Priorities for developing and evaluating data quality characteristics of road crash data in Australia

Angela Watson[1], Kirsten McKenzie[2], & Barry Watson[1]

[1]Centre for Accident Research and Road Safety-Queensland, Queensland University of Technology

[2]National Centre for Health Information Research & Training, Queensland University of Technology

Abstract

The National Road Safety Strategy 2011-2020 outlines plans to reduce the burden of road trauma via improvements and interventions relating to safe roads, safe speeds, safe vehicles, and safe people. It also highlights that a key aspect in achieving these goals is the availability of comprehensive data on the issue. The use of data is essential so that more in-depth epidemiologic studies of risk can be conducted as well as to allow effective evaluation of road safety interventions and programs. Before utilising data to evaluate the efficacy of prevention programs it is important for a systematic evaluation of the quality of underlying data sources to be undertaken to ensure any trends which are identified reflect true estimates rather than spurious data effects. However, there has been little scientific work specifically focused on establishing core data quality characteristics pertinent to the road safety field and limited work undertaken to develop methods for evaluating data sources according to these core characteristics. There are a variety of data sources in which traffic-related incidents and resulting injuries are recorded, which are collected for a variety of defined purposes. These include police reports, transport safety databases, emergency department data, hospital morbidity data and mortality data to name a few. However, as these data are collected for specific purposes, each of these data sources suffers from some limitations when seeking to gain a complete picture of the problem. Limitations of current data sources include: delays in data being available, lack of accurate and/or specific location information, and an underreporting of crashes involving particular road user groups such as cyclists. This paper proposes core data quality characteristics that could be used to systematically assess road crash data sources to provide a standardised approach for evaluating data quality in the road safety field. The potential for data linkage to qualitatively and quantitatively improve the quality and comprehensiveness of road crash data is also discussed.

Keywords: Crash data, data linkage.

Conference stream: Data and analysis.

Corresponding author:
Angela Watson
Phone: 31384570
Email: angela.watson@qut.edu.au

Introduction

Injuries resulting from transport-related incidents are a significant public health problem world-wide (WHO, 2004). It is predicted, that unless substantial gains are made in the prevention of crashes, transport-related injuries will become the third ranked global burden of disease and injury by 2020. In Australia, approximately 1600 people are killed on our roads each year. On average, the economic cost of fatal crashes is estimated at $3.87 billion, with injury crashes costing $9.61 billion (BTRE, 2009). In order to reduce the burden of transport-related injuries, there is a need to fully understand the nature and contributing circumstances of crashes and the resulting injuries. The National Road Safety Strategy 2011-2020 (ATC, 2011) outlines plans to reduce the burden of road trauma via improvements and interventions relating to safe roads, safe speeds, safe vehicles, and safe people. It also highlights that a key aspect in achieving these goals is the availability of comprehensive data on the issue. The use of data is essential so that more in-depth epidemiologic studies of risk can be conducted as well as enabling effective evaluation of road safety interventions and programs.

Before utilising data to evaluate the efficacy of prevention programs it is important for a systematic evaluation of the quality of underlying data sources to be undertaken to ensure any trends which are identified reflect true estimates rather than spurious data effects. However, there has been little scientific work specifically focused on establishing core data quality characteristics pertinent to the road safety field and limited work undertaken to develop methods for evaluating data sources according to these core characteristics.

There are a variety of data sources in which transport-related incidents and resulting injuries are recorded. These include police reports, emergency department data, hospital morbidity data, and ambulance data. However, as these data are collected for specific purposes, each suffers from some limitations when seeking to gain a complete picture of the problem. It is generally considered that no single data source is sufficient to examine the issue effectively and as a result, there is increasing interest in data linkage as a possible solution.

However, each agency and jurisdiction has different data systems with unique considerations for linkage and use. If the ultimate aim is to create an integrated national data linkage system (as researchers in the area suggest [Austroads, 2005; Holman, et al., 2008; Turner, 2008]), then it is important to understand the nature of each jurisdiction's information systems and data linkage capabilities. Given the lack of standardisation of data sources, legislation, and data linkage progress, work needs to first be undertaken at an individual jurisdiction level to inform a national (and potentially international) approach.

The aim of this paper is to outline core data quality characteristics pertinent to the road safety field that can be used to assess road crash data sources to provide a standardised approach for evaluating data quality in the road safety field. The potential for data linkage to qualitatively and quantitatively improve the quality and comprehensiveness of road crash data will also be discussed.

Framework for assessing data

To determine if a data source is capable of providing good quality information an assessment is required on any limitations of the collection in relation to its capacity to report on injury. It is also necessary to determine how these limitations may affect the accuracy and validity of conclusions that are able to be drawn from the data (Horan & Mallonee, 2003; Mitchell, Williamson, & O'Connor, 2009; WHO, 2001).

There are a variety of frameworks and guidelines with which data related to injury can be assessed, however to date these haven't been systematically defined in regards to the road safety field (e.g., ABS, 2009; Austroads, 1997; Haddon, 1970; Mitchell et al., 2009, NHTSA, 1998; WHO, 2001). For the purposes of this review, data will be discussed in terms of six core quality characteristics: relevance; completeness; accuracy; consistency; timeliness; and accessibility. These six key data quality characteristics or concepts are described below.

Relevance

Relevance is defined as how well the data meets the needs of users in terms of what is measured, and which population is represented. Relevance is important in order to assess whether the data meets the needs of policy-makers and researchers and must be useful for planning and evaluation purposes (ABS, 2009; ATC, 2011). The needs of different data users are diverse, and what one considers 'relevant' may differ from another user's view. This means that within each record, a wide range of data items is usually needed.

Mitchell et al. (2009) discusses the term usefulness, which is a characteristic that also relates to the relevance of a data collection. Usefulness refers to the ability to: (a) identify new and/or emerging injury mechanisms; (b) monitor injury trends over time; and (c) describe key characteristics of the injured population (i.e. WHO's core minimum data set for injury surveillance).

In order to address the issue of relevance, the World Health Organisation's Injury Surveillance Guidelines recommend dividing injury surveillance data into two main categories (core and supplementary) with each of these then subdivided into 'minimum' and 'optional' data. The core minimum data set (core MDS) contains the least amount of data a viable surveillance system can collect on all injuries and the core optional data set (core ODS) involves information that is not necessary to collect but may be collected, if it is seen as useful and feasible to collect. It is also suggested that the core ODS include a narrative or summary of the incident.

Supplementary data includes any additional data that a surveillance system wishes to collect on specific types of injury such as those that are transport-related. In the case of transport-related injuries, information may include details about the circumstances of an incident (e.g., speeding, fatigue) or about other people involved (even if not injured).

Another issue related to relevance is that of representativeness. In other words, to what extent the data collection represents the population of all transport-related injuries or incidents (Mitchell et al., 2009). In order to draw conclusions on the incidence and distribution of transport-related injury, the data collection would need to include all of these injuries regardless of the type of injury, where the injury occurred, or who was injured. Non-representative data may focus prevention efforts on populations that are not truly at risk and could result in a misdirection of resources (Mitchell et al., 2009).

Most data collections do not include all transport-related injuries, instead only including those that fit a particular scope that is relevant for the collection's purpose. For example, hospital admissions data would only include those transport-related injuries that were serious enough to involve admission to hospital. Data collections based on police reported incidents would also not be representative of the entire injury population, as certain transport-related injuries do not fit the definition for inclusion in these collections (e.g., if the injury does not occur on a public road).

Completeness

Strongly related to the issue of relevance is completeness. Completeness refers to the extent to which all relevant cases, all relevant variables, and all data on a relevant variable are included in the data collection (Mitchell et al., 2009). Firstly, data collections would be considered complete if they detect all cases of transport-related injury they intend to detect by definition (sensitivity) and unlikely to detect those injury events they do not intend to detect (specificity). Mitchell et al. (2009) suggest that if between 76% and 100% of the Core MDS and ODS were included in a data collection, it would rate as 'very high'.

Also, not only should the collection include variables relating to the Core MDS and/or Core ODS, these variables should have minimal missing and/or unknown data for them to be considered complete. Mitchell et al. (2009) suggest that a 'high' level of completeness would exist if less than 5% of data within a specific field is missing. In addition to missing or unknown data, a data collection can lack completeness if there are a large number of unspecified or 'other' specified classifications (Mitchell et al., 2009). Incomplete data can be due to a lack of detailed information required to assign a code or classification, a lack of appropriate codes or classifications, lack of time, or lack of skilled coders (Mitchell et al., 2009; NHTSA, 1998). The impact of incomplete data is that the data collection may not provide enough information to allow for adequate data interpretation and could lead to flawed or biased results and therefore decision making.

Accuracy

Accuracy refers to the degree to which data correctly describe the events or persons they were designed to measure (ABS, 2009). Transport-related injury data need to be accurate in several ways, some specific to a location, and others more general. Location information for engineering purposes demands a very high degree of accuracy (within metres), which is frequently not met (Austroads, 2005; Strauss & Lentz, 2009). If location information is not accurate, a problem location might go undetected, and the nature of a location-specific problem might be difficult to determine due to incomplete data.

One of the main indicators of the safety and operation of the road system is the occurrence of transport-related incidents at different levels of severity. Accurate severity information is important for prioritisation of locations, understanding transport-related incident mechanisms, and for evaluating the effectiveness of interventions or countermeasures. Importantly, police are not necessarily in the best position to judge injury severity, at the point of collection of roadside injury information, with injury severity traditionally defined and measured more comprehensively in the clinical setting.

The accuracy of a data collection, and the variable fields within them, is difficult to assess as there is often no real comprehensive or objective data by which to compare the data to a gold standard. However, the literature does suggest that accuracy may be assessed by determining if certain aspects known to enhance the accuracy of data, such as: standardised coding and/or classification (e.g., ICD, AIS); quality control procedures; and the use of technology (GPS), are present (Mitchell et al., 2009; NHTSA, 1998).

Consistency

Consistency of data refers to their ability to reliably monitor transport-related injuries over time, and compare between characteristics within a data set as well as across other relevant data (ABS, 2009). Ideally, the quality of the data should not vary over time, nor should they vary in quality, by the nature of the event/injury, where or when the event/injury occurred, or who was injured or involved. Essentially, users of the data need to be confident that any changes over time or differences between events/individuals are due to actual changes or differences, not simply due to inconsistencies in the data (NHTSA, 1998; WHO, 2001).

Inconsistencies in the data based on the characteristics of the incident or injury can also occur for a variety of reasons. Firstly, reporting policy, work practices, or coding/classification systems may vary by the location of the incident/injury. An incident occurring in a remote location may not be reported, or a lack of resources in some hospitals may lead to less detailed classification. Besides the location of the incident, certain types of incidents/injuries may be less likely to be reported or coded/classified accurately or adequately. For example, a transport-related incident involving illegal behaviour (e.g., unlicensed driving, alcohol) may not be reported to police to avoid prosecution.

One suggested way of enhancing the consistency of a data collection is the use of uniform classification systems (Mitchell et al., 2009; NHTSA, 1998; WHO, 2001). These systems should include a comprehensive set of standard coding/classification guidelines which should be readily available to personnel assigned the duty of recording, classifying or coding data collections. These personnel should also be specifically trained in the procedures and should refer to the guidelines often. Without this training and available material, personnel could base their coding or classification decisions on their own intuitions, opinions, or preconceived notions (CDC, 2001). It is also necessary that any changes to reporting, classification, and recording should be documented in detail (NHTSA, 1998).

Timeliness

Timeliness refers to the delay between the date an event occurs and the date at which the data become available (ABS, 2009). It is suggested that data should become available for use quickly, however the definition of what is ‘quick’ may vary between agencies and dependent on the purpose for which the data are to be used (Austroads, 2005). It is crucial that agencies are able to respond rapidly to emerging problems, so that the rapid processing of transport-related incident data to make it available is a key concern. For example, Logan and McShane (2006) noted that clusters of crashes could develop quickly, in just a couple of years. Unless the data become available quickly, techniques aimed at detecting emerging clusters will not be effective. Data also needs to be timely for effective evaluations of countermeasures and interventions (NHTSA, 1998). Mitchell et al. (2009) rates the timeliness of the collection, availability, analysis and dissemination as being of high importance for injury data collections. Specifically, they suggest that if data are disseminated within a month the data collection would rate as ‘very high’; one to two years as ‘high’, and more than two years as ‘low’. The NHTSA (1998) suggest that it is preferable for data to be available within 90 days. However, they highlight that some supplemental information could wait longer.

The nature of some sources of data means that not all data items can be entered into the database at once; if the data items that have been completed are withheld until each crash record is complete, timeliness will be affected. For example, blood alcohol concentration (BAC) data cannot be entered until results of the toxicology analysis are made available.

Another factor that could influence the timeliness of data availability is related to resourcing. Specifically, an insufficient number of trained personnel to input, code, analyse and/or interpret the data will likely have a negative impact on the timeliness of the data. It is also the case that the roles of the personnel involved, particularly relating to inputting and coding data, are quite diverse (i.e., police officers, nurses), with their priorities directed toward other, arguably more important, tasks (e.g., patient care). This demand on resources can increase the time taken for data to become available.

There are also trade-offs between the timeliness of the data collected and the level of detail recorded regarding a case, as well as the accuracy, completeness and consistency of the data. While the processes that may be in place for coding, recoding, checking, and cleaning of data improve the consistency and accuracy, it may also then increase the time taken for the data to become available, therefore reducing timeliness.

Accessibility

Accessibility relates to the ease with which data can be accessed, which includes ascertaining its availability and suitability for the purpose at hand (ABS, 2009).

The NHTSA (1998) suggests that data should be readily and easily accessible to policy makers, law enforcement, and for use in road safety research and analysis. The NHTSA (1998) further suggest that data should be available electronically, at a unit record level, provided that safeguards are in place to protect confidentiality and privacy. Mitchell et al. (2009) suggest that if data is accessible to users in unit record format from an internet-based interface or data warehouse, it would rates as ‘very high’ on accessibility. While it may be ideal to have free and easily accessible data, there are a number of issues that can limit accessibility.

Major barriers to accessing data relate to confidentiality and privacy. Even when names and addresses are removed, there is still concern that variables such as age and gender in combination with location and temporal variables can lead to the identification of the person/s involved. Information collected and stored by various government agencies are covered by federal and state privacy legislation. These government agencies may also have their own legislation relating to the collecting, storing and access to data. Due to these legislative requirements, there are stringent processes in place in order to access data.

Legislation, policy, and guidelines can be open to interpretation which can complicate the process of negotiating access with different agencies. Therefore, negotiation processes can be protracted where legislation, policy and guidelines are unclear. Even if the process is straightforward, completing the required documentation and having it considered by the relevant authorities can still be quite time consuming.

Another potential barrier to access relates to the concern that data will be misinterpreted or misreported. This is particularly a concern when data custodians are not confident that end-users of the data are aware of the data constraints, limitations and coding conventions. This issue may potentially be overcome by end-users and data custodians communicating better about the nature of the data, including coding information, scope and limitations, as well as by discussing the reporting of data prior to its release or publication.

A third possible barrier to access lies with the data systems themselves. Some data sets do not have relevant information in a format that is easily quantifiable. For example, data systems which compile long text descriptions or reports make extraction of specific information about an incident or its location difficult and time consuming. Even in the case of data being held in a suitable format, the software used may be difficult to navigate, except for those who are specifically trained. Data may not be easily extracted and exported into a format conventionally used by those who work with data (i.e. Excel, text delimited, SPSS, or Access).

Police collected data

At present, a primary source of data used for transport-related incidents is police collected road crash data. While the exact nature of these data collections differ from one jurisdiction to another, generally they include all crashes that are reported to police, that occur on a road, and involve a death and/or injury or substantial property damage (e.g., vehicle is towed away). These crash records usually include details relating to the crash, casualty, unit, and controller.

There are potential limitations of police reported data related to the nature of the data source. It is possible that some crashes may not be included because they are not reported to the police. There has been research about the possible limitations of police reported data (Alsop & Langley, 2001; Boufous, Finch, Hayen, & Williamson, 2008; Langley, Dow, Stephenson, & Kypri, 2003). All of these studies found that some transport-related injuries were not recorded by the police, and reporting rates varied according to a number of factors including: age, injury severity, number of vehicles involved, road user type (e.g., cyclists), whether or not a collision occurred, and geographic region. The solution may not necessarily involve any changes to the processes of reporting to police. However, it does highlight that if police data is relied on as the sole data source for understanding transport-related crashes, without the use of other data (i.e., hospital data); there is a risk that certain causes of injuries will not receive the resourcing for intervention that is commensurate with the size of the problem.

Other data sources

There are a number of other sources of transport-related injury information collected in the health sector such as admitted patient data, emergency department data and ambulance data. The data are used for a number of purposes including examination of patterns of morbidity and mortality for population health research, patient tracking through services/departments, and enumeration of diagnostic case mixes health service funding and resource allocation. While the nature of the information collected varies with each collection and across jurisdictions, the data generally include: the time and date of treatment, the nature of the injury, whether the injury was sustained via traffic or a non-traffic event, and some details about the nature of the event (including information about the mode of transport of the injured person, the mode of transport of the counterpart vehicle involved and whether the injured person was a passenger or a driver), and patient outcomes (such as length of stay, mode of separation etc).

Perhaps the biggest limitation of this sort of data is that only transport-related incidents that involve attendance or admission to hospital, or those in which an ambulance was called are included in the data collections. Some injured persons involved in transport-related incidents may not present at hospital or call an ambulance but instead attend a medical clinic for treatment. It is also possible that an injury resulting from a transport-related incident could be attributed to some other cause, as the information on the cause of an injury can be falsely reported by the patient, poorly documented by the clinical staff and/or incorrectly coded after discharge.

It should also be noted that as the primary purpose of the data collection is not for road safety research, there are other important information pertinent to the road safety field which are not included (e.g., contributing factors such as alcohol involvement, speeding, fatigue etc.). The emphasis in these data-sets is on health-specific information such as the nature of the injury, length of hospital stay and the treatment outcomes. There may be very little, and in some cases no information, regarding the location of the incident.

Based on the various purposes of these data and their potential limitations, it is generally considered that no single data source is sufficient to examine the issue of transport-related incidents and resulting injuries effectively. As a result, there is increasing interest in data linkage as a possible solution to enable a more complete understanding of the issues surrounding transport incidents and the injuries resulting from such incidents.

Data linkage

Data linkage involves the bringing together of two or more different data sources that relate to the same individual or event (NCRIS, 2008). In principle, any datasets that contain information about individuals has the potential to be linked. There are two possible methods of data linkage: deterministic and probabilistic. The deterministic method involves the linking of data sets that share a unique identifier or key, while the probabilistic method matches cases based on certain elements of data that may lead to the identification of an event and/or person. It does this by matching cases based on other indentifying variables such as name, DOB, gender, and time and date of event (NCRIS, 2008).

Potential benefits of data linkage

There are a number of suggested benefits of using linked data for research, monitoring and policy development (Austroads, 2005; Glasson & Hussain, 2008; Goldacre, 2002; Holman et al., 2008). It is possible that data linkage can result in improvements to data quality by including more cases or variables and increasing accuracy through the detection and correction of errors. It is also argued that data linkage can be cost-effective. By linking pre-existing data to provide additional information and address research questions, there is less need to collect additional data on an ad-hoc basis which can be time consuming and expensive (Goldacre, 2002). A report by Austroads (2005) suggests that investment in linked data systems for road safety would likely lead to more efficient day-to-day operations and easier access to data for decision makers. It was suggested that the linking of databases will greatly increase the value of data sets by allowing the use of data for a wider range of purposes (Austroads, 2005).

Potential barriers to data linkage

The first major barrier relates to issues of privacy and confidentiality that are outlined previously. In order to conduct a record linkage project, a researcher needs to obtain approval from multiple data custodians and human research ethics committees. The time and effort involved in this process may discourage the frequent conduct of record linkage studies. It may also be necessary to involve an appropriate third party (or possibly one of the data custodians) in the data linkage process, as access to the identifying information required for data linkage is more restricted, if not prohibited, for researchers. It is important to note, however, that processes in order to provide linked data to researchers, while safe-guarding privacy, have been established in other Australian jurisdictions as well as overseas.

Another potential barrier is the linkage process itself. The deterministic method is the most accurate method; however it involves a unique identifier being matched across data sets. Unfortunately, in the case of the data sources discussed previously, though information in different data sets may relate to the same incident, person or case, there is no system of unique identifiers across all data sets. Also, in the case of the police data, the unique identifier is often assigned to an event (i.e., the crash), while the unique identifiers within health data sets are typically assigned to a patient.

As such, the probabilistic method is required for linkage of these datasets in the absence of a shared unique identifier. However, this method relies on having specific and accurate information on the relevant linkage variables in both data sets. This method requires that enough data points can be chosen for matching purposes so that no two events or individuals will be confused, leading to a lack of specificity. Conversely, if the data matching criteria is too specific, there is a potential for an individual to not be matched despite them actually being present in both data sets (i.e. lack of sensitivity). So although this method has been utilised in the past in other jurisdictions, a limitation is that the formats used with different data sets may not be compatible, resulting in an inability for some of the data sets to communicate with each other or leading to errors in matching.

Previous data linkage research

In terms of transport-related incidents and injuries, a variety of data linkage projects have been conducted (e.g., Alsop & Langley, 2001; Boufous, et al., 2008; Cercarelli, Rosman, & Ryan, 1996; Langley, et al., 2003). Alsop and Langley (2001) used probabilistic linkage of police and hospital records in New Zealand. They found that less than two-thirds of all hospitalised traffic crash casualties were recorded in the police data. They also found that this varied based on the number of vehicles involved, the geographical location, age and injury severity. Langley, et al. (2003) conducted probabilistic linkage between hospital records and police records to specifically examine the potential under-reporting of cyclist injuries in New Zealand. The results showed that only 22% of cyclists that crashed on a public road could be linked to the police records. Of the crashes that involved a motor vehicle 54% were recorded by police. They also found that age, ethnicity, and injury severity predicted whether a hospitalised cycle crash was more likely to be recorded in the police data. Within Australia, Cercarelli, et al. (1996) linked police reports, hospital admissions and accident and emergency (A&E) department data. The researchers found that around 50% of attendances at the A&E were recorded by police, and that around 50% of cases recorded by police as being admitted to hospital were actually admitted. The researchers outline that while the discrepancy between the data sets does represent an under-reporting of cases, it also suggests that differences in coding systems may also lead to cases not being linked. Another Australian study conducted in NSW by Boufous, et al. (2008) linked hospital admissions data (Inpatient Statistics Collection [ISC]) with the Traffic Accident Data System (TADS). Using probabilistic linkage, the researchers matched 56.2% of hospitalisations coded as being as a result of traffic crash with a record in TADS. The researchers also found that the linkage rate varied according to age (i.e., lower linkage rate for younger age groups), road user type (e.g., lower linkage rate for cyclists), severity (i.e., higher linkage rates with increased severity) and geographical location.

While these studies highlight the issues of under-reporting and bias within police data systems, the barriers and limitations of data linkage were not explored either at all, or in any depth, in any of the studies conducted to date. Also, many of these studies involved the ad-hoc linkage of data as opposed to routine data linkage. It is likely that routine data linkage may involve issues (e.g., changes to data systems, inter-agency agreements) that ad-hoc project based data linkage does not and vice versa. Each jurisdiction has different data systems with unique considerations for linkage and use. If the ultimate aim, as researchers in the area suggest (Austroads, 2005; Holman, et al., 2008; Turner, 2008), is to create an integrated national data linkage system, then it is important to understand the nature of each State and Territory’s information systems and data linkage capabilities.

Research priorities

In order to improve the quality, comprehensiveness, and usefulness of transport-related injury data, there are a number of suggested priorities for future research, including: scoping existing data collections in order to assess their completeness, consistency, accuracy, accessibility and relevance; determining the barriers to and facilitators of linking transport-related injury data; and assessing whether linked data provide qualitative and quantitative advantage over non-linked data. These priorities could be addressed by: discussions with data custodians, users, and other key stakeholders; reviewing legislative and policy documents; and analysis of sample data from current traffic injury data sources. While it is important to establish whether data linkage is feasible, it is also necessary to establish whether the benefits that would be derived from linked data would be sufficient to offset the likely costs. This could be achieved by piloting data linkage (including a comparison of linked data with non-linked data) and conducting cost- benefit analysis for both routine and ad-hoc data linkage.

Summary

Data is vital to informing policies and interventions designed to reduce the burden of road trauma. This paper proposes core data quality characteristics to enable the systematic assessment of road crash data sources to provide a standardised approach for evaluating data quality in the road safety field. It is possible that linkage of key data collections has the potential to overcome the limitations of single data sources and maximize the collective benefit of data relating to road trauma. However further research needs to establish whether road safety data linkage is feasible within each jurisdiction (given differences in data linkage capabilities across jurisdictions) and whether linked data provide advantage over non-linked data, both qualitatively and quantitatively.

References

Alsop, J. and Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis and Prevention*, 33, p.353-359.

Australian Bureau of Statistics (2009). *ABS Data Quality Framework, May 2009*. Australian Bureau of Statistics: Canberra. http://www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/1520.0Main%20Features1May%202009?opendocument&tabname=Summary&prodno=1520.0&issue=May%202009&num=&view=

Australian Transport Council. (2011). *The Draft National Road Safety Strategy*. Australian Transport Safety Bureau: Canberra. http://www.infrastructure.gov.au/roads/safety/national_road_safety_strategy/files/Draft_National_Road_Safety_Strategy_ext.pdf

Austroads (1997). *A Minimum Common Dataset for the Reporting of Crashes on Australian Roads.* Report No. AP-126/97, Austroads: Sydney.

Austroads (2005). *The Prospects for Integrated Road Safety Management in Australia: A National Overview.* Report No. AP-R280/05, Austroads: Sydney.

Boufous, S., Finch, C., Hayen, A., Williamson, A. (2008). The impact of environmental, vehicle and driver characteristics on injury severity in older drivers hospitalized as a result of a traffic crash. *Journal of Safety Research,* 39, p.65-72.

Bureau of Infrastructure, Transport and Regional Economics [BITRE] (2009). *Road crash costs in Australia 2006*, Report 118, Canberra, November.

Cercarelli, L., Rosman, D., and Ryan, G. (1996). Comparison of accident and emergency with police road injury data. *The Journal of Trauma*, 40(5), p.805-809.

Connelly, L. and Supangan, R. (2006). The economic costs of road traffic crashes: Australia, states and territories. Accident Analysis and Prevention, 38, p.1087-1093.

Goldacre, M. (2002). The value of linked data for policy development, strategic planning, clinical practice and public health: An international perspective. *Symposium on Health Data Linkage: Its value for Australian health policy development and policy relevant research,* March 2002, Potts Point, Sydney, New South Wales.

Glasson, E.J., and Hussain, R. (2008). Linked data: Opportunities and challenges in disability research. Journal of Intellectual and Developmental Disability, 33(4), p.285-291.

Haddon, W. Jr. (1970). A logical framework for categorizing highway safety phenomena and activity. Paper presented at the 10th International study Week in Traffic and Safety Engineering,Rotterdam, 7-11 September.

Holman, C.D., Bass, A.J., Rosman, D.L., Smith, M.B., Semmens, J.B., Glassson, et al. (2008). A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian Health Review*, 32(4), p. 766-777.

Horan, J.M., and Mallonee, S. (2003) Injury surveillance. *Epidemiology Review*, 25, p. 24-42.

Langley, J., Dow, N., Stephenson, S., Kypri, K. (2003). Missing Cyclists. *Injury Prevention*, 9, p. 376-379.

Logan, M. and McShane, P. (2006). Emerging crash trend analysis. Proceedings of the Australasian Road Safety Research, Policing and Education Conference, Brisbane, October 2006.

Mitchell, R., Williamson, A., and O'Connor, R. (2009). Development of an evaluation framework for injury surveillance systems. *BMC Public Health*, 9, p.260.

NHTSA (1998). *Traffic Records: A Highway Safety Program Advisory.* National Highway Traffic Safety Administration, http://www.nhtsa.dot.gov/people/perform/pdfs/Advisory.pdf

NCRIS (2008). *Strategic Roadmap for Australian Research Infrastructure*. NCRIS, Department of Innovation, Industry, Science and Research: Canberra. http://ncris.innovation.gov.au/Documents/2008_Roadmap.pdf

Strauss, T. and Lentz, J. (2009). *Spatial Scale of Clustering of Motor Vehicle Crash Types and Appropriate Countermeasures*. MTC Project 2007-10, Midwest Transportation Consortium, Iowa State University Institute for Transportation: Ames.

Turner, B. (2008). Review of best practice in road crash database and analysis system design. *Proceedings of the Australasian Road Safety Research, Policing and Education Conference,* Adelaide, November 2008.

WHO (2004). *The world health report*. World Health Organisation, Geneva, 2004.

WHO (2001). *Injury Surveillance Guidelines*. World Health Organisation, Geneva, 2001.