Title
What is counted counts: An innovative linkage of police, hospital, and spatial data for transportation injury prevention.

Authors
Soltani, Shamsi
Schwarcz, Leilani
Morris, Devan
et al.

Publication Date
2022-12-01

DOI
10.1016/j.jsr.2022.08.002

HHS Public Access

Author manuscript

J Safety Res. Author manuscript; available in PMC 2022 December 10.

Published in final edited form as:

J Safety Res. 2022 December ; 83: 35–44. doi:10.1016/j.jsr.2022.08.002.

What is counted counts: An innovative linkage of police, hospital, and spatial data for transportation injury prevention

Shamsi Soltani[a,*], **Leilani Schwarcz**[a], **Devan Morris**[a], **Rebecca Plevin**[b], **Rochelle Dicker**[a], **Catherine Juillard**[b], **Adaobi Nwabuo**[b], **Megan Wier**[a]

[a]San Francisco Department of Public Health, San Francisco, CA, United States

[b]Department of Surgery, Zuckerberg San Francisco General Hospital, University of California, San Francisco, CA, United States

Abstract

Introduction: Growing research indicates transportation injury surveillance using police collision reporting alone underrepresents injury to vulnerable groups, including pedestrians, cyclists, and people of color. This reflects differing reporting patterns and non-clinicians' challenge in accurately evaluating injury severity. To our knowledge, San Francisco is the first U.S. city to link and map hospital and police injury data. Analysis of linked data injury patterns informs interventions supporting traffic fatality and injury prevention goals.

Methods: Injury and fatality records 2013–2015 were collected from San Francisco Police, Emergency Medical Services (EMS), Medical Examiner, and Zuckerberg San Francisco General Hospital (ZSFG). Probabilistic linkage was conducted using LinkSolv9.0 on match variables collision/admission time, name, birthdate, sex, travel mode, and geographic collision location.

Results: From 2013-2015, this study identified 17,000+ transportation-related injuries on public roadways in San Francisco. Twenty-six percent (n = 4,415) appeared in both police and ZSFG sources. Linked injury records represent 39% of police records (N = 11,403) and 43% of hospital records (N = 10,223). Among hospital records, 34% of cyclist, 38% of motor vehicle occupant, 61% of pedestrian, and 54% of motorcyclist records linked with a police record. Linkage rate varied by travel mode even after controlling for injury severity. Transportation-injured ZSFG-treated patients lacking police reports were more often cyclists, male, Hispanic or Black, and less often occupants of motor vehicles compared to those with injuries captured only in police reports.

[*]Corresponding author at: Department of Epidemiology and Population Health, Stanford University, 300 Pasteur Drive, Alway Building, Stanford, CA 94304, United States. shamsi@stanford.edu (S. Soltani).

Conflict of interest
None reported.

CRediT authorship contribution statement
Shamsi Soltani: Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Leilani Schwarcz:** Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – review & editing, Project administration. **Devan Morris:** Validation, Resources, Data curation, Writing – review & editing. **Rebecca Plevin:** Methodology, Resources, Writing – review & editing, Supervision. **Rochelle Dicker:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision. **Catherine Juillard:** Methodology, Resources, Writing – review & editing, Supervision. **Adaobi Nwabuo:** Writing – review & editing, Project administration. **Megan Wier:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Conclusions: Incorporating hospital and EMS spatial data into injury surveillance systems historically reliant on police reports offers trifold benefits. First, linkage captures injuries absent in police data, adding data on populations empirically vulnerable to injury. Second, it improves injury severity assessment. Finally, linked data better informs and targets interventions serving injury-burdened populations and road users, advancing transportation injury prevention.

Practical applications: Linkage closes data gaps, improving ability to quantify injury and develop evidence-based interventions for vulnerable groups.

Keywords

Probabilistic linkage; Injury surveillance; Traffic collision; Vision Zero; Local health department

1. Introduction

In the United States, transportation injury surveillance relies predominantly on police collision injury reporting. The national rate of transportation injury surpasses that of other high-income countries, calling attention to the need for injury surveillance (World Health Organization, 2018, Cherry et al., 2018). However, using police collision reports alone to measure the burden of traffic injuries underrepresents and misclassifies injuries and their severity, in particular injuries suffered by vulnerable road users (such as pedestrians and bicyclists) and people of color (Elvik & Mysen, 1999; Langley et al., 2003; Lopez et al., 2012; Sciortino et al., 2005; Tin Tin et al., 2013). These gaps in police collision data are concerning because traffic injuries are preventable via proven measures, provided data are available to accurately target intervention efforts (Centers for Disease Control and Prevention [CDC], 2019; Sauber-Schatz et al., 2016). Linked datasets have the potential to improve our understanding of transportation injury and fatality risk factors, enabling prevention measures to be tailored to locations and local environmental factors associated with injury. In 1992, the National Highway Traffic Safety Administration (NHTSA) conducted probabilistic linkage of police and medical traffic injury data for the multistate Crash Outcome Data Evaluation System (CODES) project (National Highway Traffic Safety Administration, 2021). By 1996, this improved data source informed a report to the U.S. Congress on the benefits of protective devices in motor vehicle-involved crashes—i.e. seat belts and motorcycle helmets (National Highway Traffic Safety Administration, 1996). Since CODES, entities such as New York City and the state of California have undertaken linkage projects, a step additionally encouraged by the Centers for Disease Control and Prevention (California Department of Public Health, 2017; Centers for Disease Control and Prevention, 2019; Conderino et al., 2017). In San Francisco a prior linkage focused specifically on pedestrians in local trauma and state-reported collision records (Sciortino et al., 2005). Transportation injury data linkage enables injury burden estimate improvement, evaluation of changes in crash severity over time and comparison of injury severity between injury scales (Short & Caulfield, 2016; Tainter et al., 2020; Couto et al., 2016). These examples demonstrate the feasibility and targeted prevention opportunities of linkage, but they also highlight the one-off nature and finite time period of most linkage projects. Most jurisdictions in the United States lack the funding and staff resources, institutional knowledge and/or political will to undertake these probabilistic linkage projects and incorporate them into routine surveillance activities (Milani et al., 2015).

In 2014, San Francisco became the second city in the United States to adopt Vision Zero–now embraced by over 40 cities and growing nationwide (https://www.visionzerosf.org/, Vision Zero Network, 2021). Initially developed in Sweden, Vision Zero is a road safety paradigm focused on designing a safe transportation system that ensures predictable human errors do not have fatal or severe consequences. While the overall goal is to eliminate traffic deaths and reduce severe traffic-related injury, Vision Zero also prioritizes addressing the current disproportionate harm to vulnerable groups: older adults, people of color, children and those walking in low income communities (Fleisher et al., 2016; National Complete Streets Coalition and Smart Growth America, 2019; Sauber-Schatz et al., 2016). Focused on systemic change, Vision Zero initiatives require high-quality, reliable and representative injury surveillance to inform prevention.

Our objective is to create a comprehensive transportation-related injury surveillance system linking hospital, police, medical examiner, and ambulance response data to inform Vision Zero traffic injury prevention initiatives and policy in San Francisco. The specific aims of our data linkage are fourfold. First, we seek to show the feasibility of employing a novel linkage methodology incorporating multiple datasets to provide an improved understanding of the magnitude and patterns of traffic injury and fatality. Second, we aim to demonstrate that creating linked data set leads to improved identification of injuries otherwise unreported or underreported via traditional surveillance. Notably, this linkage allows individual-level comparison of injury severity as classified in police data to both clinically-assessed trauma severity and hospital admission status. Third, we map the linked injury data to the street intersection level. Finally, the geolocated linked dataset seeks to allow San Francisco's Vision Zero program to identify locations and populations best served by injury prevention interventions, informing major capital investments. The retention of unlinked records in this dataset will improve our capture of unreported crash injuries and understanding of reporting biases in the city, relative to standard police record-based surveillance. This endeavor differs from other projects in its access to personal identifiers (e.g., name and birthdate) in both hospital and police data sources and utilization of location of injury data from ambulance providers. To our knowledge, San Francisco is the first U.S. city to complete comprehensive geolocated linkage from multiple sources.

2. Materials and methods

Injury and fatality records from calendar years 2013–2015 were collected from the San Francisco Police Department (SFPD), Zuckerberg San Francisco General Hospital and Trauma Center (ZSFG), Emergency Medical Services (EMS) and Office of the Chief Medical Examiner (OME). Records from all sources were restricted to people injured within the City and County of San Francisco where injuries were non-intentional (i.e., not assault, suicide, or homicide). The San Francisco Public Health Department (SFDPH) conducted the linkage, with the San Francisco Municipal Transportation Agency (SFMTA) funding a full-time epidemiologist within [X]DPH for this work.

2.1. Data sharing

University of California San Francisco (UCSF) Human Research Protection Program Institutional Review Board (IRB) approval #17-23497 governs the use and protection of ZSFG patient data for the purposes of this linkage and related research. Privacy of ZSFG Trauma Registry (TR), Emergency Department (ED), and EMS health information was maintained per the Health Insurance Portability and Accountability Act (HIPAA). SFDPH, UCSF, and ZSFG signed a Data Sharing Agreement and Memorandum of Understanding outlining protocols for data access, privacy protection, and intended uses which employ the linked dataset.

2.2. Data sources

Police collision reports were provided from the Crossroads Software Traffic Collision Database System, used during the project period by SFPD to manage electronic police report data. Minimum fields are defined by the California Highway Patrol 555 Crash Report form, and include detailed crash characteristics such as vehicles involved, number of parties to the collision, location, California vehicle code violation and identifiers including names and addresses of parties involved in a crash (State of California Department of California Highway Patrol, 2011). Collision reports for which an SFPD officer responded to a collision site on a public roadway in San Francisco were included, while counter reports filed after the fact (for collisions to which an officer did not respond to the scene of the incident) were excluded ensuring all injuries were classified by a medical or law enforcement professional.

ZSFG is a public hospital and the sole Level 1 Trauma Center for the City and County of San Francisco, serving 1.5 million people in San Francisco and northern San Mateo counties. The most severely injured patients are routed through a citywide coordinated trauma triage system, adhering to two levels of trauma team activation criteria which ensure that patients with injuries requiring emergent or urgent trauma care arrive to ZSFG. The trauma center serves 3,900 trauma patients annually ("About ZSFG," 2012). Patients reflected in hospital data include all road user types: pedestrians, cyclists, motorcyclists and motor vehicle occupants, with all trauma activated transportation injured people in the region routed to ZSFG. Of relevance, access to alternative Level 1 or 2 trauma centers in the region requires leaving the county and in most cases crossing a bridge. Patients injured outside of the City and County of San Francisco were excluded. Patients with traffic-related injuries that were intentional or self-inflicted (including homicides and suicides) or the result of assault were excluded. Permission to conduct this surveillance was obtained from the UCSF's Human Research Protection Program IRB.

ZSFG data were sourced from both the TR and ED. The TR conforms to the National Trauma Data Standard and captures detailed data on the most severe injuries—those requiring a trauma team response for life- or limb-threatening injury or burns ("NTDS Data Dictionary, 2015). Because these injuries are restricted to the most severe, they are not representative of the wider burden of traffic-related injury (Horan & Mallonee, 2003). Inclusion of ED records improves the generalizability of hospital injury records and supplements injury surveillance and incidence estimates by including less severe injuries. For all hospital data, injuries with an International Classification of Disease, 9th Revision

(ICD-9) external cause of injury code (e-code) indicating transportation-related injury were included (Appendix A). In addition to e-codes, hospital data offers clinical assessment of injury severity and valuable information on patient comorbidities, disability status and homelessness—variables which may be related to injury vulnerability. ED and TR data were joined by a unique hospital patient identification number.

For patients transported to ZSFG by EMS, crash geographic location information was obtained from each of three companies which provide Advanced Life Support ambulance transport: the San Francisco Fire Department, King American and American Medical Response. EMS collision location data were joined to the ZSFG dataset using prehospital run sheet number.

The OME investigates all traumatic deaths in the county and issues detailed fatality records. Their mortality data for transportation related injuries were directly linked to the dataset using police report number.

2.3. Combining datasets

For optimum matching of records between datasets, we deduplicated records and standardized matching variables to common formats and verified common coding across discrete variables (Centers for Disease Control and Prevention, 2019).

We used LinkSolv9.0 software to select candidate pairs from our hospital and police datasets using four blocking variables (first name, date of birth, primary road, and collision date), then probabilistically matched records represented in both datasets on eight fields: date of birth, Soundex of first name, Soundex of last name, sex:gender, road user type (derived from ICD-9 e-code):travel mode, collision date:hospital arrival date, and primary and secondary road. Tolerance of one typographical error was built into the match of date of birth between sources. Name fields were matched using a Soundex phonetic algorithm, which compares fields based on pronunciation rather than exact character matches. This algorithm was developed for American English and expanded for use on the Chinese Pinyin system to match a higher proportion of Asian-origin names common among the San Francisco Bay Area population. Binary sex:gender and road user type were required to match exactly. Collision date:hospital arrival date could fall within one day of one another. Primary and secondary roads were individually compared to both road names from a possible match in case of reversal— necessary for collisions occurring in intersections (Table 1). Variable match weights were calculated using five independent Markov Chain Monte Carlo algorithms, assigning higher weights to records agreeing on rare matching variable values relative to common ones.

Descriptive statistics were calculated providing demographic and injury characteristics from each data source. The continuous variable of age was reported by mean and median value. Two sample z tests were used to compare proportions between unlinked data sources, and one sample z tests for comparison of proportions from different sources referring to the same linked sample. Chi square tests were used to compare proportions between original data sources and the linked dataset and to assess associations between the potentially confounding variables considered for regression models. To explore whether injury severity confounds the odds of record linkage, we created a logistic regression model to compute linkage odds by mode with pedestrians as the reference group, adjusting for medically-designated severity category (i.e. sub-severe, severe, or critical/fatal injury). We reported odds ratios and 95% confidence intervals from our regression model. Statistical analyses were performed using R version 4.0.1.

To assess geographic vulnerability we used a binary regional proxy measure of "Communities of Concern" based on vulnerable census tracts– arrived at by a compound calculation taking into account neighborhoods with high concentrations of minority, low-income, transit-dependent, non-English speaking, disabled or older individuals ("MTC Communities of Concern Factor and Predominant Populations (Census 2010) Explorer, 2018).

Linkage allows comparison of injury severity classified in police data to both clinically-assessed Injury Severity Score (ISS), a correlate of trauma severity, and hospital admission (Senkowski & McKenney, 1999). We use the criterion of ISS > 15 as a proxy measure for critical injury, widely accepted in trauma literature (Senkowski & McKenney, 1999). Prior crash injury linkage projects have documented variation between police and medical assessments of crash-related injury severity (Cherry et al., 2018; Cryer et al., 2001; McDonald et al., 2009). We calculated the sensitivity and specificity of the police assessment of injury severity among linked records to assess this relationship in our data.

3. Results

From 2013 to 2015, there were 17,211 transportation-related injuries in San Francisco detected using this surveillance methodology, comprising over 5,000 transportation-related injuries annually. One quarter (26%, n = 4,415) of these injuries appeared in *both* police and hospital data sources (Table 2). Linked records represent 39% of total police injury records (N = 11,403) and 43% of total hospital injury records (N = 10,223) (Fig. 1).

Linkage rates (the proportion of hospital records linking to a police record) varied notably by travel mode, with 34% of injured cyclist and 38% of injured motor vehicle occupant hospital records matching to a police record (817/2,393 and 1,531/4,031, respectively), compared to 61% of injured pedestrian and 54% of motorcyclist hospital records with police record matches (1,392/2,273 and 619/1,149, respectively) (Table 2). Because injury severity significantly predicts record linkage ($p < 0.0001$), we adjusted for it in a logistic regression model (Table 3). For the model, pedestrians were the reference category, representing both the road user group with the highest linkage rate and the one most vulnerable to injury. In the adjusted model cyclists had 63% lower, motorcyclists had 33% lower, and motor vehicle occupants had 47% lower odds of linkage compared to pedestrians (Table 3).

Table 4 summarizes demographic and injury severity characteristics of linked and unlinked records by data source and overall. Injured patients treated at ZSFG who are not represented in police records showed some distinct differences from the group of people captured solely via police reporting. People with transportation injuries treated at ZSFG without associated police reports were more often cyclists, male, Hispanic or Black, and less often occupants of motor vehicles compared to those with injuries captured only in police collision reports (each $p < 0.0001$, Table 4).

Mean Linksolv-generated match probability for linked records was 99.38% (range 4.24–100%). Mean match weight for linked records was 39.64 (range 8.16–67.57), representing the likelihood ratio of observed variable values given that a record pair is matched versus unmatched. For purposes of determining travel mode in trauma data, over 99% of trauma registry records contained e-code information.

3.1. Linked data findings

Among people in the linked dataset– those who both had a pre-hospital or hospital record and a police record of injury– there were notable differences in the categorization of race/ethnicity and injury severity. Among the same 4,415 people, police reports were significantly less likely to record individuals as Hispanic (16%, $p < 0.0001$) compared to medical records (20%, Table 4). Comparing police injury severity classifications of severe and fatal injury in linked records to medical assessment of critical injury (ISS > 15), police officers were significantly more likely to classify injuries as severe or fatal than hospital staff ($p = 0.0005$). In our dataset 75% of patients with a medically-determined critical injury also appear in police injury records (n = 277/367). However, more than three in 10 nonfatal injuries with a critical ISS were missed (i.e., reported as non-severe) in police crash reports (sensitivity of police assessment 65%), and only one-quarter of nonfatal injuries that police rated as severe were confirmed as critical injuries by medical staff (27%) (Table 5). Conversely, for injury records with police severity available, there was high probability– 86%– that sub-critical injuries treated by clinicians were also designated sub-severe by law enforcement staff (Table 5). Among linked records with ISS > 15 and a police officer assessment of severity, 73% (n = 202/277) of police assessments agree with the clinical severity determination (data not shown). Among hospital records of patients *hospitalized* for their injuries, half (48%, n = 920/1,928) linked to police reports; of that group, 71% (n = 693/920) of police and hospital injury records agree in severity assessment (data not shown).

Geographic location of injury crashes was available for 70% of the total dataset (n = 12,002). Injury location data were available for 98% of linked records and 95% of police and medical examiner-only records (4,324/4,415 and 6,607/6,988, respectively), compared to 18% (1071/5,808) of hospital and pre-hospital-only records. We found a disproportionate concentration of severe and fatal injuries on street lengths in Communities of Concern (47%, n = 5,588/12,002); just 31% of San Francisco streets are located in these areas where more vulnerable populations are concentrated (Kronenberg et al., 2019).

4. Discussion

In this pilot, we developed a probabilistically linked dataset with three years of injury data from multiple sources for county-level comprehensive transportation injury surveillance. Our linked dataset provides two types of evidence that linkage is preferable to single-source injury surveillance data. First, a linked data system improves upon standard injury surveillance by guaranteeing inclusion of the most severely injured victims of collisions—comprehensively captured by the trauma registry in San Francisco's sole Level 1 Trauma

Center. Our sub-analyses confirm that this addition consists of an overlapping but distinct segment of the population relative to those captured in police collision reporting. Second, linkage leverages the relative strengths of each data source: injury severity classification using hospital data prioritizes highly predictive clinical assessment and systematically collected demographic data, whereas geographic location of injury is routinely collected in EMS and police– but not hospital– data (Baker et al., 1974; Senkowski & McKenney, 1999). These advantages highlight the ongoing need for a formal coordinated surveillance system that integrates data from all relevant transportation-related injury data sources in San Francisco for more accurate and comprehensive surveillance of injuries. Such a system informs monitoring, evaluation and targeted initiatives to prevent death and mitigate morbidity from traffic collisions.

Linked records comprise 43% of all traffic-related hospital and prehospital injury records, and 61% of the pedestrian injury subset. Notably, we did not expect 100% linkage: populations represented in hospital and police records are partially distinct. By comparison, linkages from New York City considering all modes and a prior San Francisco pedestrian injury analysis report linkage rates of 52% and 60% of hospital records, respectively (Conderino et al., 2017; Sciortino et al., 2005). In contrast to these projects, our linkage employed the near-unique identifier of patient name, indicating that the modestly lower all-mode linkage rate and comparable pedestrian linkage rate of this project more likely result from limited overlap in the populations represented in police and hospital records in our sample than unsuccessful matching.

The “value-added” of performing linkage with medical records compared to standard surveillance is demonstrated by analysis of hospital records not reflected in police collision reports. That a full quarter of critically traffic-injured ZSFG patients (ISS > 15) were not reflected in police collision reports is cause for concern. Many jurisdictions committed to data-driven injury prevention rely solely upon police collision records for traffic injury surveillance, and our findings suggest they may be missing a substantial proportion of traffic-related injury incidence. Cyclists were particularly under-represented in police data. Rates of biking and cyclist fatalities have been increasing nationally (National Center for Statistics and Analysis, 2019). Our findings highlight the potential of local surveillance efforts to use hospital and ambulance data to improve capture of injury incidence and spatial patterns of injury—both critical to inform infrastructure improvements and injury prevention measures such as protected bike lanes. Recent efforts to add variables like the prehospital run sheet number to police reporting forms would help simplify and facilitate these types of initiatives to link police and hospital data for more comprehensive injury surveillance.

Injury severity significantly predicts linkage. Our data support that the more severe the injury, the higher the likelihood that a police officer is summoned to the scene or that a patient receives ZSFG care for their injury: both are required for linkage. In the logistic regression model adjusting for injury severity, linkage odds according to travel mode shift meaningfully (Table 3). For example, unadjusted motorcycle injuries showed 90% odds of linkage relative to pedestrians. However controlling for severity, motorcycle injuries had just two-thirds the odds of linking across datasets relative to pedestrian injuries. This downward shift indicates that the modal linkage rate is driven not by comprehensive reporting but severity: relative to other modes, motorcycle collisions often occur at high speed, resulting in more severe injuries. In the reverse direction, motor-vehicle occupant odds of linkage rose from 31% to 53% after controlling for severity. Motor-vehicle occupants form the largest group of records and may reflect more comprehensive reporting of even minor injuries in the mode to police. Vehicle insurance company requirements for police reports to accompany medical and property damage reimbursement claims potentially contribute to this phenomenon.

CDC's Linking Information for Nonfatal Crash Surveillance (LINCS) guide reports only 19 U.S. state-level linkage programs as of 2017— pointing to both untapped demand for such data as well as the logistic complexity of conducting such projects (Centers for Disease Control and Prevention, 2019). Our linkage was aided by the availability of personal identifiers to which other prominent linkage projects have historically had limited or no access: specifically name, date of birth, and record identification numbers (e.g. prehospital run sheet number) (California Department of Public Health, 2017; Centers for Disease Control and Prevention, 2019; Cherry et al., 2018; Conderino et al., 2017). For instance, while there is a statewide crash injury probabilistic linkage underway in California that employs the same software, California's Crash Medical Outcomes Data (CMOD) project has not had access to first and last names as match variables which can increase the likelihood of missing a match when one exists (California Department of Public Health, 2017). In addition, CMOD does not capture data on injury severity and also limits the geographic information available for crashes to multi-county regions of the state (California Department of Public Health, 2017), while the geocoded, intersection-level crash data available from the majority of records in our data has utility for informing specific engineering improvements to intersections and streets.

EMS data are a valuable supplement to unlinked hospital records because they provide a geographic location of injury. Our linkage associated 18% of unlinked hospital records with geographic crash location via EMS records, enabling geospatial visualization of severe and critical injury crash locations that would have been missed using police collision report data alone. Mapped patterns of traffic injury are particularly helpful for informing interventions by traffic engineers, infrastructure prioritization, and education and enforcement efforts.

Many factors contribute to the possibility of a hospital injury record not matching to a police record. At least four factors may affect police record availability. First, we chose to include police collision reports where an officer physically went to the injury location, and exclude injuries reported at a police station after the fact. This decision ensured records reflect verified crash-related injuries, and may have improved the linkage rate (assuming injuries with delayed reporting to police were on balance less severe, and less likely to receive medical care). Some of these "counter reports" might have linked with a hospital injury record if included. Second, SFPD's purview and current Vision Zero San Francisco efforts exclude freeways, which fall under California state jurisdiction. Injuries sustained on city freeways are reported to the California Highway Patrol rather than SFPD. As a result, collisions occurring on freeways in San Francisco do not appear in local police data, and cannot link to a hospital record of a person injured on the freeway and not otherwise excluded based on freeway location in EMS transport data, potentially lowering the linkage rate. For context, California Highway Patrol recorded 735 severe and fatal transportation injuries San Francisco freeways over the linkage period, some of which may be represented among the 5,808 unlinked hospital records (Transportation Injury Mapping System, 2022). Thirdly, injuries not involving motor vehicles– such as solo cyclist crashes– are less likely to have associated collision reports, reducing the linkage rate. Finally, communities or individuals who experience fear, discomfort, or mistrust interacting with police may be less likely to file a police report when severely injured– likely a phenomenon of disproportionate relevance to communities of color and immigrant communities due to well-documented issues of bias in policing (Knowles et al., 1999; Wu et al., 2013). Unlinked hospital records reflect 60% non-white patients relative to 55% of people represented in unlinked police injury records, supporting the possibility that traffic injuries to people of color are underreported in our police collision report data (Table 4, $p < 0.0001$), potentially limiting the linkage rate.

Discrepancies in race and ethnicity demographic information between data sources highlight the different operational purposes of the data systems from which they are sourced. An injured person is often given the opportunity to self-identify their race and ethnicity in the hospital data system (or might have this information on file in the health record), whereas police officers make a subjective race/ethnicity determination for collision reports. This fact may explain why ethnicity designation sometimes varied significantly between police and hospital injury records referring to the same sample of people, and highlights the benefit of increased demographic accuracy within linked data.

While our linkage uses regularly-collected operational data, neither police nor hospital records' primary purpose is injury surveillance. Police collision data include vital information regarding crash circumstances and assign "fault" to parties based on vehicle code violations; however injury severity designations are limited to four levels: fatal, severe, other visible injury, and complaint of pain (State of California Department of California Highway Patrol, 2011). Hospital data are a rich source for patient injury level data, including diagnoses, days hospitalized, discharge disposition, and billing information (including important demographic data), yet lack information about injury location and cause of the collision resulting in injury. Prior local analysis indicates that police injury severity designations of *severe* and *visible* tend to underestimate the level of injury, and police report misclassification of injury level relative to objective measures is widely documented (San Francisco Department of Public Health, 2017; Amoros et. al., 2007; McDonald et al., 2009; Tsui et al., 2009; Ferenchak & Osofsky, 2022). Incomplete or inaccurate data increase the potential of "false negatives" – where records representing the same person in different data sources fail to match. There is an opportunity to improve matching in future medical and crash data linkages by including a field for the prehospital run sheet number assigned during an ambulance trip in police collision reports. This modification was made in SFPD collision reporting forms following our linkage, and was subsequently provided as a recommendation to the Model Minimum Uniform Crash Criteria, which inform law enforcement collision data collection nationally.

One immediate practical application of this dataset in San Francisco was updating the spatial analysis underlying the Vision Zero High Injury Network (HIN) – a subset of city streets on which severe and fatal crashes are disproportionately concentrated. This analysis is a critical tool, informing millions of dollars of capital investments in engineering infrastructure with a focus on pedestrian and cyclist improvements. It has also informed policy on emerging technologies, including where sidewalk delivery robots may pilot in San Francisco (San Francisco Department of Public Works, 2017). The HIN is publicly accessible alongside a rich database of street features and other environmental variables at TransBase.sfgov.org. The updated HIN utilizing the subset of linked severe and fatal injury data with geographic location of crashes now ensures that vulnerable road users and spatial locations not previously captured in police-only data analyses inform targeted safety improvements by the city (Kronenberg et al., 2019; San Francisco Department of Public Health, Program on Health, Equity and Sustainability, 2017). While this application of the data is a notable improvement, it is also still limited by a lack of injury location data for people reflected only in hospital records and not transported by ambulance. Opportunities to add injury location data to hospital records for people not transported by ambulance– potentially through addition of a field for self-reported crash location to electronic health records– would significantly improve the utility of these data for targeted local prevention efforts.

This project faces several limitations. First, the linkage focuses on ZSFG trauma data in order to capture the most severe traffic-related injuries. While we included Emergency Department injury data from ZSFG there are a host of San Francisco EDs not represented in our injury data where less severe traffic injuries may have been treated. As a result, our data do not estimate the true incidence of transportation injury occurring on the streets of San Francisco and cannot be generalized to other populations. Second, probabilistic linkage software operates based on record match weights calibrated from variables designated as important, but lacks an objective criterion against which to quantify the validity of the linked dataset. Finally, because injury surveillance is not the primary purpose of component data sources, there is potential misclassification of travel mode in both hospital and police data– particularly for relatively less-employed travel modes such as cycling and motorcycling.

Our linkage work motivates several next steps. Locally we have extended the three-year pilot of transportation injury data to ongoing practice, with a 2016–2019 linkage forthcoming and–key for longevity– regular institutional funding identified to staff this work. Future research opportunities include in-depth analysis of freeway injuries, and analyses of subsets of the linked data such as patterns of traffic injury to people severely injured, people with disabilities, or people experiencing homelessness. Linked datasets can provide particular insight to the effects of substance-involved driving and the economic burden of traffic injuries, especially with further accumulation of longitudinal data. There is potential to integrate novel data streams into injury surveillance, including those emerging from technology applications employed by road users and GPS- and wireless connectivity-equipped vehicles. In summary, we found that incorporating hospital data into police report-based surveillance adds information on populations empirically vulnerable to injury yet missing police collision reports, improves assessment of injury severity, and can be harnessed to better inform and target interventions addressing the needs of vulnerable populations– including cyclists– in support of injury and fatality prevention efforts. This work holds increased value at the local level, especially in jurisdictions which that adopted transportation fatality elimination goals and require reliable data to guide decision-making, track progress, and equitably allocate funding for injury prevention projects.

Acknowledgments

The authors wish to acknowledge the SFMTA, San Francisco Police Department, and Vision Zero Task Force for their support, including SFMTA funding for this project, and advocacy in supplying and using this data to the benefit of all those who use San Francisco's streets. We thank the San Francisco Fire Department, King American, and American Medical Response which provide EMS services to ZSFG. Mike McGlincy of Linksolv and Nina Fiore of the OME were instrumental in data linkage methodology and data acquisition, respectively. Statistical support was expertly provided by Amy Shui of the Biostatistics Core which is funded by the UCSF Department of Surgery.

Financial disclosure

Data linkage and manuscript preparation was supported through a work order from the San Francisco Municipal Transportation Agency (SFMTA) to the San Francisco Department of Public Health. The SFMTA played no role in conducting the research or decision to submit this article for publication.

Biographies

Shamsi Soltani (she/her) earned an MPH in Epidemiology at Tulane University and a BS in Neuroscience with French minor at UCLA. She draws on deep applied public health experience, including over six years as a senior epidemiologist for the San Francisco Department of Public Health. Shamsi is interested in the interplay of technology, bias, and health disparity, particularly with respect to big data. She is currently pursuing a PhD in Epidemiology at Stanford University.

Leilani Schwarcz is an epidemiologist in the Emergency Medical Division, Regional Quality Improvement Program at Public Health- Seattle & King County since 2019. Her career has focused on a broad range of interdisciplinary public health priorities, including EMS Covid-19 surveillance and epidemiologic data management, medical quality improvement evaluations in King County, development of a citywide Traffic Injury Surveillance System for San Francisco's Vision Zero, and management of statewide blood lead surveillance and biomonitoring studies of heavy metal exposures in New Mexico. Leilani has an MPH in Epidemiology and a BS in Environmental Engineering from the University of Nevada, Reno.

Devan Morris is an Integrated Business Systems Analyst at the San Francisco Department of Public Health. Devan performs geospatial and data analysis for several public health projects, including San Francisco's Vision Zero policy. He is the lead developer on transbase.sfgov.org, an innovative open source database management system designed to access, manage, and apply spatial data to inform solutions to transportation problems. He teaches beginner and intermediate GIS classes through City College of San Francisco's GIS Education Center and volunteers with BayGeo. He earned his Bachelor of Science from University of California, Davis and an Associate Degree from American River College.

Rebecca Plevin, MD received her undergraduate degree from UC Berkeley, her MD from the University of Southern California Keck School of Medicine, and completed her residency in General Surgery at the University of Washington. Dr. Plevin completed her fellowship in Trauma and Surgical Critical Care at Zuckerberg San Francisco General Hospital, University of California San Francisco (UCSF). In 2018, she became an Assistant Professor in Residence in the UCSF Division of General Surgery. Her research interests include injury prevention with particular focus on underserved/vulnerable populations.

Rochelle Dicker is Professor of Surgery and Anesthesia and Vice Chair for Critical Care at UCLA, and Co-Director of the Program for the Advancement of Surgical Equity. In 2004 while on UCSF faculty she started the Wraparound Project at San Francisco General Hospital, which became a founding member of the Health Alliance for Violence Intervention network of now over 40 nationwide hospital-based violence intervention programs. She has conducted extensive research on the effectiveness and cost effectiveness of violence intervention and has contributed to over 90 peer-reviewed publications and several book chapters.

Catherine Juillard is a surgeon interested in identifying risk factors driving disparities in susceptibility to surgically treated diseases and inequity in access to treatment, then creating ways to mitigate these disparities. She went to Stanford University for undergraduate studies, then the University of California, Los Angeles, for medical school and general surgery residency. She completed a fellowship in trauma and surgical critical care at the University of California, San Francisco, staying on as faculty until 2018. In 2018, Dr. Juillard joined the faculty at UCLA and co-founded the Program for the Advancement for Surgical Equity in the Department of Surgery.

Adaobi Nwabuo, MBBS, MPH was Injury Prevention Coordinator for the San Francisco Wraparound Project. She obtained her medical degree from the University of Lagos, Nigeria and her Master's of Public Health from the Johns Hopkins School of Public Health (JHSPH). Adaobi has worked as a physician in underserved communities in Nigeria. Through JHSPH, she conducted research assessing substance use and driving safety among adolescents. She has worked with the Baltimore City Department of Health to qualitatively assess retailers' perspectives on tobacco sales to minors. Adaobi's clinical and research interests are in sustainable and culturally acceptable healthcare services, particularly mental healthcare.

Megan Wier, MPH, is the Safe Streets Division Manager for the City of Oakland's Department of Transportation and oversees planners, engineers, electricians and traffic maintenance staff to strategically advance a safe, equitable transportation system for Oakland, California. Prior to starting that position in early 2020, Megan was the Director of the Program on Health, Equity and Sustainability at the San Francisco Department of Public Health where she worked for over a decade at the intersection of health, equity and transportation in partnership with city agencies and community stakeholders, and co-chaired San Francisco's Vision Zero Task Force to eliminate traffic deaths.

Appendix A

See Table A1.

Table A1

Transportation-related injury ICD-9-CM e-codes.

Unintentional External Cause	e-Codes
Motor vehicle traffic	E810–E819 (0.0–0.9)
Occupant	E810–E819 (0.0–0.1)
Motorcyclist	E810–E819 (0.2–0.3)
Pedal cyclist	E810–E819 (0.6)
Pedestrian	E810–E819 (0.7)
Unspecified	E810–E819 (0.9)
Pedal cyclist, other	E800–E807 (0.3); E820–E825 (0.6); E826.1,0.9; E827–E829 (0.1)
Pedestrian, other	E800–E807 (0.2); E820–E825 (0.7); E826–E829 (0.0)

Abbreviations: **ICD-9-CM** International Classification of Diseases, Ninth Revision, Clinical Modification; **e-codes** External Cause of Injury Codes.

References

About ZSFG (2012). ZSFG community wellness program. URL http://sfghwellness.org/about/about-sfgh/ (accessed 12.24.2019).

Amoros E, Martin J-L, Chiron M, & Laumon B (2007). Road crash casualties: Characteristics of police injury severity misclassification. The Journal of Trauma: Injury, Infection, and Critical Care, 62(2), 482–490. 10.1097/01.ta.0000202546.49273.f9.

Baker SP, O'Neill B, Haddon W, & Long WB (1974). The injury severity score: A method for describing patients with multiple injuries and evaluating emergency care. Journal of Trauma, 14, 187–196. [PubMed: 4814394]

California Department of Public Health (2017). About the linked Crash Medical Outcomes (CMOD) Data [WWW Document]. URL https://www.cdph.ca.gov/Programs/CCDPHP/DCDIC/SACB/Pages/EpiCenter/AboutOurData.aspx (accessed 12.24.2019).

Centers for Disease Control and Prevention (2019). Linking Information for Nonfatal Crash Surveillance (LINCS) [WWW Document]. URL https://www.cdc.gov/motorvehiclesafety/linkage/Linking-Information-Nonfatal-Crash-Surveillance.html (accessed 12.24.2019).

Cherry C, Hezaveh AM, Noltenius M, Khattak A, Merlin L, Dumbaugh E, et al. (2018). Completing the picture of traffic injuries: Understanding data needs and opportunities for road safety. Collaborative Sciences Center for Road Safety. URL https://www.roadsafety.unc.edu/research/projects/2017r4/ (accessed 06.28.2021).

Conderino S, Fung L, Sedlar S, & Norton JM (2017). Linkage of traffic crash and hospitalization records with limited identifiers for enhanced public health surveillance. Accident Analysis and Prevention, 101, 117–123. 10.1016/j.aap.2017.02.011. [PubMed: 28226252]

Couto A, Amorim M, & Ferreira S (2016). Reporting road victims: Assessing and correcting data issues through distinct injury scales. Journal of Safety Research, 57, 39–45. 10.1016/j.jsr.2016.03.008. [PubMed: 27178078]

Cryer PC, Westrup S, Cook AC, Ashwell V, Bridger P, & Clarke C (2001). Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports. Injury Prevention, 7, 234–241. 10.1136/ip.7.3.234. [PubMed: 11565992]

Elvik R, & Mysen A (1999). Incomplete accident reporting: Meta-analysis of studies made in 13 countries. Transportation Research Record, 1665, 133–140. 10.3141/1665-18.

Ferenchak NN, & Osofsky RB (2022). Police-reported pedestrian crash matching and injury severity misclassification by body region in New Mexico, USA. Accident Analysis and Prevention, 167, 106573. 10.1016/j.aap.2022.106573. [PubMed: 35085857]

Fleisher A, Wier ML, & Hunter M (2016). A vision for transportation safety: Framework for identifying best practice strategies to advance vision zero. Transportation Research Record, 2582, 72–86. 10.3141/2582-09.

Horan JM, & Mallonee S (2003). Injury surveillance. Epidemiologic Reviews, 25, 24–42. 10.1093/epirev/mxg010. [PubMed: 12923988]

Knowles J, Persico N, Todd P (1999). Racial bias in motor vehicle searches: Theory and evidence (Working Paper No. 7449). National Bureau of Economic Research.

Kronenberg C, Wier M, Reeves R, & Jacobson M (2019). Vision zero action strategy. San Francisco, California: San Francisco Municipal Transportation Agency.

Langley JD, Dow N, Stephenson S, & Kypri K (2003). Missing cyclists. Injury Prevention, 9, 376–379. 10.1136/ip.9.4.376. [PubMed: 14693904]

Lopez DS, Sunjaya DB, Chan S, Dobbins S, & Dicker RA (2012). Using trauma center data to identify missed bicycle injuries and their associated costs. Journal of Trauma and Acute Care Surgery, 73, 1602–1606. 10.1097/TA.0b013e318265fc04. [PubMed: 23032807]

McDonald G, Davie G, & Langley J (2009). Validity of police-reported information on injury severity for those hospitalized from motor vehicle traffic crashes. Traffic Injury Prevention, 10, 184–190. 10.1080/15389580802593699. [PubMed: 19333832]

Milani J, Kindelberger JH, Bergen G, Novicki E, Burch C, Ho SM, West BA (2015). Assessment of characteristics of state data linkage systems.

Metropolitan Transportation Commission (2018). MTC communities of concern factor and predominant populations explorer [WWW Document]. URL http://opendata.mtc.ca.gov/datasets/MTC::mtc-communities-of-concern-factor-acs-2012-2016-and-predominant-populations-census-2010-explorer-2018 (accessed 12.24.2019).

National Center for Statistics and Analysis (2019). Early estimates of motor vehicle traffic fatalities for the first 9 months (Jan- Sep) of 2019 (No. DOT HS 812 874), crash stats brief statistical summary. National Highway Traffic Safety Administration, Washington, DC.

National Complete Streets Coalition, Smart Growth America (2019). Dangerous by design 2019. [WWW Document]. URL https://smartgrowthamerica.org/resources/dangerous-by-design-2019/ (accessed 08.22.2022).

National Highway Traffic Safety Administration (2021). Crash Outcome Data Evaluation System (CODES) | NHTSA. URL https://www.nhtsa.gov/crash-data-systems/crash-outcome-data-evaluation-system-codes (accessed 06.28.2021).

National Highway Traffic Safety Administration (1996). Report to congress: Benefits of safety belts and motorcycle helmets (No. DOT HS 808 347). Washington, DC: U.S. Department of Transportation.

NTDS Data Dictionary 2015 [WWW Document] (2015). Am. Coll. Surg URL https://www.facs.org/-/media/files/quality-programs/trauma/ntdb/ntds/data-dictionaries/ntds-data-dictionary-2015.ashx (accessed 12.24.2019).

San Francisco Department of Public Works (2017). Autonomous Delivery Devices [WWW Document]. URL https://www.sfpublicworks.org/services/permits/autonomous-delivery-devices (accessed 12.24.2019).

San Francisco Department of Public Health, Program on Health, Equity and Sustainability (2017). Vision zero high injury network: 2017 Update – A methodology for San Francisco, CA. San Francisco, CA. [WWW Document]. URL https://www.sfdph.org/dph/files/EHSdocs/PHES/VisionZero/Vision_Zero_High_Injury_Network_Update.pdf (accessed 08.22.2022).

Sauber-Schatz EK, Ederer DJ, Dellinger AM, & Baldwin GT (2016). Vital signs: Motor vehicle injury prevention – United States and 19 comparison countries. MMWR Morbidity and Mortality Weekly Report, 65, 672–677. 10.15585/mmwr.mm6526e1. [PubMed: 27388054]

Sciortino S, Vassar M, Radetsky M, & Knudson MM (2005). San Francisco pedestrian injury surveillance: Mapping, under-reporting, and injury severity in police and hospital records. Accident Analysis and Prevention, 37, 1102–1113. 10.1016/j.aap.2005.06.010. [PubMed: 16084782]

Senkowski CK, & McKenney MG (1999). Trauma scoring systems: A review. Journal of the American College of Surgeons, 189, 491–503. 10.1016/s1072-7515(99)00190-8. [PubMed: 10549738]

Short J, & Caulfield B (2016). Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. Journal of Safety Research, 58, 1–14. 10.1016/j.jsr.2016.05.002. [PubMed: 27620929]

State of California Department of California Highway Patrol (2011). California Crash Report CHP555, Rev. 4/2011 [WWW Document]. URL https://one.nhtsa.gov/nhtsa/stateCatalog/states/ca/crash.html (accessed 12.24.2019).

Tainter F, Fitzpatrick C, Gazillo J, Riessman R, & Knodler M Jr, (2020). Using a novel data linkage approach to investigate potential reductions in motor vehicle crash severity – An evaluation of strategic highway safety plan emphasis areas. Journal of Safety Research, 74, 9–15. 10.1016/j.jsr.2020.04.012. [PubMed: 32951800]

Tin Tin S, Woodward A, & Ameratunga S (2013). Completeness and accuracy of crash outcome data in a cohort of cyclists: A validation study. BMC Public Health, 13, 420. 10.1186/1471-2458-13-420. [PubMed: 23635027]

Transportation Injury Mapping System (TIMS) (2022). Safe Transportation Research and Education Center, University of California, Berkeley.

Tsui KL, So FL, Sze NN, Wong SC, & Leung TF (2009). Misclassification of injury severity among road casualties in police reports. Accident Analysis & Prevention, 41(1), 84–89. 10.1016/j.aap.2008.09.005. [PubMed: 19114141]

Vision Zero Network (2021). Vision zero communities map. URL https://visionzeronetwork.org/resources/vision-zero-communities (accessed 07.25.2021).

World Health Organization, 2018. Global status report on road safety 2018.

Wu Y, Smith BW, & Sun IY (2013). Race/Ethnicity and perceptions of police bias: The case of Chinese immigrants. Journal of Ethnicity in Criminal Justice, 11 , 71–92. 10.1080/15377938.2012.735989.

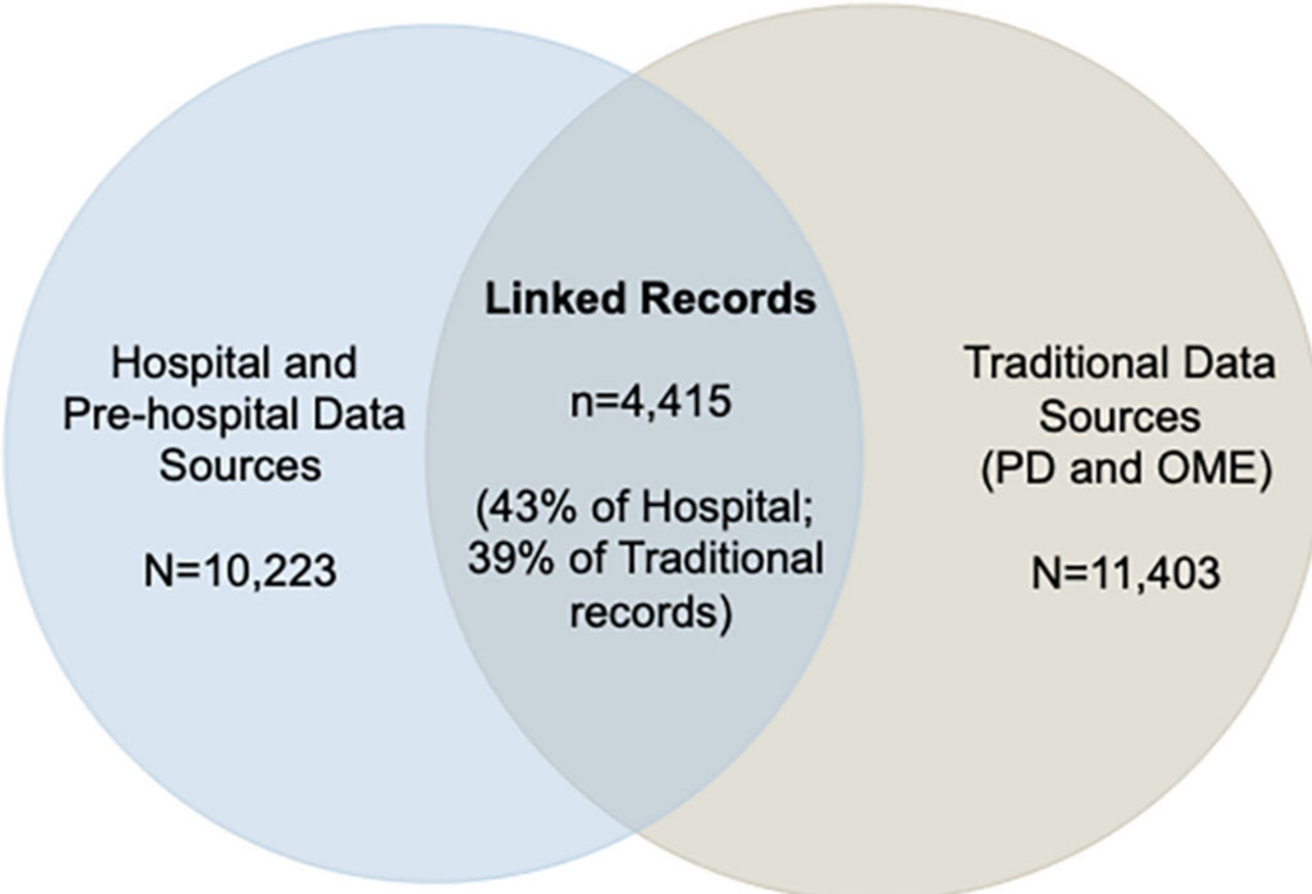

Fig. 1.
San Francisco's Transportation-Related Injury Surveillance System (2013–2015): Data Source and Linkage Status by Transportation Mode.

Table 1

Linkage match variables.

Variable	Police Report	Hospital Record
Date of Birth	Date of birth	Date of birth
Soundex of First Name[a]	First name	First name
Soundex of Last Name[a]	Last name	Last name
Sex[b]	Victim sex	Patient sex
Female		
Male		
Date of Injury	Collision date	Arrival date to hospital
Road User type	Mode taken from vehicle type	Mode taken from ICD-9-CM codes
Pedestrian		
Cyclist		
Motorcyclist		
Motor Vehicle Occupant		
Other mode/Unknown		
Primary Road (compared to ‘primary road + secondary road’)	Primary road	Primary_Rd
Secondary Road (compared to ‘primary road + secondary road’)	Secondary road	Secondary_Rd

[a] Soundex is a phonetic algorithm for indexing names by sound, as pronounced by Voice of America’s pronunciation guide (in our case with added latitude for pinyin spellings). Allows for matching despite variation in spelling.

[b] At the time of linkage, our data sources were restricted to binary sex categories.

Table 2

San Francisco's Transportation-Related Injury Surveillance System (2013–2015): Data Source and Linkage Status by Transportation Mode.

Data Source	Pedestrian records	Pedestrian linked (%)	Cyclist records	Cyclist linked (%)	Motorcyclist records	Motorcyclist linked (%)	Motor Vehicle Occupant records	Motor Vehicle Occupant linked (%)	Other/ Unknown records	Other/ Unknown linked (%)	Total Records	Total Linked Records (%)
Traditional Data Sources	**2,470**	**1,392 (56%)**	**1,760**	**817 (46%)**	**1,051**	**619 (59%)**	**5,951**	**1,531 (26%)**	**171**	**56 (33%)**	**11,403**	**4,415 (39%)**
San Francisco Police Dept. Collision Report											11,403	
Office of the Medical Examiner											95	
Hospital and Pre-Hospital Data (EMS, ED, Trauma Registry) [a]	**2,273**	**1,392 (61%)**	**2,393**	**817 (34%)**	**1,149**	**619 (54%)**	**4,031**	**1,531 (38%)**	**377**	**56 (15%)**	**10,223**	**4,415 (43%)**
Emergency Medical Services Total											4,202	
Emergency Department											9,684	
Trauma Registry											4,201	
Combined dataset	**3,351**	**1,392 (42%)**	**3,336**	**817 (24%)**	**1,581**	**619 (39%)**	**8,451**	**1,531 (18%)**	**492**	**56 (11%)**	**17,211**	**4,415 (26%)**

Abbreviations: **EMS** Emergency Medical Services, **ED** Emergency Department.

[a] Individual injury records may appear in multiple sources.

Table 3

Crude and Adjusted Odds Ratios for Travel Mode Association with Linkage from Logistic Regression Analysis[a]

	Linkage, OR (95% CI)			
Variable	**Crude**		**Adjusted**[a]	
Travel mode[b]				
Cyclist	0.46	(0.41–0.51)	0.37	(0.32–0.42)
Motorcyclist	0.90	(0.80–1.02)	0.67	(0.57–0.78)
Motor Vehicle Occupant	0.31	(0.29–0.34)	0.53	(0.47–0.59)
Other/Unknown	0.18	(0.13–0.24)	0.17	(0.13–0.23)

[a]Adjusted for Injury Severity Category.

[b]Reference group: Pedestrian; Abbreviations: **OR** Odds Ratio, **CI** Confidence Interval.

Table 4

Comparison of matched and unmatched data sources for San Francisco's Transportation-Related Injury Surveillance System (2013–2015).

Data Source	Traditional Data Sources (SFPD, OME) only - Unlinked	%	Hospital and Pre-Hospital Data (EMS, ED, Trauma Registry) only - Unlinked	%	Linked Records	%	Combined Dataset	%
Total	6,988	100%	5,808	100%	4,415	100%	17,211	100%
Year								
2013	2,402	34%**	1,767	30%	1,485	34%	5,654	33%
2014	2,192	31%**	2,116	36%	1,513	34%	5,821	34%
2015	2,394	34%	1,925	33%	1,417	32%	5,736	33%
Mode								
Pedestrian	1,078	15%	881	15%	1,392	32%	3,351	19%
Cyclist	943	13%**	1,576	27%	817	19%	3,336	19%
Motorcyclist	432	6%**	530	9%	619	14%	1,581	9%
Motor Vehicle Occupant	4,420	63%**	2,500	43%	1,531	35%	8,451	49%
Other/Unknown	115	2%**	321	6%	56	1%	492	3%
Age in Years								
Mean (Median)	40 (37)		39 (35)		41 (37)		40 (36)	
<18	386	6%	337	6%	222	5%	945	5%
18–64	5,815	83%**	5,010	86%	3,684	83%	14,509	84%
65 and up	612	9%*	461	8%	509	12%	1,582	9%
Missing	175	3%**	0	0%	0	0%	175	1%
Sex								
Female	3,031	43%**	2,008	35%	1,680	38%	6,719	39%
Male	3,519	50%**	3,800	65%	2,735	62%	10,054	58%
Missing	438	6%**	0	0%	0	0%	438	3%
ZSFG: Race/ethnicity								
Asian			1,040	18%	849	19%	1,889	11%
Hispanic			1,483	26%	875	20%	2,358	14%
Black			863	15%	541	12%	1404	8%

Data Source	Traditional Data Sources (SFPD, OME) only - Unlinked	%	Hospital and Pre-Hospital Data (EMS, ED, Trauma Registry) only - Unlinked	%	Linked Records	%	Combined Dataset	%
White			2,010	35%	1,664	38%	3,674	21%
Native American/Eskimo			12	0%	14	0%	26	0%
Native Hawaiian/Pacific Islander			33	1%	21	0%	54	0%
Other race			76	1%	60	1%	136	1%
Decline to state			1	0%	0	0%	1	0%
Unknown			78	1%	51	1%	129	1%
Not available			212	4%	340	8%	7,540	44%
SFPD: Race/ethnicity								
Asian	1,313	19%			727	16%	2,040	12%
Hispanic	969	14%**			710	16%	1,679	10%
Black	872	12%**			519	12%	1,391	8%
White	2,626	38%*			1,764	40%	4,390	26%
Other race	667	10%**			341	8%	1,008	6%
Unknown	96	1%			63	1%	159	1%
Not available	445	6%**			291	7%	6,544	38%
ZSFG: Injury Severity Score (ISS)								
Not admitted to hospital and no/zero ISS assessed (sub-severe injury)			4,034	69%	1,745	40%	5,779	34%
Admitted or ISS ≤ 15 (severe injury)			1,684	29%	2,393	54%	4,077	24%
ISS > 15 (critical or fatal injury)			90	2%	277	6%	367	2%
Not available			0	0%	0	0%	6,988	41%
SFPD: Severity								
Complaint of pain	5,497	79%			2,413	55%	7,910	46%
Other visible injury	1,371	20%			1,395	32%	2,766	16%
Severe injury	92	1%			538	12%	630	4%
Fatal	28	0%			69	2%	97	1%
Not available	0	0%			0	0%	5,808	34%
Geographic Location of Injury Data								
Available	6,607	95%**	1,071	18%	4,324	98%	12,002	70%

Data Source	Traditional Data Sources (SFPD, OME) only - Unlinked	%	Hospital and Pre-Hospital Data (EMS, ED, Trauma Registry) only - Unlinked	%	Linked Records	%	Combined Dataset	%
Unavailable	381	5%**	4,737	82%	91	2%	5,209	30%

Abbreviations: **SFPD** San Francisco Police Department, **OME** Office of the Chief Medical Examiner, **EMS** Emergency Medical Services, **ED** Emergency Department.

* $p < 0.05$ compared to unlinked records from hospital and pre-hospital data sources.

** $p < 0.0001$ compared to unlinked records from hospital and pre-hospital data sources.

Table 5

Comparing Injury Severity Scores to Police Crash Report Degree of Injury Among Nonfatal Injury Records, San Francisco's Transportation-Related Injury Surveillance System (2013–2015).

	Injury Severity Score			Diagnostic Accuracy Measures	
SFPD Crash Report Severity	**Critical (>15)**	**Severe (≤15)**	**Total**	**(Reference test is Injury Severity Score)**	
Severe Injury	137	364	501	Sensitivity	65%
Visible Injury or Complaint of Pain	75	2181	2256	Specificity	86%
Total	212	2545	2757	Positive Predictive Value (Precision)	27%