Capture-recapture: a useful methodological tool for counting traffic related injuries?

A Morrison, D H Stone

Abstract

Introduction—Although the capture-recapture technique is increasingly employed in studies of human populations to correct for under-ascertainment in traditional epidemiological surveillance, it has rarely been used in injury research.

Objectives—To estimate the completeness of official data sources on traffic related injuries (TRIs) by using the capture-recapture technique and to calculate an ascertainment corrected number of fatal and serious TRIs among Scottish young people aged 15–24 years. The appropriateness of the approach in this context is also assessed.

Method—A two sample capture-recapture technique was applied to two official sources of TRI data. Data on TRIs were obtained from the Scottish Health Service and the STATS19 dataset at the University of Essex Data Archive for 1995. Four standards (A-D) of matching were applied to fatalities and serious TRIs to allow plausible relaxation of matching standards within the context of the data collection setting. The completeness of each data source was assessed, and an ascertainment corrected number of fatalities and serious TRIs calculated.

Results—The ascertainment corrected number of TRI fatalities among 15–24 year olds using standard D was 104. This represents only a small increase in the number of fatalities using capture-recapture than when using each individual dataset. The completeness of the Scottish Health Service database for TRI fatalities was 93%. The STATS19 database was 95% complete. The ascertainment corrected number of TRI hospital admissions was 1969. The STATS19 and the Scottish Health Service databases were approximately two thirds and three quarters complete respectively for non-fatal TRIs requiring hospitalisation.

Conclusions—Injury researchers have advocated the linkage of major datasets to supplement and improve the quality of injury data. Using capture-recapture we found that routine databases enumerate TRI fatalities accurately, in contrast to injury morbidity databases that do not. Capture-recapture is a potentially useful method of evaluating the completeness of data sources and identifying biases within datasets. However, ascertainment corrected rates should be viewed with caution. A number of requirements of the capture-recapture technique are unachieved in this study of injury in the human population.

Keywords: capture-recapture; young people; epidemiology

Capture-recapture has traditionally been employed in biometrics, particularly in the estimation of animal populations. Recently, the technique has increasingly been adopted in health studies of human populations to generate more accurate rates of disease and disability. This involves estimating the number of cases in a defined population using multiple sources of information, assuming that each source alone may under-count the population. Data generated by traditional epidemiological surveillance systems are frequently criticised as being of poor quality due to under-ascertainment. For this reason, proponents argue that capture-recapture offers an efficient and cost effective alternative to conventional epidemiological surveillance and to universal enumeration which is inefficient, expensive, and often impossible.

Capture-recapture methods have been employed to estimate the prevalence of a wide range of medical conditions including diabetes, various cancers, HIV, stroke, inflammatory bowel disease, meningococcal disease, and tuberous sclerosis. The technique has also become increasingly popular for estimating “hidden populations”, such as illicit drug users, prostitute populations, and the homeless. A small number of studies have employed the capture-recapture approach in the field of injury prevention. The technique was used to ascertain the completeness of child pedestrian injury reporting in both routine public hospital discharge statistics and an active injury surveillance system in New Zealand. In Pittsburgh, USA, multiple data sources were used in an attempt to ascertain the number of
injuries sustained by adolescents in a single school district. The same researchers advocated capture-recapture as a way of monitoring the incidence of head and spinal injuries in both developed and developing countries. The technique was also adopted in Colorado in an attempt to elicit accurate incidence data for head and spinal cord injuries in the United States.22

To date, the capture-recapture technique has rarely been applied to the study of traffic related injuries (TRIs) specifically. In Northumbria, England, a capture-recapture study identified a large pool of serious childhood injuries from road traffic accidents that were not included in official data sources.23,24 Comparisons of police and hospital data (not employing capture-recapture) have been conducted.14-16 These studies concluded that not all potentially reportable accidents are recorded, and that there are biases in reporting depending upon the type of road user. Events involving casualties with less severe injuries, pedestrians, pedal cyclists, and motorcyclists are less likely to be recorded.

While proponents view capture-recapture as an under-exploited tool in injury research, there are a number of well documented arguments that endorse caution. Some of the requirements of the capture-recapture technique may be unachievable in studies of the human population. The commonest cited problems are the severity effect, biases in capture based on demographic characteristics and injury type, the relationship between input lists and the requirement for a closed population.

Cormack stresses that there is often insufficient information in capture-recapture studies to calculate accurate estimates of population size.25 With two lists, an estimate of the population size is calculable only by making an assumption about the homogeneity and independence of the input data. Others stress the need for careful source selection and piloting,26 and appropriate use of the technique depending upon the nature of the databases and the intended use of the results.27 The use of ancillary data and ad hoc information to supplement the procedure is also recommended.28

The present study pilots the two sample capture-recapture technique to estimate the completeness of data on TRIs in Scottish young people, calculate an ascertainment corrected number of TRIs, and to assess the appropriateness of the approach using two official data sources. Injury researchers have advocated the use of multiple data sources for injury research to improve and supplement existing information. In Scotland, a vast bank of health and social data is available for analysis to injury researchers including mortality data, hospital inpatient data, and police reported traffic accident data. A recently launched population based household survey in Scotland may also provide useful ancillary data for this type of study in the future. While these official data sets are not perfect, they do represent an opportunity for research with population based data and an opportunity for economies of scale. However, the existing routine databases holding information on TRIs are open to criticism of variable ascertainment.

Methods

Data on TRIs are available from two sources in Scotland: the Scottish Health Service and the Department of Transport, Environment and the Regions (DETR) (formerly the Department of Transport). The Scottish Health Service collects data on deaths (via the Registrar General for Scotland) and hospital discharges. Details of deaths coded specifically as TRIs (\textit{International Classification of Diseases}, ninth edition (ICD-9) codes E811–816, E819, and E826–829) were obtained for young people aged 15–24 years resident in Scotland in 1995. Scottish morbidity record 1 (SMR1) data were obtained from the Information and Statistics Division of the Scottish Health Service. The SMR1 is a record of non-obstetric, non-psychiatric discharges completed on discharge, death, or transfer from a Scottish hospital. Records with an ICD-9 code used in the selection of mortality data were included.

Police departments in the UK collect data on all police attended road traffic accidents occurring on public roads in their locality where at least one vehicle and one human casualty is involved. Casualties include injured drivers, riders, passengers, and pedestrians. Scottish data are submitted to the DETR for inclusion in the UK-wide STATS19 database held at the Data Archive, University of Essex.30 “Fatal” accidents are defined as cases where death occurs within 30 days of the injury event. “Serious” accidents are defined as detentions in hospital as an inpatient either immediately after the injury event, or have any of the following injuries: fractures, severe concussion, internal injuries, crushings, severe cuts and lacerations, severe shock requiring medical treatment. Most of these injury types result in hospital admission. The data are based upon information available within a short time of the accident, and do not include results of a formal medical examination. Data on all 15–24 year olds involved in an road accident in Scotland classified either as “fatal” or “serious” in 1995 were extracted. Denominator data were obtained in the form of mid-year population estimates from the Registrar General for Scotland.

A two sample capture-recapture analysis was performed to estimate the extent of “undercounting” in the two main data sources for fatal and serious TRIs among young people aged 15–24 years for the calendar year 1995. Individuals were matched using a number of key variables. Four variables were used for matching mortality records; age, sex, day of accident, month of accident (table 1). To match records of serious (but non-fatal TRIs) five key variables were used: age, sex, day of accident, month of accidents, and a location code (table 2). The police area code
(there are nine in Scotland) were matched with the hospital code of the emergency admission. For instance, Glasgow Royal Infirmary and Strathclyde police would constitute a match. Circumstances where patients may cross police authority borders for emergency treatment were taken into account. The standards used to define a match were based upon a concept used by Razzak and Luby in a capture-recapture method to estimate deaths and injuries due to road traffic accidents in Karachi, Pakistan.41 Standard A represents a perfect record match. These are progressively relaxed in standards B-D (tables 1 and 2). The formula used to calculate the ascertainment corrected number of TRIs is shown in the equation.

\[ N = \frac{(x + 1)(y + 1)}{(z + 1)} - 1 \]

Where \( x \) is the number of cases in database one, \( y \) is the number of cases in database two, and \( z \) is the number of cases common to both databases.

The estimated completeness of each database was calculated by dividing the number of injury events in each database by the ascertainment corrected number (calculated using standard D).

In the case of both fatalities and serious TRIs, relaxing the matching to standard D was considered defensible given the data collection circumstances. It is possible for a police reported injury to have occurred on the day before the hospital admission is recorded, particularly for injuries occurring in the evening. Moreover, an estimated age is frequently entered at the scene of a motor vehicle accident. Thus it is feasible to include matches where age differs by at least one year. Results for standards A-C are also presented to demonstrate the range of possible results by progressively relaxing the matching criteria. Even standard D may be regarded as fairly stringent. It does not allow for coding or computerisation errors, or an error in age estimation of greater or less than one year.

### Results

#### FATAL TRIS

A total of 97 fatal TRIs were identified in the Scottish Health Service data, representing an age specific mortality rate of 14/100 000. An identical search of the STATS19 database identified 99 fatal TRIs. This represents an age specific mortality rate of 15/100 000 for people aged 15–24 years. There were no significant differences in the demographic characteristics of the subjects between the databases. In both databases males accounted for 72% of fatalities, and in the mean age of those fatally injured was 20.2 years in the Scottish Health Service dataset and 20.3 years in the STATS19 dataset.

Matching fatalities using standard D generated 92 fatalities common to both data sets (table 3). Using the aforementioned formula the ascertainment corrected number of TRI fatalities was 104. This generated an ascertainment corrected mortality rate of 15/100 000 people aged 15–24 years due to TRIs. The estimated completeness of the Scottish Health Service mortality data was 93%. The estimated completeness of the STATS19 database for fatalities was 95%. Stricter matching criteria (standards A-C) generated ascertainment corrected numbers of TRIs ranging from 108–139.

#### NON-FATAL TRIS

A search for non-fatal, hospitalised injuries identified 1458 cases in the Scottish Health Service data, representing an age specific hospital discharge rate of 215/100 000 for young people aged 15–24 years (table 4). A total of 1290 records of “serious” (that is, injuries requiring hospitalisation) TRI casualties were found in the STATS19 database. This represents an age specific rate of 190/100 000 people aged 15–24 years. No significant differences were found in the demographic characteristics of the subjects between the databases. In both databases approximately 70% of casualties were male (69% and 72% respectively), and the mean age of those hospitalised was 19.5 years.

Matching cases using standard D produces an ascertainment corrected number of TRI related injuries of 1969. This generated an ascertainment corrected serious injury rate of 291/100 000 for young people aged 15–24 years due to TRIs. The estimated completeness of the SMR1 and STATS19 databases was
74% and 66% respectively. Stricter matching criteria (standards A–C) generated ascertainment corrected numbers of TRIs ranging from 2139–3052. Further analysis demonstrated that there was no significant difference in the mean age or sex of cases captured or not captured by both databases. However, analysis by injury type showed that cyclists appear to be under-represented in the police reported data.

Discussion

Most fatal TRI injuries were recorded by both sources, suggesting that mortality data from either source represents a reasonably complete dataset. Missing information in the Scottish Health Service data may be due to misclassification of the injury cause. Previous research shows that, even for a condition such as meningococcal disease, notifications may account for fewer than three quarters of cases. The police failed to record a small number of fatal TRI injuries in STATS19. A retrospective review of the police records revealed that one of these deaths was classified as a “serious” rather than a “fatal” injury. This case was identified in the SMRI hospital discharge database and had a hospital stay of greater than 30 days.

The completeness of information on non-fatal TRIs was lower for both sources, resulting in an ascertainment corrected number of non-fatal TRIs over 25% higher than the number available for analysis in each individual database. This is a similar rate of completeness to the routine hospital discharge statistics revealed in a survey of injuries in New Zealand. Again, non-fatal TRIs recorded by the police, but not the health service, may have been classified as “other” injuries. A number of injuries with relevant ICD codes are coded as “emergency—other” in the database (5% of all hospital discharges in this age group). Alternatively, the police may classify an injury as serious enough for hospital admission at the time of the accident, but the casualty may not be admitted to hospital. Conversely, some injuries recorded as “slight” by the police may have resulted in hospital admission. A capture-recapture analysis including non-hospitalised TRIs was not possible because there is no matchable source of population based health data.

Some TRI incidents resulting in non-fatal injuries had not been attended by the police. The STATS19 documentation suggests that an appreciable number of non-fatal TRIs are not reported to the police. Several explanations have been offered for such under-reporting. Casualties may not report an accident because they do not consider it serious enough or because insurance cover is inadequate. Another explanation (perhaps particularly pertinent to young drivers) may be that casualties, or their fellow passengers, are unlikely to call for police assistance if the driver has consumed alcohol or drugs. Reasons cited for the disproportionate number of young people injured in TRIs include their propensity to drive under the influence of alcohol and/or drugs.

There are a number of drawbacks with the use of the capture-recapture in human populations. A recent study of children with a serious injury resulting from a motor vehicle accident (applying mark-recapture) concluded that the study violated most of the requirements of the technique. The ascertainment corrected number generated in this study may be an overestimate because the population may not be “closed”. An underlying assumption of the technique is that it is applied to a “closed population”. The population of 15–24 year olds in Scotland is likely to be transient. Moreover the cases matched in each database are not homogenous to those unmatched. For instance, cyclist injuries are undercounted in the STATS19 database. Under-representation of pedestrian and cyclist injuries has been reported elsewhere. Although not specifically addressed in this study, a severity effect has been observed in other studies. The more serious the injury, the more likely it is that it will appear in a database.

A further drawback is inaccuracy based upon the relationship between data sets. An underlying assumption of capture-recapture is that the two sources are relatively independent of each other. Both a positive and negative dependency can lead to inaccuracy. Positive dependence can lead to underestimation, and negative dependence can lead to overestimation. Since the two information sources in the present study are not directly related, the ascertainment corrected number of non-fatal injuries may be an overestimate. One system does not refer individuals to the other. However, this does not discount completely some association. The police record is completed at the scene of the injury, often with the health services present. The hospital record is completed on discharge or transfer to another hospital.

The quality of the input data is also crucial to the application of capture-recapture.
relatively high quality of the input data and the discriminatory nature of available variables were fundamental to the capture-recapture technique in this study. Optimal data sources with high rates of variable completion were used in the study. The existence of a location code in both data sets was essential for matching cases of serious (but non-fatal) injury. Without this variable there were a number of cases that could have been matched to more than one record. Consideration must also be given to the data collection setting before deciding upon matching standards. In this study it was deemed justifiable to relax the matching criteria given the context of data collection.

In conclusion, the capture-recapture technique has been extensively used in other research areas, but remains an under-exploited tool in injury prevention research. This study has shown that capture-recapture method may be a useful tool for demonstrating the completeness of the data sources in Scotland and indicating where biases in capture exist within sources. This is vital contextual information in the analysis, interpretation, and reporting of the data. Accurate estimates of injury rates are vital for effective prevention. For instance, the study demonstrates that any analysis of cyclist injuries should not rely on police reported data. However, the capture-recapture technique in this human study of TRI injury did not achieve all the requirements of the approach. While we do not feel this renders the results invaluable, the ascertainment corrected rates should be viewed with caution.

We thank the staff at the Information and Statistics Division of the Scottish Health Service (particularly Adam Redpath and Martin Krievs), and at the Data Archive at the University of Essex for supplying the data used in this study. These agencies bear no responsibility for the content of this paper.

References

Carbon monoxide kills six
Six people were killed while they slept when an air conditioner absorbed carbon monoxide from a nearby furnace and pumped it throughout the house in New York State. An alarm that could have detected the poisonous gas had been turned off. Andrei Kranz discovered the bodies of his parents, his 2 year old daughter, two house guests, and a baby sitter. Kranz later told police he had disconnected his home’s carbon monoxide alarm because it buzzed repeatedly and he thought it was malfunctioning. The air conditioner had one intake vent closed and another clogged with leaves, so it could draw air only from the room it shared with the natural gas furnace, investigators said. The temperature in the neighborhood was about 60°F overnight.

Man fearing gators tapes himself to tree
A man who became lost in a Florida swamp during a trip to photograph alligators was rescued by police after taping himself high up in a tree to ensure the reptiles didn’t attack him while he slept. Gemini Wink, 26, of Louisville, Kentucky, was still taped to a tree limb 40 feet (12 m) off the ground when Hillsborough County deputies found him late Saturday night, sheriff’s deputies said on Wednesday. Wink, who was visiting friends in Tampa, set out on his alligator trek using duct tape to mark his trail. After taking shots of an alligator Wink realized he was lost and found himself in waist deep water with night falling. Fearing an alligator attack, Wink climbed a tree, secured himself with duct tape and resolved to sleep there.

Boys shocked by 13 000 volts
Associated Press reported that two boys remained hospitalized a day after being shocked by 13 000 volts as they climbed atop a power station. The boys were part of a group of seven teenagers who went swimming in the Hudson River and then climbed over two fences and scaled the power transfer station, according to Colonie police. One of the two grabbed a large, cylindrical insulator and was immediately shocked. The beam the boys were standing on became charged and shocked the other boy, throwing both about 20 feet to the ground, police said. One 14 year old was the more seriously injured, receiving second and third degree burns to his hands and arms, a head injury, and possible internal injuries. He was taken first to Albany Medical Center Hospital then transferred to Massachusetts General Hospital in Boston where he remained in serious condition today, according to a hospital spokeswoman. An investigation is continuing and the youths could face trespassing charges, police said.

Pet protection
Under a new US law, pets will be better protected when they travel on airplanes. As part of an aviation bill, airlines will have to submit monthly reports to the federal Department of Transportation when animals are lost, injured or killed, and will have to notify passengers about the conditions under which animals travel. Employee training will also have to be improved with respect to transporting animals.

Infant dies in car crash after airbag crushes her
An 11 month old girl was killed in a two car accident in North Dallas on Thursday afternoon when the driver’s side airbag deployed and crushed the child against the driver, according to a police report. The child was standing in the lap of the driver, whose name was not released, when the car ran a red light at Alpha Road and the Dallas Parkway and collided with another vehicle. The car spun out of control, hitting the median and a light pole on Alpha Road. The child was taken to the Children’s Medical Center in Dallas where she was pronounced dead. Police had not determined late Thursday whether the driver would face any charges (Dallas, 14 April 2000).
Capture-recapture: a useful methodological tool for counting traffic related injuries?

A Morrison and D H Stone

doi: 10.1136/ip.6.4.299

Updated information and services can be found at:
http://injuryprevention.bmj.com/content/6/4/299

These include:

**References**

This article cites 36 articles, 19 of which you can access for free at:
http://injuryprevention.bmj.com/content/6/4/299#BIBL

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**

Articles on similar topics can be found in the following collections

Epidemiologic studies (842)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/