Using multiple datasets to understand trends in serious road traffic casualties

Ronan A. Lyons, Heather Ward, Huw Brunt, Steven Macey, Roselle Thoreau, O.G. Bodger, Maralyn Woodford

1. Introduction

Road traffic injuries are the ninth-leading cause of death in the world (World Health Organisation, 2002) and are predicted to become the third leading cause of death and disability worldwide by 2020 (Murray and Lopez, 1996). More than a million lives are lost as a result of such injuries around the world each year and millions more are injured or disabled as a result (Department of Injuries and Violence Prevention, 2002). In Great Britain (England, Wales and Scotland) alone, 3200 people are killed and another 29,000 seriously injured from the 200,000 collisions in which at least one person is injured that occur each year (Department for Transport, 2006).

Road traffic injuries can affect all populations, regardless of age, sex, income, or geographic region (Krug et al., 2000). The economic cost of the problem in Britain is known as STATS19. It is known that not all casualties arising from road traffic collisions are reported to the police so use of STATS19 on its own is unlikely to provide an accurate reflection of the true risk of being injured on the roads. Some work has been undertaken previously to explore this issue using different sources of data, but that research has been restricted to using one additional dataset selected from a limited number of datasets (Gill et al., 2006; Stephenson et al., 2005; Tunbridge et al., 1988; Bull and Roberts, 1973; Hobbs et al., 1979; Mills, 1989; Nicholl, 1980; Pedder et al., 1981; Saunders and Wheeler, 1987; Simpson, 1996; Haynes et al., 2005).

Given these limitations, the use of multiple data sources, should in theory, improve measurement of trends in serious road traffic injuries. Such information needs to be readily available to help monitor changes in road traffic injury epidemiology, identify any associated wider public health impacts, assess the effectiveness of interventions and track progress against national road safety targets.

Our study aims to investigate the extent to which our understanding of trends in serious road traffic injuries is aided by the use of multiple datasets.
2. Methods

We used multiple sources of data to compare trends in road traffic casualties overall and by main categories of road user: vehicle occupants, motorcyclists and pedestrians. We did not have access to each of the data sources for exactly the same time periods but for most there were several years of overlap between 1996 and 2003. In Great Britain, police forces complete a set of forms for every road traffic collision reported to them which results in at least one casualty. The electronic record is known as STATS19 and is maintained locally by each police force which then submits records on a regular basis to the Department for Transport who maintain the national STATS19 dataset. STATS19 casualty data were obtained for the period 1996–2003 from local police forces and from the Department for Transport’s annual publication on casualty numbers (Department for Transport, 2004).

We used hospital admissions data from England (Hospital Episodes Statistics), Wales (Patient Episode Database for Wales) and Scotland (Scottish Morbidity Record) (1999–2003) for analysis alongside police STATS19 records. In addition, we analysed data from the Trauma Audit and Research Network (TARN) database for the period 1996–2003 (Patel et al., 2005). This dataset comprised a subset of data from 33 hospitals across England and Wales and which have been reliably reporting over this time period. The database holds information on the more severely injured casualties where, amongst other inclusion criteria, the length of stay is at least 72 h. The TARN dataset collects information (dependent on patient’s fulfilling the explicit inclusion and exclusion criteria) on the mechanism of injury, road user type and other road traffic accident groups. Unlike hospital inpatient and emergency department datasets which do not routinely contain measures of injury severity the TARN database contains Injury Severity Scores (ISS) which correlate with mortality, morbidity and hospital stay (Baker and O'Neill, 1976).

The definition of a serious road traffic related injury is not standard across datasets. The STATS19 definition of ‘serious’ injury is much broader than the term ‘hospital admission’ and includes fractures and other injuries that are treated at specialist clinics as outpatients. Analysis of emergency department (ED) data shows that about 20% of road traffic casualties attending such departments could be considered ‘serious’ in STATS19 terms and that admissions comprise about half of this category (Ward et al., 2006). However, because computerised ED data are not widely or readily available the most reasonable comparator for STATS19 serious casualties is the incidence of hospital admissions for injuries due to road traffic. It is customary to use killed or seriously injured casualty numbers (KSI) combined for analysis and these are used in government targets. However, as only about 20% of road traffic related deaths occur after admission to hospital, the number of seriously injured casualties is a more appropriate STATS19 comparator for hospital admissions than is KSI (Ward et al., 2006).

We matched ED and STATS19 data for the period 1998–2004 in an English City with only one hospital to aid understanding of differences in total numbers of serious casualties recorded by the police and health services. We applied the police definition of serious to the health service data and counted serious casualties as those admitted to hospital, presenting with fractures or requiring specialist follow-up.

We also used data from a large emergency department in Wales to study trends in road traffic casualty attendances from 2001 to 2004. No major changes in emergency department configuration in adjacent hospitals occurred during this period.

To formally make a comparison between datasets, we fitted a set of General Linear Models (GLM’s) to incorporate unique intercepts and trends for each dataset. As the datasets record information on different scales we were not interested in the intercept points, focusing instead on the trend lines.

3. Results

3.1. Trends in overall serious road traffic casualties

Table 1 shows the number of road traffic casualties in the different databases available between 1996 and 2003.

As data coverage varied by geography and year the percentage change in each dataset is shown in Fig. 1.

Our application of the GLM showed that all datasets were well fitted by a linear trend line, with four out of the five (Hospital Admissions, Welsh Hospital ED, TARN, STATS19 Fatalities) showing a flat trend (slope not significantly different from zero at the 95% confidence level). Only the STATS19 Serious Injuries dataset was found to have a trend line significantly different from zero (p<0.001), and significantly different from the trends of the other datasets.

In percentage terms the slope of the trends (change per annum) were −0.44% (95% CI −1.1% to 0.2%, p=0.2) for Hospital Admissions; −0.36% (95% CI −1.9% to 1.9%, p=0.97) for ED attendances; 0.18% (95% CI −3.8% to 4.2%, p=0.93) for TARN; and −0.48% (95% CI −3.9% to 2.9%, p=0.77) for STATS19 Fatalities. In contrast, the STATS19 Serious Injuries data indicated a decline in serious injuries of 4.3% per year (95% CI −4.7% to −4.1%, p<0.001).

Table 1

| Number of road traffic casualties in different data sources by year from 1996 to 2003 |
|----------------------------------|---------|---------|---------|---------|---------|---------|---------|
| Hospital admissions (GB)        | N/A     | N/A     | N/A     | 36,034  | 35,960  | 37,141  | 35,491  | 35,490  |
| STATS19 serious (GB)            | 44,499  | 42,984  | 40,814  | 39,122  | 38,155  | 37,110  | 35,976  | 33,707  |
| STATS19 killed (GB)             | 3598    | 3599    | 3421    | 3423    | 3409    | 3450    | 3431    | 3508    |
| TARN (33 hospitals)             | 3089    | 3072    | 2766    | 2919    | 2788    | 2876    | 3042    | 3147    |
| ED Wales (1 unit)               | N/A     | N/A     | N/A     | N/A     | 2818    | 2791    | 2798    |         |
| ED England (1 unit)             | N/A     | N/A     | 486     | 565     | 552     | 571     | 465     | 500     |

Fig. 1. Percent change in trends of road traffic casualties using data from four datasets, 1996–2003 (2001 reference year for Welsh hospital ED data; 1999 for all other datasets).
 ED data from one English hospital were also matched with local police STATS19 records for injury severity in a city in England where there is only one hospital. Of the total of 328 police-recorded serious casualties (over an 8-year period between 1996 and 2004), 38% were not found in the hospital record, 41% (135) were matched as serious in both records and 21% were classified by the police as serious but treated and discharged by the hospital (slight). Of the 2866 police-recorded slightly injured casualties during the same period, 51% were not found in the hospital record, 41% were matched as slight in both records and 8% (229) were classified by the police as slight and by the hospital as serious. The hospital recorded a further 2644 casualties not known to the police; 15% (397) of whom were seriously injured and 85% were slightly injured. Overall, the hospital recorded 761 serious casualties (135 police serious, 229 police slight, 397 unknown to police) as against 328 police-recorded serious casualties.

3.2. Trends in road traffic casualties by road user type

Differences were observed in trends of road casualties by road user category over the time period. Pedestrian casualties reduced by 4.0% per annum in STATS19, by 3.1% in hospital admissions and by 2.9% in TARN data. Motorcycling casualties showed the opposite trends with a 1.9% per annum increase in STATS19, a 4.6% increase in hospital admissions and 7.2% increase in the TARN dataset. For vehicle occupant injuries trends differed by data source with a 0.3% per annum increase in hospital admissions, a 0.3% decrease in TARN and a 3.4% decrease in STATS19.

Injury Severity Scores (ISS) were also calculated from the TARN data by category of road user, providing a general severity score for casualties. The median ISS showed little change over the period 1996–2003 (ranging between 9 and 10), indicating that there had been no trend in the severity of serious road traffic related casualties in the catchment areas of the 33 hospitals between 1996 and 2003.

4. Discussion

The results of this study indicate that accurately measuring the incidence of serious road traffic casualties is difficult and that interpretation of trends will be influenced by choices of datasets and metrics. The findings show that substantially larger numbers of serious road casualties are recorded in ED and inpatient databases than in police data in Great Britain and that there is some divergence in trends between the various health and police datasets. Several findings are consistent across datasets. All show downward trends in pedestrian casualties and an upward trend in motorcycle casualties. However, for vehicle occupant casualty, which comprise the majority of road collision casualties, the trends differ: the police STATS19 serious casualty data shows a significant declining trend. All other datasets and analyses: STATS19 killed, hospital admissions, TARN and emergency department data show no significant changes over this period. Why the trends differ needs careful consideration as there are a number of potential explanations.

A limitation of this and any study which addresses these issues is that the definition of an accident varies across data sources. In Great Britain all road traffic collisions in which at least one person is injured on a public road are meant to be recorded in the STATS19 system. Emergency department databases are well designed to capture information on the broad geographical location of the incident and the role of the injured person (driver, occupant, pedestrian). However, these databases fail to adequately define a ‘road traffic accident’ consistently, meaning that it is likely that some off-road incidents are also recorded under this heading (Ward et al., 2006).

It should be noted that hospital data will include all pedestrian and pedal cyclist casualties that occur in a public place and can be thought of as a road accident by hospital clerical staff who are not trained in the precise definitions used by the police. Such errors are likely to be repeated in hospital admissions and trauma databases. However, for vehicle occupant casualties the scale of this bias should be relatively small as the vast majority of collisions in urban societies will occur on public roads.

A far greater difficulty in trying to measure serious casualties or injuries across different datasets is the lack of a standardised definition of severity across datasets (Lyons et al., 2006; Lyons et al., 2005a; Cryer, 2005; Lyons, 2005b). Injury severity will influence whether medical treatment is sought at all and, if so, from where. Minor injuries rarely require hospital treatment and casualties with this type of injury may treat themselves at home or visit their general practitioner (GP). Previously undertaken studies that matched GP data with the police STATS19 record suggest that minor injuries are the most frequent injury outcome of road traffic accidents (Neighbourhood Road Safety, 2006) and that only 24% of people with such injuries are recorded by both GP and police datasets. Emergency department databases are well designed to capture recording procedure of ‘finished consultant episodes’ (FCE) to count events for performance and financial management. This makes data analysis quite difficult. A patient may have several FCEs whilst receiving a number of bouts of treatment for an injury, or complications arising from an injury, and so data linkage, using unique identifiers, is needed to create person-based records. These databases are useful for measuring injury incidence, but are not for determining injury severity since the international classification of disease codes (ICD–10) does not allow practitioners to capture the severity of the injury. Considerable skill is needed to interpret road traffic injury-related hospital inpatient data (Ward et al., 2006). Changes in service provision or delivery, such as a change in
more or less invasive surgery for specific injuries, or in bed availability can influence the threshold for hospital admission (Lyons et al., 2005a). Stephenson et al. (2005) have reported differential trends in admissions with differing levels of severity of traumatic brain injuries. Whilst there has been some improvement in trying to define a subset of serious injuries using ICD-10 coded data, there are still problems with the validity of this approach (Cryer, 2005; Lyons, 2005b).

Incomplete coding of the external causes of injury is a problem with hospital inpatient datasets. S and T codes (S00-T32 trauma codes denoting injury type and part of body injured [injuries and burns]) in Chapter XIX of the International Classification of Diseases, version 10 (ICD-10) are meant to be accompanied with an external cause code from Chapter XX (codes V01-V98). In practice, this is not always the case, in 1999, 9.3% of finished consultant episodes in the English Hospital Episode Statistics dataset with injury and burns codes (S00 – T32) were missing an external cause code and the percentage in 2003 was 10.4%. This change is in the opposite direction to that required to explain differential trends between health and police data due to data coding problems. However, as the number of records with a primary injury code which have missing external cause codes is similar to the number which have a transport cause (V01-V99) there is still room for a substantial systematic bias.

A further issue is that not all road casualties are brought to the attention of both the police and health sectors. Police are not called to all collisions in which a person may be injured. People may transport themselves to hospital or be conveyed by ambulance and the police may not always be informed of the incident. Where the police do not attend the collision it should be reportable to them but this does not always occur in practice for a variety of reasons.

Self-reporting rates appear to be influenced by a variety of factors. One study found that car occupants were most likely to report injury-related collisions (70%), followed by pedestrians, motorcyclists and other vehicle occupants (55–60%), with pedal cycles least likely to report (22%). In addition, reporting rates are highest on motorways (78%), followed by rural roads (68%) with urban roads having a lower rate (58%) (Simpson, 1996). Many other factors are likely to be involved, such as collisions involving motorists who are driving untaxed or uninsured vehicles or who do not want to bring themselves to the attention of the police for a variety of reasons. Even when injured casualties are reported to the police they are not always recorded in the STATS19 database; one study commented that 20% were not recorded (Hopkin et al., 1993).

Even when both police and health data contain records on the same casualties further difficulties can arise in matching process as that the same variables are not recorded in both sets of data and even when they are they may differ by categorisations or levels of detail. Names, addresses or dates of birth are not kept within the police datasets and confidentiality requirements mean that matching between datasets is often on a fuzzy or probabilistic basis and may induce false negatives as well as false positives.

The diverging trends between the police STATS19 serious casualties and the STATS19 killed and all the health databases could be real or artefactual. A reduction in injuries that were not serious enough to warrant hospital admission but still in keeping with the police definition of serious could explain a divergence between STATS19 serious and the hospital admission and TARN data (Gill et al., 2006). This would require a change in the pattern of collisions or the consequences of such collisions resulting in relatively stable numbers of slight, more severe, and fatal injuries but a decreasing number of less severe injuries. One suggestion is that modernisation of the car fleet with more recent models having substantially improved secondary safety features, such as air bags and side impact protectors, may result in reducing the severity of some injuries but not be sufficiently protective to reduce the very severe or fatal injuries. Such a hypothesis requires further research.

An artefactual cause could relate to changing thresholds for injury admissions. Recent pressures on acute bed availability in Great Britain might be expected to reduce the likelihood of admission for less severe injuries but pressures on reducing emergency department waiting times may operate in the opposite direction. Our analyses do not show such a trend in either direction. The general hospital inpatient analyses do not contain injury severity measures. However, the parallel trends between the general hospital inpatient data and TARN data, together with stable median ISS scores within the TARN dataset suggests that the trends in the general admission datasets have not been substantially influenced by any changes in admission threshold over the period studied. The trends in all hospital admissions and TARN admissions is similar for all injuries and injuries within road user types, which indicates that these are unlikely to have been influenced by changing admission thresholds over this period. One of the limitations of the TARN database is that it does not cover all hospitals in England and Wales. We have limited our analyses to the 33 core hospitals providing high quality data over a long period of time. There has been no major reconfiguration of hospitals within this core during this period. However, reorganisation of services in other hospitals, such as closure or opening of new units could influence the catchment areas for these hospitals. We think this is unlikely to have had a major effect but do not have the data from non-participating hospitals over time to measure the extent of such changes.

Another plausible explanation for the divergence between datasets is a minor change in the consistency of application of the police definition of a serious injury. As slight casualties are about 10 times more common than serious casualties, very small changes in judgement about the operational threshold for this categorisation could easily produce a large degree of change (Ward et al., 2006).

5. Conclusion

The results of this study show that trends in serious road traffic casualties differ between various police and health datasets in Great Britain. It has not been possible to deduce which dataset or datasets are the most accurate as all have strengths and weaknesses. Determining which single or combination of datasets is best suited to this task is an urgent research priority in order to be able to inform the design and implementation of future road safety strategies.

Several developments may help in this endeavour. The first is the likelihood of increasing availability of emergency department datasets which together with developments in pseudonymised data linkage will enable wide scale individual level emergency department and inpatient data to be linked, allowing better understanding of trends and any changing thresholds for admission (Ford et al., 2007). An international study to define a set of ICD-10 codes which always leads to admission has also commenced. The results of this study will help inform the use of hospital admission database metrics for the surveillance of serious road traffic casualties.

The increasing availability of the casualty postcode in the STATS19 data allied to increasing availability of electronic emergency department data should facilitate more accurate large scale police–health linkage studies.

The results of these studies should help derive the most appropriate metric for surveillance of serious road traffic casualties. Whilst the use of multiple databases has attendant costs from a societal perspective these will be far lower than the costs associated with potential misinterpretation of trends and consequent policy decisions which could result in lost opportunities to prevent road traffic injuries.
References


