Statewide Linkage of Crash, EMS, and Trauma Records

Prepared for:

Kentucky Transportation Cabinet

Prepared by:

University of Louisville

Robert Kluger, Ph.D., PE

Sajjad Karimi

University of Kentucky

Reginald Souleyrette, Ph.D., PE

Teng Wang, Ph.D., PE

1. Introduction

Crash data is a crucial source for crash analysis and prevention investigations. However, in countries like the U.S., multiple reports, including Crash data, EMS runs, and trauma registries are generated alongside crash data when an accident occurs. Each dataset is collected by distinct agencies and contains specific information about the crash and the involved patients. Linking these reports enables researchers to have a more comprehensive understanding of the crash, injuries, and safety outcomes. Consequently, this linkage allows researchers to track the entire process of the crash from beginning to end.

Although crash data linkage is a broad term that can vary based on the structure of available data, the main idea is to identify each individual involved in the crash from each dataset and link all the information related to that person together. The linkage algorithm plays a key role in finding these matches, and it varies across different research projects due to differences in data availability. Identifying and utilizing shared variables collected by agencies constitutes the initial step toward the linkage process as it determines appropriate algorithm for potential matches.

The aim of this project was to create and apply a framework that connects crash data with Emergency Medical Services (EMS) records across counties in Kentucky. The data used in the project were collected from various sources, including the Kentucky State Police (KSP), the Kentucky Board of EMS (KBEMS), and the Kentucky Injury Prevention Research Center (KIPRC).

2. Literature Review

Data linkage is the process of integrating multiple databases which report the same events or information. This procedure is frequently used to enhance comprehension areas of social complex issues and to provide guidance for the creation policies and practices in the science (Kinner et al., 2013) and healthcare (Lyons et al., 2014). In traffic safety, data linkage is becoming more crucial to understanding injury outcomes. In order to develop a more thorough understanding of the patterns and behaviors connected to motor vehicle crashes, data linkage in traffic safety seeks to integrate data from multiple sources (Cryer et al., 2001). Additional information provided by linked safety data sets can aid in understanding the causes and determinants of injury outcomes. Furthermore, patterns and trends found in linked crash data can be used to inform the development of particular safety policies and programs. Links to crash databases, in general, can be extremely helpful in supplying the data required to make decisions that will increase road safety.

Linking crash data in databases relies on the specific research question or issue at hand and the availability of data. The most commonly utilized datasets for linkages are police-reported crash data and hospital data. Police crash reports, created by law enforcement agencies, serve investigative and legal purposes, offering details about crash location, involved vehicles, individuals, road conditions, weather, and other environmental factors contributing to the incident. In the United States, crash databases adhere to the Minimum Uniform Crash Criteria (MMUCC) guidelines, featuring tables for crashes, vehicles, and individuals linked by a key field (NHTSA, 2017). When connecting to hospital datasets, the primary objective is to obtain detailed injury diagnoses and outcomes, deemed more accurate than the initial injury severity estimates found in police reports. Hospital data encompasses various datasets, such as trauma registries, inpatient diagnosis information, or billing records. Numerous studies suggest that linking police reports with hospital admission databases results in more reliable and less biased injury information compared to relying solely on police reports for motor vehicle crashes (MVC) (Amoros et al., 2006; Boufous et al., 2008a; Cryer et al., 2001; Lombardi et al., 2022). Beyond police and hospital data, other datasets employed in traffic safety data linkage efforts include emergency medical services (EMS) computer-aided dispatch, EMS patient care reports (PCR), hospital discharge data, death certificates, insurance data, injury surveillance unit data, and driver and vehicle registration data.

Numerous significant challenges are associated with individual endeavors to connect crash data, including privacy maintenance, technical obstacles, data quality concerns, and data completeness. Technical challenges within crash data linkage involve a lack of standardization in data collection and reporting, making it difficult to match data from diverse sources and establish accurate connections. As an illustration, Kudryavtsev et al. (2013) highlighted the frequent discrepancies between the dates of crashes recorded in the police database and the dates of injuries documented in the hospital database as a notable challenge during the linkage process (Kudryavtsev et al., 2013). Furthermore, the absence of data-sharing agreements and appropriate infrastructure can pose obstacles in implementing effective data linkage.

Privacy and security issues emerge as additional concerns when linking crash data. The absence of identifiers can lead to errors in the linkage process. For instance, in the United States, privacy laws such as the Health Insurance Portability and Accountability Act (HIPAA) result in data being stripped of information that could enhance linkage or, in some cases, restrict access entirely.

Studies examining data linkage in crash databases exhibit notable variability, particularly in terms of the datasets utilized, methodologies employed, and outcomes achieved. One crucial metric in these studies is the match rate, which signifies the proportion of records successfully linked within the dataset. Several factors influence the data linkage rate, encompassing data quality and completeness, the chosen matching algorithm, and the level of agreement between the two datasets. The linkage rates within crash datasets demonstrate variability across different scenarios, with variations observed based on the type of road users (Janstrup et al., 2016; Lujic et al., 2008).

Research suggests that the linkage rate for motorcycle crashes is generally lower compared to other motor vehicles (Alsop & Langley, 2001; Wilson et al., 2012). Alsop et al. found that drivers tend to have higher match rates compared to passengers (Alsop & Langley, 2001). Moreover, the linkage rate has been observed to increase with escalating levels of injury severity, with the highest rates typically associated with fatal crash records (Rosman & Knuiman, 1994; Soltani et al., 2022). Janstrup et al. demonstrated a positive correlation between the likelihood of a record appearing in both datasets and factors such as helmet and seat-belt use, the number of motor vehicles involved, alcohol presence, higher speed limits, and gender (Janstrup et al., 2016).

A low linkage rate can be ascribed to various factors, encompassing definitional issues, the organizational structure of police records, the incapacity of hospital systems to identify traffic crashes, road users' failure to report incidents to the police, and the reliability of the linking variables (Cryer et al., 2001). Wilson et al. delved into the initial probabilistic weights and thresholds of the linkage process and concluded that the method used to link data did not significantly impact the low rate of successful links. Their findings highlight that inaccuracies in spelling names or incorrect recording of other primary linking variables, such as date of birth and date of the crash, could still potentially hinder a successful match (Wilson et al., 2012).

Errors in data linkage often occur when categorizing pairs into matches and non-matches based on the size of conditional probabilities of matches or non-matches without an intermediate set of potential matches. The process involves determining a breakpoint, where records with a linkage score above it are considered matches and linked, while those below it are not linked. Two types of errors, false positives and false negatives, can arise during the matching process. False positives involve pairs mistakenly considered successful matches, while false negatives refer to pairs that were not considered matched but were indeed true matches. In probabilistic methods, selecting a higher probability breakpoint results in lower type 1 errors (fewer false positives), while choosing a lower probability breakpoint reduces type 2 errors (fewer false negatives) (Short & Caulfield, 2016). Bias may occur due to certain data types being more or less likely to be matched for systematic reasons in either the matching process or the data itself (Hosseinzadeh et al., 2022). Quantitative bias analysis serves as a general approach to assess potential biases in datasets (Janstrup et al., 2016; Tarko & Azam, 2011). To identify bias in linked datasets, it is crucial to examine the relationship between variables, use visual aids, ensure alignment with the research question, and thoroughly evaluate variables for bias both before and after linkage. A comprehensive understanding of the sample population is essential when utilizing linked data for analysis.

3. Data Sources

In this project, three primary data sources were utilized: crash data recorded by Kentucky police departments, EMS runs documented by the Kentucky Board of Emergency Medical Services (KBEMS), and trauma registries collected by the Kentucky Trauma Registry. Table 1 displays the number of initial and unprocessed records for each database separately, broken down by year. The crash data spans from 2010 to 2022; however, for the purpose of this project, only records from 2018 onwards were utilized. This decision stems from the project's emphasis on data linkage with other databases, the threshold being determined by the availability of these additional datasets. EMS records were accessible for the entirety of 2021 and 2022, while Trauma registry data ranged from 2018 to 2022.

Table 1. Annual Crash, EMS, and Trauma Data Summary (2018-2022)

Database	2018	2019	2020	2021	2022
CRASH					
Collision	158,475	157,111	119,947	131,732	130,303
Person_Collision	459,801	457,171	332,738	371,955	370,723
Vulnerable road users	1,813	1,804	1,577	1,531	1,698
Was transported	29,324	28,875	25,160	26,060	25,411
EMS					

All	-	-	-	58,982	56,273
Vulnerable road users	-	-	-	2,715	4,256
Patient transported	-	-	-	25,056	24,557
Trauma					
All	12,804	14,219	9,919	13,978	13,216
Transportation related*	4,319	4,375	3,059	3,894	3,660
Vulnerable road users	868	919	797	983	987

*Includes: Motor vehicle, motorcycle, pedestrian, bike, and other transportation related crashes.

3.1. Crash Data

Accident information comprises essential details gathered from police reports documenting incidents throughout the state. This data was acquired from the Kentucky State Police through a formal agreement known as a memorandum of understanding (MOU). The information adheres to the Minimum Model Uniform Crash Criteria (MMUCC) standards set by the National Highway Traffic Safety Administration in 2017. The data is organized into different tables (including collision, location, EMS, and person), connected by a shared crash ID field. The crash, location, and person tables were extensively utilized in the process of linking the data. Every person involved in the accident has an entry in the person table, each with a distinct identification number linked to a specific crash by its crash ID.

Table 2 presents a demographic summary of crash data from 2018 to 2022, detailing crash frequencies, injury and fatality rates, intersection involvement, and person-level characteristics. The total number of crashes peaked in 2018 but declined in 2020, likely due to reduced travel during the pandemic, before increasing again in subsequent years. The proportion of crashes resulting in injuries and fatalities followed a similar trend, with a spike in 2020 and a decline in 2022. Intersection-related crashes consistently comprised about 24-25% of all incidents.

In terms of person-level demographics, the total number of individuals involved in crashes closely mirrors the trends in crash occurrences. Age distribution patterns remained consistent, with individuals aged 31-50 consistently forming the largest group, while younger and older individuals showed relatively stable proportions. The high percentage of unknown age categories suggests data limitations in crash reporting. These insights are valuable for assessing crash trends, identifying vulnerable populations, and guiding traffic safety interventions.

Table 2. Crash Data-Demographic Summary (2018-2022)

Database	**2018**	**2019**	**2020**	**2021**	**2022**
Crash table	158,475	157,111	119,947	131,732	130,303
• Injury (at least one)	15.08%	14.91%	16.88%	16.00%	14.76%
• Fatal (at least one)	0.42%	0.45%	0.59%	0.56%	0.47%
Injury status					
• Total number of injured	34,861	33,774	29,156	30,035	7,770
• Total number of killed	732	766	781	818	192
Intersection					
• No	76.90%	75.08%	76.10%	75.26%	74.30%
• Yes	23.09%	24.92%	23.90%	24.74%	25.70%
Person table	459,801	457,171	332,738	371,955	370,723
Gender					
• Male	42.67%	43.66%	44.29%	43.97%	43.93%
• Female	37.70%	38.91%	37.03%	37.89%	37.02%
• Unknown	19.63%	17.43%	18.68%	18.14%	19.05%
Age					
• Under 18	10.38%	10.80%	9.45%	10.21%	9.91%
• 18-30	21.16%	21.54%	22.31%	21.71%	21.29%
• 31-50	23.56%	24.19%	24.53%	24.52%	24.37%
• 51 and older	21.54%	22.65%	22.16%	22.57%	22.35%
• Unknown	23.36%	20.82%	21.56%	20.97%	22.07%

Figure 1 visualizes the geographic distribution of all reported crashes across Kentucky from 2018 to 2022. The red points represent crash locations, while the yellow background delineates county boundaries. The high density of crashes along major roadways and urban centers highlights key areas of traffic incidents, emphasizing the need for targeted safety interventions and infrastructure improvements.

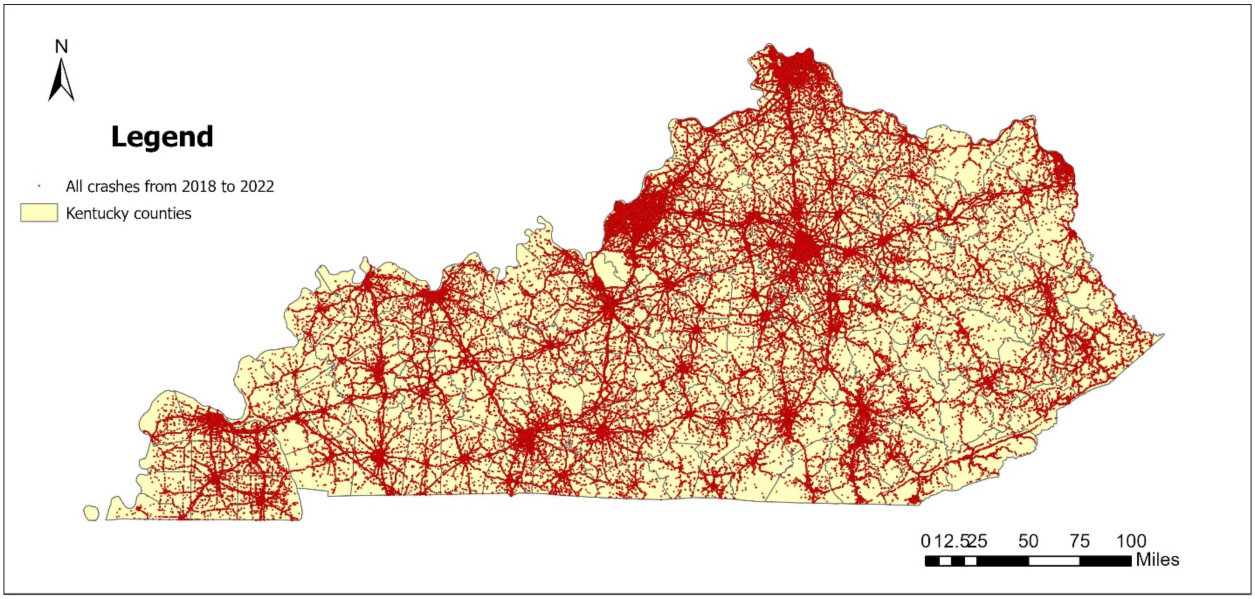

Figure 1. Spatial Distribution of Crash Locations in Kentucky (2018-2022)

3.2. EMS Data

For this research, EMS data were acquired by submitting an open records request to KBEMS, necessitating the filing of IRB protocols with the University of Louisville (U of L) and the Kentucky Community and Technical College System (KCTCS), KBEMS's parent organization. The open records request entailed querying the data from the KBEMS data repository based on the following criteria:

1. **Response Type** (*eResponse.05*) must be **"911 Response (Scene)"**.
2. **Incident Classification:**
   - **Complaint Reported by Dispatch** (*eDispatch.01*) must be categorized as a **"Traffic/Transportation Incident"**, OR
   - **Scene Incident Location Type** (*eScene.09*) must include any of the following: **Street, Highway, or Roadway**.
3. **Patient Care Report Narrative** (*eNarrative.01*) must contain at least one of the following keywords:
   - **Motor vehicle crash, Motor vehicle accident, Motor vehicle incident**

- **Car crash, Car accident, Car incident**
- **Traffic crash, Traffic accident, Traffic incident**
- **Transportation incident, Car wreck, Traffic collision**
- **Motor vehicle collision, Fender bender, Automobile accident**
- **Rollover, Hit-and-run, Traffic Incident, Transportation Incident, Truck Crash**.

Table 3 presents a summary of EMS data and patient demographics for 2021 and 2022, including gender, age distribution, injury types, and agency organizational types. The total number of EMS cases decreased slightly from 58,983 in 2021 to 56,273 in 2022. Males accounted for the largest proportion of cases, followed by females and an increasing percentage of unknown gender classifications. The 31-50 age group had the highest number of patients, with a notable proportion of cases involving older adults and unknown ages. Injury classification shows that vehicle-related injuries were the most common, though they decreased in 2022, while vulnerable road user injuries increased significantly. Among agency types, fire departments and governmental non-fire agencies handled a substantial number of EMS cases, while private non-hospital agencies maintained consistent service levels. These patterns highlight shifts in EMS response dynamics and patient demographics over time.

Table 3. EMS Data-Patient Demographics Summary (2021-2022)

Database	**2021**	**2022**
EMS Data	58,983	56,273
Gender		
• Male	24,236 (43.07%)	23,381 (41.55%)
• Female	22,527 (40.03%)	21,943 (38.99%)
• Unknown	12,220 (16.9%)	10,949 (19.46%)
Age		
• Under 18	6,804	6,698
• 18-30	13,324	12,290
• 31-50	13,719	12,883
• 51 and older	13,786	13,747

• Unknown	11,349	10,655
Category		
Intentional Injuries	109	67
Non-Vehicle Injuries (Accidents - Falls)	121	145
Non-Vehicle Injuries (Accidents - Others)	49	47
Other Injuries	15,560	16,302
Unspecified/Not Recorded	17,415	16,546
Vehicle-Related Injuries	23,013	18,910
Vulnerable Road Users	2,715	4,256
Agency organizational type		
Community, Non-Profit	355	298
Fire Department	17,760	16,945
Governmental, Non-Fire	17,264	15,667
Hospital	3,242	3,090
Private, Nonhospital	20,361	20,273

Figure 2 illustrates the distribution of EMS incidents across various stages of service, highlighting that not all EMS calls require continued medical intervention or hospital transport. The factors on the left represent the total number of EMS responses, including all dispatched units, while the numbers decrease toward the right, reflecting cases where calls were canceled, treatment was provided on scene, or transport was deemed unnecessary. This progression explains why not all EMS records can be linked to trauma or hospital datasets, emphasizing the need to account for these variations when analyzing EMS performance and patient outcomes.

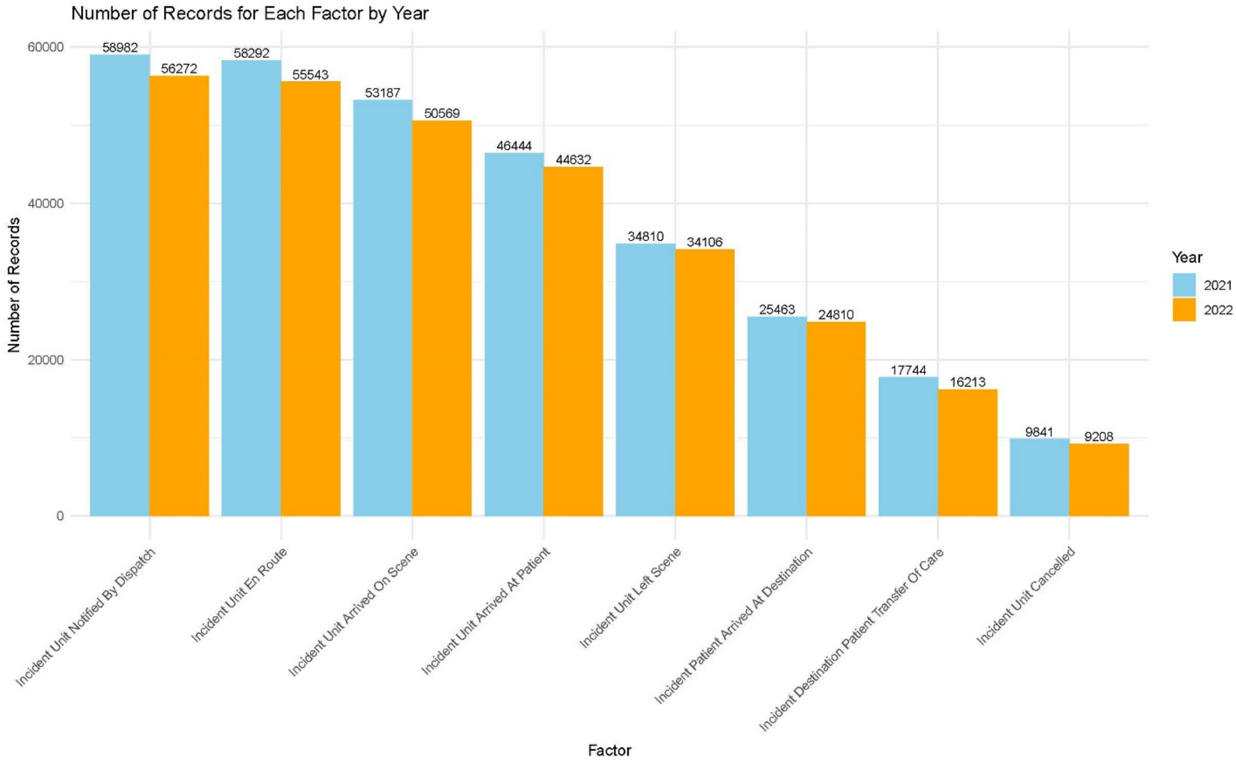

Figure 2. Distribution of EMS Incident Factors by Year (2021-2022)

3.3. Trauma Registry Data

The State Trauma Registry, managed by KIPRC under the ownership of the Cabinet for Health and Family Services (CHFS), stores information on emergency department admissions collected from trauma registries statewide. Access to Trauma Data necessitated a data sharing agreement between U of L, UK, and CHFS. Data retrieval occurs via a secure virtual machine hosted at KIPRC, accessed through a VPN.

Trauma records are categorized into motor vehicle crashes, pedestrian and bikes incidents, and various other classifications. Additionally, there are records that are either unlabeled, labeled as "other," or marked as unspecific, not elsewhere classified, or not documented within the dataset. Following linkage, a detailed review of the cases was undertaken to eliminate inaccurate matches.

4. Methodology

This project, conducted in collaboration with the Kentucky Transportation Cabinet (KYTC), focuses on linking their crash database with EMS and Trauma databases. The goal is to create an integrated dataset that allows for a more comprehensive analysis of EMS response times, patient outcomes, and crash-related injuries.

4.1. Data Preparation

The data linkage process utilizes multiple datasets, including crash reports, EMS response records, and trauma registry data. ArcGIS Pro was used to integrate and process location-based information, ensuring spatial consistency across the datasets. The extracted information includes:

- **Location Information** (Place name, street address, city, county, etc.)
- **Geographical Coordinates** (X, Y values)
- **Address Components** (House number, street name, direction, etc.)
- **Building and Sub-Address Details** (Building type, unit details)
- **Match Information** (Status, score, and match type)
- **Other Relevant Details** (Contact information, ranking, and distance metrics)

4.2. Data Cleaning and Preprocessing

To ensure accurate data matching, initial filtering steps were applied:

- Date Filtering: The EMS dataset was filtered based on the dispatch date, and the crash dataset was filtered using the collision date to ensure records were compared within the same month.
- Age Calculation: The age of each individual was computed based on the difference between the system date and the recorded date of birth.
- Incident Time Calculation: The time difference between crash occurrence and EMS dispatch was computed in minutes to enable precise temporal comparisons.

4.3. Data Linkage Process

To establish meaningful connections between the crash, EMS, and trauma records, the following criteria were used:

1. Geographic Distance Calculation: The spatial proximity between crash locations and EMS dispatch points was measured using the geodist function, with a threshold of 3 km.
2. Temporal Proximity Calculation: The difference in recorded times between the crash and EMS dispatch was calculated, with a threshold of 180 minutes.
3. Age Matching: Records were considered a match if the age difference between individuals was less than one day.
4. Gender Matching: Only records with matching gender values were linked.

4.4. Improved Matching Methodology

To enhance efficiency, a Temporal Normalization and Threshold-Based Matching approach was introduced. Unlike the previous method that required checking each crash-EMS pair individually, this approach normalizes time attributes to a common reference point, allowing for rapid vectorized comparisons. This method significantly improves processing time and reduces computational complexity. All data and reported values utilize the new method in this report.

Advantages of the New Approach:

- **Faster Processing Time**: Reduces the computational burden by avoiding direct pairwise comparisons.
- **Lower Complexity**: Uses vectorized operations instead of exhaustive record-by-record matching.
- **Improved Match Identification**: Identifies matching records based on threshold-based filtering.
- **Scalability for Large Datasets**: Handles large amounts of data more efficiently, optimizing CPU and memory usage.

Comparison of Methods for Crash Data Matching

A summary of the improvements achieved with the new approach is provided below:

Table 4. Comparison of Implemented Methods for Crash Data Matching

Criteria	Method Prior to 2022	New Method (2022-Present)
Matching Approach	Direct equality comparison for each crash pair	Temporal normalization and threshold-based matching
Processing Time	High, due to pairwise comparisons	Lower, due to vectorized operations
Computational Complexity	High	Low
Time Handling	Direct equality checks	Normalization to a common reference point
Match Identification	Manual pairwise checks	Efficient threshold-based identification
Scalability	Time-consuming for large data	Quick and efficient for large datasets
Resource Utilization	High CPU and memory usage	Optimized for better performance

5. Results

5.1. Analysis of EMS and Crash data Linkage Rates for 2021 and 2022

The following section presents an analysis of the EMS linkage rates for crash data in 2021 and 2022. The data includes the number of crashes, EMS records, linked records, and the percentage of EMS runs successfully linked to crash data for each month.

Comparison of 2021 and 2022 EMS Linkage Rates

The total number of crashes recorded in 2021 was 371,958, while in 2022, it slightly decreased to 370,726. Similarly, the total number of EMS runs recorded was 58,982 in 2021 and 56,271 in 2022. Despite the decrease in EMS records, the number of successfully linked EMS cases remained relatively stable, with 32,214 linked cases in 2021 and 31,621 in 2022. This resulted in an improvement in the overall EMS linkage rate from 54.62% in 2021 to 56.19% in 2022.

Key Observations

- The overall increase in EMS linkage rate from 54.62% in 2021 to 56.19% in 2022 suggests an improvement in the accuracy and efficiency of linking EMS data to crash reports.

- The total number of EMS records decreased slightly in 2022, but the number of successfully linked records remained nearly the same, reflecting stable data integration performance.
- The consistency in linkage rates across both years suggests that the data collection and linkage methodology have remained effective, with minor variations due to normal fluctuations in EMS reporting and crash occurrences.

Overall, the improved linkage rate in 2022 highlights progress in crash-EMS data integration efforts. Continuous monitoring and enhancements in data collection, standardization, and integration processes will be essential for maintaining and further improving these linkage rates in the future.

Table 5. Monthly Crash and EMS Data Linkage Summary (2021-2022)

	2021				2022			
Month	**CRASH**	**EMS**	**linked**	**EMS Linkage rate**	**CRASH**	**EMS**	**linked**	**EMS Linkage rate**
Jan	24888	4345	2375	54.66	27850	4026	2013	50.00
Feb	22987	3645	1833	50.29	25578	3773	2061	54.62
Mar	28091	4274	2381	55.71	28540	4221	2379	56.36
Apr	30555	4860	2648	54.49	30086	4624	2661	57.55
May	31881	5236	2825	53.95	32815	5385	2990	55.52
Jun	31867	5200	2823	54.29	29190	4775	2666	55.83
Jul	31413	5389	2999	55.65	28460	4891	2696	55.12
Aug	33460	5534	3010	54.39	31977	4915	2810	57.17
Sep	32729	4984	2801	56.20	32511	4547	2653	58.35
Oct	36688	5551	3036	54.69	34236	5314	3114	58.60
Nov	33942	4950	2742	55.39	34069	4712	2758	58.53
Dec	33457	5014	2741	54.67	35414	5088	2820	55.42
Total	371958	58982	32214	54.62	370726	56271	31621	56.19

5.2. Analysis of EMS and Trauma data Linkage Rates for 2021 and 2022

The following section presents an analysis of the linkage rates between EMS and trauma data for 2021 and 2022. The data includes the number of EMS records, trauma cases, linked cases, and the percentage of trauma cases successfully linked to EMS data for each month.

For comparison between EMS and trauma, we only considered EMS runs that ended at a hospital or healthcare center, where a patient was transported. EMS runs that did not result in a patient being transported were excluded from this comparison. However, for the linkage process itself, all EMS records were considered to prevent any possible matches from being removed. While the overall linkage analysis includes all EMS cases, the real linkage rates should be calculated only based on transported patients for a more accurate assessment of EMS-to-trauma matching effectiveness.

Comparison of 2021 and 2022 EMS-Trauma Linkage Rates

In 2021, there were 25,056 EMS records and 3,894 trauma cases, with 1,418 trauma cases successfully linked to EMS data, resulting in a linkage rate of 36.41%. In 2022, the total number of EMS records slightly decreased to 24,557, while trauma cases also slightly declined to 3,660. However, the number of successfully linked trauma cases increased significantly to 2,506, raising the overall linkage rate to 68.47%.

Key Observations

- The linkage rate for EMS and trauma data improved from 36.41% in 2021 to 68.47% in 2022, indicating a substantial enhancement in data integration processes.
- Despite a minor reduction in the total number of EMS and trauma records in 2022, the number of successfully linked cases increased significantly, suggesting improvements in data accuracy and linkage methodology.
- The increase in linkage rates across both years highlights better consistency and reliability in linking EMS and trauma records, which can contribute to more effective injury surveillance and response planning.

Overall, the improvements in EMS-to-trauma linkage rates in 2022 reflect advancements in data collection, standardization, and integration efforts. These improvements are crucial for ensuring the accuracy and completeness of injury-related data, ultimately supporting more informed decision-making in emergency response and public health planning.

Table 6. Monthly EMS and Trauma Data Linkage Summary (2021-2022)

	2021				2022			
Month	**EMS**	**Trauma**	**Linked**	**Trauma Linkage rate**	**EMS**	**Trauma**	**Linked**	**Trauma Linkage rate**
Jan	1844	260	110	42.31	1654	238	158	66.39
Feb	1468	216	89	41.20	1576	202	136	67.33
Mar	1800	267	96	35.96	1906	296	198	66.89
Apr	2053	355	131	36.90	2131	339	256	75.52
May	2196	391	143	36.57	2374	362	262	72.38
Jun	2143	389	136	34.96	2227	398	282	70.85
Jul	2333	366	126	34.43	2211	346	254	73.41
Aug	2353	375	119	31.73	2207	299	199	66.56
Sep	2097	324	138	42.59	1982	299	170	56.86
Oct	2297	331	131	39.58	2314	376	245	65.16
Nov	2038	277	95	34.30	2033	256	179	69.92
Dec	2106	293	104	35.49	2189	229	167	72.93
Total	25056	3894	1418	36.41	24557	3660	2506	68.47

5.3. Analysis of EMS and Trauma data Linkage Rates for 2021 and 2022

Table 7 provides a detailed breakdown of crash data from 2021 and 2022, categorized by transport status, injury severity, person type, gender, and age group. The table presents the number of linked and non-linked crash records along with the corresponding linkage rates, offering insights into the distribution of different crash characteristics.

5.3.1. Transport Status and Linkage Rates

The data indicate that individuals transported from crash scenes have a significantly higher linkage rate compared to those who were not transported. In 2021, 67.62% of transported individuals were linked, whereas only 4.16% of non-transported cases had a linkage. A similar trend is observed in 2022, with transported cases having a 68.63% linkage rate, compared to 4.05% for non-transported individuals. This pattern suggests that transported cases are more likely to be documented in multiple datasets, likely due to their increased severity or the involvement of medical services.

5.3.2. Injury Severity and Linkage Trends

Crash cases with more severe injuries exhibit higher linkage rates. Among individuals categorized under injury severity codes A (incapacitating injury) and B (non-incapacitating injury), linkage rates exceed 66% in both years. Cases classified under severity code C (possible injury) also maintain a relatively high linkage rate, above 60%. Conversely, cases categorized under severity code O (no injury) exhibit much lower linkage rates, with 4.77% in 2021 and 4.6% in 2022. The low linkage of non-injured individuals suggests that less critical cases may not require follow-up documentation or medical intervention.

5.3.3. Person Type and Linkage Variations

Different road user types exhibit varying linkage rates. Pedestrians and bicyclists demonstrate relatively high linkage rates, with pedestrians showing 48.92% in 2021 and 51.15% in 2022, and bicyclists having linkage rates exceeding 57% in both years. Drivers and passengers, on the other hand, exhibit much lower linkage rates, averaging around 10%, which may be due to the larger number of cases and varying injury severities. Vehicle owners, train engineers, and animal-drawn vehicle users show minimal linkage rates, indicating infrequent or inconsistent documentation in linked datasets.

5.3.4. Gender and Linkage Rates

Males and females exhibit similar linkage rates, with 9.23% for males and 10.73% for females in 2021, and 9.02% and 10.71%, respectively, in 2022. The category "Unknown" was only recorded in 2021, with no linked cases in 2022, while "Other" had minimal representation in both years. These trends indicate that gender does not significantly influence the likelihood of crash records being linked across datasets.

5.3.5. Age Group and Linkage Trends

The analysis of linkage rates across age groups reveals that younger individuals, particularly those under 18, exhibit higher linkage rates compared to older age groups. In 2021, 10.38% of cases involving individuals under 18 were linked, with a similar rate of 10.28% in 2022. The 18-35 age group also maintained a relatively high linkage rate of 10.57% in 2021 and 10.48% in 2022. However, linkage rates decline with increasing age, with individuals above 60 showing the lowest linkage rates at 4.99% in 2021 and 5.15% in 2022. This pattern may reflect differences in medical attention, reporting mechanisms, or crash severity across different age groups.

This analysis highlights key patterns in crash data, showing that transport status, injury severity, and person type strongly influence linkage rates. Severe injuries, non-motorized road users, and younger individuals exhibit higher linkage rates, likely due to increased medical attention and documentation. Conversely, minor injuries, drivers, and older individuals tend to have lower linkage rates, suggesting possible gaps in data integration. Understanding these trends is crucial for improving data linkage methodologies and ensuring comprehensive crash reporting for policy development and traffic safety improvements.

Table 7. Crash Data Summary by Transport Status, Injury Severity, Person Type, Gender, and Age (2021-2022)

		2021			2022		
Category	**Description**	**Linked**	**Not Linked**	**Linkage Rate (%)**	**Linked**	**Not Linked**	**Linkage Rate (%)**
Was Transported	No	14396	331617	4.16	13981	331440	4.05
	Yes	17819	8533	67.62	17640	8063	68.63
Injury Severity Code	Unknown	83	82216	0.1	72	81396	0.09
	O	12345	246404	4.77	11910	247163	4.6
	C	9206	6011	60.5	9012	5677	61.35
	B	8145	4143	66.28	8237	3991	67.36
	A	1975	1010	66.16	1957	925	67.9
	K	461	366	55.74	433	351	55.23
Person Type Code	Driver	22929	184058	11.08	22357	183116	10.88
	Passenger	8449	74018	10.25	8333	74984	10
	Owner	53	81310	0.07	39	80562	0.05

	Pedestrian	565	590	48.92	668	638	51.15
	Bicyclist	205	147	58.24	217	158	57.87
	Animal-Drawn	14	20	41.18	7	24	22.58
	Train Engineer	0	7	0	0	21	0
Gender Code	Male	16033	157743	9.23	15611	157473	9.02
	Female	16181	134609	10.73	16010	133437	10.71
	Unknown	0	1843	0	0	46004	0
	Other	1	45955	0	0	2589	0
Age Category	Under 18	4322	37332	10.38	4312	37622	10.28
	18-35	11887	100620	10.57	11372	97167	10.48
	36-60	10613	99573	9.63	10231	99701	9.31
	Above 60	5393	102625	4.99	5706	105013	5.15

References

Kinner, S. A., Forsyth, S., & Williams, G. (2013). Systematic review of record linkage studies of mortality in ex-prisoners: why (good) methods matter. Addiction, 108(1), 38-49.

Lyons, R. A., Ford, D. V., Moore, L., & Rodgers, S. E. (2014). Use of data linkage to measure the population health effect of non-health-care interventions. The Lancet, 383(9927), 1517-1519.

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

National Highway Traffic Safety Administration (NHTSA). (2017). *Model minimum uniform crash criteria (MMUCC) guideline* (5th ed.). U.S. Department of Transportation. https://www.nhtsa.gov/mmucc

Amoros, E., Martin, J. L., & Laumon, B. (2006). Under-reporting of road crash casualties in France. Accident Analysis & Prevention, 38(4), 627-635.

Boufous, S., Finch, C., Hayen, A., & Williamson, A. (2008). Data linkage of hospital and police crash datasets in NSW. New South Wales Injury Risk Management Research Centre.

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

Lombardi, L. R., Pfeiffer, M. R., Metzger, K. B., Myers, R. K., & Curry, A. E. (2022). Improving identification of crash injuries: Statewide integration of hospital discharge and crash report data. *Traffic injury prevention*, *23*(sup1), S130-S136.

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

Janstrup, K. H., Kaplan, S., Hels, T., Lauritsen, J., & Prato, C. G. (2016). Understanding traffic crash under-reporting: linking police and medical records to individual and crash characteristics. Traffic injury prevention, 17(6), 580-584.

Lujic, S., Finch, C., Boufous, S., Hayen, A., & Dunsmuir, W. (2008). How comparable are road traffic crash cases in hospital admissions data and police records? An examination of data linkage rates. Australian and New Zealand journal of public health, 32(1), 28-33.

Alsop, J., & Langley, J. (2001). Under-reporting of motor vehicle traffic crash victims in New Zealand. Accident Analysis & Prevention, 33(3), 353-359.

Wilson, S. J., Begg, D. J., & Samaranayaka, A. (2012). Validity of using linked hospital and police traffic crash records to analyse motorcycle injury crash characteristics. *Accident Analysis & Prevention*, *49*, 30-35.

Rosman, D. L., & Knuiman, M. W. (1994). A comparison of hospital and police road injury data. Accident Analysis & Prevention, 26(2), 215-222.

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

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.

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