Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Police-hospital data linkage for traffic injury surveillance

Ali Soltani [1 2], James Harrison [1], Courtney Ryder [1]

Joanne Flavel [3], Angela Watson [4]

[1] Injury Studies, FHMRI, Bedford Park, SA, 5042, Australia

[2] UniSA Business, University of South Australia, SA, 5001, Australia

[3] Stretton Institute, University of Adelaide, SA, Australia

[4] School of Public Health & Social Work, QUT, Qld, Australia

Email for correspondence (presenting author): ali.soltani@flinders.edu.au

Abstract

This systematic review examined the linkage between police crash data and hospital data in both developed and developing nations context. Using inclusion and exclusion criteria, relevant studies were selected from PubMed, Web of Science, Google Scholar and Scopus (N=58) using the backwards and forwards snowballing technique to identify additional relevant papers from 1994 to 2023. The research examined data from different sources; road traffic injury and fatality risk variables, and employed statistical methods. Studies varied in chronological and geographical coverage, data type, aims, incident numbers, linking methodology and software tools, link rates, and most notable discoveries. Police crash records underreported pedestrian, pedal-cyclists, and socially-disadvantaged groups' injuries and overestimated clinically serious injuries, according to the review. The research demonstrated that the probabilistic record linkage is more popular, however, more uniform categorisation criteria such as the ICD (International Classification of Diseases) might improve injury data linking.
Keywords: Data linkage; underreporting; injuries; traffic crash; systematic review.

1. Background and context of data linkage

Data linking is merging data from diverse sources to better understand phenomena, and is commonly used in epidemiological studies to obtain an overall appreciation of factors impacting on populations (Bohensky et al., 2010). Data linking uses crashes, hospital admissions, and other sources to discover trends and patterns that might guide policies and measures to avoid future crashes (Amorim et al., 2014; Kamaluddin et al., 2019). Crash data includes vehicle, road user, and crash details. Hospital data, however, focuses on crash-related injuries and health care usage (Harrison et al., 2018). Crash-hospital data links would assist practice and research, being able to reveal road traffic crash and injury trends and high-risk categories and locales, which can influence road safety regulations and actions like mobile safety enforcement cameras (Paixão et al., 2015; Thorpe & Fawcett, 2012). Research on road safety efforts or risk factors and injury severity may employ linked data (Lujic et al., 2008). Evidence suggests data linkage has improved injury identification and targeting in five Brazilian state capitals by 87% (Abulatif et al., 2018), hospital-police data links increase data completeness and quality (Mandacaru et al., 2017), and can also assist in explaining crash

causes such vehicle speed, driver behaviour, road conditions, and alcohol or drug use (Hossain et al., 2023).
Understanding road safety and its effects on public health requires linking police crash data with hospital crash admissions data. We risk underestimating road traffic crashes without this link. The absence of a linkage between these datasets makes it harder to assess injury trends, identify high-risk areas, and devise effective solutions. Thus, traffic management, infrastructure improvements, and healthcare resource allocation may be made without sufficient evidence/knowledge. The linkage of these datasets is not only about better data; it's about generating evidence-based policies to improve road crash safety and reduce the human and economic toll of road crashes.
Severity misclassification and database anomalies make data linkage challenging, especially with traffic data (Imprialou & Quddus, 2019). Data linking may reduce these misclassifications and improve road safety, reducing hospital admissions, medical expenses, and missed productivity. Linked data can influence policies and initiatives to reduce crash-related injuries, evaluate road safety programmes, and identify high-risk groups and areas for targeted interventions (Bowman & Stevens, 2004; Urie et al., 2016). Underreporting crashes and injuries as is common when using crash data alone can cause biases that can be addressed when utilising linked data (Watson et al., 2015).
This study seeks to undertake a thorough comparison of studies that have used the linkage of police crash data with hospital registry data. The research aims to shed light on a number of aspects, including study background, methodology, and data sources. This aims to identify trends in motor vehicle crash injuries and evaluate the impact of demographic, regional, and temporal variables on the linkage between police crash data and hospital data. The research questions are below:

- How is the identification and categorisation of injury patterns among people involved in motor vehicle crashes facilitated by the linkage between police-reported crash data and hospital registry data?
- When using hospital registry data and police crash data, how do past studies differ from one another when taking into account factors like data sources, methodology, software tools, research settings, and temporal coverage?
- What are the main variables that either increase or decrease the likelihood of a successful linkage between hospital registry data and police crash data?

2. Methodology

We conducted a systematic literature review on the topic of data linkage (between road crash data and hospital admission data) using the snowballing method as a search strategy. We started by selecting a "start set" of a few articles related to data linkage as our initial pool of papers. The selection of these papers was based on the research team's personal knowledge and information. We then conducted backward snowballing to identify additional relevant papers. We examined the reference lists of the papers in the start set and excluded papers that did not meet our basic inclusion criteria. We also removed papers that had already been examined based on being found earlier in the search. The remaining papers were considered as potential candidates for inclusion in our review and were further evaluated through abstract and full-text review.
Using this method, we identified a total of 58 papers on the topic of data linkage. These papers were included in our systematic review and were analysed according to our predefined inclusion and exclusion criteria. In particular, we followed the guidelines suggested by Greenhalgh and

Peacock (2005) and Vassar et al. (2016) in order to ensure that our search methodology was rigorous and systematic. The criteria of paper inclusion were:

1. Only those papers were considered that align with the scope of the research; in fact, only those studies with clear explanation of data linkage process were selected.
2. The conference papers, non-peer reviewed papers, books, book chapters, governmental or consultancy reports, research report, working papers, theses and dissertations and grey literature were all excluded.
3. Only papers written in English published from 1994 (the year that the first journal article with full details of linkage process was published) to 2023 were included.

3. Comparative review of the selected studies

Study period and characteristics

Selected papers include works from 1994 (the earliest study in the selected papers set, Rosman and Knuiman, 1994) to 2023 with diverse study themes and situations. The oldest study, Rosman and Knuiman (1994), included 18,544 participants, whereas Schneider et al. (2023) had 375. The earliest research conducted in Western Australia revealed a high degree of concordance between the two sources regarding crash type and road user classification; however, police-reported injury severity levels were found to be less reliable. Using linked police and emergency department data, the most recent data from Milwaukee, United States, was successful in improving the estimation of cyclist injury.

Variability in study duration and overlapping periods

It is possible to see that the time periods of the studies range from as early as 1991 to as recent as 2019-2023. Study duration varied, with some being one or two years, while others last longer. The 2000-2004 Wilson et al. (2012) and 2000-2009 Weijermars et al. (2016) investigations overlap. Couto et al. (2016) did not provide a time period. Reurings et al. (2011) with a time period of 15 years, from 1993 to 2008; Weijermars et al. (2016) with a time period of 10 years, from 2000 to 2009; and Lujic et al. (2008) with a time period of 8 years, from 1999 to 2006 are the longest-term studies. These studies reveal long-term traffic safety trends.

Global reach and data sources

The set of selected papers shows worldwide interest in road safety research and its effects on policy and practice. Studies ranged in location across 25 nations, mostly in high-income countries (n= 41 papers) such as the US, Australia, Canada, and France, some middle-income countries (n=9 papers) such as Brazil, China, Malaysia, and low-income countries (n=3 papers), such as Ghana, India, and Pakistan. Data linking programs can grow to encompass bigger regions within a country or whole nations, providing academics and policymakers with more extensive and diversified information. Australia and the US have national data linking systems that standardise data analysis across populations. The distribution of studies across income categories reflects each country's research resources, but income classification varies by source and year, and some studies were conducted in regions with different income levels. This significant gap between high-income countries and the rest of the world, highlights the potential for health inequities across study designs due to variations in income classification and geographical location. This can result in disparities in health outcomes between different income groups or regions, as well as biases in the study findings.

Importance of standardised coding

The WHO designed the ICD to standardise diagnostic coding, universal health data reporting, and cross-country health outcome comparison (Harrison et al., 2021). It enhances healthcare billing, resource allocation, and cost-cutting (De Coster et al., 2016; Hong and Zeng, 2023). ICD codes standardise data and allow trustworthy and genuine study findings. They also discover new health trends and patterns that traditional data collection and analysis may have overlooked, 55% of research employed the ICD standard, 45% did not.

Sample sizes and data sources

Curry et al. (2019), Couto et al. (2016), and Brubacher et al. (2019) had the most cases included with the size of 10,352,998; 40,000 and 15,022 respectively. Salifu & Ackaah (2012), Lopez et al. (2000), and Cryer et al. (2001) have the lowest participant counts. Finally, Boufous and Williamson (2006) with 13,124 participants, Lujic et al. (2008) with 17,552 participants, and Mandacaru et al. (2017) with 4,500 participants have the median of the included records in the selected papers.

A range of hospital records, discharge data, emergency department data, patient care reports, trauma registries, and death data are utilised as the medical data most often. Previous research included workers compensation, insurance, birth, death, and injury monitoring data (Boufous and Williamson, 2006; Santiago et al., 2020). EMS and trauma registry data have supplemented hospital admission data in many studies (Amoros et al., 2007; Cirera et al., 2001; Lopez et al., 2000; Sciortino et al., 2005; Stutts and Hunter, 1999). The studies share a desire to create comprehensive and accurate healthcare data systems to aid injury prevention.

Metrics and measures of road traffic crashes

Previous road traffic crash studies utilised a variety of criteria based on their objectives. Mortality rates, injury severity indicators, hospital admissions and expenses, crash rates and characteristics, and crash data completeness and correctness are employed (Cirera et al., 2001; Dandona et al., 2008). Using crash conditions and risk variables, several studies have focused on crash types or demographics (Alzaffin et al., 2023; Amoros et al., 2007; Ferreira et al., 2015). Due to the variety of road traffic crashes and study topics, selecting the most common measures is challenging. These studies demonstrate road traffic crash complexity with their many measures.

Diverse data linkage approaches

Data linking approaches include probabilistic, deterministic, manual, capture-recapture, and customised. Probabilistic methods are most commonly used, with Moore (1998), Alsop and Langley (2001), Langley et al. (2003), Benavente et al. (2006), Boufous and Williamson (2006), Gonzalez et al. (2006), Wilson et al. (2012), Mitchell et al. (2015), Paixao et al. (2015), Short and Caulfield (2016), Edwards & Gutierrez (2023), Mandacaru et al. (2017), Brubacher (2019), and Soltani et al. (2022) using this method. Watson et al. (2015), Amorim et al. (2014), and Jeffrey (2009) employed both deterministic and probabilistic approaches. Stutts and Hunter (1999), Cryer et al. (2001), Tin (2013), Kamaluddin (2019), and Tainter (2020) employed deterministic approaches. Amoros et al. (2007) utilised semi-automated record-linkage, while Hosseinzadeh et al. (2022) used the heuristic framework. Some research utilised a customised technique (Lai et al., 2006), a rule-based linkage (Tainter, 2020), a capture-recapture method (Reurings, 2011; Lateef, 2010; Weijermars, 2016), or a stepwise algorithm (Hosseinzadeh & Kluger, 2021).

Software tools and statistical analysis

The details of the selected papers show that LinkSolv and Linkage Wiz were utilised most in the research. Automatch, SPSS, SAS, Microsoft SQL, MapInfo, manual, open-source R, and self-written programmes are also utilised. Chi-square (14 studies), Logistic regression (five studies), Cramer's V (three studies), Ordered probit regression (two studies), and Poisson regression (two studies) are the most often utilised tests.

Linkage rate and data completeness

Linkage rate is the percentage of records linked across data sources to produce a more complete database. Police and hospital data have been shown to under-estimated or biasestimated. In their population-based research, Dandona et al. (2008) reported 22% under-reporting of fatal RTI, whereas Cryer et al. (2001) found that manual linkage to hospital admissions data in pedestrian crashes was 72% successful. Amorim et al. (2014) and Lai et al. (2006) found 98% and 96% linkage rates, respectively. Linkage rates vary by injury type and mode of travel. Rosman & Knuiman (1994) reported a 64% hospital-to-police link rate, although it varied by crash type. Record linkage studies employ varied terminology to describe the matching procedure, which might confuse outcome comparisons.

The differences in linkage rates across studies can be related to the methods employed for linking, the types of data being linked, and the presence of sufficient data items. Probabilistic methods affect linkage rates differently than deterministic methods, with probabilistic methods offering more flexibility and accommodating errors. The nature of records, such as differences in data collecting and data granularity between police crash and hospital admission records, can be difficult. Successful matches can be strongly influenced by the availability of unique IDs or dependable matching criteria, as well as data quality and consistency. Contextual factors, population features, and variations in data recording practises affect linkage rates. Variations in linkage terminology and definitions might cause misunderstanding and inconsistency when comparing outcomes across studies.

4. Discussion

Answering research questions

The linking of police-reported crash data with hospital registry data facilitates the identification and classification of patterns pertaining to injuries resulting from motor vehicle crashes. Previous research has indicated that the data supplied by the police tend to underestimate the number of injuries resulting from road traffic crashes (Short & Caulfield, 2016; Tin et al., 2013). The linkage of hospital and EMS data enhances precision and facilitates the identification of injury patterns. Probabilistic record linking has emerged as a promising approach to enhance the efficiency and reliability of merging diverse data sources, despite the presence of certain constraints (Sciortino et al., 2005; Santiago et al., 2020). Research comparing police reports to hospital data has demonstrated that the implementation of uniform diagnosis and injury severity coding enhances the accuracy and effectiveness of injury evaluation. Nevertheless, the utilisation of language and reporting practises may impede the process of establishing connections and doing matching research (Moore et al., 2016).

Hospital-crash data linking is employed in many studies to investigate road traffic injuries. The findings and insights of a study might be influenced by various factors, such as variations in study contexts, data sources, techniques, software tools, and temporal coverage. Research highlights the need of adopting tailored methodologies that are aligned with specific study issues and the accessibility of data. The empirical evidence suggests that the linkage of several data sources and methodology is crucial for the correct documentation of road traffic injury occurrences, despite the ongoing challenge of selecting an acceptable linkage approach. In addition, the implementation of uniform reporting practises enables researchers to facilitate comparisons between studies and establish optimal methodologies for enhancing healthcare outcomes and maximising the utilisation of data in global data linkage projects (Moore et al., 2016).

The establishment of a solid link between hospital registry data and police crash data is contingent upon various elements. The linking rates are influenced by the quality, consistency, and completeness of each dataset. The utilisation of unique IDs, precise matching criteria, and standardised categorization systems such as ICD enhances the precision of linking. The selection of appropriate methodologies, ranging from probabilistic to deterministic approaches, plays a critical role in achieving successful data connection. Variations in injury rates can be attributed to the complexity of road traffic crashes, the inclusion of specific data elements, and contextual factors such as injury categories and mode of travel. The establishment of a successful connection between various data sources is a significant opportunity for the advancement of modern data science and data mining techniques. This has the potential to greatly transform our comprehension of traffic crashes and provide valuable insights for the development of worldwide road safety regulations and interventions.

Insights into road traffic injuries require police crash data and hospital crash admissions data linkage. Combining hospital and EMS data has the potential to improve accuracy, but more research is needed. Police-reported injuries are often underestimated. Underreporting of pedestrian, motorcycle, and youth injuries raises concerns. Different data linkage methods offer benefits and drawbacks, making it difficult to choose the best course of action. Standardised diagnostic and injury severity classification improve injury assessment. Specific injuries and vulnerable road users require customised remedies. A global linking initiative reporting standard improves comparability. Modern data science methods, such machine learning and big data, can be used to investigate crash and injury causes and shape road safety policy. Developed countries like the US and EU have used data linkage for injury tracking and safety review, whereas developing nations face infrastructural and resource constraints but can gain from cooperation. Local data sources can improve global models and stakeholders' estimates in low- and middle-income countries. Data linkage has proven beneficial in some situations and has the potential to reduce traffic accidents and improve road safety worldwide.

Implications

The study findings highlight the limitations of relying solely on police-reported accident data for users of police data, such as law enforcement agencies and traffic safety authorities. A valuable opportunity to enhance the accuracy and completeness of injury information is provided by the linkage of hospital registry data. The study also emphasises the importance of implementing standardised diagnosis and injury severity classification, which permits enhanced injury assessment and more accurate insights.

Adopting standardised classification systems such as ICD enables healthcare professionals and hospital administrators who utilise hospital data to contribute to more accurate road traffic injury assessment. By aligning their data practises with law enforcement's, healthcare institutions can contribute to enhanced road safety policymaking and interventions. Users of linked records, including researchers, policymakers, and public health officials, can use the findings to enhance their understanding of injury patterns. Linked records provide a comprehensive perspective by integrating hospital and police data, mitigating the underreporting and bias inherent to individual datasets. In addition, the study's emphasis on variability in terms of methodology, data sources, and contextual factors highlights the importance of customising linkage approaches to the particular study topic and dataset characteristics. For hospital records of road injury cases for which no matching police record is found, the study's findings emphasise potential challenges in terms of data quality, consistency, and methodological choices. Users should be aware that factors such as absent or inconsistent identifiers, variations in data recording practises, and differences in terminology can impede successful linkage. Understanding the reasons for non-linking can guide efforts to improve data quality and inform strategies to enhance linkage success rates, even though not all records may link successfully.

5. Conclusion

This systematic review revealed methodological differences and findings on road traffic injury data reliability and accuracy. Despite variances in linking strategies, linkage success rates, geographical contexts, sample sizes, data sources, and analytical approaches, all analyses emphasise the importance of road safety issues. The research has shown that data linkage is essential for identifying risk variables for road traffic injuries and fatalities, highlighting the limitations of police crash records. The revelations of underreported pedestrian injuries and clinically serious injury underestimate compared to police records have major ramifications. According to Zhu et al. (2015), probabilistic record linking is preferred over deterministic techniques, demonstrating the potential to accommodate data defects for more accurate matching. The promise of standardised classification, such as ICD, in enhancing injury data linkage emerges. These findings underline the imperative to strengthen road trauma monitoring and road assessments, providing policymakers and public health professionals with useful insights for targeted strategies to reduce road injuries and fatalities. The fusion of data from multiple sources, including hospital records and EMS data, emerges as a key practise for rigorous evaluations. Using several data sources, one can find subtle trends and risk factors for informed actions and policymaking.

The conclusions of this analysis highlight the imperative to address rural road safety and cater to vulnerable road users, particularly youth, pedal cyclists, and motorcyclists. To reduce road traffic injuries worldwide, legislators and public health specialists should collaborate and create nuanced linkage schemes. The future requires addressing the intricacies of linking rates and regional distinctiveness, reconciling data gathering methods across law enforcement and healthcare, and standardising quality assessments. Data linking practises vary by country and location, especially in less researched economic contexts, necessitating in-depth research to develop flexible techniques. Equitable road safety strategies require such knowledge. Data linkage emerges as a powerful tool to improve road safety and reduce harm. Its potential can only be maximised by overcoming study constraints. Standardised data linkage practises and improved reporting are needed to help scientists replicate and politicians make evidence-based decisions. Reporting data linkage initiatives with explicit recommendations for data sources, matching criteria, and quality control would improve research integrity and the usefulness of data linkage literature findings. A concentrated effort to develop data linkage practises offers a future of data-driven road safety strategies.

The findings have significant implications for various stakeholders who rely on police data, hospital data, linked records, and non-linked records. By recognising the limitations and benefits of various data sources and linkage techniques, users can make more informed decisions, improve data quality, and contribute to a more precise injury assessment and road safety measures.

6. References

Abulatif, L. I., & Grupo Técnico de Parceiros do Projeto Vida no Trânsito (2018). Data Integation Process: an information management model for multiple databases on traffic crashes in Brazil. Processo de integração de dados - um modelo de gestão da informação para múltiplas bases de dados de acidentes de trânsito no Brasil. *Epidemiologia e servicos de saude : revista do Sistema Unico de Saude do Brasil*, *27*(2), e2017160. https://doi.org/10.5123/S1679-49742018000200018

Allison E. Curry, Melissa R. Pfeiffer, Meghan E. Carey & Lawrence J. Cook (2019). Catalyzing traffic safety advancements via data linkage: Development of the New Jersey Safety and Health Outcomes (NJ-SHO) data warehouse, *Traffic Injury Prevention*, 20:sup2, S151-S155, DOI: 10.1080/15389588.2019.1679552

Alsop, J., & Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis & Prevention*, *33*(3), 353-359. https://doi.org/10.1016/S0001-4575(00)00049-X

Alzaffin, K., Kaye, S. A., Watson, A., & Haque, M. M. (2023). A data fusion approach of police-hospital linked data to examine injury severity of motor vehicle crashes. *Accident analysis and prevention*, *179*, 106897. https://doi.org/10.1016/j.aap.2022.106897

Ameratunga, S., Hijar, M., & Norton, R. (2006). Road-traffic injuries: confronting disparities to address a global-health problem. *The lancet*, *367*(9521), 1533-1540.

Amorim, M., Ferreira, S., & Couto, A. (2014). Linking Police and Hospital Road Accident Records: How Consistent Can It Be? *Transportation Research Record*, 2432(1), 10–16. https://doi.org/10.3141/2432-02

Amoros, E., Martin, J. L., & Laumon, B. (2006). Under-reporting of road crash casualties in France. *Accident; analysis and prevention*, *38*(4), 627–635. https://doi.org/10.1016/j.aap.2005.11.006

Bambach, M. R., & Mitchell, R. J. (2015). Estimating the human recovery costs of seriously injured road crash casualties. *Accident analysis and prevention*, *85*, 177–185. https://doi.org/10.1016/j.aap.2015.09.013

Barchielli, A., & Balzi, D. (2007). Incidenti stradali in un'area urbana: danni alla salute, prestazioni e costi sanitari valutati mediante l'uso integrato di dati sanitari correnti e archivio della polizia municipale [Road traffic crashes in an urban area: linkage between municipal police data-base and routinely collected medical data to assess adverse health effects, health system activities and costs]. *Epidemiologia e prevenzione*, *31*(4), 218–224.

Benavente, M., Knodler, M. A., & Rothenberg, H. (2006). Case Study Assessment of Crash Data Challenges: Linking Databases for Analysis of Injury Specifics and Crash Compatibility Issues. *Transportation Research Record*, 1953(1), 180–186. https://doi.org/10.1177/0361198106195300121

Bianchi Santiago, J. D., Colón Jordán, H., & Valdés, D. (2020). Record Linkage of Crashes with Injuries and Medical Cost in Puerto Rico. *Transportation Research Record*, 2674(10), 739–748. https://doi.org/10.1177/0361198120935439

Bohensky, M. A., Jolley, D., Sundararajan, V., Evans, S., Pilcher, D. V., Scott, I., & Brand, C. A. (2010). Data linkage: a powerful research tool with potential problems. BMC health services research, 10(1), 1-7. https://doi.org/10.1186/1472-6963-10-346

Boufous, S., & Williamson, A. (2009). Factors affecting the severity of work related traffic crashes in drivers receiving a worker's compensation claim. *Accident analysis and prevention*, *41*(3), 467–473. https://doi.org/10.1016/j.aap.2009.01.015

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(1), 65-72.

Ceklic, E., Tohira, H., Ball, S., Brown, E., Brink, D., Bailey, P., ... & Finn, J. (2021). Motor vehicle crash characteristics that are predictive of high acuity patients: an analysis of linked ambulance and crash data. *Prehospital emergency care*, *25*(3), 351-360. https://doi.org/10.1080/10903127.2020.1771487

Ceklic, E., Tohira, H., Ball, S., Brown, E., Brink, D., Bailey, P., ... & Finn, J. (2022). A predictive ambulance dispatch algorithm to the scene of a motor vehicle crash: the search for optimal over and under triage rates. *BMC Emergency Medicine*, *22*(1), 1-11. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. (2018). Linking Information for Nonfatal Crash Surveillance A guide for integrating motor vehicle crash data to help keep Americans safe on the road. Retrieved March 3, 2023 from https://www.cdc.gov/motorvehiclesafety/pdf/linkage/cdc_lincs_guide_2019-f.pdf

Christen, P., & Goiser, K. (2007). Quality and complexity measures for data linkage and deduplication. *Quality measures in data mining*, 127-151.

Cirera, E. V. A., Plasència, A., Ferrando, J., & Arribas, P. (2001). Probabilistic Linkage of Police and Emergency Department Sources of Information on Motor-Vehicle Injury Cases: a Proposal for Improvement. *Traffic Injury Prevention*, *2*(3), 229-237. https://doi.org/10.1080/10286580108902567

Clark, D.E. (2004). Practical introduction to record linkage for injury research. *Injury Prevention*, 10(3), 186-191.

Clark, J. M., & Marshall, R. (2022). Utilising international statistical classification of diseases and related health conditions (ICD)-10 Australian modification classifications of "health conditions" to achieve population health surveillance in an Australian spinal cord injury cohort. *Spinal Cord*, 60(8), 746-756.

Conderino, S., Fung, L., Sedlar, S., & Norton, J. M. (2017). Linkage of traffic crash and hospitalization records with limited identifiers for enhanced public health surveillance. *Accident; analysis and prevention*, *101*, 117–123. https://doi.org/10.1016/j.aap.2017.02.011

Cottrill, C. D., & Derrible, S. (2015). Leveraging big data for the development of transport sustainability indicators. *Journal of Urban Technology*, *22*(1), 45-64.

Couto, A., Amorim, M., & Ferreira, S. (2016). Reporting road victims: Assessing and correcting data issues through distinct injury scales. *Journal of safety research*, *57*, 39–45. https://doi.org/10.1016/j.jsr.2016.03.008

Cryer, P. C., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P., & Clarke, C. (2001). Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports. *Injury prevention*, 7(3), 234-241.

Dandona, R., Kumar, G. A., Ameer, M. A., Reddy, G. B., & Dandona, L. (2008). Under-reporting of road traffic injuries to the police: results from two data sources in urban India. *Injury prevention*, *14*(6), 360–365. https://doi.org/10.1136/ip.2008.019638

De Coster, C., Quan, H., Finlayson, A., Gao, M., Halfon, P., Humphries, K.H., Johansen, H., Lix, L.M., Luthi, J.C., Ma, J. and Romano, P.S., 2006. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. *BMC health services research*, 6, pp.1-6.

Derrick G Lopez, Diana L Rosman, George A Jelinek, Garry J Wilkes, Peter C Sprivulis, Dhillon, G., & Torkzadeh, G. (2006). Value-focused assessment of information system security in organizations. *Information Systems Journal*, *16*(3), 293-314.

Edwards, M., & Gutierrez, M. (2023). The incidence burden of unreported pedestrian crashes in Illinois. Traffic injury prevention, 24(1), 82-88. https://doi.org/10.1080/15389588.2022.2143236

Engin, Z., van Dijk, J., Lan, T., Longley, P. A., Treleaven, P., Batty, M., & Penn, A. (2020). Data-driven urban management: Mapping the landscape. *Journal of Urban Management*, *9*(2), 140-150.

European Commission. (n.d.). Road safety: Statistics and analysis. Retrieved March 3, 2023, from https://road-safety.transport.ec.europa.eu/statistics-and-analysis/data-and-analysis_en

Ferreira, S., Falcão, L., Couto, A., & Amorim, M. (2015). The quality of the injury severity classification by the police: An important step for a reliable assessment. *Safety science*, 79, 88-93. https://doi.org/10.1016/j.ssci.2015.05.013

Gonzalez, R. P., Cummings, G., Mulekar, M., & Rodning, C. B. (2006). Increased mortality in rural vehicular trauma: identifying contributing factors through data linkage. *The Journal of trauma*, *61*(2), 404–409. https://doi.org/10.1097/01.ta.0000229816.16305.94

Gopal, D. P., Chetty, U., O'Donnell, P., Gajria, C., & Blackadder-Weinstein, J. (2021). Implicit bias in healthcare: clinical practice, research and decision making. *Future healthcare journal*, *8*(1), 40.

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, 331(7524), 1064-1065.

Gutierrez, H., Mitra, S., Neki, K., Mbugua, L. W., Balasubramaniyan, R., Winer, M., ... & Bhalla, K. (2022). Comparing estimates of road traffic deaths and non-fatal road traffic injuries in Cambodia. *Injury prevention*, *28*(4), 340-346.

Harrison, J.E., Watson, A. & Vallmuu, K. (2018). A national approach to measuring non-fatal crash outcomes, Austroads.

Harrison, J.E., Weber, S., Jakob, R. and Chute, C.G., (2021). ICD-11: an international classification of diseases for the twenty-first century. *BMC medical informatics and decision making*, 21(6), pp.1-10.

Heydari, S., Hickford, A., McIlroy, R., Turner, J., & Bachani, A. M. (2019). Road safety in low-income countries: state of knowledge and future directions. *Sustainability*, *11*(22), 6249.

Hong, Y., & Zeng, M. L. (2023). International classification of diseases (ICD). KO Knowledge Organization, 49(7), 496-528.

Hossain, A., Sun, X., Islam, S., Alam, S., & Hossain, M. M. (2023). Identifying roadway departure crash patterns on rural two-lane highways under different lighting conditions: Association knowledge using data mining approach. *Journal of Safety Research*.

Hosseinzadeh, A., & Kluger, R. (2021). Do EMS times associate with injury severity?. *Accident; analysis and prevention*, *153*, 106053. https://doi.org/10.1016/j.aap.2021.106053

Hosseinzadeh, A., Karimpour, A., Kluger, R., & Orthober, R. (2022). Data linkage for crash outcome assessment: Linking police-reported crashes, emergency response data, and trauma registry records. *Journal of safety research*, 81, 21-35. https://doi.org/10.1016/j.jsr.2022.01.003.

Hu, G., Baker, T., & Baker, S. P. (2011). Comparing road traffic mortality rates from police-reported data and death registration data in China. *Bulletin of the World Health Organization*, *89*(1), 41–45. https://doi.org/10.2471/BLT.10.080317

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. *Accident Analysis & Prevention*, *130*, 84-90.

Janstrup, K. H., Kaplan, S., Hels, T., Lauritsen, J., & Prato, C. G. (2016). Understanding traffic crash under-reporting: linking police and medical records to individual and crash characteristics. *Traffic Injury Prevention*, *17*(6), 580-584. https://doi.org/10.1080/15389588.2015.1128533

Jeffrey R. Brubacher, Herbert Chan, Shannon Erdelyi, Injury severity in police collision reports correlates poorly with requirement for hospital admission, *Journal of Transport & Health*, Volume 14, 2019, https://doi.org/10.1016/j.jth.2019.100606

Jeffrey, S., Stone, D. H., Blamey, A., Clark, D., Cooper, C., Dickson, K., Mackenzie, M., & Major, K. (2009). An evaluation of police reporting of road casualties. *Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention*, *15*(1), 13–18. https://doi.org/10.1136/ip.2008.018630

Kalkman, S., van Delden, J., Banerjee, A., Tyl, B., Mostert, M., & van Thiel, G. (2022). Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *Journal of medical ethics*, *48*(1), 3-13.

Kamaluddin, N. A., Abd Rahman, M. F., & Várhelyi, A. (2019). Matching of police and hospital road crash casualty records–a data-linkage study in Malaysia. *International journal of injury control and safety promotion*, 26(1), 52-59.

Kudryavtsev, A. V., Kleshchinov, N., Ermolina, M., Lund, J., Grjibovski, A. M., Nilssen, O., & Ytterstad, B. (2013). Road traffic fatalities in Arkhangelsk, Russia in 2005–2010: Reliability of police and healthcare data. Accident Analysis & Prevention, 53, 46-54. https://doi.org/10.1016/j.aap.2012.12.022.

Kulynych, J., & Greely, H. T. (2017). Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide. *Journal of Law and the Biosciences*, *4*(1), 94-132.

Lai, C. H., Huang, W. S., Chang, K. K., Jeng, M. C., & Doong, J. L. (2006). Using data linkage to generate 30-day crash-fatality adjustment factors for Taiwan. *Accident analysis and prevention*, *38*(4), 696–702. https://doi.org/10.1016/j.aap.2005.12.015

Langley, J. D., Dow, N., Stephenson, S., & Kypri, K. (2003). Missing cyclists. *Injury prevention*, 9(4), 376-379.

Lateef, M. U. (2010). Estimation of fatalities due to road traffic crashes in Karachi, Pakistan, using capture-recapture method. *Asia Pacific Journal of Public Health*, 22(3), 332-341. https://doi.org/10.1177/1010539509356808

Li, M. D., Doong, J. L., Chang, K. K., Lu, T. H., & Jeng, M. C. (2008). Differences in urban and rural accident characteristics and medical service utilization for traffic fatalities in less-motorized societies. *Journal of safety research*, *39*(6), 623–630. https://doi.org/10.1016/j.jsr.2008.10.008

Lombardi, L. R., Pfeiffer, M. R., Metzger, K. B., Myers, R. K., & Curry, A. E. (2022). Improving identification of crash injuries: State-wide integration of hospital discharge and crash report data. *Traffic injury prevention*, *23*(sup1), S130–S136. https://doi.org/10.1080/15389588.2022.2083612

Lopez, D. G., Rosman, D. L., Jelinek, G. A., Wilkes, G. J., & Sprivulis, P. C. (2000). Complementing police road-crash records with trauma registry data—an initial evaluation. *Accident Analysis & Prevention*, *32*(6), 771-777, https://doi.org/10.1016/S0001-4575(99)00130-X.

Lujic, S., Finch, C., Boufous, S., Hayen, A. and Dunsmuir, W. (2008). How comparable are road traffic crash cases in hospital admissions data and police records? An examination of data linkage rates. *Australian and New Zealand Journal of Public Health*, 32: 28-33. https://doi.org/10.1111/j.1753-6405.2008.00162.x

Mandacaru, P. M. P., Andrade, A. L., Rocha, M. S., Aguiar, F. P., Nogueira, M. S. M., Girodo, A. M., Pedrosa, A. A. G., Oliveira, V. L. A., Alves, M. M. M., Paixão, L. M. M. M., Malta, D. C., Silva, M. M. A., & Morais Neto, O. L. (2017). Qualifying information on deaths and serious injuries caused by road traffic in five Brazilian capitals using record linkage. *Accident analysis and prevention*, *106*, 392–398. https://doi.org/10.1016/j.aap.2017.06.018

Mannering, F. L., & Bhat, C. R. (2014). Analytic methods in crash research: Methodological frontier and future directions. *Analytic methods in accident research*, *1*, 1-22.

McDonald, G., Davie, G., & Langley, J. (2009). Validity of police-reported information on injury severity for those hospitalized from motor vehicle traffic crashes. *Traffic injury prevention*, *10*(2), 184-190. https://doi.org/10.1080/15389580802593699
Mitchell, R. J., Cameron, C. M., & Bambach, M. R. (2014). Data linkage for injury surveillance and research in Australia: perils, pitfalls and potential. *Australian and New Zealand journal of public health*, *38*(3), 275-280.
Mitchell, R. J., Senserrick, T., Bambach, M. R., & Mattos, G. (2015). Comparison of novice and full-licenced driver common crash types in New South Wales, Australia, 2001-2011. *Accident analysis and prevention*, *81*, 204–210. https://doi.org/10.1016/j.aap.2015.04.039
Moore, H. C., Guiver, T., Woollacott, A., De Klerk, N., & Gidding, H. F. (2016). Establishing a process for conducting cross-jurisdictional record linkage in Australia. *Australian and New Zealand journal of public health*, *40*(2), 159-164.
Moore, M.A. (1997). Comparison of young and adult driver crashes in Alaska using linked traffic crash and hospital data. *Alaska medicine, 39 4*, 95-102.
Nair, M. M., & Tyagi, A. K. (2021). Privacy: History, statistics, policy, laws, preservation and threat analysis. *Journal of Information Assurance & Security*, *16*(1).
Norman, R., Matzopoulos, R., Groenewald, P., & Bradshaw, D. (2007). The high burden of injuries in South Africa. *Bulletin of the World Health Organization*, 85, 695-702. https://doi.org/10.2471/blt.06.037184
Paixão, L. M. M. M., Gontijo, E. D., Mingoti, S. A., Costa, D. A. D. S., Friche, A. A. D. L., & Caiaffa, W. T. (2015). Urban road traffic deaths: data linkage and identification of high-risk population sub-groups. *Cadernos de saude publica*, *31*, 92-106.
Paixão, L. M., Gontijo, E. D., Drumond, E.deF., Friche, A. A., & Caiaffa, W. T. (2015). Traffic crashes in Belo Horizonte: the view from three different sources, 2008 to 2010. *Revista brasileira de epidemiologia = Brazilian journal of epidemiology*, *18*(1), 108–122. https://doi.org/10.1590/1980-5497201500010009
Prince, A. E., & Schwarcz, D. (2019). Proxy discrimination in the age of artificial intelligence and big data. *Iowa L. Rev.*, *105*, 1257.
Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*, 1-10.
Reurings, M. C., & Stipdonk, H. L. (2011). Estimating the number of serious road injuries in the Netherlands. *Annals of epidemiology*, *21*(9), 648–653. https://doi.org/10.1016/j.annepidem.2011.05.007
Rosman, D. L., & Knuiman, M. W. (1994). A comparison of hospital and police road injury data. *Accident Analysis & Prevention*, 26(2), 215-222. https://doi.org/10.1016/0001-4575(94)90091-4
Santoro, S., Estay, H., Avci, A. H., Pugliese, L., Ruby-Figueroa, R., Garcia, A., ... & Curcio, E. (2021). Membrane technology for a sustainable copper mining industry: the Chilean paradigm. *Cleaner Engineering and Technology*, 2, 100091.
Schneider, R. J., Willman, J., & Hargarten, S. (2023). Linking Police and EMS Records: An Approach to Strengthen Bicyclist Injury Reporting. *Transportation Research Record*, 0(0). https://doi.org/10.1177/03611981221151073
Sciortino, S., Vassar, M., Radetsky, M., & Knudson, M. M. (2005). San Francisco pedestrian injury surveillance: mapping, under-reporting, and injury severity in police and hospital records. *Accident analysis and prevention*, 37(6), 1102–1113. https://doi.org/10.1016/j.aap.2005.06.010
Short, J., & Caulfield, B. (2016). Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. *Journal of safety research*, *58*, 1–14. https://doi.org/10.1016/j.jsr.2016.05.002

Singleton, M., Qin, H., & Luan, J. (2004). Factors associated with higher levels of injury severity in occupants of motor vehicles that were severely damaged in traffic crashes in Kentucky, 2000-2001. *Traffic injury prevention*, *5*(2), 144-150.

Soltani, S., Schwarcz, L., Morris, D., Plevin, R., Dicker, R., Juillard, C., ... & Wier, M. (2022). What is counted counts: An innovative linkage of police, hospital, and spatial data for transportation injury prevention. *Journal of safety research*, *83*, 35-44., https://doi.org/10.1016/j.jsr.2022.08.002

Stutts, J. C., & Hunter, W. W. (1999). Motor vehicle and roadway factors in pedestrian and bicyclist injuries: an examination based on emergency department data. *Accident analysis & prevention*, *31*(5), 505-514, https://doi.org/10.1016/S0001-4575(99)00007-X.

Suriyawongpaisal, P., & Kanchanasut, S. (2003). Road traffic injuries in Thailand: trends, selected underlying determinants and status of intervention. *Injury control and safety promotion*, *10*(1-2), 95-104.

Tainter, F., Fitzpatrick, C., Gazillo, J., Riessman, R., & Knodler, M., Jr (2020). Using a novel data linkage approach to investigate potential reductions in motor vehicle crash severity - An evaluation of strategic highway safety plan emphasis areas. *Journal of safety research*, *74*, 9–15. https://doi.org/10.1016/j.jsr.2020.04.012

Tarko, A., & Azam, M. S. (2011). Pedestrian injury analysis with consideration of the selectivity bias in linked police-hospital data. *Accident Analysis & Prevention*, *43*(5), 1689-1695, https://doi.org/10.1016/j.aap.2011.03.027.

Tercero, F., & Andersson, R. (2004). Measuring transport injuries in a developing country: an application of the capture–recapture method. *Crash Analysis & Prevention*, *36*(1), 13-20.

Thorpe, N., & Fawcett, L. (2012). Linking road casualty and clinical data to assess the effectiveness of mobile safety enforcement cameras: a before and after study. *BMJ open*, 2(6), e001304.

Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, *34*(1), 134-153.

Tin Tin, S., Woodward, A., & Ameratunga, S. (2013). Completeness and accuracy of crash outcome data in a cohort of cyclists: a validation study. *BMC Public Health*, 13(1), 1-10. https://doi.org/10.1186/1471-2458-13-420

Torbaghan, M. E., Sasidharan, M., Reardon, L., & Muchanga-Hvelplund, L. C. (2022). Understanding the potential of emerging digital technologies for improving road safety. *Accident Analysis & Prevention*, 166, 106543.

Tsui, K. L., So, F. L., Sze, N. N., Wong, S. C., & Leung, T. F. (2009). Misclassification of injury severity among road casualties in police reports. *Accident analysis and prevention*, *41*(1), 84–89. https://doi.org/10.1016/j.aap.2008.09.005

Tulu, G. S., Washington, S., King, M., & Haque, M. (2013). Why are pedestrian crashes so different in developing countries? A review of relevant factors in relation to their impact in Ethiopia. In *Australasian Transport Research Forum 2013 Proceedings* (pp. 1-18). Australasian Transport Research Forum.

Urie, Y., Velaga, N. R., & Maji, A. (2016). Cross-sectional study of road crashes and related law enforcement efficiency for 10 countries: a gap coherence analysis. *Traffic injury prevention*, 17(7), 686-691.

Vassar, M., Atakpo, P., & Kash, M. J. (2016). Manual search approaches used by systematic reviewers in dermatology. *Journal of the Medical Library Association: JMLA*, 104(4), 302.

Vatsalan, D., Sehili, Z., Christen, P., & Rahm, E. (2017). Privacy-preserving record linkage for big data: Current approaches and research challenges. Handbook of big data technologies, 851-895.

Wang, L., Ning, P., Yin, P., Cheng, P., Schwebel, D. C., Liu, J., ... & Hu, G. (2019). Road traffic mortality in China: analysis of national surveillance data from 2006 to 2016. *The Lancet Public Health*, *4*(5), e245-e255. https://doi.org/10.1016/S2468-2667(19)30057-X.

Watson, A., Watson, B., & Vallmuur, K. (2013). How accurate is the identification of serious traffic injuries by Police? The concordance between Police and hospital reported traffic injuries. In Proceedings of the 2013 Australasian Road Safety Research, Policing and Education Conference (pp. 1-10). Australasian College of Road Safety (ACRS).

Watson, A., Watson, B., & Vallmuur, K. (2015). Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accident analysis and prevention*, *83*, 18–25. https://doi.org/10.1016/j.aap.2015.06.011

Weijermars, W., Bos, N., & Stipdonk, H. L. (2016). Serious road injuries in the Netherlands dissected. *Traffic injury prevention*, 17(1), 73-79. https://doi.org/10.1080/15389588.2015.1042577

Wilson, S. J., Begg, D. J., & Samaranayaka, A. (2012). Validity of using linked hospital and police traffic crash records to analyse motorcycle injury crash characteristics. *Accident analysis and prevention*, 49, 30–35. https://doi.org/10.1016/j.aap.2011.03.007

Zhu, Y., Matsuyama, Y., Ohashi, Y., & Setoguchi, S. (2015). When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *Journal of Biomedical Informatics*, *56*, 80-86. https://doi.org/10.1016/j.jbi.2015.05.012