

**Reporting of road traffic accidents in London:
Matching police STATS19 with hospital accident and
emergency department data.**

**Supplementary Report for St. Thomas' Hospital
Central London.**

**by Heather Ward, Sandy Robertson (UCL)
Katherine Townley and Annette Pedler (TRL Limited)**

**UPR T/043/05
Clients Project Reference Number**

UNPUBLISHED PROJECT REPORT



UNPUBLISHED PROJECT REPORT UPR T/043/05

**REPORTING OF ROAD TRAFFIC ACCIDENTS IN LONDON:
MATCHING POLICE STATS19 WITH HOSPITAL ACCIDENT AND
EMERGENCY DEPARTMENT DATA.**

**SUPPLEMENTARY REPORT FOR ST. THOMAS' HOSPITAL
CENTRAL LONDON.**

Version: Phase 3 Final

by **Heather Ward, Sandy Robertson (UCL), Katherine Townley and Annette Pedler
(TRL Limited)**

Project Record:

Client:

Reporting Levels of Road Traffic Accidents

**Street Management, Transport for London,
(Nina Webster)**

Copyright TRL Limited May 2005

This report, prepared for Transport for London (Street Management) is unpublished and should not be referred to in any other document or publication without the permission of Transport for London. The views expressed are those of the author(s) and not necessarily those of the client.

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

This report has been produced by TRL Limited, under/as part of a contract placed by Transport for London. Any views expressed are not necessarily those of the client.

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

CONTENTS

Executive summary	1
1 Introduction	3
1.1 Definition of reporting rate	3
2 Definition of the St Thomas' hospital area	6
3. Method	9
4. Results	10
4.1 The overall picture	10
4.2 Detailed breakdown of results for St. Thomas' Hospital	14
4.2.1 Hospital data for inside the area only	14
4.2.2 Lower estimates of reporting rate using All hospital data for inside and cross border casualties	16
4.2.3 Best estimates of reporting rates using Hospital data for inside and estimated unmatched cross border casualties	18
5. Interpretation of results	20
6. Other issues	22
7. Acknowledgements	22
8. References	25
Appendix 1: The data-matching process	26
A1.1 Sorting hospital records according to location in or outside the cordon	26
A1.1.1 Procedure for estimating casualties on roads running through the area and out the other side?	26
A1.1.2 Hospital data whose location was unknown.	27
A1.2 The matching process	27
A1.2.1 The combined dataset	27
A1.2.2 Data coding	28
A1.3 The matching algorithm	29
A1.4 Automated matching	30
A1.5 Manual matching	30
A1.6 Selecting subsets of the data	31

Executive summary

A previous study for Transport for London (TfL) had taken three hospitals in London and matched their road traffic casualty data with the STATS19 police records for the same areas. The hospitals were St Mary's Paddington, representing Central London, King's College Hospital Camberwell, representing Inner London, and Barnet General Hospital, representing Outer London. The results indicate that the percentage of road traffic casualties known to the police in London through the STATS19 data collection process is higher than found by other studies elsewhere in the country (see for example Simpson 1996, Ward et al 1994, Robertson and Ward 2002)) with the King's area being about 71 percent, Barnet about 70 percent and St Mary's about 87 percent. This latter figure was considered to be high even for London so a second Central London hospital, St Thomas' was chosen in order to estimate the reporting rate in this area which forms the basis of this study.

Using the same methodology and time period (2001) as in the first study it was estimated that the reporting rate for St Thomas' was about 79 percent. This is considerably higher than Barnet and King's but lower than St Mary's.

The table on the following page combines results from both studies and shows estimates of the headline percentage reporting rates to the police for the four hospital areas.

The indications are that there is a Central London effect which increases the reporting rate. Two possible explanations for this are:

- If a casualty feels well enough to continue home and once there they feel the need for treatment, they will go to the local hospital near where they live and not back to the Central London Hospital near where they were injured. Therefore, for this unquantified group of people, their injury may be in the police record for the study area and in the record of some other hospital unknown to us but not the hospital record being studied. This would have the effect of increasing the proportion of casualties known to the police relative to the number of all known casualties and thus increase the estimated reporting rate.
- The level of police presence and awareness of traffic incidents in Central London means that when a road traffic accident occurs the police are on the scene very quickly, thus making it more likely that the accident will be reported and appear in the STATS19 record.

Based on the proportion of road traffic accident casualties to all A&E attendances at the four hospitals studied, it was concluded that the St Mary's Hospital record for road traffic accident casualties was about half what might be expected based on its total number of A&E attendances. This suggests that the dataset used in the original study may not have been complete, leading to a higher than average estimated reporting rate.

The total number of road traffic casualties attending the 33 London hospitals with full time A&E departments was estimated to be about 44 700 people, there were 44 500 casualties reported to the police. The total number of road traffic casualties in London in 2001 is therefore estimated to be about 68 500.

Rate=All casualties known to police/All known casualties.

Casualty class	Area in London	Lower estimate ¹	Upper estimate ²	Current best estimate ³
All casualties				
Kings	Inner	66	80	71
Barnet	Outer	61	75	70
St Mary's	Central	77	91	87
St Thomas'	Central	71	84	79
Pedestrian casualties				
Kings	Inner	62	78	67
Barnet	Outer	72	78	70
St Mary's	Central	66	84	78
St Thomas'	Central	69	82	78
Pedal cyclist casualties				
Kings	Inner	61	75	66
Barnet	Outer	60	75	69
St Mary's	Central	89	98	96
St Thomas'	Central	N/A	N/A	N/A
TWMV casualties				
Kings	Inner	74	87	78
Barnet	Outer	65	80	73
St Mary's	Central	74	92	85
St Thomas'	Central	N/A	N/A	N/A
Car occupant casualties				
Kings	Inner	67	78	72
Barnet	Outer	58	73	68
St Mary's	Central	82	92	90
St Thomas'	Central	71	84	79

¹ This estimate includes all casualties identified as being in the area only for police data and all those that are in the area and on cross border roads for the hospital

² This estimate includes all casualties identified as being in the area only for hospital and police data

³ This estimate includes all casualties identified as being in the area only for police data and all those that are in the area plus adjusted number for those on cross border roads for the hospital

1 Introduction

During 2002, Transport for London (TfL) commissioned work to estimate the level of reporting to the police of road accident casualties in London in order to:

- Provide a base-line against which the effect of policy initiatives can be properly assessed, given that the level of reporting might be influenced and change over time. Such initiatives might include publicity campaigns, road safety engineering programmes, and congestion charging.
- Improve understanding of the extent of the accident and casualty problem.
- Provide robust information on the reporting of the severity of injuries, and
- Engender better working relationships and cooperation with Health Authorities and Hospitals.

In the original project records were used from Accident and Emergency (A&E) Departments of three representative hospitals (outer, inner and central London) to supplement those of the police to allow us to estimate of the proportion of people injured who report their injury accident to the police. From these two sets of records estimates can be made of the reporting rate to the police, and at the same time improve the estimate of the level of injury in London.

1.1 Definition of reporting rate

The reporting rate used in this study can be described as **all casualties known to the police divided by all known casualties**. Electronic records were requested from three London hospitals, out of a total of 31 hospitals with full time A&E Departments. They were chosen to represent Central (St Mary's Hospital, Paddington), Inner (Kings College Hospital, Camberwell) and Outer London (Barnet General Hospital, Barnet).

The method of calculation of reporting rate may vary between different studies. Here the reporting rate to the police is calculated as follows:-

$$(p+b)/(p+h+b)$$

Where:

p= number of casualties reported to the police only;

b= number of casualties reported to both police and hospital;

h= number of casualties reported to the hospital only.

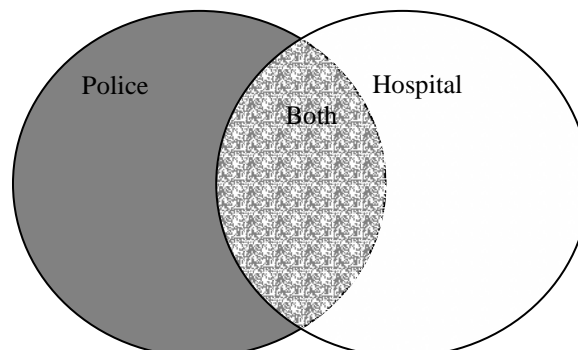


Figure 1. The overlap between police and hospital casualty records

In other words all casualties known to the police divided by all known casualties. This is shown schematically in Figure 1. The reporting rate is the shaded areas divided by the total area.

Table 1 shows estimates of the headline reporting rates to the police for each hospital area in the original study. It can be seen that the values for St. Mary's Hospital, representing Central London, are higher than for King's College Hospital, representing Inner London, and Barnet General Hospital, representing Outer London. There was some concern that there may have been a shortfall in the number of casualties reported as road traffic accidents (RTAs) at St. Mary's and a subsequent overestimation of the reporting rate.

Table 1: Estimates of the headline reporting rates to the police for the three hospital areas in the original study.

$$\text{Rate} = \text{All casualties known to police} / \text{All known casualties.}$$

Casualty class	Area in London	Lower estimate ¹	Upper estimate ²	Current best estimate ³
All casualties				
Kings	Inner	66	80	71
Barnet	Outer	61	75	70
St Mary's	Central	77	91	87
Pedestrian casualties				
Kings	Inner	62	78	67
Barnet	Outer	72	78	70
St Mary's	Central	66	84	78
Pedal cyclist casualties				
Kings	Inner	61	75	66
Barnet	Outer	60	75	69
St Mary's	Central	89	98	96
TWMV casualties				
Kings	Inner	74	87	78
Barnet	Outer	65	80	73
St Mary's	Central	74	92	85
Car occupant casualties				
Kings	Inner	67	78	72
Barnet	Outer	58	73	68
St Mary's	Central	82	92	90

To investigate this, and other possibilities, it was decided to study St. Thomas' or University College London (UCL) Hospitals as central London hospitals and calculate their reporting

¹ This estimate includes all casualties identified as being in the area only for police data and all those that are in the area and on cross border roads for the hospital

² This estimate includes all casualties identified as being in the area only for hospital and police data

³ This estimate includes all casualties identified as being in the area only for police data and all those that are in the area plus adjusted number for those on cross border roads for the hospital

rates. In the event, the UCL Hospital data was not suitable as it did not include location data so St Thomas' provided a dataset for 2001 (the same year as in the original study).

This supplementary report documents the analysis and findings using this supplementary data.

2 Definition of the St Thomas' hospital area

There are no definitive catchment areas for London hospitals. In the same way that areas for study around the other three hospitals were selected, London Ambulance Service Data was used as a guide to what was an appropriate area within which casualty records could be identified for linking and matching. The definition of this area aimed to maximise the proportion of casualties within it that were taken to St Thomas' hospital rather than any other. Figure 2 shows the London Ambulance Service data on the locations of casualties taken to St Thomas' Hospital in 2001. Figure 3 shows the boundary of the area selected for study and the location of Police Stats19 records within that area.

The boundaries agreed between UCL and TfL for St Thomas' Hospital are shown in Figure 2. The approximate area bounded by each cordon is as follows:

St. Mary's Hospital	7.9 km sq
King's College Hospital	16.2 km sq
Barnet General Hospital	44.3 km sq
St. Thomas' Hospital	10.2 km sq

By comparison with the other areas, the St Thomas' area chosen for study is compact compared with King's and Barnet but is slightly larger than St. Mary's.

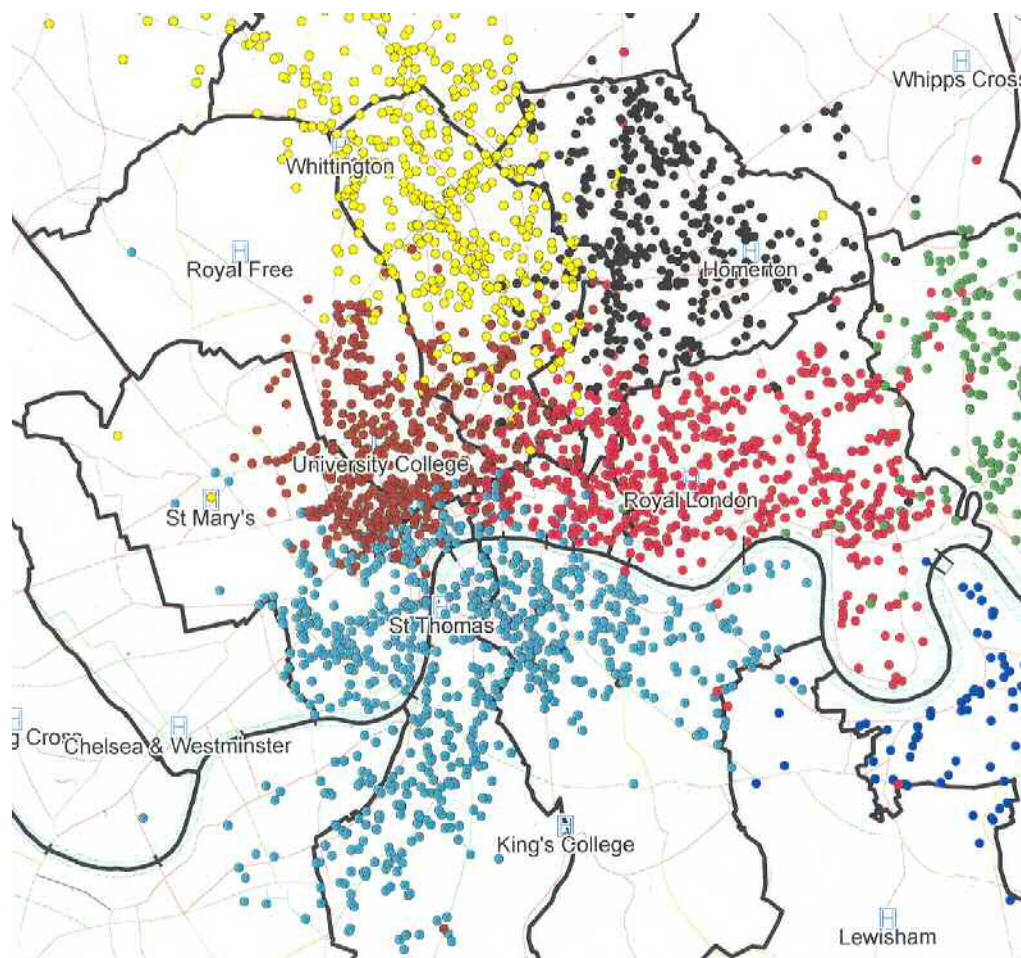
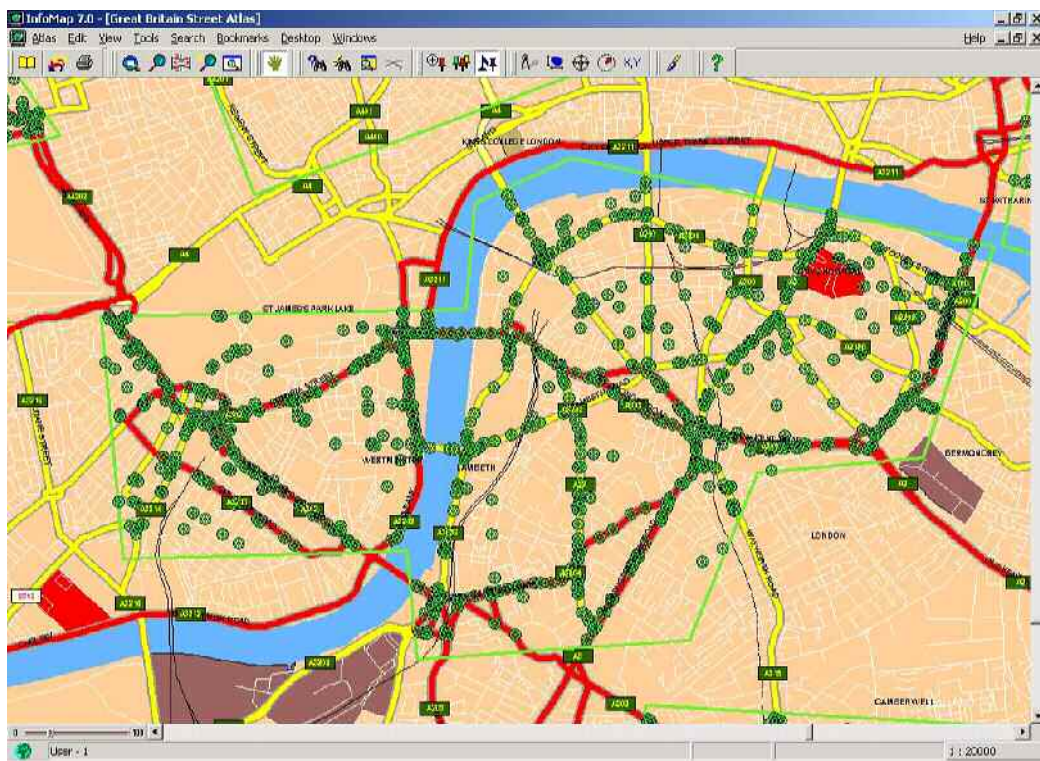


Figure 2: London Ambulance Data indicating the Location of Casualties taken to St Thomas' and four other hospitals in London



Source InfoMap

Figure 3: Location of Police Stats19 Records within the boundary (shown by the green line) of the area selected for study.

Table 2 shows the total number of casualties in the A&E datasets for St. Thomas’ compared with the other three hospitals and Table 3 shows the number of casualties in the STATS19 record for each area defined by the cordon. Whilst the area within the cordon for St Thomas’ is relatively small, the number of STATS19 casualties is some 45-50 percent higher than for St Mary’s or King’s. The number of casualties in the hospital record is some three times higher than in the total St Mary’s record and 1.5 times higher than in the King’s record. It should be noted that these totals in Table 2 are for the whole of the A&E set and not only for those inside the cordon.

Table 2: Number of casualties in hospital datasets for 2001

	Barnet	St Mary's	King's	St.Thomas'
Hospital	1739	929	1832	2865

Table 3: Number of Police STATS19 casualty records for each hospital area.

	Barnet	St. Mary's	King's College	St Thomas'
STATS19	665	862	837	1257

3. Method

Unlike the STATS19 data which includes 10 digit grid references for the location of every accident, the hospital data only records a road name, e.g. Oxford Street, Marylebone Road, Edgware Road. Whereas an area can be defined using a grid-referenced polygon for the police data this cannot be done for the hospital data. Also there is no absolute boundary for each hospital where a person might be expected to be taken by ambulance, or go to on their own, in the event of an injury.

For the St Thomas' area the hospital data needed to be sorted to ascertain whether or not each record lay within a cordon drawn around the hospital to create an area within which casualty records could be identified for linkage and matching (see Section 2). The procedure is described in Appendix 1.

The following datasets were matched with the STATS19 data as described in Section 3 of the original report and reproduced in Appendix 1

1. Hospital data for inside the area only.
2. All hospital data for inside and cross border casualties
3. Hospital data for inside and adjusted unmatched cross border casualties

This resulted in three estimates of the reporting rate. The first, using hospital data for inside the area only will give the highest reporting rate and is obviously an overestimation because only a proportion of the 'true' hospital dataset is being used. The second set of tables gives us the lowest reporting rate and is an underestimation because all the hospital data that lies on boundary roads is included and in the 'true' dataset only some of these will be in the area.

The best estimate of reporting rate lies between these two figures. Using the adjustment described in Section 2 of the original report and Appendix 1, the reader is guided to where this reporting rate might lie within the range. More work needs to be done to refine this process and an opportunity arises under work commissioned by the Department for Transport using data from other areas of England. Until this is completed the estimate is still considered to be the best estimate.

4. Results

4.1 The overall picture

The St Thomas' 2001 data did not accurately distinguish between pedal cycle casualties and two-wheeled motor vehicle casualties (TWMV). Table 4 shows how the matched records were categorised by the police and by the hospital.

In Table 4 the data shows all the police data and hospital data that were matched to the police records. Unmatched hospital data is also included for completeness. The rows represent the categorisation of casualty class recorded by the police and the columns represent the casualty class as identified by the hospital. Police data for casualty class is currently thought to be more accurate than hospital data for casualty class on the basis of the police presence at the scene of the accident. Police data is therefore used for defining class of user when it is available (i.e. for all casualties known to the police). Where there are inconsistencies between the police and hospital categorisations of road user class, these will affect casualties known to the police and casualties known only to the hospital and hence the estimate of reporting rate. Additionally the proportion of casualties in each classes is different for those known to the police and known only to the hospital for example

The largest inconsistencies for matched casualties is for the casualties identified as being TWMVs by the police. Of the 141 matched records shown in Table 4, (matched on other factors such as date, time, age and gender) 80 were categorised by the hospital as pedal cycles, 51 as other and 10 as car/van/bus. We therefore do not know how many TWMV casualties went to St Thomas's hospital as a result of their injuries. By extension, we do not know how many pedal cyclists attended hospital because of their injuries because about half of those in the hospital record are probably TWMVs.

Owing to the inconsistency of categorisation it is not possible to obtain estimates of reporting rates for PC, 'other' or TWMV that would be comparable to the other hospitals. A combined category (All others including PC and TWMV) has been created for this report, but it should be noted that this category is not comparable with the 'Other' category in earlier reports¹.

¹ Other used elsewhere is a category for motor vehicles such as milk floats, road rollers etc. it does not include TWMV or pedal cycles and therefore cannot be considered comparable.

Table 4: Classification of road user class by police and St Thomas' Hospital for casualties injured inside the area (i.e. not on the cross border roads)

Police Casualty Class	Hospital Casualty class						Total
	Pedestrian	pedal cycle	Car, van, goods, bus	Other	Unknown	Police only. Not found in hospital data set	
Pedestrian	128	1	1	1	1	126	258
PC	1	81	3	1		107	193
TWMV		80	10	51		201	342
Car etc	7	3	139	7	1	299	456
Other	1		2			5	8
Hospital only. Not found in police data set	56	75	90	23			244
Total	193	240	245	83	2	738	1501

Table 5 summarises the results for St Thomas' Hospital and Table 6a,b gives the estimates of (a) reporting rates to the police for each of the four hospital areas $((p+b)/(p+b+h))$ together with (b) percentage of records matched $(b/(p+b+h))$.

Table 5: St. Thomas' Hospital - lower and upper estimates of percentages of casualties reported to the police with the current best estimate of reporting rate for each category.

Casualty Class	Lower estimate	Upper estimate	Current best estimate
All casualties	71	84	79
Pedestrians	69	82	78
Car occupant	71	84	79
0-15 years	63	78	75
16-24 years	66	79	74
25-59 years	71	84	79
Over 60 years	75	87	84
Males	71.0	84.7	79.6
Females	70.4	81.8	78.7
Fatal	100	100	100
Serious	60	80	73
Slight	73	85	81

Table 6a, b: (a) Estimates of the headline reporting rates to the police for each hospital area together with (b) percentage of records matched.

Casualty Class	Lower estimate ¹	Upper estimate ²	Current best estimate ³
(a) Estimates of percentages of all casualties reported to the police (reporting rate)			
Kings	66	80	71
Barnet	61	75	70
St Mary's	77	91	87
St Thomas'	71	84	79
(b) Estimates of all casualties known to both/all known casualties (matching rate) expressed as a percentage			
Kings	29	22	31
Barnet	31	30	36
St Mary's	27	21	31
St Thomas'	29	35	33

¹ This estimate included casualties identified as being in the area only for police data and all those that are in the area and on cross border roads for the hospital

² This estimate included casualties identified as being in the area only for hospital and police data

³ This estimate included casualties identified as being in the area only for police data and all those that are in the area plus adjusted number for those on cross border roads for the hospital

Table 7: Estimates of the headline reporting rates to the police for each of the four hospital areas.*Rate=All casualties known to police/All known casualties.*

Casualty class	Area in London	Lower estimate ¹	Upper estimate ²	Current best estimate ³
All casualties				
Kings	Inner	66	80	71
Barnet	Outer	61	75	70
St Mary's	Central	77	91	87
St Thomas'	Central	71	84	79
Pedestrian casualties				
Kings	Inner	62	78	67
Barnet	Outer	72	78	70
St Mary's	Central	66	84	78
St Thomas'	Central	69	82	78
Pedal cyclist casualties				
Kings	Inner	61	75	66
Barnet	Outer	60	75	69
St Mary's	Central	89	98	96
St Thomas'	Central	N/A	N/A	N/A
TWMV casualties				
Kings	Inner	74	87	78
Barnet	Outer	65	80	73
St Mary's	Central	74	92	85
St Thomas'	Central	N/A	N/A	N/A
Car occupant casualties				
Kings	Inner	67	78	72
Barnet	Outer	58	73	68
St Mary's	Central	82	92	90
St Thomas'	Central	71	84	79

¹ This estimate includes all casualties identified as being in the area only for police data and all those that are in the area and on cross border roads for the hospital

² This estimate includes all casualties identified as being in the area only for hospital and police data

³ This estimate includes all casualties identified as being in the area only for police data and all those that are in the area plus adjusted number for those on cross border roads for the hospital

4.2 Detailed breakdown of results for St. Thomas' Hospital

4.2.1 Hospital data for inside the area only

Hospital data for those casualties that occurred on roads that are completely within the study area only. Table 9 shows casualties by road user class. Because of inconsistencies in coding vehicle class data the 'All others' category has been used. The 'All others' user class is not comparable to the 'Other' user class reported for the other hospitals.

Table 8: Overall frequencies and upper estimate of reporting rate

All casualties: St. Thomas', 2001				
Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate
738	244	519	1501	84%

Table 9: By Road User class

Road User Class	Casualties by road user class: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate
Pedestrian	126	56	132	314	82%
Car, van, goods, bus user	299	90	157	546	84%
All others including PC and TWMV	313	98	230	641	85%
All	738	244	519	1501	84%

Table 10: By Age

Age Band	Casualties by age: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
0 to 15	36	18	26	80	78
16 to 24	98	50	89	237	79
25 to 59	491	163	366	1020	84
Over 60	48	13	38	99	87
Unknown	65	0	0	65	100
All	738	244	519	1501	84

Table 11: By Gender

Gender	Casualties by gender: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Male	496	152	348	996	85
Female	242	92	171	505	82
All	738	244	519	1501	84

Table 12: By Severity

Severity	Casualties by severity: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Fatal	5	0	3	8	100
Serious	62	46	122	230	80
Slight	671	193	394	1258	85
Unknown	0	5	0	5	0
All	738	244	519	1501	84

4.2.2 Lower estimates of reporting rate using All hospital data for inside and cross border casualties

Table 13: Overall numbers and lower estimate of reporting rate

All casualties: St. Thomas', 2001				
Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
738	518	519	1775	71

Table 14: By Road User Class

Road User Class	Casualties by road user class: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Pedestrian	126	115	132	373	69
Car, van, goods, bus user	299	185	157	641	71
All others including PC and TWMV	313	218	230	761	71
All	738	518	519	1775	71

Table 15: By Age

Age Band	Casualties by age: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
0 to 15	36	37	26	99	63
16 to 24	98	97	89	284	66
25 to 59	491	355	366	1212	71
Over 60	48	29	38	115	75
Unknown	65	0	0	65	100
All	738	518	519	1775	71

Table 16: By Gender

Casualties by gender: St. Thomas', 2001					
Gender	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Male	496	344	348	1188	71
Female	242	174	171	587	70
All	738	518	519	1775	71

Table 17: By Severity

Casualties by severity: St. Thomas', 2001					
Severity	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Fatal	5	0	3	8	100
Serious	62	121	122	305	60
Slight	671	391	394	1456	73
Unknown	0	6	0	6	0
All	738	518	519	1775	71

4.2.3 Best estimates of reporting rates using Hospital data for inside and estimated unmatched cross border casualties

Table 18: Overall numbers and best estimate of reporting rate

All casualties: St. Thomas', 2001				
Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Best estimate Reporting rate per cent
738	328	519	1585	79

Table 19: Overall numbers and best estimate of reporting rate by road user class

Road User Class	Casualties by road user class: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Pedestrian	126	73	132	331	78
Car, van, goods, bus user	299	119	157	575	79
All others including PC and TWMV	313	136	230	679	80
All	738	328	519	1585	79

Table 20: Overall numbers and best estimate of reporting rate by age

Age Band	Casualties by age: St. Thomas', 2001				
	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
0 to 15	36	21	26	83	75
16 to 24	98	66	89	253	74
25 to 59	491	225	366	1082	79
Over 60	48	16	38	102	84
Unknown	65	0	0	65	100
All	738	328	519	1585	79

Table 21: Overall numbers and best estimate of reporting rate by gender

Casualties by gender: St. Thomas', 2001					
Gender	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Male	496	216	348	1060	80
Female	242	112	171	525	79
All	738	328	519	1585	79

Table 22: Overall numbers and best estimate of reporting rate by severity

Casualties by severity: St. Thomas', 2001					
Severity	Casualties known only to the Police	Casualties known only to the hospital	Casualties known to both	All known casualties	Reporting rate per cent
Fatal	5	0	3	8	100
Serious	62	69	122	253	73
Slight	671	254	394	1319	81
Unknown	0	5	0	5	0
All	738	328	519	1585	79

5. Interpretation of results

The results from matching St Thomas' Hospital with STATS19 data indicate that the reporting rate to the police is about 79 percent of all casualties. This is higher than at both King's College Hospital (71 percent - Inner London) and at Barnet General (70 percent - Outer London) but lower than at St Mary's (87 percent – Central London). From this extra analysis of St Thomas' data it is suggested that:

- the level of reporting to the police in Central London is generally higher than that in Inner and Outer London, and
- the reporting rate calculated from St Mary's Hospital data still seems on the high side.

The analysis of the St Thomas' data throws further doubt on the data from St Mary's. St Mary's A&E department has on average just over 63 000 attendances each year whilst St Thomas' has about 108 000 attendances

(http://www.performance.doh.gov.uk/hospitalactivity/data_requests/index.htm). The percentage of road traffic accident attendances to the total attendances is 1.07% for St Mary's and 2.65% for St Thomas'. Repeating the calculation for King's and Barnet the percentages are 2.12 and 1.53 respectively. Taking the average percentage road traffic casualties to all attendances at King's, Barnet and St Thomas' we would expect the number of road traffic accident casualties to be registered by St Mary's to be nearer to 2000 per annum than the 929 reported. This suggests an incomplete dataset was used in the original analysis of St Mary's.

The lack of identifiable pedal cyclist and TWMV casualties in the St Thomas' hospital record is a disappointment because these are two important road user groups. However, the data for the pedestrian casualties indicates that their injuries are generally well reported with a reporting level of about 78 percent. Even if the lower estimate is used, then the reporting rate to the police is still high at 69 percent. The best estimate is similar to that found at St Mary's but higher than that for the inner and outer London hospitals. It is difficult to assess differences between hospitals in the number of pedestrian casualties and their reporting rate in the absence of exposure data for this group (and the pedal cycle and TWMV riders if we had their casualty information at the hospital).

Possible explanations for the high Central London reporting rates:

1. If a casualty feels well enough to continue home and once there if they or their GP consider there is a need for hospital treatment, they are more likely to go to their local hospital and not back to the central London hospital near where they were injured. Therefore, for this unknown group of people, their injury will be in the police record and in the record of some other hospital unknown to us but not the hospital record being studied.

Lowdell et al (2002) in their study of injuries and accidents in London found that pedestrians were likely to be injured in the borough where they live, and pedal cyclists in the borough where they live or in an adjacent one. Drivers and riders of two wheeled motor vehicles are likely to be injured further from their home borough. Having said this there are substantial numbers, not quantified in the report, of road user casualties coming from elsewhere in London and outside. The borough with the highest number of road accident casualties is The City of Westminster with 2737 casualties in 2001, followed by Lambeth with 1928 (TfL 2002). Casualties from large parts of both these boroughs are likely to be taken to St Thomas'.

St Mary's hospital is situated in the City of Westminster and St Thomas', whilst in Lambeth, is just across the river from Westminster and is the nearest hospital to the southern part of it. Westminster has the most accidents of all London boroughs, has the lowest rate of injury to its own residents, and has the largest mix of casualties from other parts of London. An assumption that also may be made is that Westminster has the largest mix of casualties among people not living in London.

2. The level of police presence and awareness of traffic incidents in Central London means that when a road traffic accident occurs the police are more likely to be on the scene very quickly, thus making it more likely that the accident will be reported and appear in the STATS19 record.

6. Other issues

In the previous report, an estimate was made of the number of people injured in road traffic accidents for the whole of London. This was achieved by using the relatively conservative estimate of 70 percent as the reporting rate to the police. This, together with the number of casualties reported to the police but absent from the hospital record, led to an estimate of 63 000 casualties on London's roads in 2001.

Subsequently the total number of casualties has been re-estimated using the hospital activity data for London to estimate the number of A&E attendances for the groups of central, inner and outer London hospitals. For Central London there were about 398 000 attendances at A&E in 2000-01, the figures for inner London are 735 000 and for outer London 1.2m (see Table 23). Using the percentages of road traffic casualties to all attendances calculated using our representative hospitals (excluding St Mary's), it is estimated that about 44 700 people attended the 33 A&E departments in London hospitals as road traffic casualties in 2000-01.

Table 23: Estimates of numbers of road traffic accident casualties at London hospitals in 2000-01

Area of London	Percentage of rta/all attendances from 3 sample hospitals	First attendance at A&E (DH data)	Estimated number of rta casualties for area
Central	2.65	397891	10555
Inner	2.12	735248	15590
Outer	1.54	1205096	18554
London		2338235	44700

Table 24: Estimates of the proportion of all casualties in both police and hospital record for three London hospitals.

Hospital area	Casualties known only to police	Casualties known only to hospital	Casualties known to both police and hospital	All known casualties	Casualties known to both / all known casualties
King's (inner)	605	214	232	1051	0.221
St Thomas' (central)	738	328	519	1585	0.327
Barnet (outer)	324	280	341	945	0.361
All areas	1667	822	1092	3581	0.305

There were about 44 500 casualties reported to the police but as shown in Table 24 this represents those known only to the police (i.e. do not appear in the hospital record) plus a proportion of those that appear in the hospital record. Similarly for the 44 700 people

estimated to be in the hospital records across London there will be those known only to the hospital plus a proportion that appear in the police record as well. From Table 24 it is estimated that about 30.5 percent appear in both sets so the 89 200 people in the hospital and police records together represent about 130.5 percent of the number of people injured. This leads to an estimate of about 68 500 injured people on London's roads in 2001. This is within the range of 63 000 to 73 000 previously reported

Still of interest is to investigate whether there is a better way of estimating the number of casualties occurring on cross-border roads given they make up a substantial proportion of the total number of casualties in the hospital dataset. It has not been immediately obvious how this might be done without more precise locational data from hospital records but work for the Department for Transport on investigating trends in serious accidents may afford an opportunity to look at data from other smaller areas of England and through this a more refined method may be investigated.

Lowdell et al's (2002) work estimated that there were about 4300 hospital admissions each year for residents of London injured in road traffic accidents. In the course of the work for the DfT, the relationship between the number of casualties in the police record and those admitted to hospital will also be investigated. This may provide a methodology for providing an additional estimate of the more serious injuries.

7. Acknowledgements

The data was provided by Chris Lacey, A&E Consultant and Julie Herring, Emergency Department Service Manager at St Thomas' Hospital.

8. References

- Department of Health Hospital activity statistics, A&E Attendances.
(http://www.performance.doh.gov.uk/hospitalactivity/data_requests/index.htm)
- Lowdell C., Fitzpatrick J., Wallis R., Mindell J., and Jacobson B (2002) Too high a price. Injuries and accidents in London. Health of Londoners Programme Sept 2002.
- Pjsoft (1999) InfoMap ver.7.0,
- Robertson S. and Ward H. (2002) Levels of reporting to the police of road traffic accident casualties 1996-2000. Unpublished report, centre for Transport Studies, UCL.
- Simpson H.F. (1996) Comparison of hospital and police casualty data: A national study. TRL Report 173. TRL. Crowthorne.
- Transport for London (2002) Accidents and Casualties in greater London during 2001. Fact Sheet - London Accident Analysis Unit Quarterly summary – Year 2001 (Final) May 2002
- Ward H, Cave J, Morrison A, Allsop R , Evans A, Kuiper C and Willumsen L (1994) Pedestrian activity and accident risk. AA Foundation for Road Safety Research., Basingstoke.
- Ward H., Robertson S., Lester T., and Pedler A. (2002) Under-reporting of road traffic accidents in London: Matching police STATS19 with hospital accident and emergency department data. Unpublished TRL report PR/T/069/2002. TRL Limited, Crowthorne

Appendix 1: The data-matching process

A1.1 Sorting hospital records according to location in or outside the cordon





Sorting the hospital data is made difficult because only vague descriptions of location are given. Using InfoMap a list was extracted of all road lying within the cordon and each of the hospital records checked manually using the location information. This resulted in the records being partitioned into four sets:

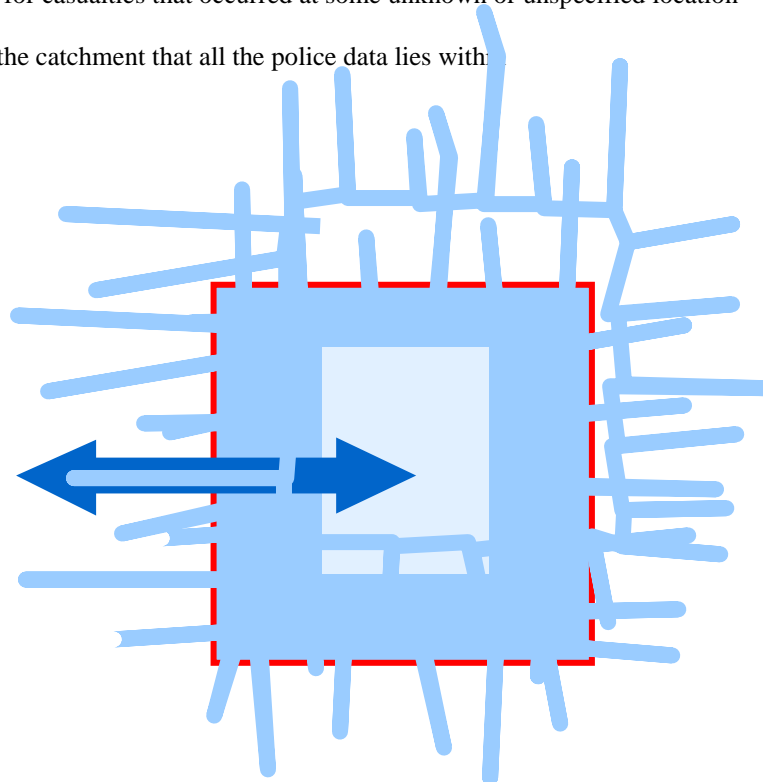
1. Definitely in the area
2. Definitely not in the area
3. On a road that passes through the area and continues outside it, or on a road that starts (or finishes) in the area and crosses the boundary
4. Unknown location

A1.1.1 Procedure for estimating casualties on roads running through the area and out the other side?

Figure A1 illustrates the issues associated with set 3 in the list. Here we do not know whether the casualty was injured on the part of the road that is in the area or the part of the road that is outside the area. The size of these boundary data will differ according to the road characteristics of the cordon area. This has implications for the size of the hospital dataset and hence the denominator used in the calculation of reporting rate.

Figure A1. Schematic diagram of roads within and roads crossing the boundary to the study area.

-  Hospital data for casualties that occurred on roads with their entire length inside the catchment
-  Hospital data for casualties that occurred on roads with some length inside the catchment and some length outside the catchment
-  Hospital data for casualties that occurred at some unknown or unspecified location
-  Boundary of the catchment that all the police data lies with



For the STATS19 records we know they are all in the area (or at least we are assuming that the grid reference is correct) so these issues do not arise here.

If we use only the data that are known to be in the area for matching with STATS19 data we will be ‘discarding’ hospital casualties that lie on the roads passing through the area even though we cannot identify these records individually. We would therefore overestimate the reporting rate. See Figure A1.

However, if we include all the boundary data as being within the hospital set we will be overestimating the size of the hospital dataset and thereby underestimate the reporting rate.

The actual number of casualties known only to the hospital will lie somewhere between the values described above. In this report we have included the high and the low estimates of reporting rate together with an estimate of the actual value.

For the purposes of this report, we have made an estimate of the proportion of casualties that are known to the hospital only (unmatched hospital data in the boundary area) on roads which cross the boundary of the study area. This estimate is calculated separately for each hospital area and is based on the number of matched casualties known to be in the study area divided by the number of casualties in the hospital dataset that are known to be wholly within the study area. Then the number of casualties in the cross border dataset are divided by this proportion to give the total number of casualties we might expect to have occurred within the study area on the roads that cross the borders of the study area. To get to our final estimated number we then have to remove the casualties that are deemed to be outside the boundary but as we don’t know which these are we exclude, and put to one side, the ones that already match then remove the others according to a random selection process.

This estimate is the best that can be provided within the scope of this report and in terms of the time available for the analysis. We have made the assumption that the reporting and other casualty characteristics for casualties occurring within the study area are the same as those occurring on the roads crossing the boundaries of the study areas. This assumption is not very robust as the characteristics of roads that cross the study area boundaries can be markedly different from those within the study area in terms of size, traffic flow, composition and numbers of casualties. An example of this would be the M1 in the Barnet study area. On the other hand, there are some roads which cross the study area that would be more typical of the roads found entirely within the study area.

A1.1.2 Hospital data whose location was unknown.

A similar approach could be applied to casualties known to the hospital only whose location was unknown. Time constraints did not permit a similar treatment at this stage. There were relatively few of these and it was concluded that the difference in the resulting estimate of numbers would not be great as there were few matches for casualties where the hospital location was unknown.

A1.2 The matching process

The matching process was undertaken in two stages with an automated and a manual component. The police data and hospital data (having been pre-processed to a consistent format) were placed on the same page of a spreadsheet which allowed automated and visual comparison of the two data sets. This section will describe the data sets used, and the matching process.

A1.2.1 The combined dataset

The data from the hospitals and from the police were converted to a compatible format for the matching exercise. This included parsing of date and time data to a form that could be used in a spreadsheet. These are shown below.

Police data was converted to:

Refnum	Sex(p)	Age(p)	Cas Class(p)	Severity (p)	Date	Police =1	Time	East	North	Location	Postcode	Match	Date
--------	--------	--------	-----------------	-----------------	------	--------------	------	------	-------	----------	----------	-------	------

Hospital data was converted to:

Date	Age(h)	sex(h)	Class(h)	Severity(h)	Hospital= 2	RTA Date	RTA Time	Location of Accident	In area	Ref number
------	--------	--------	----------	-----------------	----------------	----------	----------	----------------------------	---------	---------------

There are two date fields in each of the data sets. The data field at the end of the police and the one at the beginning of the hospital data share the same spreadsheet column and for matched data become a common field. A further set of fields converted the police and hospital codes to a common set of codes that were used in the analyses. These fields were:

PHB	Age(hp)	Severity(ph)	Class(ph)	Year code	sex(hp)
-----	---------	--------------	-----------	-----------	---------

The field PHB identifies whether the line of data is known to police only, to hospital only or to both. The year code is to allow year by year comparison in the future.

A1.2.2 Data coding

Data coding was key to the success of the work. The data codings are shown in Appendix 3 of Ward et al (2002). Some of the key data coding issues are described below.

A1.2.2.1 Severity

The severity coding used was the STATS19 levels of killed, seriously injured and slightly injured. The hospital descriptions of severity were in the form of disposal codes or outcomes and these were mapped onto the STATS19 severity codes. Each hospital had a different set of disposal codes (or discharge outcomes) which were individually mapped on to the severity codes mapped.. These can be found in Appendix 2 of Ward et al (2002)

A1.2.2.2 Codes for vehicle class

The coding used for the class of a road user casualty by the police and by the hospital is different so a set of categories that encompassed both coding systems was required. The following categories encompassed both sets and have been used throughout the phases of the study:-

Pedestrian

Pedal cycle

Two Wheeled Motor Vehicle

Car, bus , Heavy Goods Vehicle,Light Goods Vehicle (etc

Other (horse, milk floats etc)

Unknown

Police user class data were taken as definitive for tables. It was assumed that the vehicle type was more likely to be correctly identified by police at the scene of the crash than by hospital A&E staff. The hospital user codes are shown in Appendix 3 of Ward et al (2002)

A1.2.2.3 Age band codes

Hospital age data were taken as definitive for tables. Where these were not available the police age data were used. The following age bands were used

0-15

16-24

25-59

60+

Unknown

A1.2.2.4 Codes for area

The locations of the hospital casualties were coded in terms of whether they were on roads that were within the study area. These data were used to allow selection for data for different analyses.

The categories were:-

- On roads wholly within the study area.
- On roads that were partially within the study area.
- On roads whose location was unknown.
- On roads outside the study area.

A1.2.2.5 Codes for date and time

In some cases for police data, the time given has the same minutes and hours shown where the hours are greater than 23. This means that for the purposes of the analysis, the actual time of the incident is unknown and also cannot be handled by the matching algorithm. Such times are therefore coded in this analysis as 12:12. The use of 12:12 as an unknown time value provides a mid point during the day and allows the data to be included in the automated matching process. In some cases the time of incident recorded by the police is rounded to the nearest hour. The use of the value 12:12 also indicates that the value is likely to be a missing one and not a rounded one.

For the hospital data, both the time of registration at the hospital and the time of the incident were recorded. Where the date of the incident date was missing it was replaced by the date of registry at the hospital. Where the incident time was missing this was set to 00:12 so as to be 12 hours from missing police incident time thus avoiding a spurious time match.

A1.3 The matching algorithm

Experiments were undertaken using 1 month of data to determine the effectiveness of various algorithms. 5 fields were used in the algorithm:

Date of incident

Date and time of incident (allowed incidents spanning midnight to be included)

Age

Gender

User class

A number of runs using different combinations of mandatory matches were undertaken and the total number of matches

Number of matching variables	Mandatory matching variables	Matches	Of which false positives
5	All	12	0
4	Date	24	5 False 2 uncertain
4	Date & age	20	2 uncertain
4	Date & gender	20	2 uncertain
4	Date age & class	20	2 uncertain
4	Date & time	12	0
4	Date or time, and age & gender	20	2 uncertain

Given the high proportion of false positives obtained where 4/5 variables with same date were used it was decided not to investigate the effect of 3 matches. The algorithm finally selected for the auto matching was at least 4 of the above fields matching with the following fields mandatory.

- age
- gender
- date or time

A1.4 Automated matching

The data sheet consisted of two sets of columns of data one for the police and one for the hospital data with one common column for the date field. When two lines of data matched they were placed in a single line. The matching process moved lines of hospital data to the same line as the matching police data. Codes for age, gender, and whether a given line contained data from police, hospital or both etc were calculated in further fields from which pivot tables were used to produce results tables.

The automated matching process used an algorithm written in visual basic for application running within an excel spreadsheet. The automated matching process was run using all police and all hospital data including those hospital data that had been provisionally identified as being outside the study area. This allowed a double check on the exclusion of hospital data. The automated matching process did not take account of the plain English description of locations.

Codes for matches were inserted in the matches field on the combined data page. A value was inserted to indicate a match, a blank is left to indicate no match. In the auto sort routine, a number was also inserted where additional lines also matched the specific police data. These were used to help inform the manual matching process.

4 auto match 4 matching variables: - (Date or time) and (age and gender)

5 auto match 5 matching variables (date, time, age, gender, class)

The matching process was run on all police and all hospital data supplied (this included data which had been provisionally identified as outside the area so as to provide a double check on location data).

A1.5 Manual matching

After the automated matching process, a manual matching process was undertaken on the output of the automated matching process.

For the manual matching process there was an element of qualitative judgement applied to the matches based on the experience of the people undertaking the matching process. Possible additional matches and false matches were identified by one individual and the decisions on whether to include these were made by two or in difficult cases by 3 individuals.

The fields used in the manual matching process were:-

Date

Time

Age
Gender
User class
Location (plain English)

In some cases the following codes were included in the matched spreadsheet

6 Manual match where the matchers had high confidence.

7 Manual match where the matchers had less high confidence.

This process was undertaken on the matched datasheet. The datasheet used earlier to identify the location was not updated. Where matches were made on data that had been previously been identified as being outside the area, a careful examination of location information was undertaken. In some cases additional data that were within the area were identified. Matching lines of data were moved manually within the spreadsheet.

The manual matching required some qualitative assessment. In a number of cases the original data had to be revisited especially where there were inconsistencies of date. An example would be where the date of incident was recorded by the hospital as being on one day and the date of registration at the hospital was recorded as being on the day before, as was the police record of the accident. Such a combination of factors would imply a match. other types of inconsistency included misuse of the 12/24 hour clock system so tie of incident could sometimes appear to be 12 hours apart.

A1.6 Selecting subsets of the data

The original matching exercise was undertaken on the full dataset, but for a number of the analyses a sub set of the data was required, (for example hospital casualties on roads that were entirely from within the study area only). Where subsets of the data were required, the unwanted hospital data was identified by sorting on the "in area" field then removing the unwanted hospital data and saving the spreadsheet under a new name.