Evaluating Research on Data Linkage to Assess Underreporting of Pedestrian and Bicyclist Injury in Police Crash Data (19-05465)

Sarah Doggett; David R. Ragland, PhD, MPH; Grace Felschundneff
University of California, Berkeley
Safe Transportation Research & Education Center

Introduction and Definitions of Underreporting

Traffic safety decisions are almost universally based on information from police collision reports, which may underrepresent bicycle and pedestrian collisions. This review examined ten studies that used data linkage to explore potential underreporting of pedestrian and/or bicyclist injury in police collision reports.

Different definitions of reporting level exist and are defined in Table 1 using the terminology illustrated in the figure below. Reporting level can be measured using the capture-recapture method, which was developed to estimate animal populations based on how many animals are captured and then later recaptured.
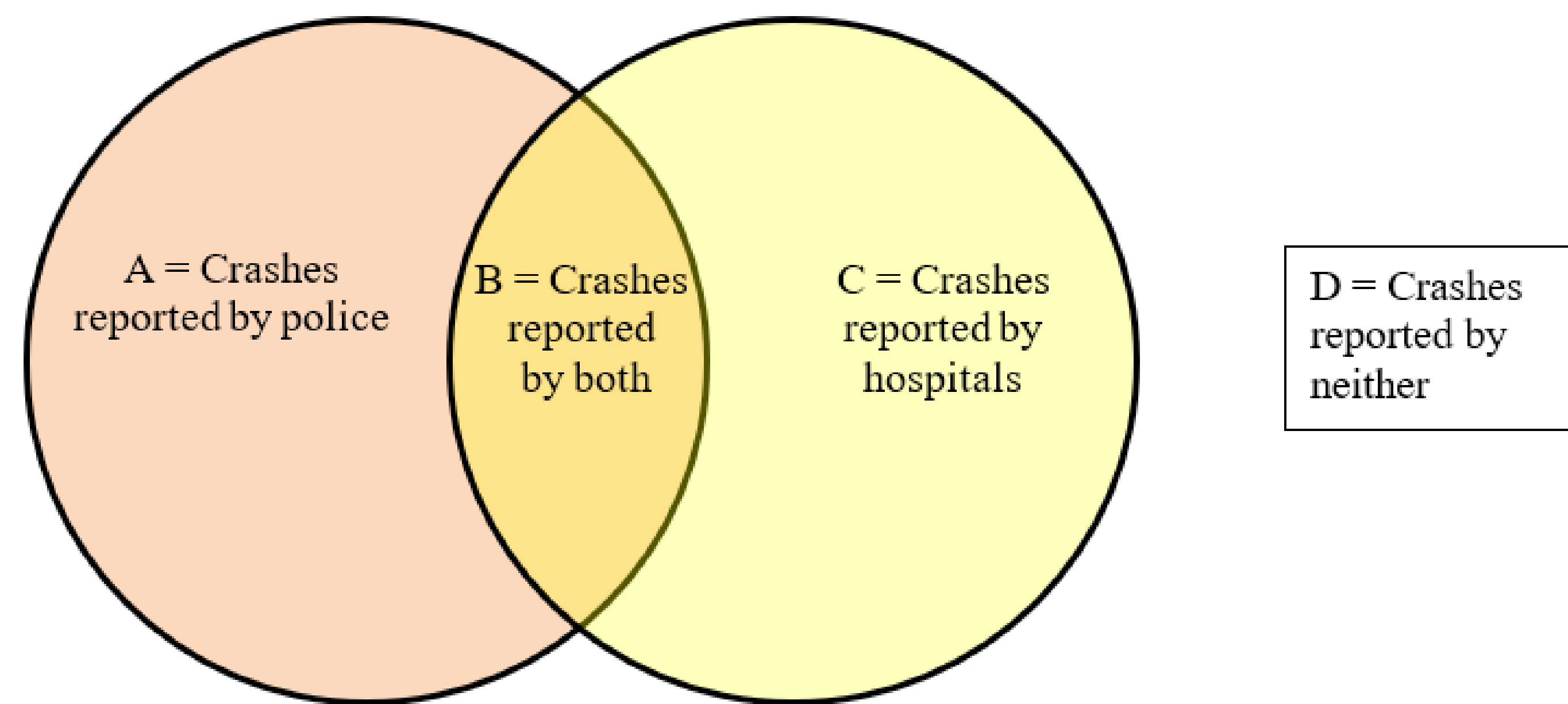

A = Crashes reported by police
B = Crashes reported by both
C = Crashes reported by hospitals
D = Crashes reported by neither

Table 1: Definitions of Reporting Level

Definition	Formula	Question Addressed
1: Hospital Link Rate	B/(B+C)	Proportion of cases reported in hospital records for which a police crash report is found
2: Police Link Rate	B/(A+B)	Proportion of cases reported in police crash reports for which a hospital record is found
3: Police Ascertainment Rate	(A+B)/(A+B+C)	Proportion of identified cases accounted for in police crash reports
4: Hospital Ascertainment Rate	(B+C)/(A+B+C)	Proportion of identified cases accounted for in hospital records

Some crashes may be reported in both datasets but cannot be linked due to errors in data recording—these crashes are *misreported* but cannot not be distinguished from underreported crashes using current linkage techniques. There are also crashes that are not reported to either the police or to hospitals (D in Figure 1), due to lack of serious injury or unwillingness on the part of those involved in the crash to contact authorities. To identify these unreported crashes, other datasets such as those collected by insurance companies can be analyzed. One of the studies addressed this, and found that relying solely on police and hospital injury data can result in a significant underestimation of the actual number and severity of crashes.

Table 2: Summary of Data Linkage Articles Relevant to Pedestrian and Bicyclist Safety

Source	Study Period	Study Location	Focus	Definition Used	Findings/Conclusions
Conderino, Fung, Sedlar & Norton, 2017	2009-2015	New York City	General	Hospital Link Rate	• 50% of hospital reports involving a pedestrian crash linked to a police report • 45% of hospital reports involving a bicyclist crash linked to a police report
Janstrup et al., 2016	2003-2007	Funen, Denmark	General	Capture-Recapture	• Compared with car occupants, pedestrians are more likely to appear in both police and hospital databases; bicyclists are more likely to appear in either • Only 7% of bicycle crashes resulting in slight injury and only 15% of bicycle crashes resulting in severe injury are reported by the police
Langley, 2003	1995-1999	New Zealand	Bicyclist	Hospital Link Rate	• Only 22% of bicycle crashes on public roads could be linked to a crash report • When limited to bicycle crashes on public roads involving motor vehicles, 54% could be linked to a crash report
Lujic, Finch, Boufous, Hayen & Dunsmuir, 2008	2000-2001	New South Wales	General	Hospital Link Rate	• 69% of road traffic crashes were linked to police records • Drivers were most likely to have their hospital records linked to police records (83%) • 46% of bicyclist crashes and 75% of pedestrian crashes were linked to police records • Authors hypothesized that underreporting for cyclists is due to "ambiguity of…laws and regulations" and the fact that cyclists are "less likely to cause property damage"
Sciortino, Vassar, Radetsky & Knudson, 2005	2000-2001	San Francisco	Pedestrian	Police Ascertainment Rate	• Police reports underestimate the number of pedestrian injuries by 21% (e.g., reporting level is 79%) • African-American pedestrians were less likely than white pedestrians to be linked to a police report • Women were more likely than men to be linked to a police report
Short & Caulfield, 2016	2005-2011	Ireland	General	Hospital Link Rate, Police Link Rate, Hospital Ascertainment Rate, Police Ascertainment Rate	• For pedestrian injuries, 28.9% of police records were matched with a hospital record; 44.3% of hospital records were matched with a police record • For bicyclist injuries, 24.8% of police records were matched with a hospital record; 8.2% of hospital records matched with a police record • Police Ascertainment Rate was 73.4% for pedestrians and 26.4% for bicyclists • Hospital Ascertainment Rate was 47.7% for pedestrians and 80.2% for bicyclists
Stutts & Hunter, 1999	1995-1996	Various locations in California, New York, and North Carolina	Pedestrian and Bicyclist	Hospital Link Rate	• 70% of bicyclist injuries reported by hospitals did not involve a motor vehicle • 64% of pedestrian injuries reported by hospitals did not involve a motor vehicle • 31% of bicyclist and 53% of pedestrian injuries occurred in non-roadway locations (e.g., sidewalks, parking lots, trails) • Police crash reports capture less than 33% of serious bicyclist injuries
Tarko & Azam, 2011	2003-2008	Indiana	Pedestrian	Police Link Rate	• Pedestrians struck on a state road, at a Y intersection, or on a divided roadway were less likely to be included in both police and hospital databases • Pedestrians struck while crossing a road, as opposed to walking along or standing outside of the roadway were much more likely to be included in both databases • Authors hypothesized that more severe injuries were more likely to appear in both databases
Tin Tin, Woodward & Ameratunga, 2013	2006-2011	New Zealand	Bicyclist	Capture-Recapture	• Police reports were linked to insurance, hospital, and mortality records • 13% of hospital reported crashes and 64% of hospital reported collisions were linked to police records • 39% of police reported crashes and 43% of police reported collisions were linked to hospital records • When compared with self-reported data from the cyclists, the entire linked dataset had a sensitivity of 63.1% and specificity of 93.5%
Watson, Watson & Vallmuur, 2015	2009	Queensland	General	Hospital Link Rate	• The study used discordance rates between police data and hospital records to measure underreporting of crashes to/by the police • The discordance rate was 44% for pedestrians and 93% for bicyclists (e.g., a reporting level of 56% and 7% respectively) • Authors hypothesized that bicyclist injuries are not reported to the police because injuries are generally less serious, are less likely to have insurance implications, and are more likely to involve young people, who generally have high discordance rates

Potential Implications

• Findings from these ten studies indicate that there is a severe underreporting problem in datasets based solely on police collision reports.
• Underreporting can result in forecasting of incorrect estimates of crash and fatality rates, and identification of erroneous factors responsible for crashes. When road safety is evaluated based on data other than the actual number of crashes that occurred, there is a tendency to misinterpret trends in crash reporting as trends in traffic safety.
• Inaccurate crash data can result in improper prioritization of funding and resources, and under- or overestimation of injury severity risk.

Limitations of Existing Research

• Few of the studies provide an estimate of the error around their reporting level estimates, thus it is possible that underreporting of bicycle and pedestrian crashes is not as extensive as these studies suggest.
• Because linkage and analysis processes are often separated due to privacy concerns, researchers using linked datasets may be unable to determine the extent of bias that linkage errors have introduced into their research.
• Missing data is common. When data for key variables are missing, cases that are linked are less likely to be representative of the study population because they have fewer common values such as unusual zip codes.

Suggestions for Improvement

• Standardize definitions of variables used for linking at the national or international level.
• Develop a clear and systematic method of reporting linkage methodology.
• Use a consistent definition of reporting level so that results can be compared. While hospital and police ascertainment rates are the most theoretically accurate definitions of reporting level, researchers may consider use of hospital link rate instead because it is the most common definition.

Conclusion

• Due to different definitions of reporting level, periods of study, and study locations, it was difficult to directly compare the studies. Even among the six studies using the hospital link definition, estimates of reporting levels ranged from 44 to 75 percent for pedestrian crashes, and from 7 to 46 percent for bicycle crashes.
• To account for underreporting bias when making traffic safety decisions based on police data, the estimated reporting level and the uncertainty of the estimated reporting level must be known.