Police-reporting of Pedestrians and Bicyclists Treated in Hospital Emergency Rooms

Jane C. Stutts and William W. Hunter
University of North Carolina
Highway Safety Research Center
Chapel Hill, NC 27599-3430

INTRODUCTION

A number of studies conducted in the U.S. and in other countries as well have shown that official crash records significantly underestimate the numbers of pedestrians and bicyclists injured in collisions with motor vehicles. Some of the studies have compared overall numbers of injured pedestrians and/or bicyclists reported by hospital or emergency room sources with numbers reported on official motor vehicle crash databases over the same time period. Others have attempted to link individual records reported on medical and crash data files.

In the U.S., one of the earliest such efforts was reported by Barancik and Fife (1985), drawing from a probability sampling of emergency room visits to 42 hospitals in northeastern Ohio. The authors were able to identify police crash reports for only 55 percent of the patients treated for injuries received in a motor vehicle crash; among those hospitalized, 74 percent were matched. Separate information for pedestrians and bicyclists was not reported. In a more recent study focusing on bicyclists treated in a cross-section of hospital emergency rooms in North Carolina, Stutts et al. (1990) identified police reports for 60 percent of the cases where a motor vehicle was involved. And in Western Australia, 69 percent of pedestrians and 74 percent of bicyclists admitted to hospitals were linked to official police records (Rosman and Knuiman, 1994).

Studies that have compared overall numbers of hospital or emergency room reported cases with official crash statistics have shown similar reporting discrepancies. In a study conducted in New Zealand, numbers of hospital admissions over a ten-year period were compared to numbers of police-reported road accident victims (Morrison and Kjellstrom, 1987). The proportion of police reported to hospital reported road accident cases declined over the ten-year study period from .66 to .43 overall, from .74 to .48 for pedestrians, and from .21 to .15 for bicyclists. The lower proportion for bicyclists likely reflects the inclusion of bicycle only events in the sample, even though these are not legally required to be reported in New Zealand. A similar study by Dutch researchers yielded higher overall proportions — .78 for pedestrians and .82 for bicyclists (Maas and Harris, 1984). In a followup study, however, Harris (1990) noted that these proportions had declined to less than .70 by the late 1980s. Lower proportions were also reported in two other European studies that looked only at bicyclists: a German study reported proportions of .30 for hospitalized bicyclists and .20 for bicyclists receiving outpatient treatment only (Hautzinger et al., 1993), while an earlier British study had reported .24 for bicyclists receiving either inpatient or outpatient treatment (Bull and Roberts, 1973).

These studies all point to a significant underreporting of injured pedestrians and bicyclists in official crash statistics, especially for those less seriously injured. The extent of underreporting varies substantially across the various studies cited, and likely reflects differences in study methodologies as well as differences in reporting requirements and record keeping practices in the various jurisdictions. The study by Rosman and Knuiman (1994) examined factors associated with police reporting of all road casualties in Western Australia admitted to hospital. None of the studies, however, systematically examined and reported on differences between pedestrian and bicyclist injury cases that are and are not reported on official crash databases, which is the intent of this paper. It is clear that injury severity is a factor: hospitalized victims are more likely to be reported than those receiving only emergency room or outpatient treatment. But are there other differences as well? Are older victims more likely to be reported? young children? those injured on public roadways? those injured during the daytime? persons under the influence of alcohol or drugs?

The answers to these questions are important because they provide insight to potential biases that may result from relying on official police-reported crash data when estimating numbers of injured pedestrians and bicyclists and when developing and targeting traffic safety programs and countermeasures aimed at these populations. Knowledge of reporting biases can also be used to adjust the crash data to more accurately reflect both the numbers of pedestrians and bicyclists being injured in collisions with motor vehicles and the most salient characteristics of these events.

The current paper draws from a larger study of injured pedestrians and bicyclists treated at hospital emergency rooms. The study included pedestrians and bicyclists injured in motor vehicle and non-motor vehicle events, occurring in roadway as well as non-roadway locations. As part of the larger study, an attempt was made to link the emergency room reported cases to state motor vehicle crash data. This paper reports on the results of that matching process, along with an analysis of factors associated with the occurrence of a match. The purpose of the paper is twofold: (1) to provide additional information on the likely level of underreporting of pedestrian- and bicycle-motor vehicle collisions on state crash files, and (2) to examine whether cases that are reported differ in any systematic ways from those that are not reported, i.e., whether important biases exist in the police-reported databases.

METHODS

Data on injured pedestrians and bicyclists were collected at eight hospital emergency rooms in three states over approximately one-year time periods. The hospitals included three in the Buffalo, New York area, three in south-central California (in Modesto, Oxnard and Goleta, near Santa Barbara), and two more in eastern North Carolina (in Wilmington and Greenville). Staff at the participating hospital emergency rooms assisted in completing a two-page data form for each injured pedestrian or bicyclist treated. A total of 2,509 cases were identified, including 1,443 pedestrians and 1,066 bicyclists. Of these totals, 36 percent of the pedestrians and 30 percent of the bicyclists were reported to have been injured in collisions with motor vehicles, and are the primary focus of the current analysis. (See Stutts and Hunter, in press, for a more detailed description of the methodology and resulting database.)

In addition to the emergency room data, computerized motor vehicle crash data were also obtained from each of the three participating states, corresponding as closely as possible to the time period during which the emergency room data were collected. An attempt was then made to link the individual emergency room cases to cases on the state crash files. Since no names, addresses, social security numbers, or other unique identifiers were included on the files, the process of matching cases was carried out based on the following available variables: date and time of injury event; age and sex of the pedestrian or bicyclist (and for North Carolina, date of birth and race were also available); and location (city and county) of injury event. To reduce the number of potential matches needing to be checked, the crash files were restricted to cases occurring in counties or reporting districts that might reasonably be serviced by one of the participating hospital emergency rooms. For each state, a printout of all the cases and relevant matching variables was generated for both the emergency room and crash files, ordered by the date of the injury event. The process then became one of checking case-by-case to determine whether each reported emergency room incident was contained on the state crash file listing.

For those cases where a match was uncertain, the hard copy of the emergency room report was checked for additional information that might facilitate a decision. For example, in some cases the narrative might provide an approximate age if no exact age was available, or note that the time of the injury event was only approximate. Cases were coded at four levels: match, probable match, possible match, and no match. For a definite match to occur, the crash date, victim age and sex, county or city, and approximate time would all need to be in agreement. A probable match might have one of these items (usually the time or location, but not age) in disagreement or missing, while possible matches would generally have two or more "disagreements" or missing pieces of information. Overall, 48 percent of the emergency room cases were matched to the crash files, 43 percent were designated as non-matches, six percent as probable matches, and three percent as possible matches.

Although an attempt was made to match all of the emergency room reported cases, virtually none of the cases that did not involve a motor vehicle were identified on the state crash files, so that the focus of the current analysis is only on pedestrian-motor vehicle and bicycle-motor vehicle events. An analysis file was created containing only those cases where a match was possible, i.e., where emergency room and statewide crash data were available for the same time periods. In addition, only those cases where a definite match or no match decision was made were included in the analysis (i.e., "probable" and "possible" match cases were excluded from the analysis file).

Data analysis initially consisted of two-way crosstabulations of potential variables of interest with match status. Based on these descriptive results, two-level variables were defined that identified specific factors that might be associated with differential reporting on state motor vehicle crash files. Results of these analysis are reported in terms of the relative probability of a match occurring on the appropriate state crash file. For example, if 60 percent of hospitalized cases are matched, but only 40 percent of non-hospitalized cases, the relative probability (or relative risk) of a match for hospitalized cases is .60 / .40 = 1.5, or one-and-a-half times that of non-hospitalized cases. Estimates of relative risks and their 95 percent confidence limits were obtained from the Cochran-Mantel-Haenszel test statistics generated for each crosstabulation, along with a chi-square test of the overall significance of the association.

All variables found to be significant at the .10 level or higher were then included in a multivariate data analysis using multiway contingency table analysis. Adjusted relative probabilities and 95 percent confidence intervals were again generated from the Cochran-Mantel-Haenszel test statistics. All analyses were carried out using Statistical Analysis System (SAS) software version 6.12 (SAS, Inc., Cary NC).

RESULTS

A total of 388 pedestrian- and 255 bicycle-motor vehicle crashes were available for inclusion in the analyses, with 56 percent of the pedestrian and 48 percent of the bicycle cases matched to state motor vehicle crash file cases. Table 1 summarizes the results of initial descriptive tables examining the percentage of matched cases across levels of the variables available for analysis. Frequency counts (Ns) are also shown in the table. For several of the variables, the Ns were sufficiently small at some levels so that Chi-square was not a valid test of association. Thus, the primary purpose of this initial table is to provide an overall feel for the data and to suggest variable levels or factors that might be more closely examined in a multivariate analysis.

The results in Table 1 show higher percentages of matched or police-reported cases for pedestrians who are male, Black, injured on the roadway, injured during the early morning hours, impaired by alcohol, and either admitted to the hospital or killed. There was also a *lower* level of matching for pedestrians in the 15-19 and 20-24 year age groups. For bicyclists, higher police-reported rates are found for Blacks, those injured in the mid-morning to early afternoon hours (10 am - 2 pm) and those admitted to the hospital or killed. Interestingly, in contrast to pedestrians, a lower rate of matching is shown for bicyclists impaired by alcohol.

Drawing from these initial descriptive results, a series of two-level variables was created and crosstabulated with match status. Results for pedestrian-motor vehicle cases are summarized in Table 2. The results show the calculated relative probabilities of a match, i.e., the likelihood that an emergency room case with a given characteristic will be reported on a state motor vehicle crash file compared to cases without that characteristic. A relative probability of one (1.00) indicates that a particular characteristic (young age, male, Black, etc.) is not associated with any increase or decrease in the likelihood that an event will be reported. Numbers less than one indicate a decreased likelihood of reporting, while those greater than one indicate an increased likelihood of reporting. For a factor to be statistically significant, its 95 percent confidence interval must not include 1.00.

Table 2 results show that pedestrians treated in hospital emergency rooms who are aged 15-24 are significantly *less* likely to be reported on state motor vehicle crash files than are pedestrians of all other ages (p=.008). In contrast, pedestrians who are hospitalized or killed are 1.4 times more likely to be reported (p=.001), those struck in the roadway 1.7 times more likely to be reported (p=.002), and those impaired by alcohol 1.3 times more likely to be reported (p=.049).

Similar results for bicyclists are summarized in Table 3. These results show that only injury level, as reflected by the emergency room disposition variable, is significantly associated with reporting status. Bicyclists who are hospitalized or killed as a result of their injuries are 1.4 times more likely to be reported on police crash files than bicyclists who require only emergency room treatment (p=.024).

Since several of the bi-level variables shown in Table 2 were associated with reporting status of pedestrian-motor vehicle events, a multivariate contingency table analysis was carried out to examine the effects of each of the variables independently while controlling for the effects of the remaining variables. Once event location, emergency room disposition, and age were controlled for in the multivariate analysis, the alcohol impairment variable was no longer significant. The other factors, however, remained significant with only slight adjustments to their associated relative probabilities. The adjusted probabilities and their associated p-values are reported in Table 4.

Table 4 shows that, even after adjusting for other potential confounders, the location where a pedestrian is injured with respect to the roadway, the severity of the injury, and the age of the pedestrian significantly impact whether or not that event is reported on official motor vehicle crash files. Pedestrians who are struck while in the roadway are 1.7 times more likely to be reported on state motor vehicle crash files than are those struck in non-roadway locations, and those who are injured seriously enough to require hospitalization are 1.3 times more likely to be reported compared to those who receive only emergency room treatment. In addition, one age group, young persons ages 15-24, are significantly less likely to be reported, even after controlling for the location and severity of their injury event.

No similar multivariate analysis was carried out for the bicycle-motor vehicle crashes, since only the single variable, emergency room disposition, was found to be significantly associated in the univariate analysis.

DISCUSSION

The results of this study confirm a significant underreporting of pedestrian- and bicycle-motor vehicle injury events on state crash files. In a sample of cases collected at eight hospital emergency rooms in three states, only 56 percent of the pedestrians and 48 percent of the bicyclists were successfully linked to cases reported on their respective state motor vehicle crash files. Even among crash events occurring on the roadway, only 59 percent of the pedestrians and 50 percent of the bicyclists were matched.

This finding, while disheartening, was not unexpected, since a number of studies in the U.S. and foreign literature have shown a similar underreporting of crash events in general and pedestrian and bicycle events in particular. The finding that emergency room disposition was a significant factor in reporting status was also anticipated, since studies based on hospitalized samples have consistently yielded higher linkage rates or reporting percentages than those based on emergency room or general population samples.

In the current analysis, no variables in addition to emergency room disposition were found to be associated with police reporting of **bicycle-motor vehicle** crashes. This finding is basically “good news,” since it suggests no new biases that need to be taken into account when analyzing police-reported bicycle-motor vehicle crash data. It might even be argued that the overrepresentation of more serious injury cases is justified, since it targets limited resources to areas of greatest need. Also, knowing that the data overrepresents more serious injury cases, appropriate adjustments can be made during data analysis.

Two variables in addition to injury level were found to be associated with the likelihood of a **pedestrian-motor vehicle** crash being reported. These were the location of the event with respect to the roadway and the pedestrian being a teen or young adult aged 15-24. Pedestrian events that occurred in parking lots, driveways, and other off-road locations were significantly less likely to be identified on official crash files than were those occurring in the roadway. This finding is important, since results from the larger study (Stutts and Hunter, in press) showed that 12 percent of all pedestrian-motor vehicle collisions resulting in injury serious enough to require emergency room treatment occur in non-roadway locations. If these events are less likely to be reported, then the number of non-roadway cases on state motor vehicle crash files should be inflated.

The analysis also suggests that pedestrian crashes involving young persons ages 15-24 may be underreported, even when other factors such as injury level and event location are taken into account. This is an interesting finding which may simply reflect an anomaly in the data, inadequate control for potential confounding variables, or a genuinely lower reporting rate for this age group. It may be the case, for example, that young adults are less inclined to draw the attention of law enforcement authorities, more inclined to downplay their injuries, etc. Further research is needed to clarify the significance of this result.

Additional research might also be able to address other important questions that could not be answered with the existing data. Small sample sizes in the current study, for example, precluded more detailed analyses of specific factors or variable levels, such as the relative likelihood that an event occurring at a certain non-roadway location (driveway, sidewalk, etc.) is reported. It would also have been useful if more specific pedestrian- and bicycle-motor vehicle crash type information had been available for analysis. As it was, we were limited to a relatively small set of variables collected by the participating hospital emergency rooms. Of particular interest would be certain “combination” events, such as young children injured by backing vehicles in driveways or teenagers struck by cars while rollerblading or skateboarding. These are situations that the medical community recognizes to be a problem, but which are likely to be underreported by official crash statistics.

A number of caveats should be considered in assessing the usefulness of the current study. The results are based on data from eight hospital emergency rooms in three states, and are not necessarily representative of the nation as a whole. Also, reporting requirements varied among the states. In particular, the State of California does not require reporting of collisions that occur on private property, whereas both New York State and North Carolina require reporting of all collisions resulting in personal injury or property damage amounting to $1,000 or more, regardless of where they occur. Although each of the three states had a lower match rate for non-roadway versus roadway pedestrian crashes, the difference was greatest in California, where only 3 or the state’s 16 reported non-roadway cases were successfully matched. Thus, the choice of data collection states can affect the outcome of this type study.

A final caveat pertains to the matching process itself. As described in the methodology section, matching was carried out on a case-by-case bases using the date, time and location of the crash event and patient age, gender, and other identifying information from both files. A definite “match” or “no-match” decision was made in 91 percent of the cases, and only these cases were included in the analysis. However, it should be noted that computerized algorithms have been developed to facilitate the linkage of large databases (as part of the Comprehensive Crash Outcome Data Evaluation System (CODES) and similar projects), and it is possible that this type of automated approach might have affected the outcome of the study. Also, whatever approach is utilized will be limited by the quality of available data—to the extent that either the emergency room or the police records have inaccurate or missing data for needed variables, the matching process will be compromised and the results may be biased.

Despite these caveats, it is clear that official crash statistics are not capturing a significant portion of pedestrian and bicycle motor vehicle crashes -- both those occurring on the public roadway and those occurring in non-roadway locations. A recommendation, therefore, might be to conduct a more detailed follow-up investigation of those emergency room cases that are not matched to the police files to determine whether they should, indeed, have been reported, and the reasons for non-reporting. The results of such a study could provide input for improving current reporting practices. Consideration should also be given to promulgating policy to require official reporting of traffic events occurring in non-roadway locations, at the state as well as national level.

With more widespread use of The International Classification of Diseases Cause of Injury or “E-codes” (U.S. Department of Health and Human Services, 1991) by hospitals to identify the underlying cause of a patient’s injury, larger scale databases will become available for comparison and even merger with police crash files. Whereas the current study relied on special data collection procedures at the participating hospital emergency rooms, pedestrian-motor vehicle and bicycle-motor vehicle cases can be readily identified on a hospital’s computerized records if E-coding is used. Currently fewer than 10 states require E-coding, and this only for patients who are hospitalized. But the potential clearly exists for larger scale studies to address some of the issues raised by the current research without reliance on time-consuming and costly primary data collection activities.

Finally, with its focus on pedestrian-motor vehicle and bicycle-motor vehicle crashes, the current study has ignored the even larger numbers of pedestrians and bicyclists injured in falls and other non-collision events not involving a motor vehicle. As noted elsewhere in this paper, only 36 percent of the pedestrians and 30 percent of the bicyclists treated at the participating hospital emergency rooms were injured in collisions with motor vehicles. The remainder were injured in non-collision events -- falls in the roadway, falls in parking lots, trips on sidewalks, slips on curbs, etc. While E-codes can be used to identify bicycle only falls, they are not suited for identifying “pedestrian only” events. And since neither of these events are typically captured on state motor vehicle crash files, special data collection efforts may continue to be needed in order to address the full range of situations posing injury risks to pedestrians and bicyclists.

REFERENCES

Barancik, J.I. and Fife, D. Discrepancies in Vehicular Crash Injury Reporting: Northeastern Ohio Trauma Study IV. Accident Analysis & Prevention, 17(2), 1985, pp. 147-154.

Bull, J.P. and Roberts, B.J. Road Accident Statistics—A Comparison of Police and Hospital Information. Accident Analysis & Prevention, 5, 1973, pp. 45-53.

Harris, S. The Real Number of Road Traffic Accident Casualties in the Netherlands: A Year-Long Survey. Accident Analysis & Prevention, 22(4), 1990, pp. 371-378.

Hautzinger, H., et al. Dunkelziffer bei Unfallen mit Personenschaden [Number of Undetected Casualty Accidents]. West Germany, 1993.

Maas, M.W. and Harris, S. Police Recording of Road Accident In-Patients. Accident Analysis & Prevention, 16(3), 1984, pp. 167-184.

Morrison, P. and Kjellstrom, T. A Comparison of Hospital Admissions Data and Official Government Statistics of Serious Traffic Accident Injuries. New Zealand Medical Journal, 1987, pp. 517-520.

Rosman, D.L. and Knuiman, M.W. A Comparison of Hospital and Police Road Injury Data. Accident Analysis & Prevention, 26(2), 1994, pp. 215-222.

Stutts, J.C. and Hunter, W.W. Injuries to Pedestrians and Bicyclists: An Examination and Comparison of Hospital and State Motor Vehicle Crash Data. Chapel Hill: University of North Carolina Highway Safety Research Center, in press.

Stutts, J.C., Williamson, J.E., Whitley, T. and Sheldon, F.C. (1990). Bicycle Accidents and Injuries: A Pilot Study Comparing Hospital- and Police-Reported Data. Accident Analysis & Prevention, 22(1), pp. 67-78.

U.S. Department of Health and Human Services. The International Classification of Diseases, 9th Revision, Clinical Modification, Fourth Edition, Volume I. Washington, D.C.: DHHS, Publication No. (PHS) 91-1260, 1991.

Table 1. Results of descriptive crosstabulations of potential study variables with match status.

Variable	Pedestrian-MV Total N	Pedestrian-MV % Matched	Bicycle-MV Total N	Bicycle-MV % Matched
Age				
0-9	106	51.9	30	40.0
10-14	46	69.6	56	51.8
15-19	49	40.8	27	59.3
20-24	27	44.4	24	33.3
25-44	94	61.7	83	44.6
45-64	30	60.0	28	60.7
65+	34	58.8	4	50.0
Gender				
Male	250	58.4	206	47.6
Female	137	51.1	49	49.0
Race				
White	184	53.8	130	43.1
Black	132	61.4	83	56.6
Hispanic	50	52.0	33	48.5
Other	13	53.9	6	50.0
Location of Event				
Roadway	323	59.4	220	50.0
Sidewalk	4	100.0	12	50.0
Parking Lot	27	33.3	6	50.0
Driveway	10	20.0	0	- -
Other (trail, park, etc.)	4	25.0	2	0.0
Time of Event				
6 am - 10 am	27	66.7	19	36.8
10 am - 2 pm	43	48.8	37	51.4
2 pm - 6 pm	91	58.2	71	42.3
6 pm - 10 pm	89	53.9	62	43.6
10 pm - 2 am	20	40.0	13	30.8
2 am - 6 am	7	57.1	1	0.0
Alcohol Involvement				
Unknown / Presumed Not Impaired	309	55.3	234	49.2
Tested, Impaired	29	72.4	7	14.3
Tested, Not Impaired	36	41.7	29	58.6
Not Tested, Impaired	14	64.3	14	42.9
ER Disposition				
Treated and Released	221	47.5	181	43.6
Admitted to Hospital	146	65.1	59	57.6
Transfer/Other	7	57.1	4	50.0
Fatal	13	84.6	5	80.0
Total	388	55.7	255	47.8

Table 2. Factors associated with police-reporting of pedestrian-motor vehicle crashes.

Factor	Relative Probability of Match	95% C.I.
Age 15-24	0.71**	(0.55, 0.91)
Male	1.14	(0.95, 1.38)
Black	1.15	(0.96, 1.38)
On-Roadway	1.67**	(1.20, 2.33)
Daytime (6 am - 6 pm)	0.97	(0.77, 1.22)
Alcohol Impaired	1.29*	(1.00, 1.67)
Hospitalized/Killed	1.40**	(1.17, 1.67)

* $p \leq .05$
** $p \leq .01$

Table 3. Factors associated with police-reporting of bicycle-motor vehicle crashes.

Factor	Relative Probability of Match	95% C.I.
Age 15-24	0.98	(0.73, 1.32)
Male	0.97	(0.70, 1.34)
Black	1.28	(0.98, 1.66)
On-Roadway	1.06	(0.65, 1.71)
Daytime (6 am - 6 pm)	1.21	(0.84, 1.75)
Alcohol Impaired	0.68	(0.39, 1.17)
Hospitalized/Killed	1.38*	(1.04, 1.81)

* $p \leq .05$

Table 4. Adjusted relative probabilities of a match and associated p-values for factors associated with police-reporting of pedestrian-motor vehicle crashes.

Factor	Relative Probability of Match	95% C.I.	P-value
On-Roadway Location	1.68	(1.18, 2.38)	.004
Admitted to Hospital/Killed	1.30	(1.09, 1.56)	.003
Age 15-24	0.72	(0.56, 0.93)	.011