Comparative Analysis of Pedestrian Injuries using Police, Emergency Department, and Death Certificate Data Sources in North Carolina, U.S., 2007–2012

Laura S. Sandt[1], Scott K. Proescholdbell[2], Kelly R. Evenson[3], Whitney R. Robinson[3], Daniel A. Rodríguez[4], Katherine J. Harmon[1], and Stephen W. Marshall[3]

Abstract

Pedestrian safety programs are needed to address the rising incidence of pedestrian fatalities. Unfortunately, most communities lack comprehensive information on the circumstances of pedestrian crashes and resulting injuries that could help guide decision-making for prevention program development and implementation. This study aimed to evaluate how three commonly available data sources (police-reported pedestrian crashes, emergency department [ED] visits, and death certificates) define and capture pedestrian injury data, and to compare the distribution of pedestrian injuries and fatalities across these data sources. Existing state-wide data sources in North Carolina, U.S.A.,—police-reported pedestrian crashes, ED visits, and death certificates—were used to perform a descriptive analysis of temporal and demographic pedestrian injury severity distributions for a 6-year period (2007–2012). After excluding non-relevant cases, there were 12,646 police-reported pedestrian crashes, 17,369 pedestrian-injury-related ED visits, and 993 pedestrian-related death certificate cases. Pedestrian injury distributions appeared similar across the three data sets in relation to pedestrian sex, age, and temporality. Police data (which represented crashes rather than all pedestrians involved in a crash) likely underrepresented pedestrian injury incidence, while ED data (which represented ED visits, with multiple visits per person possible) likely overrepresented pedestrian injury incidence. The study provides a better understanding of the discrepancies between pedestrian injury data sources and key considerations when using police, ED, and death certificate data for surveillance or injury prevention efforts.

Over the past decade (2008–2017), the motor vehicle fatality rate in the U.S. has declined by 8% (*1*). In 2017, the rate of motor vehicle fatalities per 100 million vehicle miles traveled (VMT) fell to 1.16, one of the lowest fatality rates on record since the National Highway Traffic Safety Administration (NHTSA) first began collecting U.S. VMT data in 1921. Unfortunately, pedestrians have not benefitted from the recent safety gains that drivers have experienced. In 2018, the pedestrian fatality rate per 100,000 U.S. population was 1.9, an increase of nearly 50% since 2009 (*2*). Furthermore, although driver exposure (measured in VMT) is regularly estimated, pedestrian exposure is not routinely measured—and may not be well-represented by rates of fatalities per population—so it is not known if actual pedestrian rates are increasing, staying constant, or even declining over time.

Pedestrian fatalities account for 16% of all traffic fatalities in the U.S. and as much as 24% (Louisville, KY) to 60% of all traffic fatalities in urban cities (San Francisco, CA) (*3*). For this reason, pedestrian safety is a rising priority for many states and cities across the U.S. Researchers and community leaders alike have acknowledged the importance of safe, walkable communities in

[1]UNC Highway Safety Research Center, Chapel Hill, NC
[2]State Injury Epidemiologist, Division of Public Health, NC Department of Health and Human Services, Raleigh, NC
[3]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC
[4]Department of City and Regional Planning, Institute for Transportation Studies, University of California, Berkeley, Berkeley, CA

Corresponding Author:
Laura S. Sandt, sandt@hsrc.unc.edu

achieving health, economic, and environmental goals (4). Calls for a reduction in pedestrian fatalities (as part of "Vision Zero" or "Toward Zero Deaths" strategic plans or other initiatives) are being made in U.S. states and cities as diverse as Arkansas and New York City. Similarly, many municipalities, schools, and communities are developing Safe Routes to School (SRTS) plans and programs with the goal of creating safer walking conditions for people of all ages—and children in particular—and encouraging active transportation among the next generation of roadway users (5). As these pedestrian safety plans and programs evolve, there is a great need for epidemiologic information about pedestrian crash and injury trends and risk factors to help guide decision-making and prioritize fiscally constrained investments in the most effective way. Such data can also be used as outcome data to evaluate these programs (6, 7).

Unfortunately, a major limitation for many state and local officials is the lack of accurate local crash and injury data available to inform and evaluate injury prevention planning efforts. Crash data from police reports may provide details on the location and circumstances of the crash but provide very little information in relation to outcomes of the crash, such as the nature, location, and severity of pedestrians' injuries. Hospital data, such as hospital or emergency department (ED) discharge data, can provide a lot of information on the pedestrians' injuries but offer little context about the when, where, or why the injury took place. Linking data sources such as these can provide a more comprehensive understanding of both the context of pedestrian crashes and the resulting injury outcomes (8). From 1992 to 2013, NHTSA worked with 15 states to establish data linkage programs as part of an effort called Crash Outcome Data Evaluation Systems (CODES). A study utilizing linked Wisconsin CODES data found that pedestrians had higher fatality and inpatient hospitalization rates per person-trip and per person-mile, as compared with motor vehicle occupants (9). A follow-up study determined that black/African-American and Hispanic/Latino pedestrians were disproportionately at risk of injury (10). Although CODES had the capacity to continue to address important pedestrian safety questions, NHTSA financial support was terminated in 2013 (11). While some CODES states have continued their data linkage activities, many have encountered considerable barriers, including lack of funding, staff, institutional support, and access to clinical data (12). Further, there are few additional publications using linked data sources (such as CODES projects) to specifically address pedestrian injury issues (13–17).

Because of the lack of routinely linked data, most studies have been limited to using a single data source to describe pedestrian crash trends and injury outcomes, or to evaluate or plan interventions. Data sources include police crash reports; ED, hospital discharge, or trauma registries; and death certificates and other sources of mortality data (18–28). Assessments of the performance of non-linked data sources, and comparisons between data sources with respect to how they report pedestrian crash frequencies and distributions, have rarely been made. Thus, little is known about the relative "completeness" of each source and key differences between the data sources that could affect the validity of studies using them. Two recent studies that have attempted to quantify differences in reporting among police crash report/hospital data have found considerable underreporting in police crash data for pedestrian injuries, with 20%–60% of pedestrian injuries reported in hospital data not reported in police crash data (29, 30).

In addition to underreporting in police crash data, another concern is the misclassification of injury severity. Police crash data typically report injury severity using the five-level scale KABCO (K = fatal injury, A = serious injury, B = moderate/minor injury, C = possible injury, and O = no injury). KABCO designations are usually made at the scene of the crash by police officers with limited medical training. Evidence suggests that police officers may misclassify life-threatening internal injuries with few visible signs as "minor" while misclassifying non-life-threatening lacerations with bleeding as "severe" (29, 31, 32). However, few published studies have compared indices of injury severity across police/hospital data sources for pedestrian injuries.

Study Purpose

For communities that do not have access to linked databases or the resources to combine such data sources, key questions remain: what can the available unlinked data sources reveal about the nature of pedestrian injuries, what are the differences and limitations of each data source, and what is the potential impact of using one data source over another for surveillance practices or decision-making? One goal of this study was to evaluate how three commonly available data sources—police-reported crash data, ED data, and death certificate data—define and capture the total number (absolute incidence) of pedestrian injury or fatality "cases" in a defined geographic area—the state of North Carolina (NC) over a specified time period (January 1, 2007–December 31, 2012). A second goal was to compare the distribution of pedestrian injuries and fatalities by demographic (age and sex), temporal, and seasonal variables that are commonly collected in all three data sources. The motivation for the second goal was the possibility that, despite that there is no single comprehensive data source for pedestrian injury, if the degree of case under-ascertainment is consistent across demographic and temporal variables
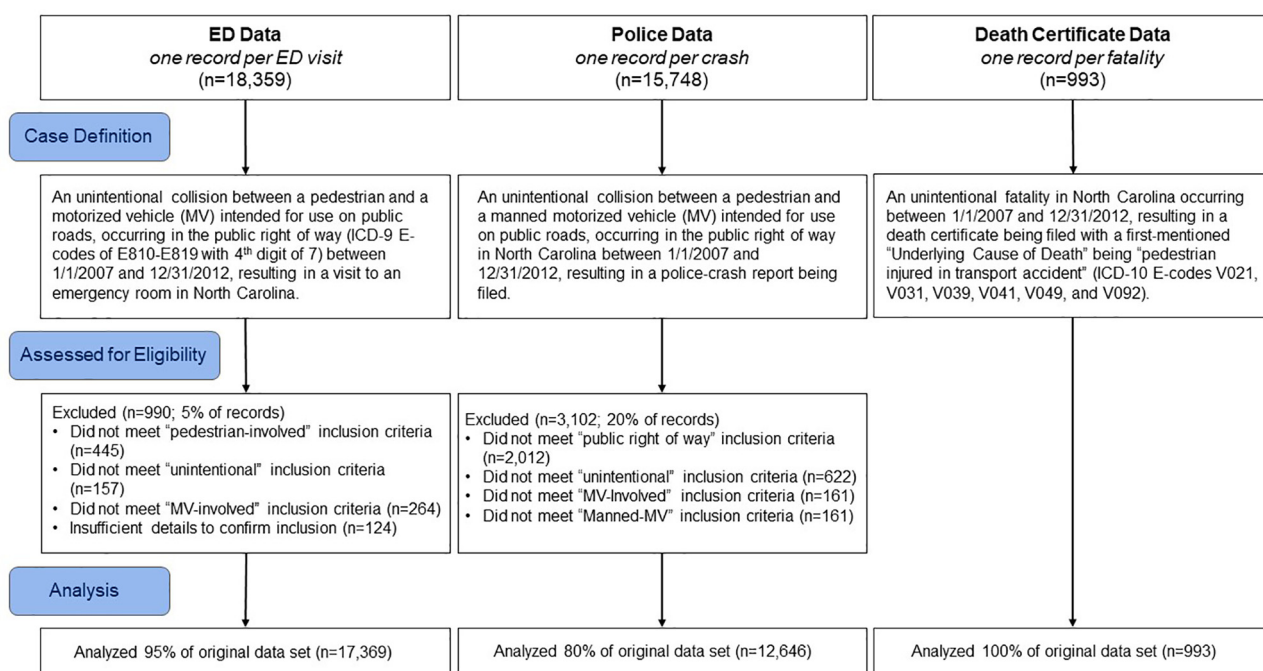
ED Data *one record per ED visit* (n=18,359)	Police Data *one record per crash* (n=15,748)	Death Certificate Data *one record per fatality* (n=993)

Case Definition

An unintentional collision between a pedestrian and a motorized vehicle (MV) intended for use on public roads, occurring in the public right of way (ICD-9 E-codes of E810-E819 with 4th digit of 7) between 1/1/2007 and 12/31/2012, resulting in a visit to an emergency room in North Carolina.	An unintentional collision between a pedestrian and a manned motorized vehicle (MV) intended for use on public roads, occurring in the public right of way in North Carolina between 1/1/2007 and 12/31/2012, resulting in a police-crash report being filed.	An unintentional fatality in North Carolina occurring between 1/1/2007 and 12/31/2012, resulting in a death certificate being filed with a first-mentioned "Underlying Cause of Death" being "pedestrian injured in transport accident" (ICD-10 E-codes V021, V031, V039, V041, V049, and V092).

Assessed for Eligibility

Excluded (n=990; 5% of records) • Did not meet "pedestrian-involved" inclusion criteria (n=445) • Did not meet "unintentional" inclusion criteria (n=157) • Did not meet "MV-involved" inclusion criteria (n=264) • Insufficient details to confirm inclusion (n=124)	Excluded (n=3,102; 20% of records) • Did not meet "public right of way" inclusion criteria (n=2,012) • Did not meet "unintentional" inclusion criteria (n=622) • Did not meet "MV-Involved" inclusion criteria (n=161) • Did not meet "Manned-MV" inclusion criteria (n=161)	

Analysis

Analyzed 95% of original data set (n=17,369)	Analyzed 80% of original data set (n=12,646)	Analyzed 100% of original data set (n=993)

Figure 1. Approach used to screen police-reported, emergency department, and death certificate pedestrian fatality or injury cases for eligibility in the study.
Note: ED = emergency department; ICD = International Classification of Diseases.

then single data sources are still potentially useful for program evaluation.

A priori, it was hypothesized that: 1) pedestrian injury and fatality frequencies would be higher in the ED data relative to the police-reported data; 2) ED and police data would have similar seasonal/temporal distributions but that demographic distributions may differ because of factors influencing event reporting; and 3) death certificate data would be consistent with police data but not with emergency room data. Hypothesis 1 was suggested by the reasoning that ED data records represent visits, and one injured pedestrian may generate multiple visit records, whereas the police-reported records represent crashes, and multiple pedestrians may be injured in a single crash. Hypothesis 2 reflects the observation that police records may under-ascertain pedestrian injuries involving young children, since these are less likely to occur in the public right-of-way and may be excluded from police report databases, while ED data may over-ascertain child-involved injuries because of higher levels of care-seeking by parents and childcare providers. Hypothesis 3 was based on the rationale that a police report is very likely to be filed in the event of a fatal crash (or a crash where a serious injury could lead to a fatality), but a fatally injured pedestrian would likely not be taken to the ED if the victim was already deceased.

Methods

Data Sources, Case, and Injury Definitions

This study drew from three existing statewide data sources in NC: police-reported pedestrian crash data (fatal and non-fatal crashes), ED data (fatal and non-fatal patient visits), and death certificate data (fatal events only). Figure 1 illustrates the process used to define cases and exclude non-relevant records in each of the datasets.

Police-Reported Pedestrian Crashes

Police-reported pedestrian crash data were obtained from the NC Division of Motor Vehicles (DMV). Data originate from the Crash Report Form DMV-349, which is completed by law enforcement officers to report motor vehicle crashes (MVCs) in NC. For a crash to be reportable, it must meet at least one of the following criteria (*33*):

• The crash resulted in a fatality, or

- The crash resulted in a non-fatal personal injury, or
- The crash resulted in total property damage amounting to $1,000.00 or more, or
- The crash resulted in property damage of any amount to a vehicle seized, or
- The vehicle has been impounded by police (e.g., abandoned vehicles).

Additionally, reportable MVCs must occur on a public trafficway (33).

All police-report crashes between January 1, 2007, and December 31, 2012, where one or more of the units involved in the crash was classified as a pedestrian on the DMV-349 form were identified. Police-reported data were run through a quality control process which included reviewing each crash report form and full narrative and applying the Pedestrian Bicycle Crash Analysis Tool (PBCAT) to add additional data to each record in relation to the nature of the crash, or crash type. PBCAT is a commonly applied crash typing framework used to classify the pre-crash actions and locations of the pedestrians and drivers involved in the crash based on the crash narrative and other form information (34).

Emergency Department Visits

Statewide ED data were obtained from the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT). NC DETECT is NC's statewide syndromic surveillance system and is one of the most comprehensive near real-time statewide ED systems in the U.S. (35). Staff at the Carolina Center for Health Informatics (CCHI) in the University of North Carolina (UNC) Department of Emergency Medicine review and monitor the quality of the data and develop and manage NC DETECT. NC DETECT captures all visits to a 24/7 civilian acute-care, hospital-affiliated NC ED.

International Classification of Diseases, 9th Revision (ICD-9-CM) E-codes (external cause of injury codes) were used to identify all ED visits between January 1, 2007, and December 31, 2012, for pedestrian-involved MVCs, E810-E819 (7). While E-coding data is not mandated for NC hospitals, E-codes are available for nearly 90% of NC ED records with one or more injury diagnoses (defined as an ICD-9-CM diagnosis code between 800 and 999) (36). Once the data were acquired, additional review of the chief complaint data field was conducted to ensure that included records met the intended definition of "pedestrian." Cases were excluded if there was definitive or strong evidence in the chief complaint data field that the visit was not an unintentional MVC involving a pedestrian on a public roadway. Excluded scenarios included ED visits by injured bicyclists who

had collided with pedestrians, pedestrians struck intentionally in domestic violence situations, "pedestrians" who fell out of golf carts, and so forth.

Death Certificates

Death certificate data were provided by the NC Department of Health and Human Services. Data included all motor vehicle transportation (MVT) unintentional death cases from January 1, 2007, to December 31, 2012, with the "underlying cause of death" being pedestrian injured in transport accident, defined as ICD-10 codes V021, V031, V039, V041, V049, and V092. No cases were excluded from death certificate data as there was no additional means to verify this information. Death certificate data are subject to an extensive set of data quality checks by NC Vital Records, a unit within the NC Division of Public Health, and are a very complete source of data for pedestrian and other traffic-related fatalities.

Statistical Analysis

This study used descriptive analyses to characterize injury distributions among demographic, temporal, and seasonal variables that were common to all three sources. For key variables, chi-square tests of homogeneity were used to compare distributions across data sources, with a null hypothesis that the datasets examined were homogenous with respect to event frequency or rate. Available demographic data included sex and age of the pedestrian. Age data were coded in 5-year groups for ED data because of patient privacy concerns; as such, the completed years of age data available in police and death certificate data were aggregated into the same 5-year categories. To calculate crash rates per person-years for each age group, 2007–2012 bridged-race population estimates from the National Center for Health Statistics (NCHS) were used (37). Additionally, police data included the age of the driver involved in the crash. Race/Hispanic ethnicity were available for police and death certificate data, but not for ED data, so were included in analyses.

Available temporal and seasonal variables included the date and time (hour) of police-reported crash, ED arrival date and time, and date and time of death in the death certificate data. The ED dataset included date and time (categorized in 2-h blocks) when the patient arrived at the emergency room. In comparison, the police-reported crash data provided an estimate of the actual time of the crash, based on police officer investigation. Time of ED arrival necessarily will lag the time and date of the actual crash. In some cases, as evidenced by the chief complaint notes, the patient presented to the ED

Table 1. Distribution of Pedestrian Injury from DC-Reported Deaths, Police-Reported Crashes, and ED-Reported Patient Visits in North Carolina, 2007–2012

Injury category[a]	DC death		Police-reported crashes		ED visits	
	Status	n (%)	Injury level	n (%)	Disposition	n (%)
Fatality	Died	993 (100%)	K	953 (7.5%)	Died	203 (1.2%)
Serious injury	na	na	A	961 (7.6%)	Admitted to hospital	1,656 (9.5%)
					Admitted to ICU	119 (0.7%)
					Admitted for psych care	7 (0.0%)
					Transferred between hospitals	675 (3.9%)
Other injury	na	na	B	4627 (36.6%)	Discharged from ED	13,551 (78%)
			C	4904 (38.8%)	Left AMA	123 (0.7%)
					Left without advice	97 (0.6%)
					Observation	54 (0.3%)
					Other	97 (0.6%)
Non-injury	na	na	O	702 (5.6%)	na	na
Missing	na	na	na	499 (3.9%)	na	787 (4.5%)
Total	na	993 (100%)	na	12,646 (100.0%)	na	17,369 (100%)

Note: AMA = against medical advice; DC = death certificate; ED = emergency department; ICU = intensive care unit; K = fatal injury; A = serious injury; B = moderate/minor injury; C = possible injury; O = no injury; na = not applicable.
[a]Injury category was defined by the research team. It is acknowledged that some of the injuries determined by the police to be serious (A) may not be admitted or transferred. Likewise, some of the injuries determined by the police to be less serious (B, C) or non-injury (O) may be transferred, admitted, or die.

more than a week after the incident that caused the injury. Nevertheless, the police-reported crash times were categorized into the same 2-h blocks as the ED data, without a lag function (since the average time from crash to ED arrival is unknown). Events occurring at typical "peak times"—times when the volume of vehicle and pedestrian commuter traffic is the highest—were defined as those from 6:00 a.m. to 9:59 a.m. and 4:00 p.m. to 7:59 p.m.

The ED and death certificate datasets contained additional information about the injury diagnoses, and the police-reported data contained numerous variables in relation to conditions at the time and location of the crash (such as lighting, weather, roadway facilities, traffic conditions, etc.). However, since these variables were not available for comparison across all three data sources, they were not included in this analysis.

For police-reported data, the analysis was conducted at the crash-event level, using the characteristics of the first pedestrian harmed in each crash event (assumed to be the most severely injured person) if multiple pedestrians were involved. It therefore undercounts pedestrians involved in multi-pedestrian crash events. For ED data, analyses were conducted at the patient visit level, using the characteristics of each person involved in any visit to the ED after a crash. Whereas police-reported data tend to undercount injured pedestrians, the ED data tend to over-count injured pedestrians seen in the ED, since multiple visits for treatment of injuries from a single crash event will generate multiple visit records per patient. For

ED data, there is no publicly available unique patient identifier to facilitate analyses at the level of the patient (rather than the visit).

Table 1 provides an overview of the injury levels that were assumed, for this study, to be approximately equivalent. Fatal injury events were examined separately from non-fatal injury events. Within the police report database, pedestrian injury is coded using the KABCO scale, which is a measure of the injury level of the victim at the crash scene based on police officer judgment when investigating the crash. In the ED data, there is no readily available equivalent injury scale. Rather, it was chosen to categorize ED data based on the patient disposition: Died, Admitted to the Intensive Care Unit (ICU) (reflecting more serious injuries requiring immediate medical attention), Admitted (reflecting injury requiring medical attention), Transferred (to another ED), Discharged or Left (reflecting less serious injuries), Observation (typically for follow-up care), and Other. It is acknowledged that some of the injuries determined by the police to be serious (A) may not be admitted or transferred through an ED. Likewise, some of the injuries determined by the police to be less serious (B, C) or non-injury (O) may be transferred, admitted, or die. These categorizations are clearly not equivalent, but they represent the closest approximations of injury outcome possible within the limitations of the individual datasets, and they represent a reasonable working assumption that was used to generate outcome categorizations from a population health perspective.

Results

After data cleaning (Figure 1) there were 17,369 ED pedestrian injury visits, 12,646 police-reported crashes, and 993 pedestrian deaths reported in death certificate data for the period 2007–2012 in NC.

Injury Severity

Table 1 provides the distribution of injury level among the different data sources. The 953 police-reported fatalities accounted for 7.5% of the police-reported crashes. The 203 ED pedestrian deaths (1.2% of all ED visits) represent 20% of the death certificate fatalities and 21% of the police-reported fatalities in the 6-year period.

In the police-reported crashes, 7.6% reflected a disabling or incapacitating injury (which plausibly may have required hospital admission), slightly less than the 14.1% of ED patients that were actually admitted or transferred to another facility. Of all ED visits, 79.3% were discharged (i.e., less serious injuries), while 75.4% of police crash reports indicate that the injury was possible or non-incapacitating.

Demographics

Injury distributions by sex and age (Table 2) were compared across all three data sources. The distribution of injuries appeared similar across the data sets in relation to pedestrian sex. In police reports, death certificates, and ED data, male pedestrians were involved in over 60% of the injury events, and the proportion increased to over 70% for the fatal events.

Although details on drivers were not available in the ED or death certificate data, an analysis of driver characteristics from the police reports shows that most of the drivers involved in pedestrian crashes were male. The proportion of male drivers increased as the injury severity increased, with males accounting for 67.8% of drivers in fatal crashes. For all injury levels, drivers aged 20–39 years were involved in the greatest proportion of crashes relative to other age categories.

In comparing the distribution of pedestrian age (in 5-year groups) for fatal events, the police and death certificate data were never more than a percentage point different. A test of homogeneity confirmed that any differences in the age distributions among the two data sets were not statistically significant. In contrast, the ED fatality data—which represents only a portion of actual fatalities—showed a larger proportion of young pedestrians (0–9-year-olds) and older adults (75 years or older), and fewer people age 20–39 than the death certificates or police reports.

The distribution of age in the ED data was significantly different than the police-reported data for severe and other injury events as well. For severe crashes, the ED data contained a slightly higher proportion of visits involving young, middle-age, and older adult pedestrians. But for the "other injury" events, the ED data contained higher proportions of visits involving adults age 20–39 years.

Pedestrian crash rates per 100,000 person years were calculated for each 5-year age group using National Vital Statistics System (NVSS) bridged-race population estimates (Figure 2). For all data sources (at all levels of injury), pedestrian crash rates were highest among 20 to 24-year-olds. The distribution of pedestrian crash frequency by age group (Table 2) was generally consistent with the distribution of crash rates by age group for severe and other injury crashes, meaning crash frequencies were also highest among young adult pedestrians. However, for fatal crashes, middle-age adults (between 40 and 59 years of age) had the highest crash frequencies.

Temporal and Seasonal Trends

Table 3 provides a summary of the temporal and seasonal distributions. Most police-reported crashes and patient ED visits (between 58.2% and 68.7%) took place during "off-peak" commute times. However, the police data reflected slightly more events taking place during peak commute times for non-fatal crashes than the ED data. This may be a reflection that the time of the crash is inherently different than the time at which a patient arrives at an ED. For injuries and fatalities, most police- and ED-reported events occurred on a weekday (65.6%–76.4%) compared with a weekend (23.6%–34.4%).

There were strong similarities between the datasets in relation to other temporal and seasonal distributions (Figure 3). For example, both police and ED datasets indicated an increase in "other injury" (or less severe) events in almost every year from 2007 to 2012. In police, death certificate, and ED data, the calendar quarter from October to December had the highest frequency of pedestrian injury and fatality events, ranging between 28.3% and 31.4% of all crashes/patient visits (Table 3). For "other injury" crash events, the January and February months indicated a larger difference between the ED and police-reported frequencies than other months. The difference in distribution between the two data sets was significant ($p \leq 0.01$) for "other injury" crashes but not for serious or fatal crashes (Figure 3).

Comparing ED, Police, and Death Certificate Datasets

Through the data cleaning and analysis, several discrepancies were identified—particularly between ED and police-reported data—that may have impacted the differences in injury frequencies and distributions observed across the data sets (Table 4).

**Table 2.** Driver and Pedestrian Demographics of Police, ED, and Death Certificate Reported Pedestrian Injury Events Occurring from 2007 to 2012 in North Carolina

	Fatal crashes				Severe injury events				Other injury events			
	Police-reported crashes[a]		Death certificates[c]		Police-reported crashes[a]		ED visits[b]		Police-reported crashes[a]		ED visits[b]	
	n	%	n	%	n	%	n	%	n	%	n	%
Pedestrian age												
0–19 years	103	11.0	115	11.6	201	21.2	588	23.9	2,283	24.3	3,236	23.2
20–39 years	334	35.7	345	34.8	354	37.4	836	34.0	3,479	37.0	6,039	43.4
40–59 years	354	37.8	369	37.2	289	30.5	688	28.0	2,708	28.8	3,613	26.0
60+ years	145	15.5	162	16.3	103	10.9	345	14.0	942	10.0	1,034	7.4
Total	936	100.0	991	100.0	947	100.0	2457	100.0	9,412	100.0	13,922	100.0
Pedestrian sex												
Male	684	71.8	720	72.5	656	68.3	1621	66.0	5,722	60.1	8,410	60.4
Female	268	28.2	273	27.5	304	31.7	835	34.0	3,794	39.9	5,508	39.6
Unknown	0	0.0	0	0.0	0	0.0	1	0.0	0	0.0	3	0.0
Total	952	100.0	993	100.0	960	100.0	2457	100.0	9,516	100.0	13,921	100.0
Driver age[d]												
5–19 years	66	7.4	NA	NA	90	9.4	NA	NA	651	6.8	NA	NA
20–39 years	392	43.8	NA	NA	375	39.0	NA	NA	3,239	34.0	NA	NA
40–59 years	288	32.1	NA	NA	272	28.3	NA	NA	2,510	26.3	NA	NA
60+ years	150	16.7	NA	NA	149	15.5	NA	NA	1,746	18.3	NA	NA
Total	896	100.0	NA	NA	961	100.0	NA	NA	9,531	100.0	NA	NA
Driver sex[d]												
Male	598	67.8	NA	NA	512	59.3	NA	NA	4,365	55.2	NA	NA
Female	284	32.2	NA	NA	352	40.7	NA	NA	3,536	44.8	NA	NA
Total	882	100.0	NA	NA	864	100.0	NA	NA	7,901	100.0	NA	NA

Note: ED = emergency department; NA = not available.

[a]There were 499 police records missing the pedestrian injury classification; an additional 702 were excluded because of lack of injury sustained in the crash; 164 were missing data on sex, and 150 were missing data on age.

[b]There were 787 records missing the ED disposition classification; one record was also missing age.

[c]Two death certificate records were missing age.

[d]2,029 police records were missing information on driver sex, and 2,016 records were missing information on driver age.
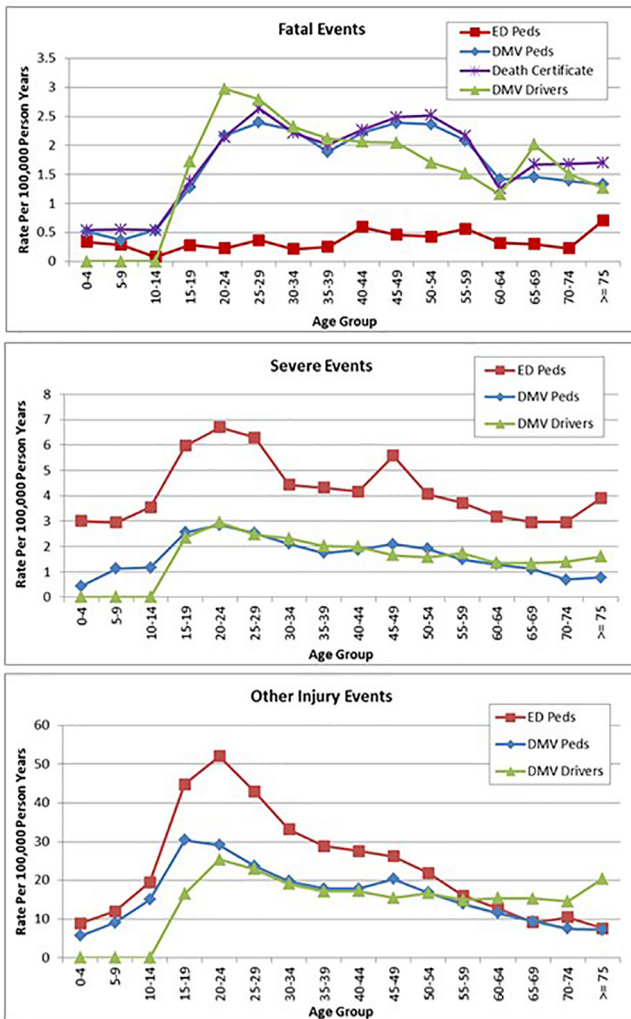
Figure 2. Police-reported rates of pedestrian crashes by pedestrian and driver age compared with emergency department and death certificate reported pedestrian age for North Carolina crashes occurring from 2007 to 2012.
Note: DMV = Division of Motor Vehicles; ED = emergency department.

While the frequency of events reported in the police and ED data did not always align perfectly because of the nature of the database structures, the distribution of events by sex, seasonal, and temporal variables was remarkably similar between police and ED data. Studies seeking to assess the relative distributions by sex, seasonality, or temporal variables could consider either ED or police data as an appropriate data source. However, the distribution of events by age group in ED and police data did not align closely. As evidenced by Figure 2, the choice of data source when calculating pedestrian injury rates could largely impact the conclusions drawn from the data, particularly for pedestrians at either end of the age spectrum, where ED and police reporting differences were most pronounced.

Discussion

The results reported here confirm previous studies that have noted over-representation of males and vulnerable age groups in pedestrian crashes, and distributions by day of the week and seasonality. For temporal/seasonal distributions, it is important to note that these factors do not necessarily reflect higher risk (e.g., on weekdays, at off-peak times, in October–December, etc.), but rather are likely to be markers of "exposure," indicating when pedestrian activity, vehicle activity, or both is higher. Likewise, the distribution of injuries by age reflects, in part, differences in pedestrian exposure to traffic by age. As mentioned earlier, it is a major limitation that the U.S. lacks a system for the surveillance of pedestrian exposure.

Table 3. Seasonal and Temporal Attributes of Police, ED, and Death Certificate Reported Pedestrian Injury Events Occurring from 2007 to 2012 in North Carolina

	Fatal events				Severe injury events				Other injury events			
	Police-reported crashes[a]		Death certificates		Police-reported crashes[a]		ED visits[b]		Police-reported crashes[a]		ED visits[b]	
	n	%	n	%	n	%	n	%	n	%	n	%
Year												
2007	158	16.6	189	19.0	186	19.4	425	17.3	1,500	15.7	1,968	14.1
2008	155	16.3	175	17.6	165	17.2	375	15.3	1,545	16.2	2,207	15.9
2009	145	15.2	143	14.4	133	13.8	360	14.7	1,462	15.3	2,223	16.0
2010	161	16.9	156	15.7	140	14.6	427	17.4	1,535	16.1	2,362	17.0
2011	156	16.4	156	15.7	148	15.4	413	16.8	1,656	17.4	2,573	18.5
2012	178	18.7	174	17.5	189	19.7	457	18.6	1,833	19.2	2,589	18.6
Month												
Jan–Mar	206	21.6	219	22.1	232	24.1	577	23.5	2,213	23.2	2,952	21.2
Apr–Jun	186	19.5	210	21.1	208	21.6	538	21.9	2,212	23.2	3,395	24.4
Jul–Sept	262	27.5	272	27.4	233	24.2	622	25.3	2,291	24.0	3,632	26.1
Oct–Dec	299	31.4	292	29.4	288	30.0	720	29.3	2,815	29.5	3,943	28.3
Day												
Weekday	642	67.4	651	65.6	680	70.8	1,713	69.7	7,277	76.4	10,000	71.8
Weekend	311	32.6	342	34.4	281	29.2	744	30.3	2,254	23.6	3,922	28.2
Time of day												
Peak	298	31.3	NA	NA	383	39.9	840	34.2	3,997	41.9	5,261	37.8
Off-peak	655	68.7	NA	NA	578	60.1	1,617	65.8	5,533	58.1	8,661	62.2
Total	953	100.0	993	100.0	961	100.0	2,457	100.0	9,531	100.0	13,922	100.0

Note: ED = emergency department; NA = not available.
[a]There were 499 records missing the pedestrian injury classification; an additional 702 were excluded because of lack of injury sustained in the crash; one record was missing time of day.
[b]There were 787 records missing the ED disposition classification.
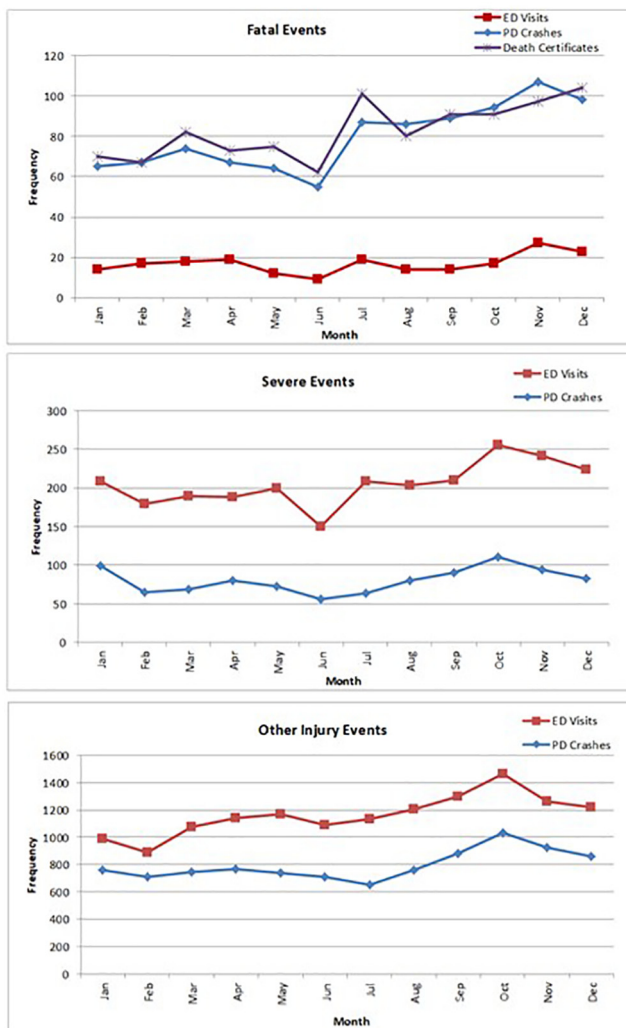
Figure 3. Police-reported rates of pedestrian crashes by month compared with emergency department and death certificate reported month for North Carolina crashes occurring from 2007 to 2012.
Note: DMV = Division of Motor Vehicles; ED = emergency department; PD = police department.

Police records are a commonly used data source among transportation safety practitioners and researchers. In comparison with the death certificate data, police records appear to undercount fatalities among pedestrians who are under the age of 10, or 75 years or older. This is an important limitation of police data, though the differences in this study were small. As another study indicated, many crashes involving children and older adults occur in driveways and parking lots rather than near on-road facilities; for this reason, they may be excluded from some police reporting (*38*).

The police data also underrepresent the total number of people injured in MVCs. Examples of how underreporting occurs are provided in Table 4, but the degree to which it occurs is not well understood. Many pedestrian injuries and fatalities may not meet the criteria for reporting outline in NC DMV-349. A prior NC study, performed in the late 1990s, linked police and ED cases (*39*). The incidents captured in the ED data but not captured in police data included falls, crashes not involving motor vehicles, crashes involving motor vehicles that did not occur on a public roadway, and crashes that did not meet the police criteria for reporting a crash. That study estimated that police records represented only about 56% of pedestrian injury incidents that were included in the ED data.

Death certificate data on pedestrian fatalities was largely consistent with police data and it appears that either would be a reliable data source for calculating fatality counts and rates by age. Death certificate data included E-codes on the circumstances of injury, but, overall, provided a limited picture on the nature of where

Table 4. Sources of Discrepancies Between ED and Police-Reported Pedestrian Injury Events

Source	Discrepancy	Example
Location of the event	Police-reported crashes reflect events that occurred on NC roadways to which a NC police officer responded. ED data can include crashes that occur in NC or outside of NC, provided that the patient presents to a NC hospital.	If a pedestrian is injured in a crash that occurred in South Carolina, but the nearest hospital is in NC, then the ED but not the police data would include the event.
Criteria used to report a crash	ED and police data have different criteria for reporting an event. Police may not report a crash if it did not involve property damage above $1,000 (and a clear injury).	If a pedestrian does not appear injured at the scene of the crash and no police report is filed, but later decides to go to the ED, then the event is only captured in the ED data.
Database structure: crash event versus patient visit	ED data may capture multiple patient visits for one injury, and multiple patients involved in the same crash event.	If a person presented at an ED, was transferred to another location, admitted for care, and then returned for observation, they would have multiple records in the ED database, but only one crash recorded in the police database.
Lack of data to support inclusion/exclusion based on a comparable case definition	Police-reported data, by its case definition, was limited to cases where the vehicle that struck the pedestrian was being driven by a person other than the pedestrian. This detail was not available within the ED data to use as an exclusion criterion.	If a person steps out of a car that is not in park and it rolls over them, injuring their foot, they would be included in the ED dataset but not in the police dataset. There were at least 88 ED cases where the complaint data says a car "ran over" or "rolled over" a person's foot or leg, but there is no clear indication of whether the car was actively being driven by a person other than the pedestrian involved.
Inaccurate E-code data, leading to improper inclusion/exclusion	E-codes were used to determine if an ED case involved a pedestrian and occurred in the public right of way (versus a private roadway); however, E-code data were missing for 3% (471 of 18,359) of the data and there was some evidence that E-codes were inconsistent or incompatible with information provided in the chief complaint field.	Though the E-codes were for crashes occurring on public roadways, there was still evidence of a few events occurring in private parking lots based on chief complaint notes. Similarly, a review of chief complaint data showed that 2% of cases clearly did not involve a pedestrian (but rather a bicyclist) and another 3% indicated that the crash could have involved a pedestrian or bicyclist. There may be relevant "pedestrian" cases that were given a bicycle E-code, in which case they would never appear in the ED dataset used.

Note: ED = emergency department; NC = North Carolina.

or when fatalities occur. Death certificates also represent only a fraction (8%) of all injurious pedestrian crashes.

Study Strengths and Limitations

A strength of this study was that it made use of three existing, high-quality population-based datasets to draw comparisons and examine pedestrian injury distributions in NC. It is important to maximize the completeness of information on the nature of pedestrian crashes available to decision-makers and health practitioners, so that they can make informed policy decision and develop appropriate intervention programs. The study detailed a process for reconciling inconsistencies in the coverage of cases from the ED and police data sets, such as those involving intentional injuries (assaults with a vehicle), and those not involving motor vehicles (such as falls and crashes involving toy and non-motorized vehicles), so that police and ED data sources can be meaningfully compared. Rather than rely exclusively on E-codes in ED data, an extensive review of chief complaint data and other ED data available was performed to assess the

eligibility of each case for inclusion in the analysis. This effort was time-intensive but led to a dataset that more clearly met the intended case definition and provided a better understanding of the reasons for potential discrepancies both within ED data variables and between ED and police datasets.

The study had limitations as well, mainly because of the nature of the data available for use. This study demonstrated that it is not possible to develop a perfectly comparable case definition, and, in part, this reflects that some data were missing, unavailable, or inaccurate, particularly in the ED data. There is a need to consistently monitor and improve the quality and consistency of ED data entry so that it can be used to for epidemiologic studies. For example, only 20% (of the original 18,359 ED records) contained chief complaint data that clearly indicated that a pedestrian was involved. In 55% of the ED records there was chief complaint information that made no specific reference to a pedestrian, so an assumption had to be made based exclusively on the presence of a pedestrian-related E-code. Future efforts to link the police and ED data sets should address this issue, potentially by accessing more detailed data (such as electronic triage notes). Another limitation is the age of the datasets used for analysis. This study used 2007–2012 data, because these were the only years in which all data sources were readily available to the researchers. An ongoing study by Harmon et al. comparing crash and ED visit data has yielded similar results, with the number of pedestrian crash injury-related ED visits exceeding the number of police-reported crashes for the years 2010–2015 (personal communication).

A limitation of the analysis approach was that it relied on case-only data, so crash and injury distributions reflected frequencies but cannot be interpreted to reflect crash rates or risks. The association of case frequencies with certain variables may reflect exposure (or opportunity to be involved in a crash), such as the amount of walking/driving a person does, or the characteristics of the roadways where travel occurs, rather than actual risk per se. Unfortunately, such data were not routinely available, which is staggering in view of the magnitude of the problem of pedestrian injury (approximately 6,000 deaths and 137,000 non-fatal injuries in the U.S. annually) (40). The authors fully support future efforts to obtain better measures of exposure to help assess pedestrian risks.

Conclusion

Overall, both police- and ED-reported injury events had comparable distributions in relation to pedestrian sex, seasonal, and temporal factors related to the injury event. This provides some evidence that communities (and possibly researchers) lacking one of the data sources could rely

on the other to obtain an approximate sense of what groups are involved in pedestrian crashes and when they occur. Similarly, police reports of fatal crashes aligned very closely to death certificate data. In relation to pedestrian age distributions, there were larger discrepancies between police and ED data—particularly among the youngest and oldest pedestrians—and more research is needed to determine the extent to which under/overcounting occurs and how the data sets can be improved.

Police-reported crash data typically provide a great deal of information in relation to crash location characteristics, pre-crash actions, and characteristics of the drivers involved in the events. They also provide a more reliable source of data on the actual time and date that the injury event occurred. However, police data provide only rudimentary information on the injuries and their severity. In contrast, ED data can provide a more complete description of injury and medical outcomes, and they may better capture the incidence of pedestrian injury events than the police dataset but lack details on the nature and circumstances of the crash event. Since police-reported pedestrian crashes have been geo-located statewide in NC since 2007, they could also be linked to additional spatial data—such as roadway inventories, land use data, and socio-demographic characteristics associated with spatial features—that could provide insights into pedestrian exposure to different facility types or built environment features. Both police and ED data provide relevant and complementary insights in relation to the nature of pedestrian crashes and injury outcomes and, ideally, their ongoing linkage should be part of routine surveillance to support injury prevention planning efforts.

Data Accessibility Statement

NC DETECT is a statewide public health syndromic surveillance system, funded by the NC Division of Public Health Federal Public Health Emergency Preparedness Grant and managed through collaboration between the NC Division of Public Health and the University of North Carolina Chapel Hill Department of Emergency Medicine's Carolina Center for Health Informatics.

References

1. National Highway Traffic Safety Administration. *2017 Fatal Motor Vehicle Crashes: Overview*. Publication DOT HS 812 603. U.S. Department of Transportation, Washington, D.C., 2018.
2. Insurance Institute for Highway Safety. *Fatality Facts 2018: Pedestrians*. https://www.iihs.org/topics/fatality-statistics/detail/pedestrians. Accessed April 30, 2020.
3. National Highway Traffic Safety Administration. *Pedestrians*. Publication DOT HS 812 681. U.S. Department of Transportation, Washington, D.C., 2019.
4. Federal Highway Administration. *Safer People, Safer Streets: Summary of U.S. Department of Transportation Action Plan to Increase Walking and Biking and Reduce Pedestrian and Bicyclist Fatalities*. U.S. Department of Transportation, Washington, D.C., 2014.
5. National Center for Safe Routes to School. *Creating Healthier Generations: A Look at 10 Years of the Federal Safe Routes to School Program*. U.S. Department of Transportation, Washington, D.C., 2015.
6. Dimaggio, C., and G. Li. Effectiveness of a Safe Routes to School Program in Preventing School-Aged Pedestrian Injury. *Pediatrics*, Vol. 131, No. 2, 2013, pp. 290–296.
7. Muennig, P. A., M. Epstein, G. Li, and C. DiMaggio. The Cost-Effectiveness of New York City's Safe Routes to School Program. *American Journal of Public Health*, Vol. 104, No. 7, 2014, pp. 1294–1299.
8. National Highway Traffic Safety Administration. *Linking Traffic Records Data Systems*. Publication DOT HS 812 056. U.S. Department of Transportation, Washington, D.C., 2014.
9. McAndrews, C., K. Beyer, C. E. Guse, and P. Layde. Revisiting Exposure: Fatal and Non-Fatal Traffic Injury Risk across Different Populations of Travelers in Wisconsin, 2001–2009. *Accident; Analysis and Prevention*, Vol. 60, 2013, pp. 103–112.
10. McAndrews, C., K. Beyer, C. Guse, and P. Layde. Linking Transportation and Population Health to Reduce Racial and Ethnic Disparities in Transportation Injury: Implications for Practice and Policy. *International Journal of Sustainable Transportation*, Vol. 11, No. 3, 2017, pp. 197–205.
11. Kindelberger, J., and J. A. Milani. *Crash Outcome Data Evaluation System (CODES): Program Transition and Promising Practice*. Publication DOT HS 812 178. U.S. Department of Transportation, Washington, D.C., 2015.
12. Milani, J., J. Kindelberger, G. Bergen, E. J. Novicki, C. Burch, S. M. Ho, and B. A. West. *Assessment of Characteristics of State Data Linkage Systems*. Publication DOT HS 812 180. National Highway Traffic Safety Administration, Washington, D.C., 2015, p. 77.
13. Ballesteros, M. F., P. C. Dischinger, and P. Langenberg. Pedestrian Injuries and Vehicle Type in Maryland, 1995–1999. *Accident; Analysis and Prevention*, Vol. 36, No. 1, 2004, pp. 73–81.
14. Sciortino, S., M. Vassar, M. Radetsky, and M. M. Knudson. San Francisco Pedestrian Injury Surveillance: Mapping, Under-Reporting, and Injury Severity in Police and Hospital Records. *Accident; Analysis and Prevention*, Vol. 37, No. 6, 2005, pp. 1102–1113.
15. Tarko, A., and M. S. Azam. Pedestrian Injury Analysis with Consideration of the Selectivity Bias in Linked Police-Hospital Data. *Accident; Analysis and Prevention*, Vol. 43, No. 5, 2011, pp. 1689–1695.
16. Dasu, R. *Traffic Safety Reports: Pedestrian Injuries in California 2007–2013*. California Department of Health, Sacramento, 2017.
17. Roe, M., H. S. Shin, A. Blatt, and K. Majka. *The New York City Pedestrian Safety Study & Action Plan: Technical Supplement*. New York City Department of Transportation, 2010.
18. Haleem, K., P. Alluri, and A. Gan. Analyzing Pedestrian Crash Injury Severity at Signalized and Non-Signalized Locations. *Accident; Analysis and Prevention*, Vol. 81, 2015, pp. 14–23.
19. Quistberg, D. A., E. J. Howard, B. E. Ebel, A. V. Moudon, B. E. Saelens, P. M. Hurvitz, J. E. Curtin, and F. P. Rivara. Multilevel Models for Evaluating the Risk of Pedestrian-Motor Vehicle Collisions at Intersections and Mid-Blocks. *Accident; Analysis and Prevention*, Vol. 84, 2015, pp. 99–111.
20. Slaughter, D. R., N. Williams, S. P. Wall, N. E. Glass, R. Simon, S. R. Todd, O. S. Bholat, S. Jacko, M. Roe, C. T. Wilson, D. A. Levine, G. Marshall, P. Ayoung-Chee, H. L. Pachter, and S. G. Frangos. A Community Traffic Safety Analysis of Pedestrian and Bicyclist Injuries Based on the Catchment Area of a Trauma Center. *The Journal of Trauma and Acute Care Surgery*, Vol. 76, No. 4, 2014, pp. 1103–1110.

21. Wheeler-Martin, K., S. J. Mooney, D. C. Lee, A. Rundle, and C. DiMaggio. Pediatric Emergency Department Visits for Pedestrian and Bicyclist Injuries in the US. *Injury Epidemiology*, Vol. 4, No. 1, 2017, p. 31.

22. Chong, S. L., L. W. Chiang, J. C. Allen, E. W. Fleegler, and L. K. Lee. Epidemiology of Pedestrian-Motor Vehicle Fatalities and Injuries, 2006–2015. *American Journal of Preventive Medicine*, Vol. 55, No. 1, 2018, pp. 98–105.

23. Hickox, K. L., N. Williams, L. F. Beck, T. Coleman, J. Fudenberg, B. Robinson, and J. Middaugh Middaugh, and Centers for Disease Control and Prevention (CDC). Pedestrian Traffic Deaths among Residents, Visitors, and Homeless Persons–Clark County, Nevada, 2008–2011. *MMWR. Morbidity and Mortality Weekly Report*, Vol. 63, No. 28, 2014, pp. 597–602.

24. Jang, K., S. H. Park, S. Chung, and K. H. Song. *Influential Factors on Level of Injury in Pedestrian Crashes: Applications of Ordered Probit Model with Robust Standard Errors*. UC Berkeley Safe Transportation Research and Education Center, Berkeley, CA, 2010.

25. Statter, M., T. Schuble, M. Harris-Rosado, D. Liu, and K. Quinlan. Targeting Pediatric Pedestrian Injury Prevention Efforts: Teasing the Information through Spatial Analysis. *The Journal of Trauma*, Vol. 71, No. 5, 2011, pp. S511–S516.

26. Rice, T. M., R. B. Trent, K. Bernacki, J. K. Rice, B. Lovette, E. Hoover, J. Fennell, A. Z. Aistrich, D. Wiltsek, E. Corman, C. L. Anderson, and J. Sherck. Trauma Center-Based Surveillance of Nontraffic Pedestrian Injury among California Children. *The Western Journal of Emergency Medicine*, Vol. 13, No. 2, 2012, pp. 139–145.

27. McElroy, L. M., J. J. Juern, A. Bertleson, Q. Xiang, A. Szabo, and J. Weigelt. A Single Urban Center Experience with Adult Pedestrians Struck by Motor Vehicles. *WMJ: Official Publication of the State Medical Society of Wisconsin*, Vol. 112, No. 3, 2013, pp. 117–122; quiz 123.

28. Agran, P. F., C. Anderson, D. Winn, R. Trent, L. Walton-Haynes, and S. Thayer. Rates of Pediatric Injuries by 3-Month Intervals for Children 0 to 3 Years of Age. *Pediatrics*, Vol. 111, No. 6, 2003, pp. e683–e692.

29. Cherry, C., A. M. Hezaveh, M. Noltenius, A. Khattak, L. Merlin, E. Dumbaugh, D. Ragland, and L. Sandt. *Completing the Picture of Traffic Injuries: Understanding Data Needs and Opportunities for Road Safety*. Publication CSCRS-R4. Collaborative Sciences Center for Road Safety, Chapel Hill, NC, 2018.

30. Doggett, S., D. R. Ragland, and G. Felschundneff. *Evaluating Research on Data Linkage to Assess Underreporting of Pedestrian and Bicyclist Injury in Police Crash Data*. UC Berkeley Safe Transportation Research and Education Center, CA, 2018.

31. Farmer, C. M. Reliability of Police-Reported Information for Determining Crash and Injury Severity. *Traffic Injury Prevention*, Vol. 4, No. 1, 2003, pp. 38–44.

32. Burdett, B., Z. Li, A. R. Bill, and D. A. Noyce. Accuracy of Injury Severity Ratings on Police Crash Reports. *Transportation Research Record: The Journal of the Transportation Research Board*, 2015. 2516: 58–67.

33. Traffic Records Branch, North Carolina Division of Motor Vehicles. *North Carolina Crash Report Instruction Manual*. North Carolina Department of Transportation, Raleigh, 2018.

34. Harkey, D. L., S. Tsai, L. Thomas, and W. W. Hunter. *PBCAT - Pedestrian and Bicycle Crash Analysis Tool Version 2.0*. Publication FHWA-HRT-06-090. Federal Highway Administration, McLean, VA, 2006.

35. Schinasi, L., A. E. Waller, A. Ising, and J. E. Tintinalli. *Medicine Carolina Center for Health Informatics Report: Overview and Analysis of NC DETECT Emergency Department Visit Data: 2009*. Carolina Center for Health Informatics, Chapel Hill, NC, 2011.

36. Harmon, K. J., A. E. Waller, C. Barnett, S. K. Proeschold-bell, S. Marshall, and A. J. Dellapenna. *Carolina Center for Health Informatics Report: Overview and Analysis of NC DETECT Emergency Department Visit Data for Injuries: 2010*. Carolina Center for Health Informatics, Chapel Hill, NC, 2012.

37. National Center for Health Statistics, Centers for Disease Control and Prevention. *Bridged-Race Population Estimates Data Files and Documentation*. https://wonder.cdc.gov/bridged-race-population.html. Accessed July 14, 2020.

38. Mueller, B. A., F. P. Rivara, and A. B. Bergman. Factors Associated with Pedestrian-Vehicle Collision Injuries and Fatalities. *The Western Journal of Medicine*, Vol. 146, No. 2, 1987, pp. 243–245.

39. Stutts, J., and W. Hunter. Police Reporting of Pedestrians and Bicyclists Treated in Hospital Emergency Rooms. *Transportation Research Record: Journal of the Transportation Research Board*, 1998. 1635: 88–92.

40. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. *Pedestrian Safety*. https://www.cdc.gov/motorvehiclesafety/pedestrian_safety/index.html. Accessed April 30, 2020.