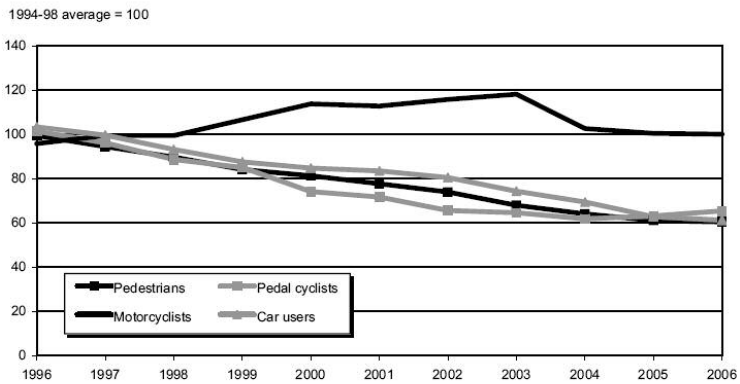

Figure 1-1. Trend of road casualties in Great Britain by road user type (Department for Transport et al., 2007)

In 2006, 1.2% of all road user casualties suffered fatal injuries. Both pedestrians and motorcyclists had above average rates of fatalities with 2.2% and 2.6% of those injured being killed respectively (Department for Transport et al., 2007). In total, of all the fatally injured casualties on Britain's roads in 2006, 675 (21%) were pedestrians. Of all the killed or seriously injured casualties, 7,051 (22%) were pedestrians.

Apart from being a large group of the road casualties in Great Britain, there are a number of reasons for investigating pedestrians in HES and STATS19. Arguably, compared to car occupants, the injury epidemiology and characteristics of pedestrians are less well understood, because of the lack of large pedestrian-focused accident studies. Also, new European Union pedestrian regulations have meant that the design of cars has changed, so it is important to find any corresponding change in pedestrian casualties.

Pedestrian casualties are often the most vulnerable members of society. Children have a high exposure to traffic as pedestrians, especially on their journey to and from school. Elderly people, who may not have another form of transport available to them, are also exposed and at greatest risk of serious injuries if they are involved in an accident. Additionally, these are the two groups of pedestrians who are less well equipped for the road-crossing task.

Finally, previous studies (Ward, Lyons, and Thoreau, 2006) have found that the difference in the number of casualties recorded in HES and STATS19 is greatest for vulnerable road users. Vulnerable road users include pedal cyclists and pedestrians; the relationship between HES and STATS19 for pedal cyclists has already been explored (Department for Transport et al., 2007). For the above reasons, this project will look at pedestrian accidents in HES and STATS19 to explore how a large medical database can be used for accident research.

2 Literature review

The primary purpose of this literature review was to locate previous research which had used both STATS19 and hospital data to investigate some aspect of traffic accidents. This includes describing the differences between the two datasets, as well as any studies which have used them together. Also included in Appendix A is a review of trials which have linked hospital data to accident data.

2.1 Method

A comprehensive review of international research literature relating to the comparison and connection of Police and hospital road traffic accident statistics was carried out.

Searches of the TRL Knowledge Base were conducted in order to identify literature to be reviewed. The Knowledge Base comprises a number of databases, including the Transport Research Abstracting and Cataloguing System (TRACS). This is the main catalogue of transport research publications held both in the TRL library and elsewhere. It contains bibliographic references and abstracts of English and foreign language articles from journals, books and research reports. It is the English language version of the worldwide ITRD (International Transport Research Documentation) database and contains abstracts from publications in the USA, Australia, Scandinavia, the Netherlands and Canada, in addition to UK material. In order to find relevant material, searches for records that specifically mention STATS19 and Police statistics in relation to hospitals were performed. These were done using keyword searches of "STATS 19", "Hospital and Police statistics", "Hospital Episode Statistics", and "HES" amongst others.

Further to this, the Department for Transport's (DfT) online material was searched, and a general internet-wide search was conducted using a variety of search engines and keywords. These were similar to those mentioned above but also included many others such as "linking Police and hospital data", "hospital admissions and STATS19", "SHIPS", "under-reporting", "traffic accident Police reporting", and "STATS19 and HES".

Throughout this search further studies were identified by browsing between publications using their lists of references.

2.2 Differences between STATS19 and HES

Variations between the STATS19 and HES datasets arise due to their different purposes and collection methods. STATS19 was designed to provide detailed statistical reports and highlight trends of road casualties over time, and relies heavily on Police continuity. Hospital Episode Statistics stem from a system used to provide administrative and medical data to the health service, where any information about road traffic accidents is simply a useful by-product (Department for Transport et al., 2007). The table below is an extensive list of the ways in which STATS19 and HES differ and why (Department for Transport et al., 2007).

Table 2-1. Differences between HES and STATS19 (Department for Transport et al., 2007)

Topic	STATS19	HES
Coverage	Confined to accidents on the public highway involving injury reported to the Police, including slight injuries where patients are not	Covers only patients admitted to hospitals, which currently excludes attendance at A&E only. Includes accidents not

Topic	STATS19	HES
	admitted to hospital. Excludes confirmed suicides; death from natural causes; injuries to pedestrians with no vehicle involvement (e.g. falls).	reported to the Police. ICD codes allow non-traffic accidents to be excluded, but only where they are known to have taken place off the highway. Accidents are assumed to have occurred on the public highway unless another place is specified. Excludes patients not admitted to hospital, e.g. treated in A&E, by GPs etc, or not requiring professional treatment.
Fatalities	Almost all fatalities within 30 days of the accident are likely to be recorded.	Many fatalities take place at the scene of the crash or in A&E, and are never admitted to hospital. Fatalities can be identified in HES, but deaths may be later than 30 days after the accident. These are identifiable if they are still in hospital when they die.
Injury details and severity	Fatal, serious or slight only. 'Serious' accidents include all admitted to hospital or any of the following injuries: fractures, concussion, internal injuries, crushings, non-friction burns, severe cuts, severe general shock requiring medical treatment and any injuries causing death 30 or more days after the accident. Note that severity is judged by the Police, who are not medical experts and may misclassify serious injuries as slight or vice versa.	Very detailed coverage of medical diagnosis using ICD codes. Intensive investigations are likely to lead to much more accurate severity diagnosis, however, ICD codes have no measure of severity.
Duplicate Records	Unlikely	Records created for each 'finished admission episode'. The same patient can be counted more than once if they are discharged from hospital and then readmitted.
Accident details	Extensive, including all vehicles involved, location of accident.	Few – confined to identifying vehicle type(s) or pedestrian involved.
Patient details	Age (may be estimated), sex, home address postcode in about 75 per cent of cases.	Likely to be fuller and more accurate, including home postcode in most cases.

Further to these differences, care should be taken when considering the overall reliability and compatibility of data sourced from across the country and through time. Where the STATS19 database is cross-checked and validated at a local and national level, the health databases are essentially a descriptive record for health audit purposes and are not internally validated as such (Ward et al., 2006). Also changes in procedures, coding schemes, organisation and funding have caused fluctuations in HES data over the years, as is warned on the HES website (HESonline, 2008).

2.3 Comparison of the HES and STATS19 datasets

A large amount of work has previously been carried out on comparing Police (STATS19) data to information collected from hospitals, although the source of hospital data used in studies has varied considerably. Stone (1984) was one of the first to identify the need to improve Police records of road traffic accidents by adding hospital treatment data, and developed a method to do this via computer.

Overall differences and trends for various road user types have been discussed and analysed by the DfT (2006 and 2007), the DfT et al. (Road Casualties of Great Britain - 2007) and Goldacre et al. (2006), but until recently individual records had never been one-to-one matched. The Office for National Statistics have now performed one-to-one matching of the STATS19 and HES datasets, the results of which are presented in Department for Transport et al., 2008. These results are discussed in section 2.3.3. Other research into linking Police statistics with a variety of hospital databases is discussed in Appendix A.

2.3.1 Overall variation between the datasets

The general opinion is that since 1996 Police statistics (STATS19) have shown a reduction in the number of people killed or seriously injured, whereas hospital data (HES) has not and has remained fairly constant. This differs when looking at individual road user types and is discussed in section 2.3.2 of this report. Goldacre et al. (2006) quantifies the overall reduction and notes a fall in killed or seriously injured road users from 85.9 to 59.4 injuries per 100 000 total population when looking at Police held STATS19 statistics from 1996 to 2004. And in the same time frame HES data actually shows an increase from 90.0 to 91.1 hospital admissions per 100 000 total population. There are many possible reasons for the variation between the datasets and, without linking individual records and maintaining these links over time, no absolute answers could be found. This general trend is shown in Figure 2-1 as presented by Goldacre et al.

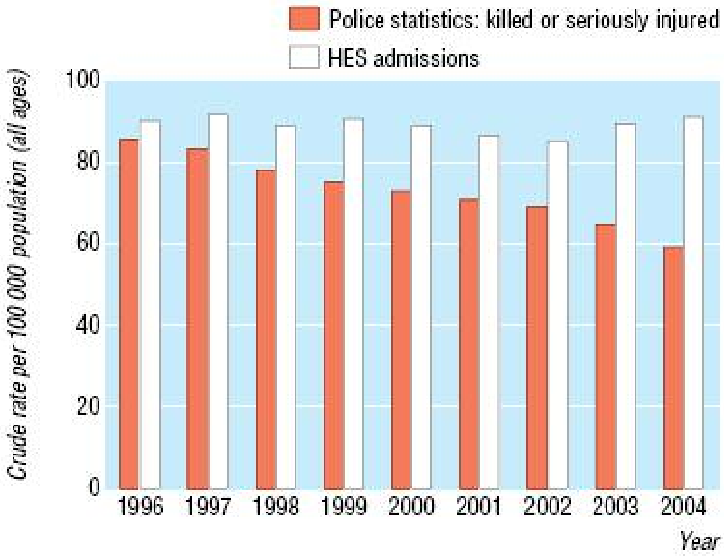

Figure 2-1. Traffic injury rates from HES and Police statistics for people killed or seriously injured (all ages and road user types) – Goldacre et al., 2006.

One of the main factors that has been discussed in nearly all reports in this field is under-reporting. Ward et al. (2006) have produced an entire study into the under-reporting of road casualties and used HES and STATS19 to illustrate the problem, and Simpson (1996) also highlights this problem in her data: "National road accident and casualty statistics are based on those reported to the Police but many potentially reportable accidents are not reported" (Simpson, 1996). Other contributing factors include:

- Changes in types and severities of road traffic injuries, which will affect the two datasets differently. For example, a reduction in less severe injuries would be more evident in Police records as they are still coded as "serious" here, but may not even appear in hospital data due to only being treated in A&E. Attempts to identify the magnitude of this problem are hindered by under-reporting affecting more or less severely injured casualties differently (Department for Transport, 2006).
- Hospital practices and procedures for recording data change over time and are difficult to monitor. Changes in the way in which casualties are treated also affect this, as is pointed out by the DfT et al. (2007) in showing how the number of overnight admissions has increased steeply over the last 4 or 5 years. The statistics are further complicated by the recent introduction of 'payment by results' which has naturally encouraged the more accurate and fuller recording of injuries. This, along with other changes made to the coding and IT systems used, will all affect any trends seen in the statistics over time.
- Police practices also change over time and varying emphasis within Police training can encourage/discourage the ways in which injuries are recorded, depending on Police priorities at the time. As well as varying over time, Police procedures vary between forces, also making analysis of this factor difficult.
- As part of the issue with under-reporting, differences between datasets have partially been attributed to accidents occurring off-road and so have no Police involvement, yet they are still included as traffic accidents in HES data due to incorrect ICD-10 coding. The DfT (2006) comparison of HES and STATS19 looks at this in more detail and backs up the theory by pointing out that the users of vehicles most associated with off road use (such as cyclists) show the greatest variation when comparing hospital and Police statistics, although it is also

possible that these accidents are less likely to be recorded in STATS19 if they occurred on the road.

As stated by the DfT (2006) "The relative importance of these factors is likely to vary among different road user groups, but the relative significance of each cannot be determined from currently available information. The matching of individual HES admissions to individual STATS19 records (1 to 1 matching) may provide a better understanding of the impact of these three factors in each road user group." Hence again emphasising the need to properly link data from the two sources before any real conclusions can be drawn.

Ideally the factors listed above need to be investigated further as a greater understanding of them is the key to explaining the trends seen in each data source over time. This could be done by liaising with the health service and Police and by further detailed analysis of variation in both datasets with time (whilst bearing in mind changes in procedures/policy that have been made).

2.3.2 Road user types

Different road user types exhibit different trends when their hospital and Police accident statistics are compared. Most reports that have studied these two sources of data have also analysed how each road user group is represented within them. Stone (1984) did this very early on but concentrated on how successful the linking was for each group rather than the reasons as to why it varied between them. Simpson (1996 and 1997) was one of the first to produce a very comprehensive study into variation with age, sex, road user type and even the vehicle configuration of the accident (e.g. types of vehicle involved in a collision). The only drawbacks with the studies were that the hospital data used was from a survey of A&E wards and not HES and that the data used is over ten years old now. Therefore the results can't be used to predict exactly how modern HES data will behave when similarly analysed, but nonetheless they give a good indication. The same is true of the Keigan, Broughton, and Tunbridge (1999) study of Scottish hospital data and Police statistics, although this study does discuss the relative success of linking records for each road user group and could be used as a sign of how successful linking could be when using HES instead.

The reports that have looked at the representation of different road user groups within HES statistics and compared them to STATS19 are mainly those written by the DfT. The 'Road Casualties of Great Britain 2006' report (Department for Transport et al., 2007) compares the overall number of injuries for each road user group in Police and hospital databases, and then goes on to offer possible explanations for discrepancies between the two. It is only done for the years 2005/6 though, and does not look at the trends in each road user group over time. Another DfT report on comparing HES and STATS19 (Department for Transport, 2006) does cover this topic, and plots the variation of each road user type in both datasets from 1996 to 2003 (Figure 2-2 and Figure 2-3 show examples of two of these).

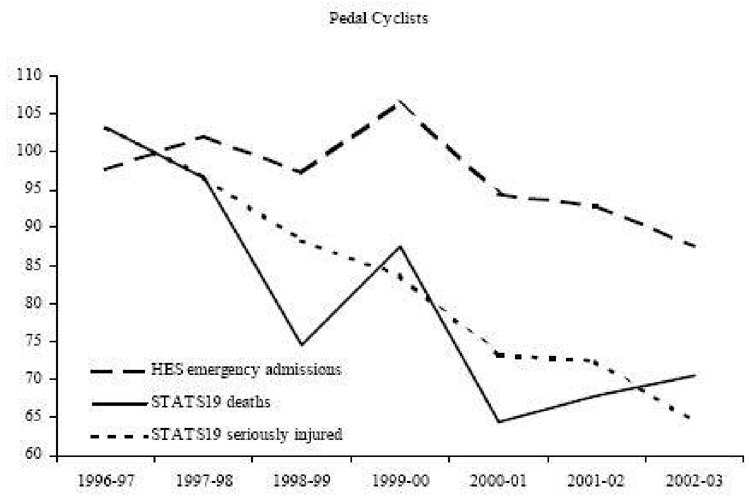

Figure 2-2. A comparison of HES and STATS19 data for pedal cyclists (100=1996/7 average)

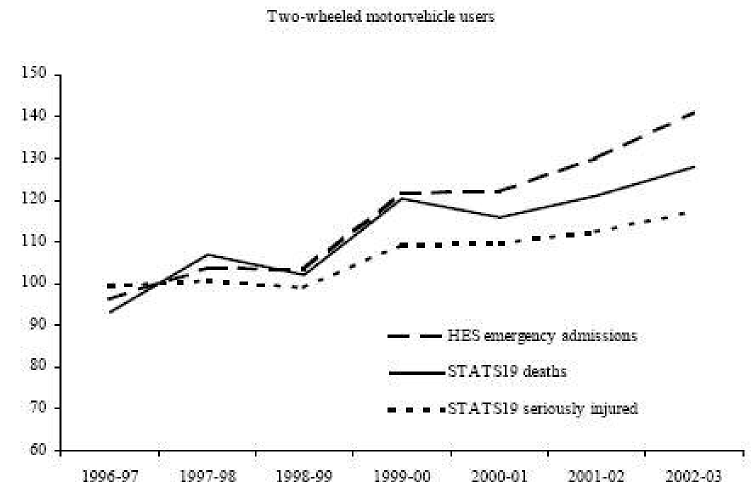

Figure 2-3. A comparison of HES and STATS19 data for two-wheeled motor vehicle users (100=1996/7 average)

Figure 2-2 and Figure 2-3 clearly highlight the problems with this data and why further work is needed in this area. Although the trends in HES and STATS19 are similar, the magnitudes of the differences are different for each road user type. It is also apparent that the divergence of HES and STATS19 appears to be getting worse as time goes on and without monitoring this problem, the reasons cannot be understood and action cannot be taken. Of particular interest on this topic are the trends shown by car occupants in recent years. Police collected STATS19 data has been showing a reduction in the number of seriously injured car occupants, whereas hospital (HES) data has not. As STATS19 data in particular is used for a large range of analyses which affect road safety, it is important that different sources of accident data are compared to ensure that any analysis used is representative of the accidents that are actually occurring on the roads.

In every study that has compared hospital and Police statistics, vulnerable road user groups have been highlighted as those that show the most variation, regardless of where the data is sourced from. In particular pedal cyclists are mentioned in nearly all reports on the subject along with possible explanations as to why this is the case. The general consensus is that mis-coding of cycling accidents that have occurred off road could contribute, along with under-reporting of accidents that involve child cyclists or those where the damage to personal property is minimal, including single vehicle accidents. Despite this the DfT et al. (2007) rightly recognises the importance of this user group due to the potential for serious injury or death within it. As a result this report includes analysis of injured pedal cyclists' ages, injury locations, collision partners and lengths of stay in hospital within HES and STATS19 records. It highlights how analysis of pedal cyclist injuries can become very confusing and difficult to fully explain. For example, when pedal cycle accidents that do not involve a collision partner are excluded, the data seems to show proportionally more pedal cyclists in STATS19 than HES, but this is not the case for all age groups and changes when looking at different locations of accident. The issue with cycling is that it involves all age groups and comes in many different forms (road, mountain, social, exercising etc.) and so it is very difficult to analyse any data collected without taking the accident scenario and journey purpose into account. Again, only overall numbers from each of the datasets were used and HES and STATS19 records are not linked one-to-one, but due care is taken in the report and the need for further work is recognised again.

The same applies to other vulnerable road users as an understanding of the way in which they sustain injuries (and the levels of under-reporting of these) is pertinent to their prevention. More detailed analysis of the HES and STATS19 data for pedestrians and two-wheeled motor vehicles could be carried out in the same way as was done for pedal cyclists by the DfT. It could be helpful to use the previous findings on these user groups from other studies that do not use HES to aid this analysis of STATS19 and HES together.

2.3.3 Results of linking HES and STATS19

The results of the Office of National Statistics linking of HES and STATS19 were published in Road Casualties Great Britain 2007 on 25th September 2008 (Department for Transport et al., 2008). The linking was performed using the fields common to both databases, which were: age of casualty; gender of casualty; home postcode of casualty; date; casualty type; casualty class; and local authority (of the accident in STATS19, and the patient's home in HES). Slightly and seriously injured casualties in STATS19 were selected, and non-fatal casualties selected in HES, for the years 1995 to 2004.

Each match was given a confidence level of low, medium, high, or very high, depending on how many fields matched. The proportion of casualties matched for all levels of confidence was 52%, and the proportion matched with very high confidence was approximately 15%. The proportion of casualties matched did not vary over the 6 year period.

The paper concentrates on the characteristics of the casualties matched with very high confidence. The most significant finding was that approximately 40% of the casualties matched in STATS19 and HES were recorded as being slightly injured. A decrease in the proportion of matches correctly classified as serious in STATS19 was observed, from 63% in 1999 to 58% in 2004. The paper also compares the characteristics of the records matched with very high confidence, and unmatched records.

The paper also performs a short analysis of the pedestrian casualties in STATS19 and HES in 2006/07, looking at the 5,525 seriously injured pedestrian casualties in STATS19 and the 7,688 pedestrian casualties admitted to hospital in England, regardless of whether the records are matched or not. This analysis compares the age and gender of the pedestrians in the two datasets, then briefly investigates the length of stay in hospital and the body regions injured.

2.4 Future work

2.4.1 Further analysis of HES and STATS19

The benefits of linking individual records (1 to 1 matching) across Police and hospital datasets has been seen in a number of reports dating back to the 1980's and yet this has only recently been done for the STATS19 and Hospital Episode Statistics datasets. Overall comparisons have been made in the past and discussed here, but as mentioned previously there is great difficulty in drawing firm conclusions from this. A Scottish Government (2006) report on how comparable Police road casualty statistics are to those from other sources summed this up in saying:

"The work done so far cannot answer the question of whether there has been a systematic change in the degree of under-reporting. Any conclusions drawn from a simple comparison of 'STATS19' and 'hospitals' figures would be misleading."

As mentioned earlier, the DfT began direct matching of HES and STATS19 to create a linked database because "the matching of individual STATS19 records to individual HES records (1 to 1 matching) may provide a better understanding of the differences between the trends. For instance, 1 to 1 matching could show whether, for particular road user groups, there has been a change over time in the number of injuries classed as less serious in HES and serious under STATS19. This would indicate a decline in the number of less seriously injured road users, which may be attributed to the increase of safety features introduced over recent years." (Department for Transport, 2006)

In the same DfT report comparing STATS19 data with Hospital Episode Statistics, the following actions are suggested as a result of their analysis:

> "1. Undertake regular comparison of STATS19 and HES data on road accident casualties.
>
> 2. Consult with the Department of Health to obtain available information on the reliability and consistency of coding of HES data and recent trends in data collection practices.
>
> 3. Attempt one-to-one matching of STATS19 and HES records to try to establish the trend of more seriously injured casualties within STATS19."

The regular analysis of both datasets is easily practicable and would be invaluable to obtain the trends that occur over time in each of them. It would allow an assessment to be made of the scale of discrepancies between the two and identify if action is required (analysis so far shows that it is). Point two above recommends consultation with the Department of Health, which could also be useful but probably only when done in conjunction with Police force consultation. A greater understanding is needed of how both datasets are populated, and not only how data is intended/designed to be collected, but how it actually is in the real world. This should be one of the very first steps as only when what the data really represents is understood can effective analysis be undertaken.

2.4.2 Vulnerable Road Users (VRUs)

The greatest discrepancies between Hospital Episode Statistics and Police recorded STATS19 data appear to lie in the vulnerable road user groups. As mentioned earlier, the linking of the datasets which has now been performed could help provide an in-depth understanding of the actual injuries that result from different traffic accident scenarios. It could also provide a much needed modern understanding of under-reporting. This literature review has shown that VRUs have been highlighted in almost every study of the subject for being the most under-represented in STATS19 data when compared to hospital admissions. The complicated nature of comparisons of the datasets has already been mentioned in this report and the need for further work has been highlighted. In particular the need for further work in understanding how injured cyclists are coded in hospitals and dealt with at accidents by Police has become apparent, along with more general analysis of other VRU groups. A possibility for work in this area would be to take older analysis that has matched VRU hospital and Police data, investigate the reasons which account for the large discrepancies, and apply these to newer work on HES and STATS19. This could involve using previous knowledge to attempt to manipulate the new data into accounting for discrepancies.

2.4.3 Feedback to Police and hospital data collection

A system to continually relate Police STATS19 records to their corresponding Hospital Episode Statistics could be beneficial to both of the data collection systems by generating invaluable feedback. It could allow some standardisation of hospital admission data across the country which in turn would encourage appropriate funding levels. For Police forces it could aid them in correctly assessing injury severity and understanding the implications of the accidents they attend. It could also alert them to the number of accidents that go un-reported or that are not properly recorded by the attending Police officer. If the importance of the STATS19 database was also related back to these officers then it may increase the levels of reporting, particularly in VRU groups where this is especially needed. Obviously to do this, continual comparisons of the hospital and Police statistics is required and the need for further work highlighted.

2.5 Summary of literature review

There has been a huge amount of work done in this field for many years and any plans for future studies should be made with careful consideration of this. It should be carried out in a way so as not to duplicate any work already done, which may include studies from different locations and using different data sources. Bearing this in mind, the possibility of furthering the work done on vulnerable road user statistics should be investigated. Nearly every piece of literature that compares Police and hospital data has highlighted this group (in particular pedal cyclists) and any research that can improve the understanding of how different sets of data are related would be of benefit.

The DfT has very recently created a matched database of STATS19 and HES records, but has currently only performed a very brief analysis of this. Apart from this, their reports include overall comparisons for different road user groups and ages (including a review of cyclists), the trends seen in these groups over time and various other factors that have contributed. Other studies from a variety of sources have looked at hospital and Police data in more detail, but not by using HES data. Instead SHIPS data, A&E surveys, TARN statistics or other sources have been used, but mostly with fairly similar findings.

As a result of this literature review, general recommendations for further work include:

- Continual analysis of the representation of vulnerable road user groups with STATS19 and HES, with the trends for these groups being tracked over time. There has been a lot of this done before and so any future work should be focused on attempting to explain the trends and discrepancies seen rather than simply pointing them out. For pedal cyclists especially this should be done in a way that complements analysis done by the DfT already.
- General research into the two data sources, including continuing to compare them over time in a way that does not duplicate the activities of recent reports by the DfT. This could also involve looking at the way in which data is collected in the real world and how hospital staff and Police play a part shaping the data collected.

- Feedback to Police forces and hospital authorities could be provided using any findings from further research or analysis of previous work. This would benefit the hospital authorities because it would help them allocate funding appropriately, and increase the Police knowledge of the accidents they attend, and where accidents or injury severity may not be correctly recorded.

From the findings of this literature review, it was decided that this project would focus on the gaps in the research of VRUs, with particular focus on pedestrians, their accident characteristics and injuries. Because the linking of STATS19 and HES was not completed at the start of this project, this analysis was performed without linking the datasets. Dependant on the success of the linking, this work could be enhanced by the addition of analysis on the linked set of data.

3 Methodology

The aim of this project was to examine how a large medical database can be used for accident research. As an example of the methodology required and the results which can be obtained, data from HES and the Police was used to outline in detail the characteristics of pedestrian crashes and the nature, type, and severity of injuries which were sustained. By analysing data collected by hospitals on pedestrian casualties, and the data collected by the Police (STATS19), it is possible to develop a more sophisticated toolkit for accident research, where studies can be carried out not just on the types of accidents occurring but also the injuries and treatment received, which is identified after they have left the scene.

While the purpose of STATS19 is to record traffic accidents, recording details of traffic casualties is only a small part of the HES. As such, only limited accident analysis has been performed using HES in the past. This project provided the opportunity to investigate how HES could be used in accident investigation, and how it relates to STATS19.

This project has explored the methodology of creating a link between hospital and accident research professionals, by working with the South East Public Health Observatory (SEPHO) to gain access to HES data. This included establishing a link between the two organisations, then finding a topic on which the two organisations could work together to investigate the benefits of the link.

The topic chosen was pedestrian accidents, and the particular groups of pedestrians who are at a greater risk of severe injury. From the HES data provided to TRL which records patients admitted to hospital across England, it was possible to see what analysis could be carried out if this link was upheld. In addition to this information, STATS19 data already held at TRL was analysed to determine how it differed to the information in HES, and how the two databases could be used together.

3.1 South East Public Health Observatory (SEPHO)

The South East Public Health Observatory is one of nine health observatories in England and Wales, and is the lead observatory for work related to transport (SEPHO, 2008). One of the aims of this project was to build a working relationship between SEPHO and TRL. As part of this relationship, SEPHO provided TRL with anonymous HES data, and helped with the interpretation of this data. The data was held on the secure TRL network, and the data analysis was carried out by the TRL authors.

3.2 Hospital Episode Statistics (HES)

Hospital Episode Statistics are compiled by the Department of Health and record details of all hospital admissions, finished consultant episodes (FCEs) and hospital discharges for England. Data of this type has been collected since 1989, with its main purpose being to ensure correct funding of hospitals from their Primary Care Trust (PCT) (Department for Transport et al, 2007). HES contains data such as age, sex, dates of admission and discharge, diagnoses, operations and procedures, place of residence and ethnicity, with approximately 12 million new records being added each year. Information regarding the diagnosis of injury and its causation is coded using the 'International Statistical Classification of Diseases, Injuries and Causes of Death' (ICD), of which the latest version ICD-10 has been used since 1995. Injuries sustained in road traffic accidents can easily be identified when coded in this way. It should be noted that HES do not include details of any casualties treated in Accident and Emergency (A&E) that are not admitted to hospital (Department for Transport, 2006).

3.2.1 Gaining access to the data

Due to previous communications with the deputy director at SEPHO (now the director of the National Obesity Observatory), a link with SEPHO was readily available. He put TRL in contact with various employees at SEPHO who work on data entry and analysis of the HES database. Through this link it was possible to carry out the work in this project.

In order to gain access to the HES data, a declaration had to be signed in order to ensure that the data was used correctly, and that patient confidentiality was upheld. It was also agreed that the members of SEPHO who helped with the gaining of the data would be acknowledged in the report.

3.2.2 Understanding the data

3.2.2.1 International Statistical Classification of Diseases, Injuries and Causes of Death (ICD)

In HES, injuries are recorded in 14 fields (7 before April 2002), which contain information about a patient's illness or condition (HESonline, 2008). The first of these fields contains the primary diagnosis and the other fields contain secondary/subsidiary diagnoses, in addition to the primary diagnosis. This means that a maximum of 13 different injuries can be recorded for each patient. The codes are defined in the International Statistical Classification of Diseases, Injuries and Causes of Death (World Health Organisation, 1992). HES records currently use the tenth revision (ICD-10). Prior to April 1995, the ninth revision was used (ICD-9). Diagnosis codes start with a letter and are followed by two or three digits.

The ICD10 codes are recorded in HES in both their 3-character and 4-character formats. The first 3 characters of the ICD10 code provide the core classification of the injury, whereas the first 4 characters of the code provide a more specific injury description. An example of this would be a 3-character code of "S01 – an open wound of the head" (World Health Organisation, 1992). When split into its 4-character codes it could be any of the following:

- S01.1 – Open wound of eyelid and periocular area
- S01.2 – Open wound of nose
- S01.3 – Open wound of ear
- S01.4 – Open wound of cheek and temporomandibular
- S01.5 – Open wound of lip and oral cavity
- S01.7 – Multiple open wounds of head
- S01.8 – Open wound of parts of head
- S01.9 – Open wound of head, part unspecified

3.2.2.2 Operation codes

There are twelve fields (four prior to April 2002), which contain information about a patient's operations in HES. The first code contains the main (i.e. most resource intensive) procedure. The other fields contain secondary procedures. The codes are defined in the Tabular List of the Classification of Surgical Operations and Procedures (Surginet, 2008). The current version is OPCS4. Procedure codes start with a letter and are followed by two or three digits. The third digit identifies variations on a main procedure code containing two digits. A single operation may contain more than one procedure.

3.2.2.3 Deprivation

Deprivation is recorded in the HES database using the index of multiple deprivation for each super output area.

Super output areas (SOAs) were designed to improve the reporting of small area statistics in England and Wales (Office of National Statistics, 2008). A range of areas were developed that are a consistent size and whose boundaries do not change. These SOAs are built from the output areas used for the 2001 Census. There are three layers of SOA:

1. Lower layer – Minimum population 1,000, mean 1,500. Built from groups of OAs (typically 4 to 6) and constrained by the boundaries of the standard table wards used for 2001 Census outputs.
2. Middle layer – minimum population 5,000, mean 7,200. Built from groups of lower layer SOAs and constrained by the 2003 local authority boundaries used for 2001 Census outputs.
3. Upper layer – to be determined, minimum size approximately 25,000.

The 32,482 lower layer SOAs (HESonline, 2008) in England and Wales were generated by a computer programme which merged OAs taking into account measures of population size, mutual proximity and social homogeneity. The boundaries were released to the public in February 2004.

The 6,780 middle layer SOAs were generated firstly by a computer, as with the lower layer, then local authorities and other local agencies were invited to propose changes to the draft boundaries in order to establish SOAs that better met local needs.

This project uses the index of multiple deprivation (IMD) decile group for deprivation analyses; this uses IMD overall ranking to identify which one of ten groups a super output area belongs to, from most deprived through to least deprived. The IMD overall ranking is made by combining the seven IMD domain scores using the following weights (HESonline, 2008):

- Income (22.5%)
- Employment (22.5%)
- Health Deprivation and Disability (13.5%)
- Education, Skills and Training (13.5%)
- Barriers to Housing and Services (9.3%)
- Crime (9.3%)
- Living Environment (9.3%)

The SOA with a rank of 1 is the most deprived and 32,482 the least deprived area on this overall measure in the HES dataset.

3.3 STATS19

STATS19 data is comprised of the details of road traffic accidents attended by the Police in Great Britain. In theory the Police are required to attend every road traffic accident that involves an injury and whilst on scene officers fill out a series of standard forms. Details of the nature of the accident, the location, a crude classification of injuries and the overall accident severity are all collected. Officers make a judgement, often without further information from hospitals, and record the severity of the injured casualties and the overall accident as ‘slight’, ‘serious’ or ‘killed’. This data is then collected, collated and analysed by the Department for Transport (DfT).

STATS19 is, in principle, the national database in which all traffic accidents that result in injury to at least one person are recorded, although it is acknowledged that some injury accidents are missing from the database and a few non-injury accidents are included (Ward, 2006). The database primarily records information on where the accident took place, when the accident occurred, the conditions at the time and location of the accident, details of the vehicles involved and information about the casualties. Approximately 50 pieces of information are collected for each accident (Department for Transport, 2007).

Between 1999 and 2004 a system of recording contributory factors was trialled in 15 Police forces around the country. This led to the national adoption of the contributory factor system in 2005. (Department for Transport, 2007).

The accidents that are recorded in STATS19, including the contributory factors, are summarised annually in the DfT "Road Casualties Great Britain" (RCGB) series.

STATS19 consists of four forms, which record the circumstances of the accident, details of the vehicles involved, details of the casualties, and the contributory factors. Up to six contributory factors can be recorded for each accident, from a list of 77 options. These 77 contributory factors are grouped into nine categories: road environment contributed, vehicle defects, injudicious action, driver/rider error or reaction, impairment or distraction, behaviour or inexperience, vision affected by, pedestrian only, and special codes.

Each contributory factor chosen for the accident can be given a confidence of "very likely" or "possible", and can be allocated to a particular person or vehicle involved in the accident. The analysis of contributory factors in this report combines the "very likely" and "possible" factors into one dataset.

3.4 Methodology of analysis

3.4.1 HES

In HES there were 82,908 admissions of pedestrian casualties and 80,116 patients in the original dataset, which shows that some of the patients were admitted more than once. However, some of these admissions were duplicated in the dataset due to coding errors, once these were removed, there were 82,811 unique admissions for the 80,116 patients. This is summarised in Table 3-1.

Table 3-1. Number of admissions and patients in HES

Number of admissions in original dataset	82,908
Number of patients in original dataset	80,116
Number of unique admissions, after duplicate records removed	82,811
Number of patients, after duplicate records removed	80,116

These admissions were then broken down by their accident type classification, using their 4 digit causation code. From these codes the admissions which were described as non-traffic or unspecified non-traffic were eliminated for comparison with STATS19. This resulted in 72,878 admissions for analysis. These 72,878 are shown in black in Table 3-2.

Table 3-2. Accident types

Cause code	Casualty class	Striking vehicle	Accident type	Number of admissions
V010	Pedestrian	Pedal cycle	Non-traffic	676
V011	Pedestrian	Pedal cycle	Traffic	1,135
V019	Pedestrian	Pedal cycle	Unspecified whether traffic or non-traffic	722
V020	Pedestrian	2/3 wheel vehicle	Non-traffic	358
V021	Pedestrian	2/3 wheel vehicle	Traffic	2,749
V029	Pedestrian	2/3 wheel vehicle	Unspecified whether traffic or non-traffic	342
V030	Pedestrian	Car/pickup/van	Non-traffic	3,301
V031	Pedestrian	Car/pickup/van	Traffic	56,114
V039	Pedestrian	Car/pickup/van	Unspecified whether traffic or non-traffic	5,004
V040	Pedestrian	Heavy transport vehicle	Non-traffic	452
V041	Pedestrian	Heavy transport vehicle	Traffic	3,997
V049	Pedestrian	Heavy transport vehicle	Unspecified whether traffic or non-traffic	564
V090	Pedestrian	Other/non-specified vehicle	Non-traffic	1,347
V091	Pedestrian	Other/non-specified vehicle	Unspecified non-traffic	1,243
V092	Pedestrian	Other/non-specified vehicle	Traffic	1,106
V093	Pedestrian	Other/non-specified vehicle	Unspecified traffic	1,145
V099	Pedestrian	Other/non-specified vehicle	Unspecified transport	2,556

3.4.2 STATS19

The pedestrian casualties used for analysis in STATS19 were selected to be only those who were killed or seriously injured as only these pedestrians could have been admitted to hospital and therefore be in the HES dataset. In the same 1998 to 2007 time period there were 64,233 pedestrian accidents in England recorded in STATS19. This consisted of 64,253 vehicles and 65,526 pedestrian KSI casualties. Contributory factors were also analysed for STATS19. Up to six contributory factors can be coded for any accident, and in total there were 29,229 contributory factors recorded for these accidents. The casualties in this dataset were only those of fatal or serious severity, the 65,526 casualties consisting of 6,000 fatalities and 59,526 seriously injured pedestrians.

3.4.3 Overview of demographics

Both HES and STATS19 were interrogated in order to present an overview of the demographics of pedestrian casualties in England. From this a matrix was identified based on the casualties' characteristics and the typology of their crashes. This enabled a number of important casualty groups to be defined, which could be further investigated using the in-depth injury information. Two of these groups were investigated in this report to show what could be done with the data; the methodology for this is detailed in section 3.4.6.

3.4.4 In-depth review of principle trauma

This analysis used the ICD-10 diagnosis codes to analyse the most frequent primary injuries, and injury regions received by pedestrians in HES. Also investigated, was the duration of stay of pedestrians, and type of operations received by pedestrians as classified in OPCS4. ICD-10 and OPCS4 are described in section 3.2.2 of this report. Some of the injuries and the relevant anatomy are explained in the glossary at the end of the report.

3.4.5 Changes in trauma over time

The large amount of data present in the HES dataset from 1998-2007 enabled some variations in pedestrian injuries over time to be investigated. This involved looking at the most frequent 4-character ICD codes in 1998/1999 and 2006/2007 and comparing the injuries received in order to assess whether pedestrian injuries have altered over time. These injuries were then split into injury regions and also compared between these two periods.

3.4.6 Important casualty groups

From the overview of demographics of the pedestrian casualties in HES and STATS19, two important casualty groups were investigated in more detail for this report in order to demonstrate how this could be done. The in-depth trauma and various other characteristics of these casualty groups were analysed using the HES and STATS19 data, the key results of which are shown in section 4.4.

4 Key findings

This section of the report presents an overview of the two datasets, and compares the datasets where possible. The data analysed for this report included all pedestrian casualties contained in HES from April 1998 to March 2007 in England, and all the killed or seriously injured pedestrian casualties in STATS19 from April 1998 to March 2007 in England. The period of April to March is referred to as a financial year.

It is expected that the majority of pedestrian casualties recorded in HES should be present in STATS19. This is because of the definition of a "serious" casualty in STATS19, which includes "detention in hospital as an in-patient, either immediately or later" (Department for Transport, 2004). Most of these should be in HES, which contains all patients "admitted" to hospital. The reverse is not true: there are likely to be many pedestrians in STATS19 which would not appear in HES. This could be because they had a serious injury which did not require admission to hospital, or because they died at the scene so were not admitted to hospital.

4.1 Overview of demographics of pedestrians

Table 4-1 shows the demographics of pedestrian casualties in England from interrogating the HES and STATS19 datasets. The matrix gives the largest groups, or greatest difference from the average found when analysing the different characteristics of the database. Some of the cells of the matrix could not be filled in due to one characteristic being recorded in only HES or only STATS19 as shown in the headings of the rows and columns. From this, two of the interesting groups were selected for this report to investigate further as an example of what could be done with the data. These were 10-15 year olds and females over 70 years, which are highlighted red in the table and the results of the analysis of these two groups are shown in section 4.4.

Table 4-1. Matrix showing overview of demographics of pedestrians in England

	Admission date (STATS19 and HES)	Vehicle type (STATS19 and HES)	Gender (STATS19 and HES)	Age (STATS19 and HES)	Deprivation (HES only)	Location (STATS19 only)	Injury severity (STATS19 only)
Admission date (STATS19 and HES)	Most in 1998/99 (14% in STATS19)						
Vehicle type (STATS19 and HES)	Slight decrease in car/pickup/vans and slight increase in HGVs over time	Biggest group: car/pickup/van (84% in HES, 86% in STATS19)					
Gender (STATS19 and HES)	No change over time	No apparent relationship	Most Frequent: Males (63% in HES, 61% in STATS19)				
Age (STATS19 and HES)	Decrease from 15% to 10% in 5-9 year olds over time	Older pedestrians hit more frequently by pedal cycles than younger, vice versa for car/pickup/vans	10-15 for both genders, and 70+ females (8% in HES)	10-15 year olds (19% in both datasets)			
Deprivation (HES only)	Proportion in most deprived 10% was greatest in 2001/02 (22%)	No apparent relationship	No apparent relationship	30% of under 9 years olds were in the most deprived 10% of SOAs	21% in most deprived 10%		
Location (STATS19 only)	No apparent relationship	58% of 2/3Wheelers and 56% of Heavy transport vehicles or buses on A roads	No apparent relationship	65% of under 9s on unclassified roads		A roads 42% and unclassified 36%	
Injury severity (STATS19 only)	No apparent relationship	Fatal injuries increase for HGV collisions and decrease for car/pickup/vans	65% of fatalities were male	Fatal injuries were more frequent for those over 40		55% of fatal casualties on A-roads, 22% of fatal casualties on unclassified roads	9% Fatalities

4.2 In depth review of principle trauma suffered by pedestrians

4.2.1 ICD analysis

The primary injuries coded using the 4-character codes were analysed and are shown in Figure 4-1. The most frequent primary injury recorded was an unspecified injury to the head which 9,051 pedestrians had. 6,987 pedestrians had a fracture to the shaft of the tibia. The next two top injury categories were fractures to the tibia as well, with 3,526 pedestrians receiving fractures to the lower end of the tibia and 3,054 fractures to the upper end of the tibia.

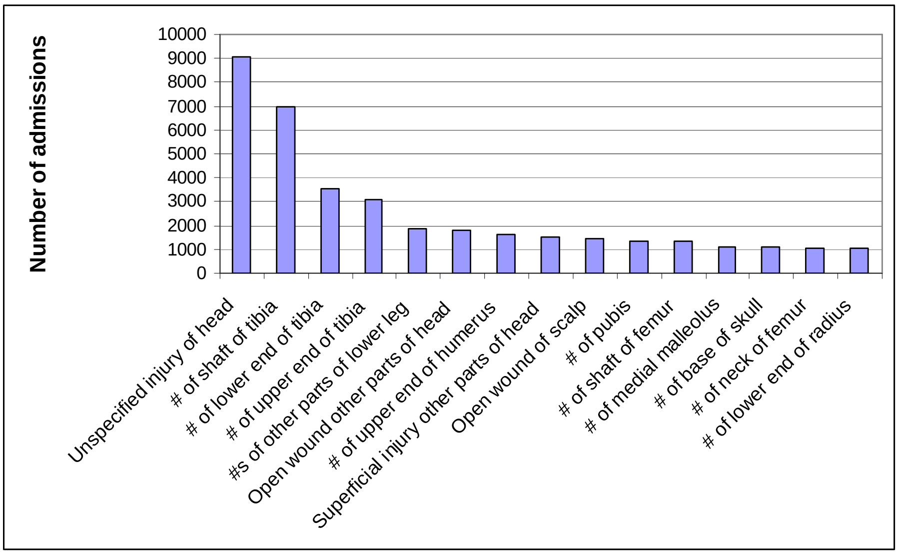

Figure 4-1. Most frequent primary injuries (using 4-character code) in HES

4.2.2 Injury vs. age and gender

The percentage of pedestrians with head injuries decreased with increasing age, as can be seen in Figure 4-2. Hip and thigh injuries were fairly constant for all ages until 59 years of age, after which the percentage of the age group with injuries in that region increased from 5% for 50-59 to 16% for those over 90. Knee and lower leg injuries had the opposite trend, decreasing from 30% for 60-69 year olds to 23% of over 90 year olds. Injuries in the shoulder and arm region were particularly low for those aged up to 9 years, but were then fairly constant for all other age groups.

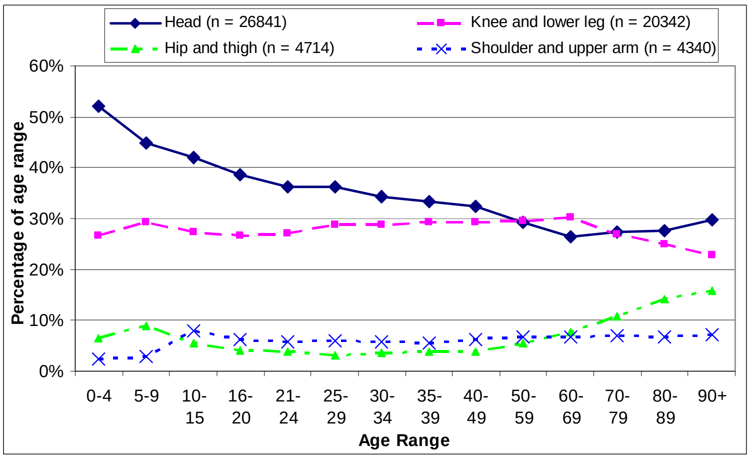

Figure 4-2. Relationship between the most frequent injuries and age of pedestrians, as a percentage of pedestrians in each age range in HES

4.2.3 *Injury vs. vehicle type*

Knee and lower leg injuries were the most common injury regions for all vehicle types apart from 2/3 wheel motor vehicles for which wrist and hand injuries were slightly more frequent. Pedestrians hit by heavy transport vehicles received the highest proportion of injuries to multiple body regions, the abdominal region and the shoulder. Pedestrians struck by pedal cycles received the highest proportion of ankle and foot, and thorax injuries.

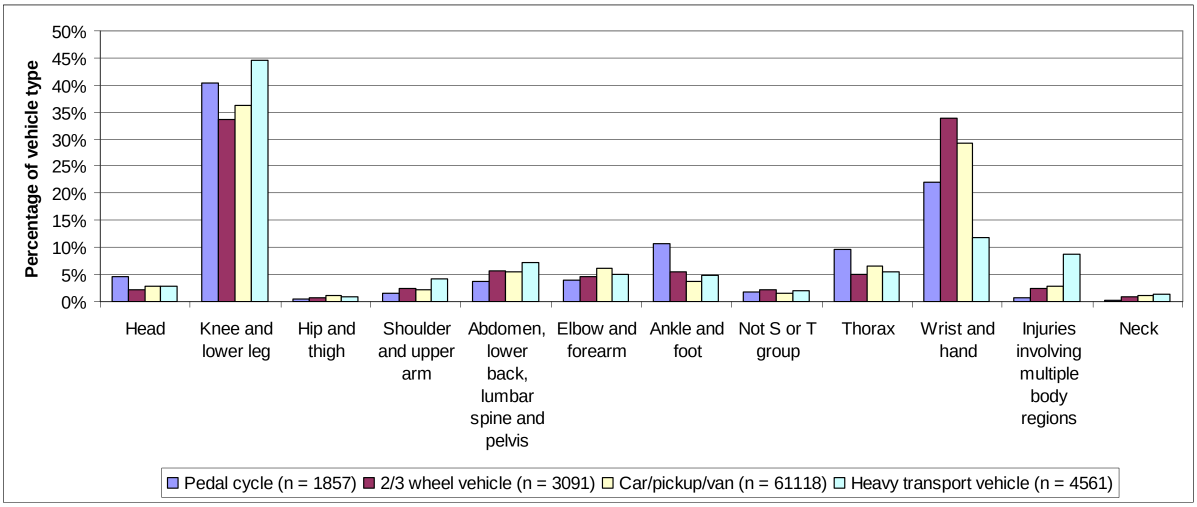

Figure 4-3. Pedestrian injuries caused by different vehicle types in HES

4.2.4 *Injuries related to time in hospital*

The duration of stay for pedestrians peaked with 25% of pedestrians staying for 1 day. The next two most frequent lengths of stay were 0 or 2 days both accounting for 11% of pedestrians.

Figure 4-4 shows a box plot which shows how primary injuries related to the length of time spent in hospital. The figure gives the 10 injuries with the highest mean duration of stay, received by at least 100 pedestrians. The injuries are coded using the 4 character ICD code, the descriptions of which are given in Table 4-2. The central horizontal line within the bars gives the median duration of stay, and the bars themselves give the upper and lower quartiles. The lines extending from the bars contain approximately 99% of the pedestrians. The circles and stars outside these lines are outliers. The body regions themselves are sorted by the mean duration of stay, descending from the left. The longest mean duration of stay was 68 days for those pedestrians with fractured cervical vertebra. This large mean duration was due to two pedestrians who received this injury and were in hospital for 1,082 and 2,878 days.

The majority of other injuries which led to long durations of stay were fractures of the legs. There was a large spread in the duration of stay of the pedestrians suffering these injuries.

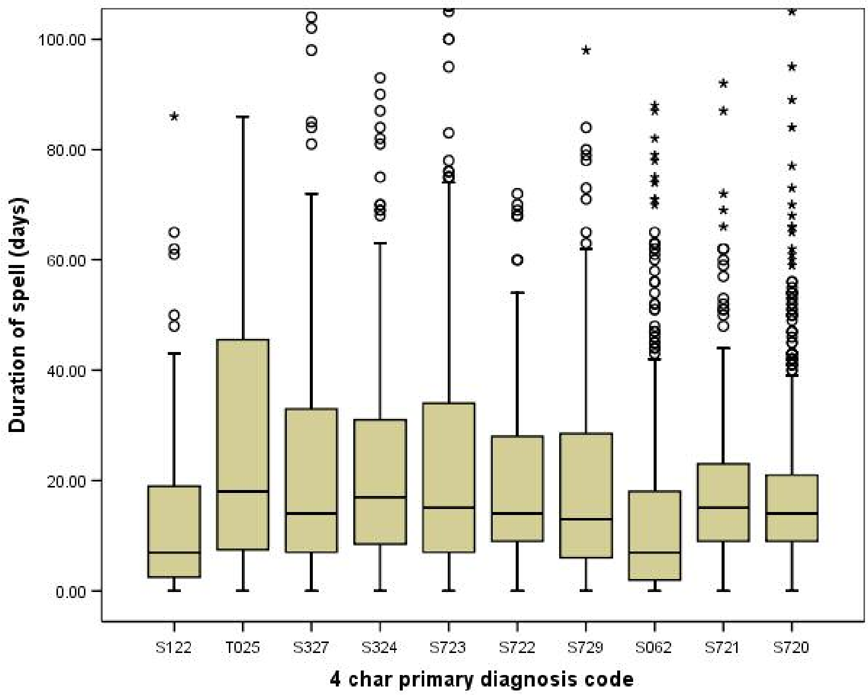

Figure 4-4. Mean duration of stay for injuries suffered by at least 100 pedestrians in HES

Table 4-2. Descriptions of 4 character primary diagnosis codes in HES

4char code	Injury description	No. of pedestrians	Mean stay duration (days)
S122	Fracture of other specified cervical vertebra	101	67.5
T025	Fractures involving multiple regions of both lower limbs	100	33.9
S327	Multiple fractures of lumbar spine and pelvis	150	25.3
S324	Fracture of acetabulum	308	24.4
S723	Fracture of shaft of femur	1335	21.9
S722	Subtrochanteric fracture of femur	154	20.9
S729	Fracture of femur, part unspecified	336	20.3
S062	Diffuse brain injury	924	19.1
S721	Pertrochanteric fracture of femur	379	18.6
S720	Fracture of neck of femur	1067	17.5

4.2.5 Operations and procedures

Table 4-3 shows the most frequent operations and procedures recorded for the pedestrians in the HES dataset. These are the primary operations for each pedestrian. The majority of pedestrians had no primary operation or procedure recorded. Of those that did, the most frequent operations were those involving bone fractures.

Table 4-3. Most frequent primary operations and procedures for pedestrians in HES

3-char code	Operation / procedure description	No. of pedestrians	% of pedestrians
-	No operation	38133	52.3%
W19	Primary open reduction of fracture of bone and intramedullary fixation	4728	6.5%
W26	Other closed reduction of fracture of bone	4371	6.0%
W20	Primary open reduction of fracture of bone and extramedullary fixation	4224	5.8%
W24	Closed reduction of fracture of bone and internal fixation	2991	4.1%
X48	Immobilisation using plaster cast	1925	2.6%
S41	Suture of skin of head or neck	1560	2.1%
S57	Exploration of other skin of other site	1560	2.1%
S42	Suture of skin of other site	655	0.9%
W25	Closed reduction of fracture of bone and external fixation	631	0.9%
U05	Diagnostic imaging of central nervous system	570	0.8%

4.3 Changes over time

Table 4-4 and Table 4-5 compare the most frequent 4-character ICD codes recorded in 1998/1999 and 2006/2007. In the tables for 2006/2007, the percentage point difference to the 1998/1999 figures for that injury are presented in brackets.

Although the most frequent injuries have changed very little, the changes in the proportion of pedestrians receiving these injuries paint an interesting picture. With the exception of “fractures of other part of lower leg”, all of the injuries which have increased in proportion are relatively minor, while those that have decreased are fractures and other serious injuries. This suggests that pedestrians received less severe injuries in 2006/2007 compared to those in 1998/1999.

Table 4-4. Most frequent 4-character ICD codes, 1998/1999 in HES

Injury description	No. of pedestrians	% of pedestrians
Unspecified injury of head	1228	13.8
Fracture of shaft of tibia	941	10.6
Fracture of lower end of tibia	416	4.7
Fracture of upper end of tibia	377	4.2
Fracture of upper end of humerus	226	2.5
Fracture of shaft of femur	224	2.5
Fractures of other parts of lower leg	222	2.5
Open wound of other parts of head	204	2.3
Fracture of pubis	187	2.1
Intracranial injury, unspecified	180	2.0

Table 4-5. Most frequent 4-character ICD codes, 2006/2007 in HES

Injury description	No. of pedestrians	% of pedestrians
Unspecified injury of head	791	10.0 (-3.8)
Fracture of shaft of tibia	646	8.2 (-2.4)
Fracture of lower end of tibia	367	4.7 (-)
Fracture of upper end of tibia	298	3.8 (-0.4)
Open wound of other parts of head	233	3.0 (+0.7)
Fractures of other parts of lower leg	215	2.7 (+0.2)
Superficial injury of other parts of head	198	2.5 (+0.9)
Open wound of scalp	179	2.3 (+0.8)
Fracture of upper end of humerus	157	2.0 (-0.5)
Superficial injury of head, part unspecified	150	1.9 (+0.9)
Fracture of pubis	140	1.8 (-0.3)
Fracture of shaft of femur	100	1.3 (-1.2)
Intracranial injury, unspecified	40	0.5 (-1.5)

Figure 4-5 shows the change over time of two of the injuries which were shown to have reduced in frequency between 1998/1999 and 2006/2007: these were the fractures to the shaft of the tibia and femur. This graph shows that there has been a steady decline in these injuries over the last few years.

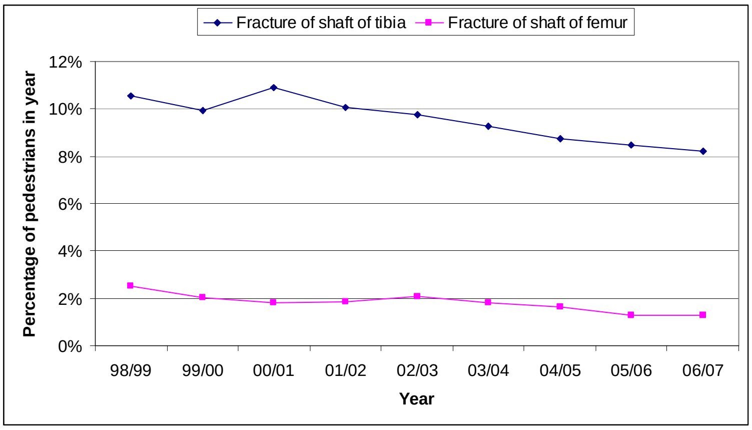

Figure 4-5. Change in time of pedestrians receiving tibia and femur fractures in HES

Figure 4-6 compares the distribution of the primary injuries for the pedestrians admitted in 1998/1999 and 2006/2007. There was a small decrease in the proportion of head & neck and lower limb injuries, and a slight increase in the proportion of upper limb and thorax injuries.

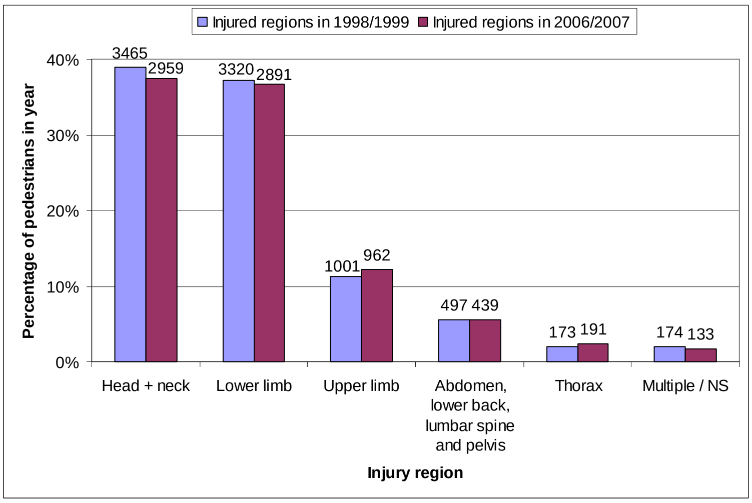

Figure 4-6. Distribution of injury regions in 1998/1999 and 2006/2007 in HES

4.4 Important casualty groups

This section gives an example of the sort of in-depth analysis which can be performed on individual casualty groups. These two groups were identified as being interesting in the matrix of pedestrian casualty demographics (Table 4-1).

4.4.1 10-15 year old pedestrians

4.4.1.1 HES

Both the STATS19 and the HES datasets showed that large numbers of 10-15 year olds are involved in accidents as pedestrians. Using the HES data, this section outlines the injuries of 14,044 pedestrians in this age range.

Figure 4-7 shows the distribution of age and gender of pedestrian casualties aged 10-15 years. There were 8,603 male and 5,421 female pedestrian casualties in the sub-sample.

The distribution of male casualties peaked at 11 years, while the female peaked at 12 years.

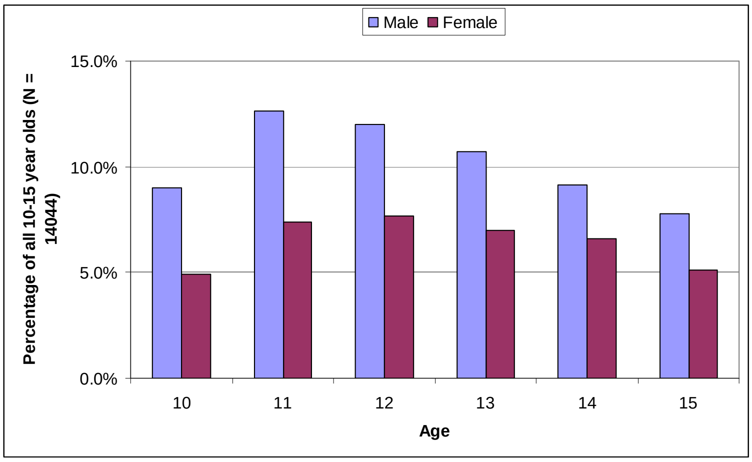

Figure 4-7. Age and gender of pedestrians aged 10-15 years in HES

Table 4-6 lists the most frequent injuries of 10-15 year old pedestrians. These were calculated using the 4-character ICD code of the primary diagnosis of each pedestrian. With the exception of the humerus fracture, the list is completely made up from head and leg injuries.

Table 4-6. Most frequent injuries of 10-15 year olds in HES

Injury	Frequency	Percentage of 10-15 year olds
Unspecified injury of head	2364	16.8
Fracture of shaft of tibia	1208	8.6
Fracture of lower end of tibia	1195	8.5
Fracture of upper end of humerus	551	3.9
Superficial injury of other parts of head	365	2.6
Fractures of other parts of lower leg	359	2.6
Open wound of other parts of head	313	2.2
Fracture of medial malleolus	293	2.1
Fracture of shaft of femur	291	2.1
Fracture of base of skull	261	1.9

4.4.1.2 STATS19

STATS19 can be used to describe the characteristics of the accidents involving 10-15 year old pedestrians, which lead to the injuries described in HES.

The location of the accidents involving 10-15 year olds was similar to that for all pedestrians: the majority of the accidents occurred on A roads, followed by unclassified roads. Figure 4-8 shows the time of day that these accidents occurred. There were two significant peaks with respect to time of day, which coincide with travelling to and from school. Accidents between 08:00 and 08:59 accounted for nearly 10% whereas those between 15:00 and 15:59 accounted for 18% of accidents.

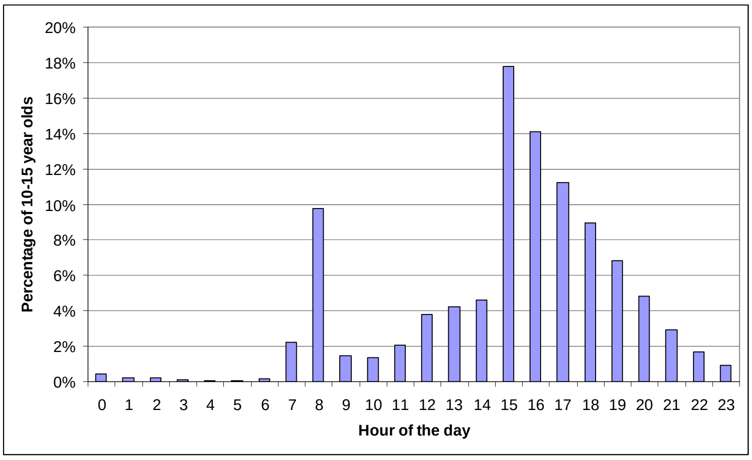

Figure 4-8. Time of day for 10-15 year old pedestrian accidents in HES

When the day of the week of the accidents was compared, there was a decrease in the proportion of accidents which happened at weekends compared to pedestrians of all ages. This is more evidence of the importance of travelling to and from school when considering child pedestrian casualties.

When the pedestrian's movement was analysed, a greater proportion of 10-15 year olds were recorded as "Crossing from driver's nearside, masked by parked or stationary vehicle" (15 %) than for all pedestrians (10 %).

There are few major differences in the contributory factors associated with 10-15 year old pedestrian casualties compared to all casualties. The biggest change was in the reduction of the proportion of factors recorded as "Impaired by alcohol" for 10-15 year olds. Apart from this, the "Failed to look properly" and "Careless, reckless or in a hurry" factors were recorded more frequently for 10-15 year olds than for all pedestrians.

STATS19 records which pedestrian casualties were school pupils on a journey to or from school. Of the 10-15 year old pedestrians, 31% were school pupils on their way to or from school. This emphasises the importance of the journey to or from school when considering ways to reduce child pedestrian casualties.

4.4.2 70+ year old female pedestrians

4.4.2.1 HES

When the distribution of the age of pedestrian casualties relative to gender in STATS19 and HES was examined, it was observed that the number of female casualties increased after the age of 70. This section explores the injuries of these 6,081 pedestrians as recorded in HES.

Figure 4-9 shows the age distribution of the female pedestrians aged 70 and older which were recorded in the HES dataset. The number of casualties increased until the age of 76, then began to fall around the age of 83.

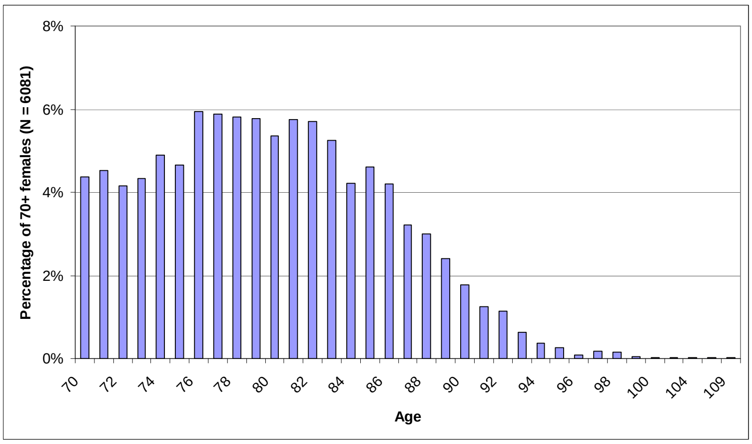

Figure 4-9. Age of female pedestrian casualties in HES, aged 70 and older

The ten most frequent injuries recorded for elderly female pedestrian casualties are shown in Table 4-7. Eight of these ten most frequent injuries were lower limb injuries, mostly fractures of the upper femur, pelvis (hip) and tibia.

Table 4-7. Most frequent injuries of 70+ year old female pedestrians in HES

Injury	Frequency	Percentage of 70+ year old females
Fracture of upper end of tibia	564	9.3
Fracture of neck of femur	455	7.5
Unspecified injury of head	385	6.3
Fracture of shaft of tibia	288	4.7
Fracture of pubis	286	4.7
Fracture of lower end of radius	194	3.2
Fracture of upper end of humerus	192	3.2
fractures of other parts of lower leg	182	3.0
Pertrochanteric fracture of femur	139	2.3
Traumatic Subdural Haemorrhage	128	2.1

4.4.2.2 STATS19

The types of accidents which caused the injuries suffered by the female pedestrians aged 70 and over were analysed using STATS19. The road class of the accidents had a similar distribution to that observed for pedestrians of all ages, with the majority of the accidents occurring on A roads, followed by unclassified roads.

Figure 4-10 shows the time that the accidents occurred which involved female pedestrians aged 70 or over. This peaked from 10:00-11:59, which was different to the distribution for all pedestrians. This is likely to be because these pedestrians were likely to be retired and not walking to and from work, so the rush-hour peaks seen for all pedestrians were not present for this subgroup.

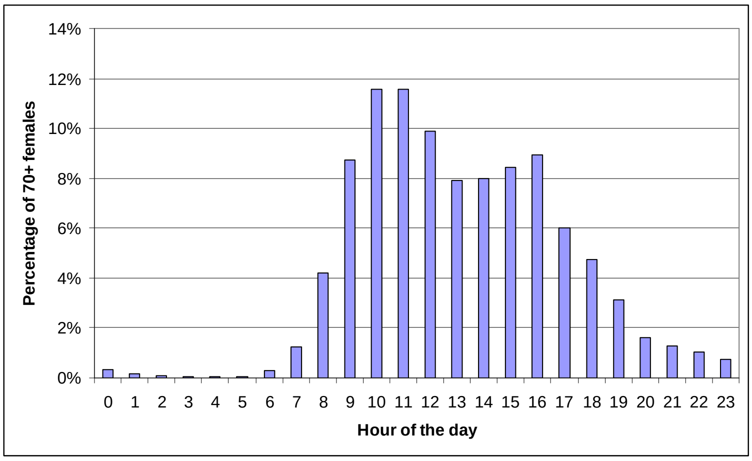

Figure 4-10. Time of day of 70+ female KSI pedestrian accidents in STATS19

The day of the week on which the accidents occurred was also analysed. The most interesting difference to the distribution for all pedestrians was the small proportion of accidents which occurred on a Sunday: 6 % of 70+ female pedestrian accidents, compared with over 10 % of all pedestrian accidents. The reasons for this difference are not known.

When the pedestrians' movement was analysed, a greater proportion of women aged over 70 years were recorded as "Crossing from driver's offside" (30 %) than for all pedestrians (23 %).

The contributory factors "Failed to look properly" and "Failed to judge vehicle's path or speed" were both recorded more frequently for 70+ female pedestrians than for all pedestrian casualties. "Careless, reckless or in a hurry" was recorded less frequently.

5 Discussion

5.1 Using a large medical database for accident research

This project has explored the methodology of using a large medical database for accident research, and given examples of how it can be used, along with the Police accident database STATS19, to investigate a group of road traffic casualties (in this case pedestrians).

The South East Public Health Observatory (SEPHO) provided TRL with an anonymised version of the Hospital Episode Statistics database for admissions in England from April 1998 to March 2007, which included details of pedestrian casualties. This was stored on the TRL secure network, and the analysis was performed by the authors with SEPHO available to answer any questions and aid with interpreting some of the data. A subset of the Police STATS19 database was also analysed, which contained details of accidents in England from the same time period, with at least one killed or seriously injured pedestrian.

The major difference between the HES database and other accident research databases which include details of injuries (such as the Co-operative Crash Injury Study) is the method used to record the injuries. The HES database uses ICD to code injuries, while other studies often used the Abbreviated Injury Scale (AIS). There are various methods available to convert ICD codes into AIS codes, which have their own advantages and disadvantages, but this project focused on what could be done with the ICD codes as they are recorded in the HES database.

The major advantage of the HES data is the combination of the huge amount of data it contains (the details of every patient who is admitted to hospital in England), and the level of detail in the coding of injuries and operations. This enables it to be used to look at changes in the frequency of individual injuries or operations over time. Unfortunately, from an accident research point of view, the HES data alone cannot be used to relate the injuries to the circumstances of the accident.

5.2 Key findings

In order to explore how HES could be used for accident research, the characteristics of the pedestrian casualties contained in the dataset were investigated. Following on from the identification of interesting groups of casualties using STATS19 and HES (presented in the matrix of pedestrian demographics), the rest of the analysis focused on things which were only possible using the HES database, such as the injuries of the pedestrians and their duration of stay in hospital. This helps to show how the HES database can be used, and what its limitations are. The key findings from this analysis are discussed in the following sections.

5.2.1 Matrix of pedestrian demographics

Correlating different pedestrian characteristics in the HES and STATS19 pedestrian datasets led to the creation of a two-dimensional matrix (Table 4-1). This gives details of any interesting relationships between the two variables investigated. Some of those relationships which included use of the HES data are discussed below.

Creating a matrix, either 2-dimensional as this one or multi-dimensional, can help to identify interesting groups of casualties. Amongst others, this matrix identified children aged 10-15 years and female pedestrians aged 70 years or older as interesting groups, which were then explored in more detail.

5.2.1.1 Deprivation

The HES data shows that the rate of pedestrian injuries is much higher in areas which are more deprived with over 20% of pedestrian casualties in the most deprived 10% of SOAs. This is more pronounced for casualties under the age of 9, where 30% were in the most deprived 10% of areas. These seem like very high proportions; however it is very difficult to separate the effect of deprivation from the location of the accident. Most accidents will occur near the home of the pedestrian, so pedestrians living in the most deprived areas are likely to have accidents in inner cities, where a lot of deprived areas are, and where there is more exposure of pedestrians to traffic.

5.2.1.2 Admission date

The most startling change over the 9 years of HES data is the way that the distribution of the age of the pedestrian casualties has altered. Specifically, the proportion of pedestrians aged 5-9 years has dropped from 15% to 10%, and the proportion of pedestrians aged 40-49 years has increased from 6% to 9%. It is not clear why these changes have occurred. The proportion of 5-9 year olds is different in STATS19 and HES, which could mean that the reduction in 5-9 year olds is caused by something which affects the HES data only. However, a drop in the proportion of 5-9 year old pedestrian casualties is also seen in STATS19. This is evidence that it is a real effect, possibly due to a reduction in the exposure of children to traffic, for example if fewer children walk to school.

Other changes over the 9 years are a reduction in the proportion of casualties who are in impacts with cars, and a reduction in the proportion of casualties who live in the most deprived 10% of areas.

5.2.2 Injury discussion

The most valuable part of the HES dataset is the injuries it records for pedestrians, data which is not available in any other database on such a large scale.

5.2.2.1 ICD analysis

Using the 4 character ICD codes, the most frequent injury is “unspecified injury of head”; however, we do not know the severity of this injury due to the ICD coding system not including a measure of injury severity. This also means that the severity of different injuries can not be compared.

The next four most frequent injuries are all fractures of the lower leg, and the majority of the top ten most frequent injuries are head and leg injuries. This agrees with previous studies on smaller samples of pedestrians (Cuerden *et al* 2007, Ashton and MacKay 1979).

5.2.2.2 Injury vs. age and gender

The relationship between the age of the pedestrian and the proportion of injuries in the four most frequently injured regions was investigated. This showed that the proportion of head injuries decreases with age, and the proportion of hip and thigh injuries increases with age. The rate of hip and thigh injuries increases most above the age of 60, which would coincide with the decreasing bone density and strength of older people, especially women. This would help to explain why the number of female pedestrian casualties increases above the age of 70.

The rate of head injuries was greatest for young children, which is likely to be because they will receive a more direct contact to the head from the front of the vehicle, because of their height. From the age of 16 and older, the rate of head injury remains relatively constant. It might be expected that elderly pedestrians would also see an increase in head injuries because they are generally less tolerant to injury. However, it may be that these pedestrians are being seriously injured at lower impact speeds (receiving leg fractures), which may balance their reduced tolerance to head injuries. Some knowledge of impact speed would be required to determine whether this was true.

5.2.2.3 Injury vs. vehicle type

Knee and lower leg injuries were the most common injury regions for all vehicle types apart from 2/3 wheel motor vehicles for which wrist and hand injuries were slightly more frequent. This is possibly due to 2/3 wheel motor vehicles being more likely just to clip a pedestrian on the arm as they pass them whereas for other vehicle types it is likely the pedestrians were crossing in front of the vehicles and were struck by the bonnet onto their lower legs. Pedestrians hit by heavy transport vehicles received the highest rate of injuries to multiple body regions, the abdominal region and the shoulder. This is to be expected, as pedestrian impacts with heavy goods vehicles are more likely to have severe consequences, for example the possibility of being run over by the wheels of a heavy goods vehicle is greater than for a car.

5.2.2.4 Duration of stay

The duration of the stay in hospital is one way in which the HES data can be used to estimate the cost to the hospital of different injuries. Overall the duration of stay in hospital of the pedestrian casualties in HES is a very skewed distribution, with a large number of pedestrians staying for only one day, and a very small number of pedestrians remaining in hospital for very long periods of time. This distribution is a similar shape when individual injuries are investigated. For this reason, the relationship between individual injuries and the duration of stay in hospital has been investigated using the mean stay in hospital, and box-plots showing the distribution of the length of stay for different injuries.

Looking at individual injuries shows that a fracture of the cervical spine leads to the longest mean duration in hospital, but this is mainly due to one pedestrian who received this injury and remained in hospital for almost eight years. Apart from this spinal injury, multiple fractures of the lower legs led to the longest mean stay in hospital. The remainder of the ten longest mean durations of stay in hospital are mostly made up of fractures to the femur.

There are a number of limitations to using this method of determining which injuries are most costly to the hospitals. The first is that it only takes into account the primary injury, and not any other injuries sustained. Secondly, it does not take into account other costs, such as operations and procedures in the hospital, post-hospital care, and the effect on quality of life. These are things which could be investigated, but would require additional information to that in the HES data.

5.2.2.5 Operations

The HES dataset also contains details of the operations and procedures undertaken on every patient. These are recorded using the Tabular List of the Classification of Surgical Operations and Procedures, version OPCS4. The majority of the pedestrians in the HES dataset did not have any operations or procedures recorded. Of those that did, the most frequent operations were reductions of fractured bones.

One possible use of the information on operations and procedures in the HES data for accident research would be to calculate the overall cost of the treatment, which would be possible if the cost of each individual operation and procedure could be obtained.

5.2.3 Changes over time

The HES data provided to TRL covers a period of nine years, and the large number of pedestrians recorded in each year has enabled some of the changes over time to be investigated. Some injuries, such as fractures to the shaft of the tibia and fibula, show a small but steady decline over the ten year period. This could be evidence that improved car design in recent years has reduced the rate of these injuries.

5.2.4 Pedestrians aged 10-15 years

The HES and STATS19 datasets contain large numbers of 10-15 year old pedestrian casualties. Analysis shows that although they are involved in more accidents, they generally receive fewer injuries and do not remain in hospital as long when compared to all the pedestrian casualties. They seem to receive more head injuries, especially the "unspecified injury of head" codes in the ICD system which are mostly minor head injuries. These differences are likely to be because children have a higher biomechanical tolerance than older adults.

A large proportion of these pedestrian casualties (31%) were injured on their way to and from school. This leads to a peak in the number of 10-15 year old pedestrian casualties at 8-9 am and 3-4 pm, the times at which children will be travelling to and from school. This highlights the importance of considering the school journey when looking at pedestrian safety.

5.2.5 Female pedestrians aged 70+ years

The number of female pedestrian casualties aged 70+ is large in both the HES and STATS19 datasets. It was expected that this would be because of the lower biomechanical tolerance in older women, related to lower bone densities. They were found to receive more injuries and stay in hospital for a longer duration than other pedestrians. They also received a larger proportion of fractures to the upper end of the tibia and the neck of the femur compared to other pedestrians, making these the two most frequent injuries in elderly women. This is evidence that the lower bone density of these women makes them more vulnerable to certain types of leg fracture, such as those to the femur and pelvis.

Also of interest is the age distribution of these pedestrians. The number of pedestrian casualties rises to a peak at the age of 76, remains relatively constant until the age of 84, then begins a steady decline. With an aging population, the size of these groups may increase and potentially the age range could also increase.

5.3 The future of HES and accident research

This project has shown that with no alterations, a large medical database such as HES can be a useful addition to the accident research 'toolkit'. It provides in-depth information on the injuries of casualties on a macro scale, including details on every patient admitted to hospital in England. Road traffic casualties can be easily categorised, and the road user type and the vehicles involved in the accident can also be determined. The large scale of the database also makes it possible to investigate changes in the injuries received in road traffic accidents over time.

As it stands currently, the injury coding system used in HES is incompatible with that used in other accident studies. It is possible to convert the ICD codes used in HES to AIS codes which are used in other accident research studies, but a certain amount of information would be lost. The merits of performing this transformation in the future would have to be carefully considered.

Unfortunately, the HES database on its own is missing information about the accident itself which would make it invaluable to accident researchers in England. For example, for pedestrian accidents, if it contained information on the vehicle involved (such as make, model and year of registration), new pedestrian safety features on cars could be correlated with any change in the injuries received by pedestrians.

However, the Office of National Statistics has recently linked the accidents recorded in the Police STATS19 database with the patients admitted to hospital and recorded in HES. This would enable the injuries of the casualties and the circumstances of the accident to be investigated on a macro scale for the very first time in this country. This combined dataset has the potential to become one of the most useful sources of accident data in England at the moment as long as the link is updated frequently, preferably on a yearly or more frequent basis.

Further work in analysing either this combined dataset or the data separately, could involve looking at the costs of injuries received in pedestrian accidents in order to prioritise them for introduction of countermeasures to reduce the frequency of these injuries. Also this work could involve further analysis of the mechanisms of injury to pedestrians struck by differing shapes of vehicles which could then be used to influence the design of new vehicles in order to prevent/reduce the severity of these injuries in the future.

6 Conclusions

This project explored the possibilities of using a large medical database for accident research purposes. TRL successfully collaborated with the South East Public Health Observatory (SEPHO) who provided Hospital Episode Statistics (HES) data for pedestrian casualties admitted to hospitals in England from April 1998 to March 2007.

The HES database, along with the Police STATS19 database, was used to investigate the demographics of pedestrian accidents. This showed how the two databases can be used in their current form to determine casualty groups which may be of interest.

Because it is a medical database, the strength of the HES database is the amount of detail recorded on the injuries received, and the operations undergone by each casualty. These were investigated for the pedestrians in the HES dataset. This shows that the most frequently recorded injuries for pedestrians are head and leg injuries, which agrees with previous studies on smaller samples of pedestrians.

The large number of casualties recorded in the HES database (every patient admitted to hospital in England) means that the injuries due to road traffic accidents can be investigated on a scale not previously available in this country. This enables changes over time to be investigated; for example, fractures of the tibia and fibula are seen to have declined over time. This points towards an improvement in car design having a positive impact on the injuries received by pedestrians in impacts.

In an unaltered form, the HES database is limited to reporting the frequency of injuries and cannot be used to determine the causes of the injuries, beyond the types of vehicle which were involved. There are two developments which could increase the usefulness of the HES data. The first is transforming the ICD codes into AIS codes, making it compatible with other accident studies, although the effect that the transformation would have on the data would need to be understood. The second development is the linking of the HES data to STATS19, which has been performed by the Office of National Statistics. This would enable injuries to be related to the circumstances of the accident, meaning that the HES data could be used to investigate injury causation. Further investigation of this data could indicated the most costly injuries and ideas for the future design of vehicles.

Acknowledgements

The work described in this report was carried out in the Accident Research Group of the Transport Research Laboratory. The authors are grateful to Richard Cuerden who carried out the technical review and auditing of this report.

TRL would also like to acknowledge the contributions of Harry Rutter (Deputy Director of SEPHO at the start of the project), Steve Morgan, Jo Watson and Bryony Tatem of the South East Public Health Observatory who supplied the data and information required for the analysis of the HES data, and provided help and advice throughout.

References

Ashton S J, MacKay G M (1979), *Some characteristics of the population who suffer trauma as pedestrians when hit by cars and some resulting implications*, Proceedings of the 4th IRCOBI conference 1979.

Broughton, J., Keigan, M., & James, F. J. (2001). *Linkage of hospital trauma data and road accident data.* TRL Report TRL518. Crowthorne: TRL Limited.

Cuerden R W, Richards D C, Hill J (2007), *Pedestrians and their survivability at different impact speeds*, Proceedings 20th International Technical Conference on the Enhanced Safety of Vehicles 2007.

De Mol, J., & Boets, S. (2003). *Optimisation of traffic accident statistics, Interim Report*. Brussels: Belgian Public Planning Service Science Policy.

Department for Transport. (2004). *STATS20: Instructions for the completion of road accident report form STATS19*. Retrieved 12 September 2008 from http://www.dft.gov.uk/pgr/statistics/datatablespublications/accidents/casualtiesgbar/

Department for Transport. (2006). *Road accident casualties: a comparison of STATS19 data with Hospital Episode Statistics.*

Department for Transport, Scottish Executive, & National Assembly for Wales. (2007). *Road Casualties Great Britain 2006.*

Department for Transport, Scottish Executive, & National Assembly for Wales. (2008). *Road Casualties Great Britain 2007.*

Goldacre, M. J., Gill, M., & Yeates, D. G. (2006). *Changes in safety on England's roads: analysis of hospital statistics.* BMJ, doi:10.1136/bmj.38883.593831.4F (published 23 June 2006)

HESonline. (2008). http://www.hesonline.nhs.uk

Keigan, M., Broughton, J., & Tunbridge, R. J. (1999). *Linkage of STATS19 and Scottish hospital in-patient data — analyses for 1980–1995.* TRL Report TRL420. Crowthorne: TRL Limited

Lujic, S., Finch, C., Boufous, S., Hayen, A., & Dunsmuir, W. (2008). *How comparable are road traffic crash cases in hospital admissions data and Police records? An examination of data linkage rates.* Australian and New Zealand Journal of Public Health.

NHTSA. (1996). *The Crash Outcome Evaluation System (CODES).* NHTSA technical report for the U.S. Department of Transportation.

Office of National Statistics. (2008). Super Output Areas, http://www.statistics.gov.uk/geography/soa.asp

Oxford University Press. (1998). *Concise Medical Dictionary, 5th Edition*. Oxford University Press.

Rosman, D. L. (1999). *The Western Road Injury Database (1987-1996): ten years of linked Police, hospital and death records of road crashes and injuries*. Department of Public Health, Road Accident Prevention Research Unit, The University of Western Australia.

Rosman, D. L., & Knuiman, M. W. (1993). *A Comparison of Hospital and Police Road Injury Data.* Department of Public Health, University of Western Australia.

Scottish Government. (2006). *Comparison of Police "Stats19" Road Casualty Statistics with some other figures for Scotland.* Retrieved 5 December 2008 from Scottish Government website, www.scotland.gov.uk/Publications/2006/11/22093058/14

SEPHO. (2008). *South East Public Health Observatory website*. Retrieved 24 September 2008 from www.sepho.org.uk

Simpson, H. F. (1996). *Comparison of hospital and Police casualty data: a national study.* TRL Report TRL173. Crowthorne: TRL Limited.

Simpson, H. F. (1997). *National hospital study of road accident Casualties.* TRL Report TRL272. Crowthorne: TRL Limited.

Stone, R. D. (1984). *Computer Linkage of Transport and Health Data.* TRRL Laboratory Report LR1130. Crowthorne: TRL Limited.

Surginet (2008). *OPCS4 code search engine.* Retrieved 8 December 2008 from http://www.surginet.org.uk/coding/opcs4.php

TARN (contributed to by Nee, P., Woodford, M., & Zoltie, N.). (2008). *Anatomy*.

Ward, H., Lyons, R., & Thoreau, R. (2006). *Under-reporting of Road Casualties – Phase 1.* A report for the Department for Transport, with the Centre for Transport Studies - University College London and the Centre for Health information, Research and Evaluation – Swansea University.

Ward, H., Robertson, S., Townley, K., & Pedler, A. (2005). *Reporting of road traffic accidents in London: Matching Police STATS19 with hospital accident and emergency department data.* TRL Report TRL242. Crowthorne: TRL Limited.

World Health Organisation. (1992). *International Statistical Classification of Diseases and Related Health Problems – Tenth Revision: Volume 1.* World Health Organisation.

Glossary of terms and abbreviations

Intracranial injury is any injury which occurs within the skull (Oxford University Press, 1998).

Diffuse brain injury – an injury to the brain which does not occur in one specific spot in the brain

Subdural haemorrhage – is a form of traumatic brain injury in which blood gathers between the dura (the outer protective covering of the brain) and the arachnoid (the middle layer or the meninges). Subdural bleeding usually results from tears in the veins that cross the subdural space. (Oxford University Press, 1998).

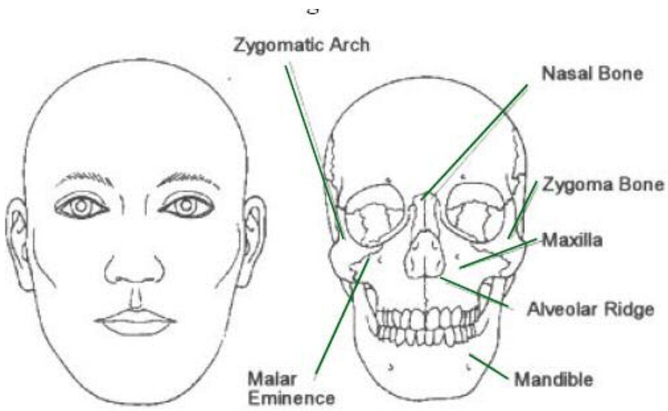

Figure G- 1. The bones of the face (TARN, 2008)

Lumbar spine – consists of the lumbar vertebrae which are the five bones of the backbone that are situated between the thoracic vertebrae and the sacrum, in the lower part of the back. (Oxford University Press, 1998). (see Figure G- 2)

Cervical spine – consists of the cervical vertebrae which are the seven bones making up the neck region of the backbone. The first vertebra supports the skull which rotates on the second vertebra, enabling the head to turn. (Oxford University Press, 1998). (see Figure G- 2)

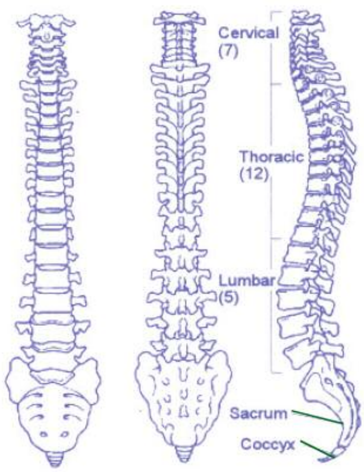

Figure G- 2. The spine (TARN, 2008)

Pelvis – the bony structure formed by the hip bones, sacrum and coccyx that protects the organs of the lower abdomen and provides attachment for the bones and muscles of the lower limbs. (Oxford University Press, 1998). (see Figure G- 3)

Pubis - a bone which forms the lower front part of each side of the hip bone (Oxford University Press, 1998). (see Figure G- 3)

Acetabulum – one of the two deep sockets either side of the hip bone where the head of the femur fits

Femur – (or thigh bone) is a long bone between the hip and the knee. The neck of the femur is the narrowed top end of the bone which carries the head of the femur. The shaft of the femur is the long central length of the bone (Oxford University Press, 1998). (see Figure G- 3)

Pertrochanteric fracture – a fracture of the great trochanter of the femur, which is at the top of the femur near the head.

Subtrochanteric fracture – a fracture of the femur between the lesser trochanter (which is at the base of the greater trochanter) and a point approximately 5 cm further down the femur.

Tibia – the shin bone: the inner larger bone of the lower leg. The shaft of the tibia is the long central section. (Oxford University Press, 1998). (see Figure G- 3)

Fibula – the long thin outer bone of the lower leg. (Oxford University Press, 1998). (see Figure G- 3)

Malleolus – either of the two protuberances on each side of the ankle: the lateral malleolus at the lower end of the fibula or the medial malleolus at the lower end of the tibia. (Oxford University Press, 1998).

Humerus - the bone of the upper arm. The head of the humerus articulates with the scapular at the shoulder joint, and the lower end articulates with the ulna and the radius of the forearm. (Oxford University Press, 1998). (see Figure G- 3)

Radius – the outer and shorter bone of the forearm. It partially revolves about the ulna, permitting pronation and supination of the hand. The head of the radius articulates with the humerus. The lower end articulates with the wrist bones and ulna. (Oxford University Press, 1998). (see Figure G- 3)

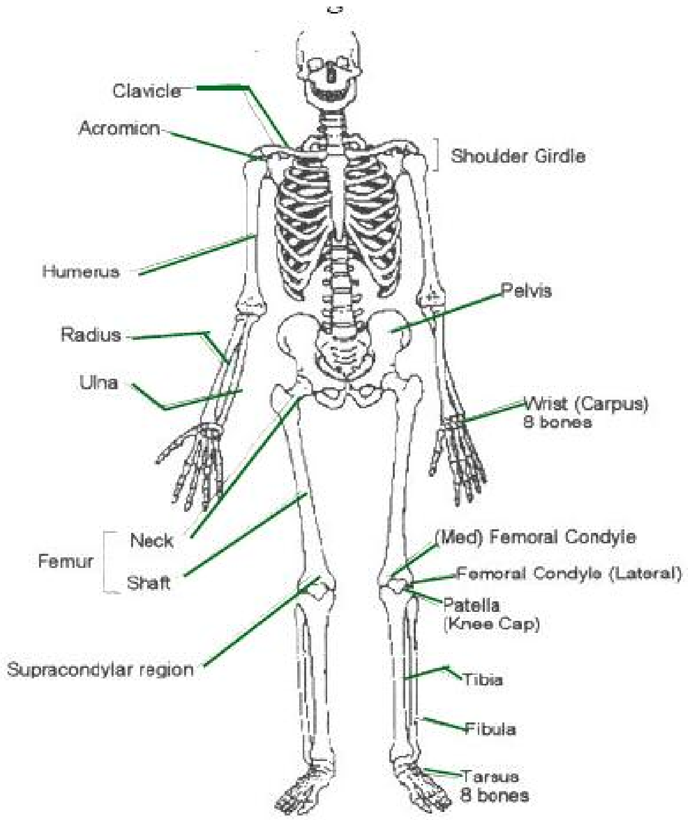

Figure G- 3. The skeleton (TARN, 2008)

Appendix A – Literature review

A.1 Linking data sources and the success rate

Stone (1984) was one of the first to identify the need to improve Police records of road traffic accidents by adding hospital treatment data, and developed a method to do this via computer. He used Scottish hospital in-patient records (SHIPS), and created an algorithm for linking individual records to Police STATS19 information. This method of linking has provided a base for much of the similar work done since. For example Simpson (1996) used information collected by survey clerks working in A&E departments and successfully matched this to appropriate STATS19 records. Broughton et al. (2001) linked information on injured road users from the Trauma Audit and Research Network (TARN) with individual STATS19 records. Also, Ward et al (2005) took road traffic casualty data from central London hospitals and created links to Police records by making assumptions on their catchment areas also. And Keigan et al (1999) used an effective method to link SHIPS records to their corresponding STATS19 entries. A smaller report by the Scottish Government (2006) has since updated this and included hospital data up to 2004 as well as tying in data from the 'General Register Office for Scotland' (GROS) and Scottish household surveys on road traffic accidents. It also acts as an excellent summary of previous work carried out (not just in Scotland) and the future research that has been planned/recommended. The methods used within the reports mentioned above illustrate the potential to match HES and STATS19 in a similar way, which has in fact not been done before.

Efforts to link information held on road traffic accident with that held be hospitals can be seen elsewhere around the world, such as the 'Crash Outcome Data Evaluation System' (CODES) in the USA that even includes insurance claim details in its database (NHTSA, 1996). Also the Australian equivalent of STATS19, the 'Traffic Accident Data System' (TADS) has successfully been matched to their 'Inpatient Statistic Collection' (ISC) in a way that HES and STATS19 are hoped to be matched (Lujic et al., 2008). The importance of connecting Police held accident details with hospital held injury details was recognised long before this report in Australia though, and a great deal of work can be found on the methods and outcomes of doing this. In particular Rosman and Knuiman (1993) emphasize the need for links to be made and analyse linking success rates for different road user types, and then later Rosman gives an excellent overview of the last ten years of Australian linking and their findings (Rosman, 1999).

The algorithm developed by Stone (1984) that allows computerised matching of Scottish hospital in-patient data with individual STATS19 records was implemented again by Simpson (1996) to compare these two data sources. The variables used to match were date of accident or hospital attendance; casualty severity; age; sex; and road user type, and a tolerance level was assigned to each variable to allow for discrepancies between datasets. Using this method, Simpson (1996) successfully linked 46 percent of hospital road casualty cases to their STATS19 records. Keigan et al. (1999) used the same algorithm, with similar tolerances, and Scottish in-patient data (SHIPS) and this time found a higher linking success rate with between 68 and 72 percent of SHIPS records able to be linked to their STATS19 counterparts for the years 1980 to 1995. A similar success rate of 74 percent was achieved when matching TARN records to STATS19 data by Broughton et al. (2001) by using the same method as was used for SHIPS data and by employing similar tolerances. The resulting report also provides details of each tolerance used and how the percentage of records successfully matched will alter when changing these tolerance levels.

Sample areas used by Ward et al. (2005) complicate the matter of matching records somewhat as the casualty data used was taken from a variety of central London hospitals and catchment areas. This meant some manual matching and manipulation of records was required alongside the automatic computer matching. The full details of this can be found in appendix 1 of the Ward et al. report. The result of this study was a generally higher rate of linking for central London hospitals than for other studies of broader regions of the UK. It ranged from around 70 percent of hospital admissions up to 87 percent in one particular hospital. Ward et al attribute these high levels to the proximity of hospitals and greater involvement of Police in accidents that occur in central London.

A.1.1 STRADA model

STRTADA stands for Swedish Traffic Accident Data Acquisition, and it is the national system used in Sweden to record traffic accidents. In this system, the Police and the hospitals produce individual reports on the accident, which are then sent to a central database. Here the reports are combined, creating one report combining all the information provided by the Police and the hospital. In this system it has been observed that only about 50% of the accidents are recorded by the Police compared to those recorded by the hospitals. (De Mol and Boets, 2003)