Statewide Linkage of Crash, EMS, and Trauma Records

Prepared for:

Kentucky Transportation Cabinet

Prepared by:

University of Louisville

Robert Kluger, Ph.D., PE

Aryan Hosseinzadeh

University of Kentucky

Reginald Souleyrette, Ph.D., PE

Teng Wang, Ph.D., PE

Executive Summary

In this report, we describe the process and outcome of a data linkage effort between the Kentucky State Crash Database, Kentucky Emergency Medical Services Information System, and the Kentucky State Trauma Registry. The result shows linked crash rate (linked crashes/total crashes) varies 0% to 23.9%, county-level injured persons match rate (linked individuals/total injured crash-involved individuals) varies from 0% to 57.3% and county-level patient care reports match rate (linked individuals/total patient care reports) varies from 0% to 75%. A variable-level analysis was conducted to show which variables were more likely to be present in the linked data set compared to the individual data sets. The project team recommends investigation into additional data sets for inclusion in the linkage activities moving forward, updating query language for improved linkage rates, and investigation into low-linkage rate counties.

Contents

List of Tables

List of Figures

Introduction

Crashes are one of the leading causes of preventable death in the United States, carrying a severe burden on public health and wellness. Police-reported crash data are key to the systematic evaluation of the highway crash problem (Hosseinzadeh et al., 2020) However, additional datasets can help explain factors associated with variance in crash outcomes and indicate how safety may be addressed. Emergency Medical Services (EMS) and Hospitals both collect data about victims of traffic injuries (Hosseinzadeh et al., 2021). Both include specifics of the injury through diagnoses and narratives (Burch et al., 2014; Hosseinzadeh and Kluger 2021a; Hosseinzadeh and Kluger 2021b). The objective of this project was to build and apply, a framework to link crash data to EMS records and trauma records on a statewide, county-by-county basis in Kentucky. Data were obtained from Kentucky State Police (KSP), the Kentucky Board of EMS (KBEMS), and the Kentucky Injury Prevention Research Center (KIPRC).

Literature Review

Record linkage is the process of linking data from different sources. There are three methods applied to link data: manual, deterministic, and probabilistic. Manual linkage is described as “a process that requires human labor and involves visually comparing two (or more) datasets and determining whether each individual episode is the same across datasets” (Dean et al., 2001). Manual linkage is infeasible for processing large amounts of data, e.g., statewide crash data. Deterministic linkage “involves linking records based on an exact agreement of the selected match variables,” such as personal identifiers (Karmel et al., 2010). The deterministic technique necessitates the presence of unique identifiers, such as social security number, in both databases, which is often not the case, especially in datasets that are available to the public, which has often been stripped of identifiers. Probabilistic linkage is defined as “linking records in two (or more) files and is based on the probabilities of agreement and disagreement between a range of match variables” (Karmel et al., 2010). Probabilistic linkage employs models to determine likely matches.

A commonly used record linkage approach is the Bayesian probabilistic method (Conderino et al., 2017; McGlincy, 2004, 2006; Short and Caulfield, 2016; Watson et al., 2015; Winkler, 2002). Several existing software suites can guide users through implementing the Bayesian record linkage approach (Cook et al., 2015). Bayesian record linkage has also been used in the transportation safety context. A study in Dublin, Ireland, utilized Bayesian record linkage to link crash data with both hospital records and injury insurance claims based on age, gender, time, road user type, collision type, crash severity, and county. Their findings indicated a substantially lower linkage rate among bicyclist and motorcyclist injuries (Short and Caulfield, 2016). Conderino et al. (2017) used Bayesian record linkage to link crash and hospital data in New York

City, New York. It was found that 52% of total trauma records were linked to a crash by using date, time, gender, age, role, collision type, injury body location, and injury occurrence (Conderino et al., 2017). Milani et al. (2015) noted that the complexity of the Bayesian approach to probabilistic record linkage was one of the barriers to implementation in states across the US (Milani et al., 2015).

In the United States, Crash Outcome Data Evaluation System (CODES) was a national effort led by the National Highway Traffic Safety Administration (NHTSA) to link hospital records with crash data (Cook et al., 2015). Each participating state was responsible for implementing linkage. Numerous studies utilized the linked datasets to investigate healthcare costs related to specific circumstances such as demographics (Shen and Neyens, 2015), aggressive driving (Chitturi et al., 2011), barrier and median-crossing crashes (Conner and Smith, 2014), seatbelt usage (Han et al., 2017), and motorcycle crashes (Olsen et al., 2014). CODES datasets have also been used to evaluate the quality of police reporting of injuries compared to injury severity ratings by medical professionals. Burdett et al. (2015) found significant differences between KABCO injury severity and Maximum Abbreviated Injury Scale (MAIS) in Wisconsin (Burdett et al., 2015). Burch et al. (2014) found consistency between distributions of injury reports in Maximum Abbreviated Injury Scale (MAIS) between Utah and Maryland crash data among injured persons involved in crashes, while KABCO injury severity varied (Burch et al., 2014). In the United States, the focus has been to link hospital data with crash data, primarily through CODES (Cook et al., 2015), while only a few studies were identified by the authors that linked EMS data with crash data.

Regarding the studies across the world, a study in Portugal linked EMS, crash, and hospital data (Amorim et al., 2014) and used it to assess the quality of injury severity classification by the police using MAIS and length of hospital stay from the hospital data (Couto et al., 2016; Ferreira et al., 2017; Ferreira et al., 2015). The method was also used to assess the length of the prehospital impact on crash injury (Ferreira et al., 2019). A study in Queensland, Australia, linked hospital and crash datasets and found that motorcyclists, bicyclists, males, younger demographics, and injuries occurring in remote locations were more likely to be unlinked (Watson et al., 2015).

Table 1 shows a summary of the crash-related data linkage in previous studies and their linkage rate results. The linkage rate among studies in the literature varies from 29.8% to 74%. Most of the literature employed police-reported crash data and either EMS dispatch data or hospital data. Utilizing four crash-related data sets provides an opportunity to track and monitor crash injuries in each phase of the emergency.

Table 1 Summary of crash-related data linkage literature

Study	Objective	Method	Datasets	Linkage rate	Geographical context
(Moore, 1998)	Comparison of young and adult crashes	MINICODES software (Probabilistic method)	- Police-reported crash data - Hospital data	69% of MVC-related hospital data	Alaska, US.
(Stutts and Hunter, 1999)	Pedestrian and bicyclist crash analysis	Deterministic	- Police-reported crash data - Hospital data	California: 43%*, 45%** New York: 42%*, 56%** North Carolina: 66%*, 67%** *of Bicycle MVC-related hospital data **of Pedestrian MVC-related hospital data	California, US. New York, U.S. North Carolina, U.S.
(Cryer et al., 2001)	Investigating if hospital admission data linked to police MVC reports results in less biased information for the injury prevention policymaker and planner than police MVC reports alone.	Manual method	- Police-reported crash data - Hospital data	50% of MVC-related hospital admissions were found in the linked dataset	England
(Alsop and Langley, 2001)	Exploring under-reporting of motor vehicle traffic crashes	Automatch software package	- Police-reported crash data - Hospital data	63% of the total MVC-related hospital data	New Zealand
(Langley et al., 2003)	Exploring linkage rate of cyclists and the factors associated with the cyclist linkage rate	Automatch software package	- Police-reported crash data - Public road data	22% of cyclist crashes on public roads linked to a crash report	New Zealand
(Sciortino et al., 2005)	pedestrian injury surveillance	Matching thresholds	- Police-reported crash data - Hospital data	59% of the pedestrian MVC-related hospital data	California, US.
(Benavente et al., 2006)	Analysis of Injury Specifics and Crash Compatibility Issues	Probabilistic method	- Police-reported crash data - Hospital data	46% of MVC-related hospital admitted patients	Massachusetts, US.
(Boufous and Williamson, 2006)	Investigating factors affecting work-related traffic crashes	Probabilistic method	- Police-reported crash data - workers compensation data	46% of MVC-related work compensation claims	Australia
(Amoros et al., 2006)	Exploring under-reporting of road crash casualties	Semi-automated record-linkage procedure	- Police-reported crash data - Hospital data	37% of the total MVC-related hospital data	France

(Gonzalez et al., 2006)	Exploring factors affecting mortality in rural areas	Probabilistic algorithm	- Police-reported crash data - Patient Care Reports - Hospital data	73% of the total MVC-related patient care reports	United States
(Lojic et al., 2008)	How comparable are road traffic crash cases in hospital admissions data and police records?	Linkage Wiz software	- Police-reported crash data - Hospital data	45% of the total MVC-related hospital data	Australia
(Tarko and Azam, 2011)	Investigating linked data selection bias in pedestrian crashes	Probabilistic method	- Police-reported crash data - Hospital data	51% of the MVC crashes matched with hospital records	Indiana, U.S.
(Wilson et al., 2012)	Validity of using linked hospital and police traffic crash records to analyse motorcycle injury crash characteristics	Automatch software	- Police-reported data - Hospital data	46% of the hospital data, 60% of serious injuries and 41% of moderate	New Zealand
(Kudryavtsev et al., 2013)	Evaluating reliability of police and healthcare data	Manual	- Police-reported crash data - Hospital data	162 matched fatality cases among 217 police records (74%) and 237 healthcare data (68.3%)	Russia
(Tin Tin et al., 2013)	Completeness and accuracy of cyclist crash outcome Data	deterministic	- Police-reported crash data - Hospital data - Insurance data - Mortality record	13% of hospital reported crashes and 64% of hospital reported crashes were matched with police records, 39% of police reported crashes and 43% of police reported crashes were matched with hospital records	New Zealand
(Mitchell et al, 2015)	comparison of novice and full-licensed driver common crash types	Choice maker software (Probabilistic method)	- Police-reported crash data - Hospital data	54% of MVC-related hospital admitted patients	Australia
(Watson et al., 2015)	Estimating under-reporting of road crash injuries	Linkage Wiz software (Combination of both deterministic and probabilistic approaches)	- Police-reported crash data - Hospital data - EMS data - Injury surveillance unit data	54% of MVC-related hospital admitted patients 29% of MVC-related EMS dispatch data 36% of MVC-related injury surveillance unit:	Australia
(Paixao et al., 2015)	Exploring motor vehicle crash death in high-risk population subgroup	Link Plus (Probabilistic approach	- Police-reported crash data - Mortality information system	1,072 resulted in initial match but manual review	Brazil

				showed 311 of them are true matches	
(Short and Caulfield, 2016)	Linking police data with hospital and injury claims data	Probabilistic approach (Bayesian)	- Police-reported crash data - Hospital data - Injury claims -	61% of the total MVC-related hospital data	Ireland
Janstrup et al., 2016)	Understanding traffic crash under-reporting	Deterministic approach	- Police-reported crash data - Hospital data	23% of the total MVC-related hospital data 34% of the MVC crashes matched with hospital records	Denmark
(Conderino et al., 2017)	Linking traffic crash and hospitalization	LinkSolv 9.0 (probabilistic approach)	- Police-reported crash data - Hospital data	52% of the total MVC-related hospital record	New York, U.S.
(Kamaluddin et al., 2018)	Matching of police and hospital road crash casualty records to investigate underreporting	Deterministic and probabilistic using Microsoft SQL	- Police-reported crash data - Hospital data	4% of MVC-related hospital records matched with police-reported crash data	Malaysia
(Tainter et al., 2020)	Data linkage approach to investigate potential reductions in motor vehicle crash severity	Iterative approach	- Police-reported crash data - EMS data	58% of the total MVC-related EMS data	Massachusetts, US.
(Ceklic et al., 2021)	Investigating MVC characteristics that are predictive of high acuity patients	Linkage Tool (v2. 1.5, Emory University, US.)	- EMS data - Police-reported crash data	62% of MVC-related EMS record matched with police-reported crash data	Australia

Data Sources and Management

This section will outline which and how datasets were obtained, and what fields were used in the data linkage approach. All datasets obtained were from 2018-2019.

Crash Data

Crash data consists of key information collected on police reports filed for crashes across the state. Crash data were obtained from the Kentucky State Police under a memorandum of understanding (MOU). The data are formatted following Minimum Model Uniform Crash Criteria (MMUCC) standards (National Highway Traffic Safety Administration, 2017) with three tables (crash, vehicle, and person) linked by a unifying crash ID field. Both the crash and the person tables were used extensively in the data linkage.

Each crash record has a unique crash ID field and contains information about crash time, location, type, and more. In 2018, a total of 157,351 crash records were obtained. Table 2 outlines all fields present in the crash table. The specific fields used in the data linkage are in bold font.

Table 2 Fields available in crash table dataset

Master File #	Mile Post
Collision Date	Motorcyclist
Collision Time	Commercial Vehicle
Latitude Decimal Number	Young Driver
Longitude Decimal Number	Mature Driver
Weather Code	Pedestrian

First Aid Scene Indicator	Bicyclist
Time Notified	Distracted
Time Arrived	Aggressive
Time Roadway Opened	Impaired
Directional Analysis	Unrestrained
Time Last Left	Intersection
Year	Lane Departure
KABCO	Roadway Departure
KTC_RT	Median Crossover

For every individual involved in the crash, there is a record in the person table. Each person has a unique ID and is mapped to an individual crash through the crash ID. For 2018, a total of 458,546 crash–person records were obtained. Table 3 outlines all fields present in the person table. The specific fields used in the data linkage are in bold font.

Table 3 Fields available in a crash-person table dataset

Master File #	Injury Location Code
Unit Number	Position In/On Vehicle Code
Person Number	Restraint Use Code
Person Type Code	Trapped Code
Birth Date	Ejection From Vehicle Code
Death Date	Ejection Path Code
Age at Collision Time	Suspected Drinking Indicator
Gender Code	Year

Injury Severity	

Figure 1 shows the distribution of crashes in Kentucky. Note the larger clusters of crashes in Jefferson (Louisville), Fayette (Lexington), and northern Kentucky counties (Campbell, Kenton, and Boone).

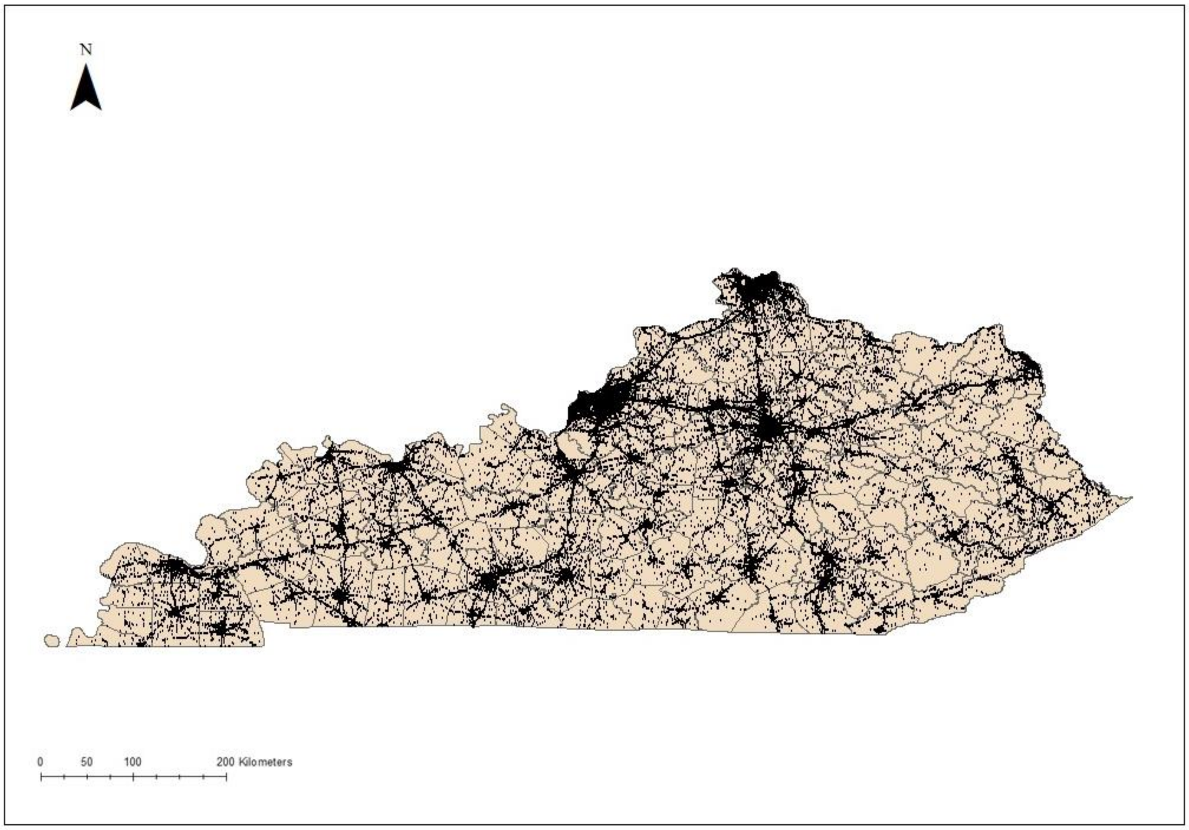

Figure 1. Distribution of crashes in Kentucky

EMS Data

EMS data contain a wide range of information about the EMS response to 911 calls. Each record represents a patient care report (PCR) filed by the team that responded to the emergency. KBEMS collects the data from EMS agencies across the state, standardizes it, and stores it in a state database called KEMSIS. The KEMSIS database follows National EMS Information System (NEMSIS) standard and contains 11 Tables:

- Table 1: EMS responded agency information
- Table 2: Patient medical examinations outcome
- Table 3: Injury automated collision notification
- Table 4: Patient medications given
- Table 5: Patient general body assessments
- Table 6: EMS response description
- Table 7: Scene information and status
- Table 8: EMS times
- Table 9: Vitals information
- Table 10 & 11: Patient examination information

In this study, EMS data were obtained through an open records request to KBEMS which required IRB protocols to be filed with the University of Louisville (U of L) and Kentucky Community and Technical College System (KCTCS), the parent organization of KBEMS. In the open records request, the following criteria were used to query the data from the KBEMS data repository:

1) Response Type (eResponse.05) matches 911 Response (Scene)
2) Complaint Reported by Dispatch (eDispatch.01) matches Traffic/Transportation Incident **OR** Scene Incident Location Type (eScene.09) contains any Street, Highway, Roadway.
3) Patient Care Report Narrative (eNarrative.01) contains one of the following keywords:
    a. 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

For 2018-2019, a total of 57,083 records were requested. Under the HIPAA privacy rule requirements for de-identification, personally identifiable information was stripped from the dataset. Appendix A2 shows the full queries requested from EMS dataset.

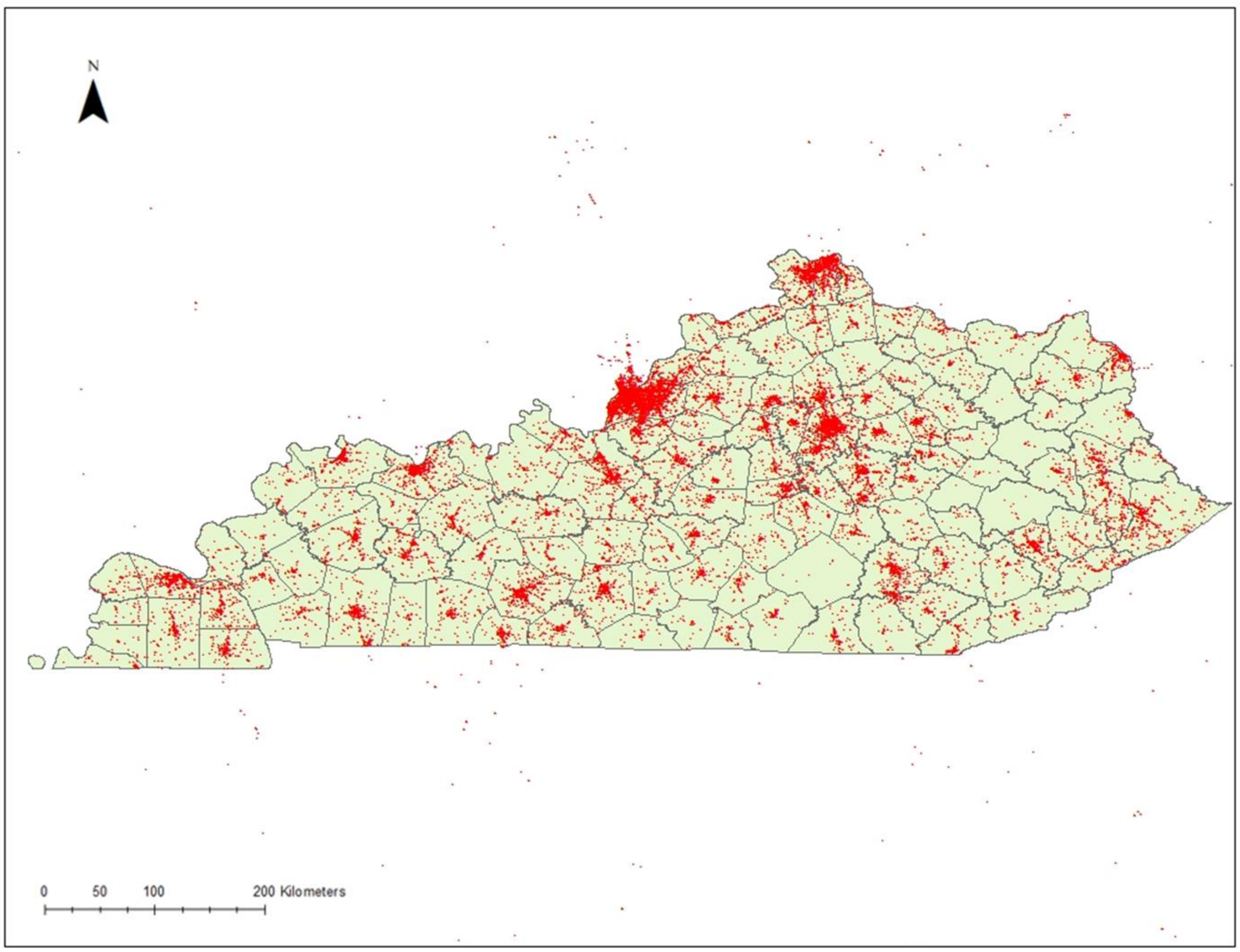

Figure 2. Distribution of EMS runs in Kentucky

Figure 2 shows the density of EMS runs at the county-level. Note the pronounced differences between counties. Jefferson County (26.72 per sq.mi) and Fayette County (13.07 per sq.mi) are the only counties with a density of EMS over 10. At the other extreme, 78 counties (out of 120) recorded less than 1 EMS run per square mile.

Trauma Registry Data

The State Trauma Registry is owned by the Cabinet for Health and Family Services (CHFS) and maintained by KIPRC. It contains data on emergency department admissions reported by trauma registries across the state.

The acquisition of Trauma Data required the signing of a data sharing agreement between U of L, UK and CHFS. Data is accessed through a secure virtual machine housed at KIPRC through a VPN. Table 4 outlines all fields present in the trauma data.

Table 4 Fields available in a trauma dataset

Date of Birth	Hospital Arrival Date & Time
Age	Temperature
Race	Alcohol Use
Gender	Drug Use
Incident Date & Time	Emergency Department Discharge Disposition
Injury Zip Code	Comorbid Condition
Airbag Deployment	Injury Diagnosis
EMS Notify Date and Time	Total ICU Level of Service
EMS Arrival Date and Time	Total Vent Days
EMS Left Date and Time	Hospital Discharge Date and Time
Transport Mode	AIS Severity

EMS Pulse Rate	Trauma Type
EMS Respiratory Rate	Cause Code
EMS Glasgow Coma Scale	Injury Detail
Inter Facility Transfer	Death in Emergency Department
Injury Severity Score	Trauma Type
Admit Service	Blood Alcohol Level
Injury Details	Position in the Vehicle
International Classification of Diseases, Tenth Revision (ICD-10)	ICD-10 Procedure

For 2018-2019, 12,803 trauma records are available in the dataset. Among them, 2979 records labeled as motor vehicle crashes, 267 pedestrian and 167 bikes. Also, there are 734 unlabeled records, 1217 records labeled as "other", 32 records labeled as unspecific, 12 not elsewhere classified and 7 not documented in the dataset that could possibly be related to motor vehicle crashes. However, due to the fact that the cause of the injury could be reported as "not-motor vehicle crashes" but still be related to motor vehicle crashes, the other causes of injury were not filtered out. A closer examination of the cases was conducted after linkage to filter out incorrect matches. Appendix A3 shows the queries requested from trauma dataset.

Method

Data Management and Preparation

MySQL was used in this project for data management, and datasets were stored in a relational database. R studio software was used for data management and statistical analysis (R Core Team, 2019). ArcGIS was used for mapping and spatial analysis. Moreover, although PCR data included latitude and longitude of the events, crash data used the addresses. The Google Maps platform (geocoding API[1]) was employed to provide latitude and longitude of crash locations. The addresses were prepared in a single field to be readable by the Google API. Of 158,332 addresses (Jan 2018 to September 2019) representing all EMS runs, 150,662 were successfully geocoded (geocoding rate: 95.1%). The remaining 7760 records were returned as "NA" or the coordinate found was out of the study area and clearly wrong. For the rest of 7760, the google spreadsheet geocoding add-in tool (Awesome Table) was used. Using this tool successfully geocoded 6540 addressed in the study area (successful geocoding rate: 84.2%). With limiting the data to transportation-related EMS runs and 2018, the number of EMS runs entered to the linking process was 57,083.

Data Linkage

EMS runs and crash incidents are linked through location, time, age, and gender. Incidents reported within a three-kilometer distance and a 3-hour time window, for individuals with the same age and gender in the EMS PCR and crash reports database were considered to be matching pairs. Loops in R studio software were used to compare every two pairs in the crash and EMS data to find candidate matches. Figure 3 shows the algorithm used for this task.

[1] https://developers.google.com/maps/documentation/geocoding/overview

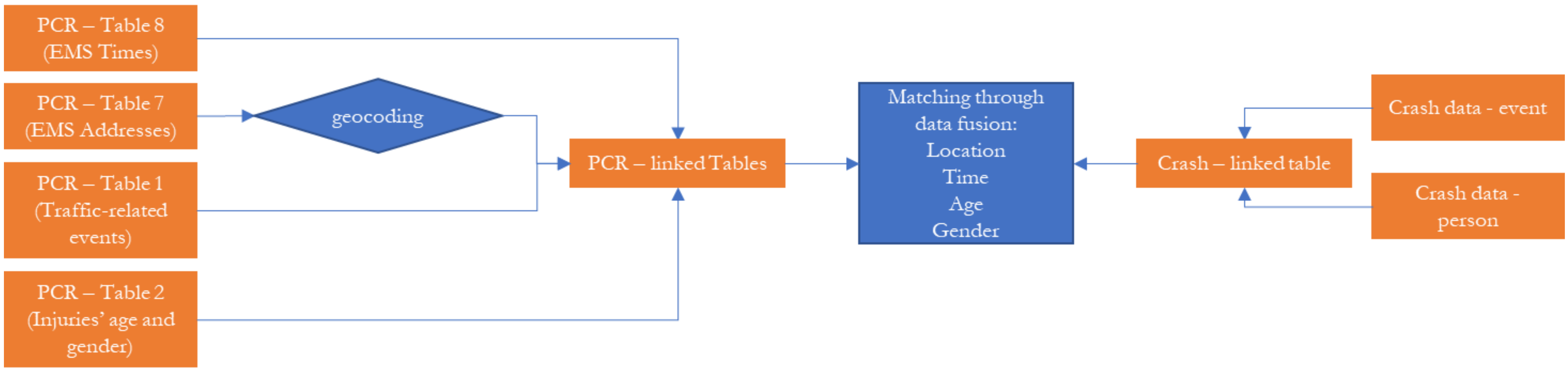

Figure 3. The algorithm applied to link PCR data and crash data

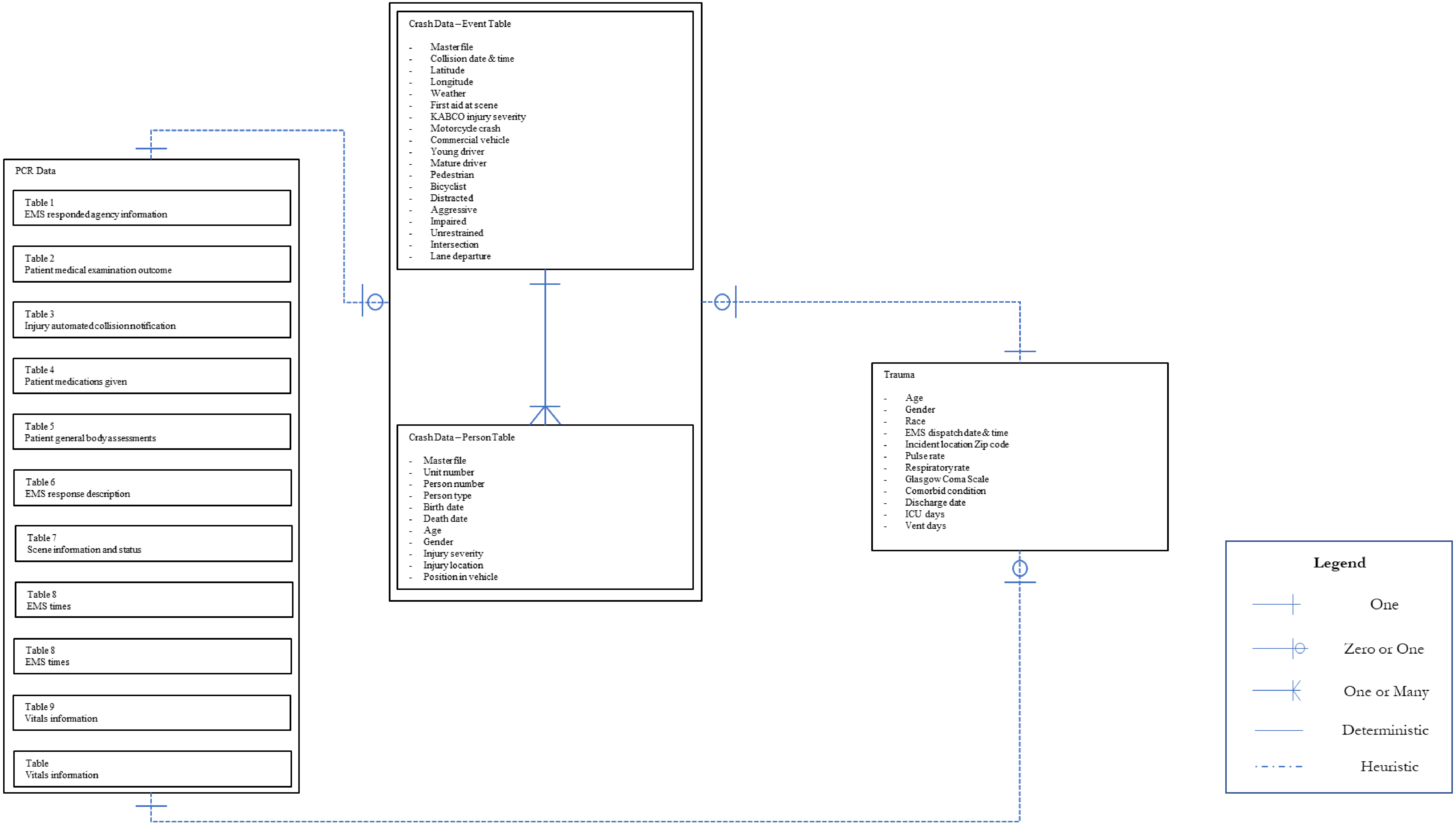

Figure 4. Entity relationship diagram of available datasets

Results

Figure 4 shows the entity-relationship diagram of datasets used in this project and relationships among them. A unique match is the favorable result (i.e., one crash-person linked with one EMS PCR). There were a few duplicate matches (i.e., one crash-person linked with two and more EMS runs, or one EMS run linked with two and more crash- persons), but these were not considered for further analysis in this project.

Police-reported Crash-EMS Linkage - State and County-Level Results

Key metrics tracked include the total number of records in each linked database and the rate at which a match was obtained for each database. These metrics were calculated for the entire state, as well as on a county-by-county basis. Table 5 shows the linkage rates of matched records on a state-level basis.

Table 5 Linkage percentage of crash-events/crash-person/EMS runs

Metric	Description	State-level Outcome	Map
% of linked crash records	# of linked crash IDs (matched with EMS runs) / # of all crash IDs	8.4%	Figure 5
% of linked crash-person records	# of linked crash-person IDs (matched with EMS runs)/ # of all crash-person IDs	5.5%	Figure 6
% of linked injured crash-person records	# of linked injured crash-person IDs (matched with EMS runs)/ # of all injured crash-person IDs	44.7%	Figure 7
% of linked EMS runs	# of linked EMS runs (match with crash-person table) / # all EMS runs	44.9%	Figure 8

Figure 5 shows the county-level crash data match rate (Linked Crashes/Total Crashes). Note that the match rate varies from 0 to 23.9% across counties. Most crashes do not require EMS so the low percent of total crashes linked is expected.

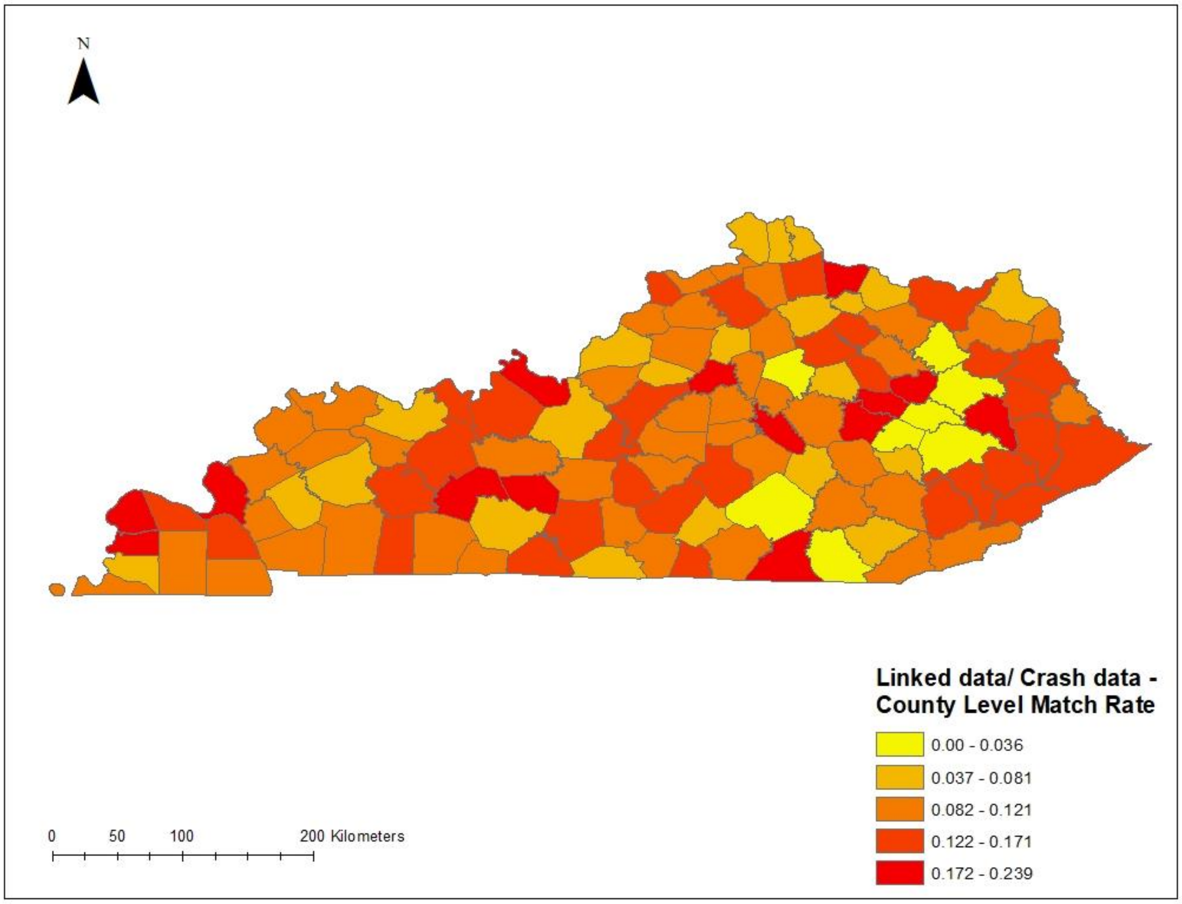

Figure 5. County-level crash data match rate

Figure 6 shows the county-level crash-person data match rate (Linked Crash-persons/Total Crash persons). Note that the match rate varies from 0 to 17.2% across counties.

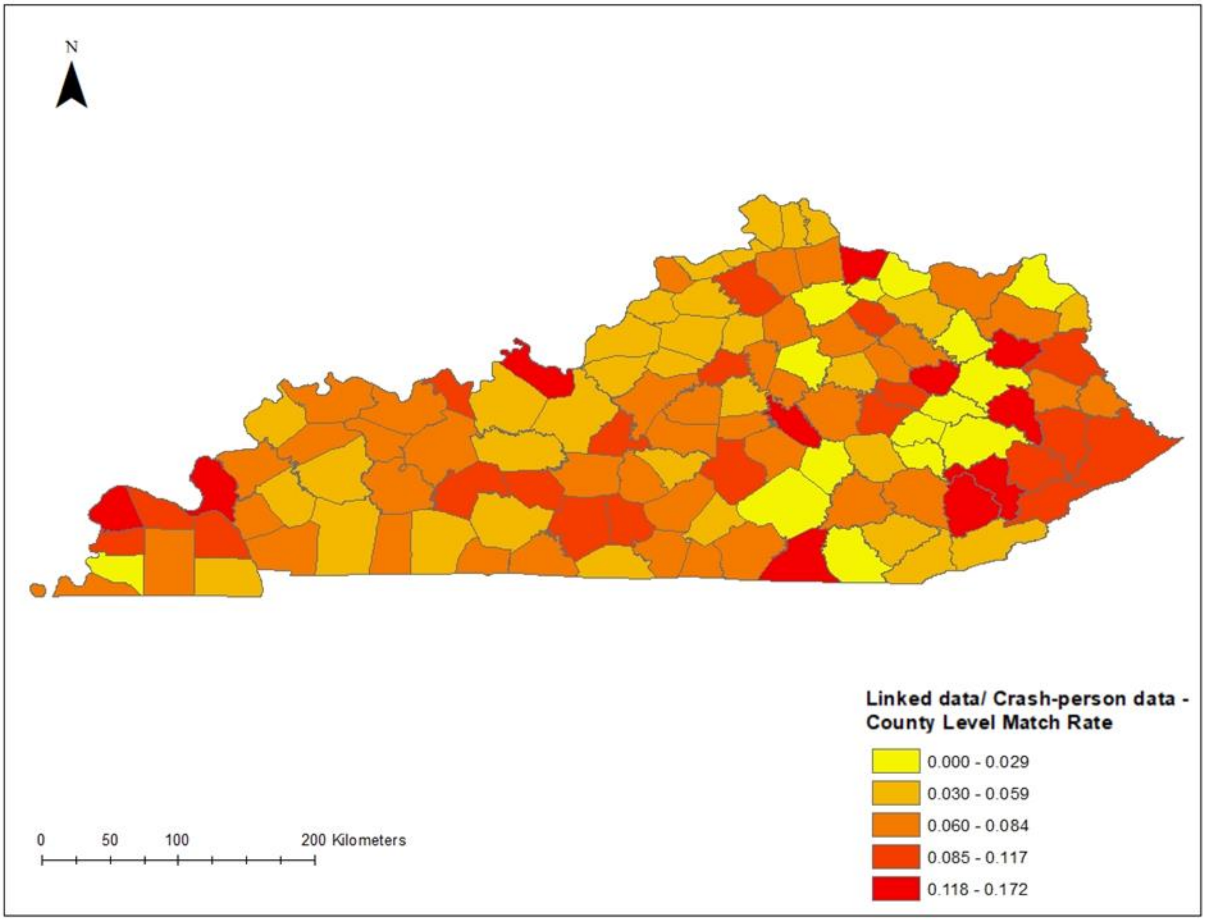

Figure 6. County-level crash-person data match rate

Figure 7 shows the county-level injured persons match rate (Linked Individuals/Total Injured Crash-involved Individuals). The match rate varies from 0 to 57.3% across counties.

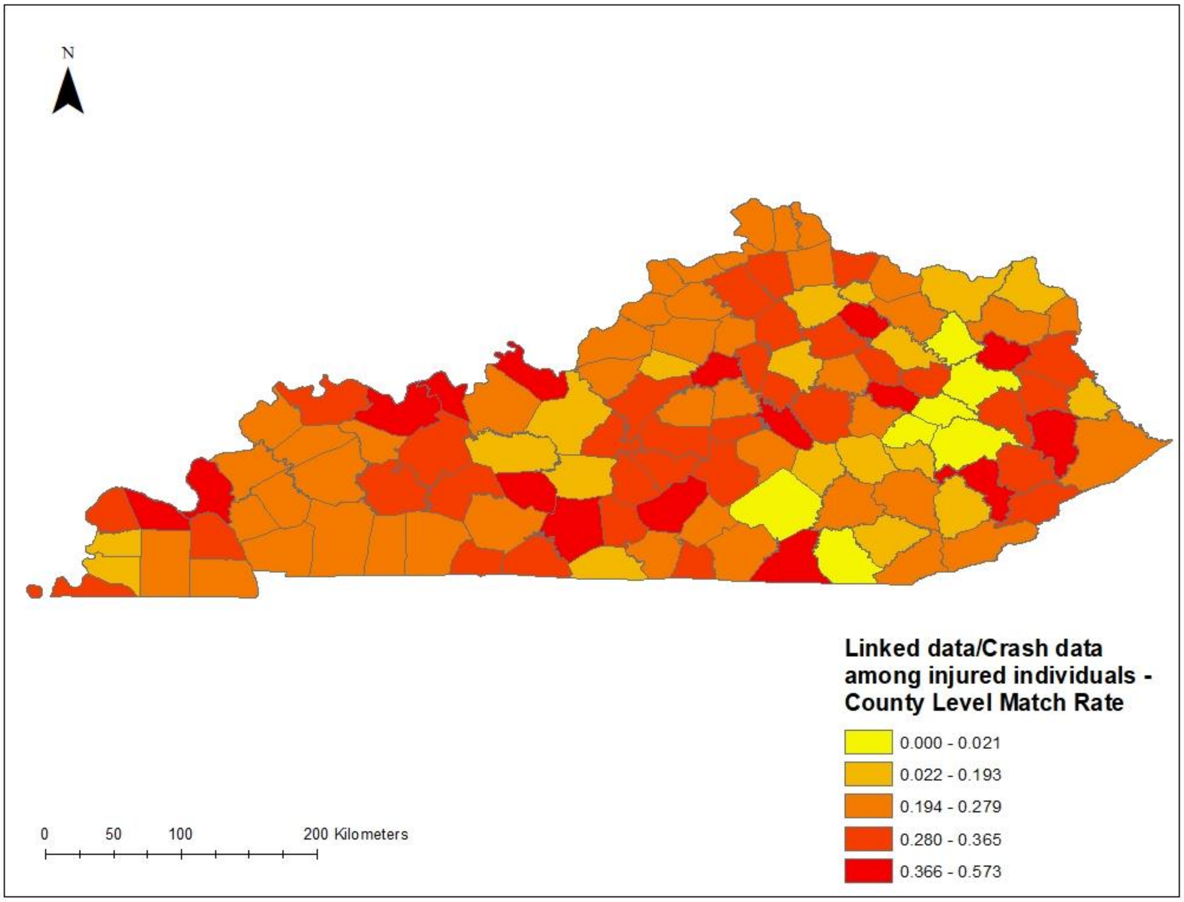

Figure 7. County-level injured persons match rates

Figure 8 shows the county-level PCR match rate (Linked Individuals/Total Patient Care Reports). The match rate varies from 0 to 75% across counties.

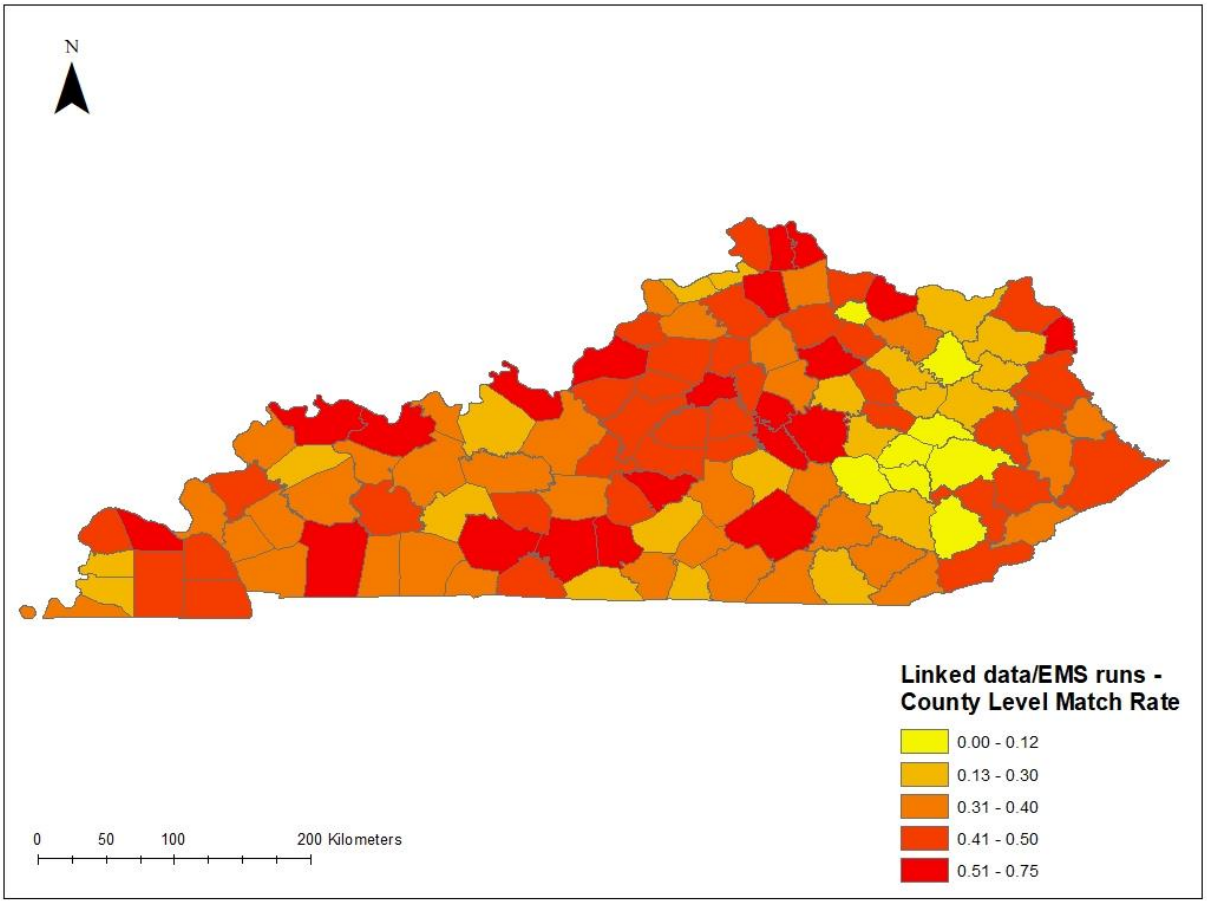

Figure 8. County-level PCR data match rate

Several observations can be made regarding the linkage success rate. While one would not expect every crash to match to an EMS patient care report, it should be expected that most EMS patient care reports should be assigned to a crash-involved individual, given how the EMS runs were queried.

Lower rates of crash linkages can be explained through several characteristics. First, and foremost, not all crashes require an EMS response. Of those that do require an EMS response, fatal crashes where this is not an opportunity to provide care also do not have patient care reports filed. Finally, it is possible that the query used excluded some cases. For example, if an EMS agency doesn't define a motor vehicle crash correctly, it might not end up in the EMS runs dataset based on the search parameters defined.

Variable-level Analysis of Match Rates

This section investigates differences between the linked datasets and the original datasets in terms of variable distributions. Table 6 displays characteristics of several variables among the linked data, crash data, and PCR data.

Table 6 descriptive comparison of records in linked data, crash data and PCR data

	Linked dataset (n = 25,664)		Crash data (n = 157,351)		PCR data (n = 57,083)	
	Avg	sd	Avg	sd	Avg	sd
Age	38.23	20.17	37.91	19.69	40.36	21.01

	Linked dataset (n = 25,664)	Crash data (n = 157,351)	PCR data (n = 57,083)
Gender			
Male	47.99%	52.83%	54.16%
Female	52.01%	47.17%	45.84%
Injury severity			
O	37.87%	90.19%	-
C	33.05%	5.42%	-
B	22.61%	3.39%	-
A	5.29%	0.78%	-
K	1.15%	0.20%	-
Pedestrian			
Yes	2.59%	0.77%	1.48%
No	97.41%	99.23%	98.52%
Bicycle			
Yes	0.62%	0.18%	-
No	99.38%	99.82%	-
Intersection			
Yes	35.44%	25.93%	-
No	64.56%	74.07%	-
Suspect of Drinking			
Yes	4.66%	2.18%	-
No	95.34%	97.82%	-

Although the average age in the linked data and crash data are almost the same, injuries transferred to the hospital averaged approximately two years older. More males were involved in the crashes.

However, more females were transferred to the hospital, and more females were available in the linked data. Moreover, more than 90 percent of the incidents in crash data are labeled as no-injury crashes. In comparison, this percentage for linked data is less than 40 percent. It's expected to have fewer no injury crashes in the linked data since the probability of request for an EMS would decrease for cases without injuries. The percentage of pedestrian and bicycle crashes is more than three times that of the linked data. More intersection crashes are also available in the linked data, probably because intersection crashes tend to be more severe than other crashes and involve more people (since there are usually multiple cars) leading to more opportunity for injury. Suspected of drinking cases were found to be more likely to be linked.

At the county level, there are different reasons for low match rates PCR data. First, these are the counties with very low numbers of crash/EMS runs, sometimes just because of the small size of the county. For example, the match rate in Roberson County is only 9 percent. However, one should consider that only 11 EMS runs met the query criteria in this county in 2018. In some counties, the match rate is suspiciously low. For example, for Lee and Wolfe counties, no traffic incident EMS runs were reported in 2018. According to Appendix 1, Wolfe County had 448 crashes and 92 injuries during that time period. One recommendation from this finding is to reevaluate the query used and to further investigate how possible errors in reporting may have led to this issue.

Some counties with even relatively high numbers of EMS runs produced poor linkage results. For instance, in Leslie County, among 176 EMS runs, only nine were matched by PCR (5 percent). Pulaski (75 percent), McCracken (66 percent), and Meade (59 percent) counties have the highest PCR data match rates (Although in Pulaski, only 4 EMS runs were recorded in 2018).

Police reported Crash - EMS Runs -Trauma Linkage

The police reported crash - EMS runs -trauma linkage was conducted between the linked dataset and trauma data. Date of birth, age, gender and race of the injured individuals in the linked data matched with the ones in the trauma data. Also, crash date and time in crash data matched with hospital admission time and a window of 12 hours have been used as the threshold. Incident date and time and EMS times reported in trauma data were also used; however, this filed is not reported for most of the crashes thus were not helpful extensively. Incident zip code in trauma data were the only location specific field to use for the linkage and matched with zip code reported in crash data.

After performing the initial linkage a few steps were conducted to validate the linked data. First, a couple of fields such as position in the vehicle in both crash data and trauma data were compared. Second, the based on the location of the hospitals that the injured individuals were transported, the cases with high transported distance (more than 100 km) from scene to the hospital were gone under close attention to make sure these cases are true matches. The third step focused on injury details description. Text mining approach was used to make sure all the records, regardless of injury cause listed in another field, are actually related to motor vehicle crashes. The fourth step was a manual random check to ensure there is no systematic error in the matched dataset and figuring out the reasons for unmatched pairs. A detailed elaboration on the reasons of unmatched pairs were provided in the next section.

As a result of the matching process, the final linked crash, EMS runs, and trauma data is included 235 records. Table 7 shows the attributes of the linked dataset and the descriptive information of the fields.

Table 7. Descriptive statistics of some of variables in crash-EMS runs-trauma registry linked data

Attributes	Frequency	Percentage
Injury severity		
K	8	3.4%

A	93	39.6%
B	75	31.9%
C	51	21.7%
O	8	3.4%
Pedestrian	23	9.7%
Bicyclist	2	0.8%
Gender		
Male	138	58.7%
Female	97	42.3%
Transport Mode		
Ground ambulance	205	87.3%
Helicopter	26	11.1%
Private/public vehicle	3	1.2%
NA	1	0.4%
Admit Service		
Trauma	148	58.4%
Neurosurgery	9	3.5%
Orthopedics	30	11.8%
Medicine	12	4.7%
Others/NA	36	15.3%
Position in the car		
Driver	147	58.1%
Front Passenger	27	10.6%
Back Passenger	6	2.3%
Not specified/ NA	73	28.8%

Attribute	**Average**	**S.D.**	**Min**	**Max**
Age	43.1	21.8	1	96
Injury Severity Score	11.9	10.11	1	66

The matching process of the trauma data was conducted separately with police-reported crash dataset and EMS runs. 246 crash-trauma data and 286 EMS-trauma records were available in the these linked datasets.

Discussion, Recommendations and Conclusions

The objective of this project included building and applying a framework to link crash data to EMS records and trauma records on a statewide, county-by-county basis in Kentucky. Data were obtained from Kentucky State Police (KSP), the Kentucky Board of EMS (KBEMS), and the Kentucky Injury Prevention Research Center (KIPRC). The results section and appendixes outlined the linkage performance at the state and county levels.

There are some suspicious results in which further investigation into the data is needed. For example, although there were 191 individuals involved in crashes, including 25 injury individuals in Lee County in 2018, there were no EMS runs reported during the same period. Additional suspicious results such as Pulaski County (6527 crash-person records, 930 injured crash-person records and only 4 EMS runs in 2018) and Rowan County (2797 crash-person records, 341 injured crash-person records and 3 EMS runs in 2018). These warrant a deeper look into the queries made for EMS data, the methods implemented, and more.

Non-matched Records

The manual review provides an opportunity to ascertain performance of the linkage algorithm. Overall, more than 100 records were reviewed manually to investigate the quality of the matching algorithm and further fine tune the parameters. Specifically, we reviewed non-matches and how inconsistencies lead to a lower match rate.

a) *Data incompleteness*

Some variables play a vital role in the linkage process as strong identifiers such as age, exact date of birth and gender. However, data incompleteness in these attributes causes the linkage serious problems. Data incompleteness in some of the important attributes is provided in Table 8.

Table 8. Incompleteness percentage in some of the important attributes

Attribute	No. of incomplete records	Total No. of Records	Incompleteness percentage
Age (EMS runs)	10,487	57,082	18.37%
Gender (EMS runs)	10,427	57,082	18.26%
Date of Birth (Crash data)	72,260	458,545	15.70%
Age (Crash data)	72,260	458,545	15.70%
Gender (Crash data)	55,343	458,545	12.10%

b) *Incomplete or Inconsistent Formatting of Text Fields*

Due to the formatting of addresses in the EMS data, geocoding was implemented to determine the latitude and longitude in EMS data. The addresses sometimes are incomplete or imprecise resulting in geocoding failures. For example, there is a pair of records in trauma data, and linked crash-EMS run data in which all the indicators matched except the location. After a careful deeper look at the attribute, it can be realized the issue is how precise the recorded address was in the EMS data. The address was "KY-194, Pikeville, KY 41501" which could be the span 30 kilometers of a road. Formatting of addresses was also a notable issue.

c) *Data Entry Error*

Another case found was two pairs of matches in linked crash-EMS runs and trauma in which the birthday of the injured individual may have been recorded incorrectly. While all the other attributes matched and insinuated the pair records were related to a specific injured individual, the birthday in EMS data was "10/3/1986", while it was recorded as "10/3/1987" in trauma data. It is not possible to fully correct for data entry errors, though it is possible to implement checks and relax the parameters of the matching algorithm to catch the most common suspected errors. The most common entry errors must first be identified to account for this.

d) *Transported with the helicopter or private/public vehicle*

Some of the true matches that were not matched successfully through the linkage scheme are related to the fact that the injured individuals in cases transported with helicopter or private/public vehicles. So, these cases are not in EMS runs data then cannot find in the crash-EMS runs linked dataset previously matched. Therefore, it's not available in the crash-EMS runs-trauma linked data. 67.52% (8,645/12,803) of the records used ground ambulance for the EMS transport mode and the rest of 32.48% of the records were used other methods of transport. EMS data is a critical part of the linkage methodology, and the gap will lead to lower success rates in matching.

e) *The transported from the referred facility*

Some EMS runs included inter-facilities transfers (transfers between hospitals). In these cases, the time between the crash and hospital admission might be several days even since the EMS run is still associated with a crash. In these cases, it's difficult to ensure the matches are accurate. Only 39.1% (5018/ 12803) of the records were transported straight from the scene to the hospital.

f) *Recorded as motor vehicle crashes but it's not*

Some cases in trauma records are recorded as MVC in trauma records but may not be classical cases included in other datasets. Digging into the injury detail description shows this phenomenon. For example, one record was recorded as "Ped vs. dump truck while working". This will count as an unsuccessful match of an MVC-related trauma record even though matching this type of case is not among the objectives of this analysis.

g) *Reporting*

Gaps in reporting varied among datasets. Follow up with data managers indicated that several agencies are failing to fully report data to their respective systems, particularly within KEMSIS and the Trauma Registry. For example, Rowan County reported three total EMS runs that were valid to be included within the linkage.

h) *Categorization and Capture of Data*

When querying the data sets from their original sources it is possible that the query did not capture how certain counties or agencies recorded information. A review of the consistency and quality of reported data may help to ensure each field is operating as intended.

Recommendations

Based on the project outcomes the team recommends the following steps be taken to further the findings of this project.

1. Additional quality checks into counties with low linkage rates relative to expected. Subsequent adjustments to the algorithm to improve linkage rates.
2. Modeling of expected linkage rates for key benchmarks based on county characteristics
3. Identifying new data sources for inclusion in this database to improve linkage rates or data coverage.
4. Continue data linkage efforts moving into 2022

References

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

Amorim, M., Ferreira, S., Couto, A., 2014. Linking police and hospital road accident records: how consistent can it be? Transportation Research Record 2432, 10-16.

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

Benavente, M., Knodler Jr, M.A., Rothenberg, H., 2006. Case study assessment of crash data challenges: Linking databases for analysis of injury specifics and crash compatibility issues. Transportation research record 1953, 180-186.

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

Burch, C., Cook, L., Dischinger, P., 2014. A comparison of KABCO and AIS injury severity metrics using CODES linked data. Traffic injury prevention 15, 627-630.

Burdett, B., Li, Z., Bill, A.R., Noyce, D.A., 2015. Accuracy of injury severity ratings on police crash reports. Transportation research record 2516, 58-67.

Chitturi, M.V., Ooms, A.W., Bill, A.R., Noyce, D.A., 2011. Injury outcomes and costs for cross-median and median barrier crashes. Journal of safety research 42, 87-92.

Conderino, S., Fung, L., Sedlar, S., Norton, J.M., 2017. Linkage of traffic crash and hospitalization records with limited identifiers for enhanced public health surveillance. Accident Analysis & Prevention 101, 117-123.

Conner, K.A., Smith, G.A., 2014. The impact of aggressive driving-related injuries in Ohio, 2004–2009. Journal of safety research 51, 23-31.

Cook, L.J., Thomas, A., Olson, C., Funai, T., Simmons, T., 2015. Crash Outcome Data Evaluation System (CODES): An Examination of Methodologies and Multi-State Traffic Safety Applications.

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.

Dean, J.M., Vernon, D.D., Cook, L., Nechodom, P., Reading, J., Suruda, A., 2001. Probabilistic linkage of computerized ambulance and inpatient hospital discharge records: a potential tool for evaluation of emergency medical services. Annals of emergency medicine 37, 616-626.

Ferreira, S., Amorim, M., Couto, A., 2017. Risk factors affecting injury severity determined by the MAIS score. Traffic injury prevention 18, 515-520.

Ferreira, S., Amorim, M., Couto, A., 2019. The prehospital time impact on traffic injury from hospital fatality and inpatient recovery perspectives. Journal of Transportation Safety & Security, 1-21.

Ferreira, S., Falcão, L., Couto, A., Amorim, M., 2015. The quality of the injury severity classification by the police: An important step for a reliable assessment. Safety science 79, 88-93.

Han, G.-M., Newmyer, A., Qu, M., 2017. Seatbelt use to save money: Impact on hospital costs of occupants who are involved in motor vehicle crashes. International emergency nursing 31, 2-8.

Hosseinzadeh, A., Karimpour, A., Kluger, R. and Orthober, R., 2020. A framework to link crashes to emergency medical service runs and trauma admissions: for improved highway safety monitoring and crash outcome assessment. In Transportation Research Board. 99th Annual Meeting Transportation Research Board.

Hosseinzadeh, A. and Kluger, R., 2021a. Do EMS times associate with injury severity?. Accident Analysis & Prevention, 153, p.106053.

Hosseinzadeh, A. and Kluger, R., Data Linkage for Traffic Safety in Jefferson County, Kentucky. In International Conference on Transportation and Development 2021b (pp. 243-250).

Hosseinzadeh, A., Haghani, M. and Kluger, R., 2021. Exploring Influencing Factors on Crash-related Emergency Response Time: A Machine Learning Approach (No. TRBAM-21-00614).

Karmel, R., Anderson, P., Gibson, D., Peut, A., Duckett, S., Wells, Y., 2010. Empirical aspects of record linkage across multiple data sets using statistical linkage keys: the experience of the PIAC cohort study. BMC health services research 10, 41.
Kentucky State Police. Kentucky collison for the Public. Available: http:// crashinformationky.org (accessed 21.07.06).
Kudryavtsev, A.V., Kleshchinov, N., Ermolina, M., Lund, J., Grjibovski, A.M., Nilssen, O. and Ytterstad, B., 2013. Road traffic fatalities in Arkhangelsk, Russia in 2005–2010: Reliability of police and healthcare data. Accident Analysis & Prevention, 53, pp.46-54.
McGlincy, M.H., 2004. A Bayesian record linkage methodology for multiple imputation of missing links, ASA Proceedings of the Joint Statistical Meetings. Citeseer, pp. 4001-4008.
McGlincy, M.H., 2006. Using test databases to evaluate record linkage models and train linkage practitioners. Proceedings of the 29th American Statistical Association, Survey Research Method Section, Seattle, WA, 3404-3410.
Milani, J., Kindelberger, J., Bergen, G., Novicki, E., Burch, C., Ho, S., West, B., 2015. Assessment of characteristics of state data linkage systems.
National Highway Traffic Safety Administration, 2017. MMUCC Guideline: Model Minimum Uniform Crash Criteria Fifth Edition. (Report No. DOT HS 812 433). Washington, DC: National Highway Traffic Safety Administration.
Olsen, C.S., Thomas, A.M., Cook, L.J., 2014. Hospital charges associated with motorcycle crash factors: a quantile regression analysis. Injury prevention 20, 276-280.
R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
Sciortino, S., Vassar, M., Radetsky, M., Knudson, M.M., 2005. San Francisco pedestrian injury surveillance: mapping, under-reporting, and injury severity in police and hospital records. Accident Analysis & Prevention 37, 1102-1113.
Shen, S., Neyens, D.M., 2015. The effects of age, gender, and crash types on drivers' injury-related health care costs. Accident Analysis & Prevention 77, 82-90.
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.
Tin, S.T., 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.
Tainter, F., Fitzpatrick, C., Gazillo, J., Riessman, R. and Knodler Jr, M., 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, pp.9-15.
Watson, A., Watson, B., Vallmuur, K., 2015. Estimating under-reporting of road crash injuries to police using multiple linked data collections. Accident Analysis & Prevention 83, 18-25.
Winkler, W.E., 2002. Methods for record linkage and bayesian networks. Technical report, Statistical Research Division, US Census Bureau.

Appendix

Table A1 Number of crashes, EMS runs, number of linked records, and match rate at the county level for Kentucky

County name	# of crash-person records	# of injury crash-person records	# of EMS runs	# of linked EMS run records - crash records	# of linked EMS run records - crash records among injured individuals	# of linked crash-persons/ # of crash - persons	# of linked injury crash-persons/ # of all injury crash-persons	# of linked EMS runs/# of EMS runs
Adair	737	114	209	59	44	0.080	0.386	0.282
Allen	1461	215	247	118	76	0.081	0.353	0.478
Anderson	1678	298	341	193	138	0.115	0.463	0.566
Ballard	489	120	162	72	39	0.147	0.325	0.444
Barren	4757	813	815	430	314	0.090	0.386	0.528
Bath	566	122	167	38	23	0.067	0.189	0.228
Bell	2218	326	322	122	91	0.055	0.279	0.379
Boone	19579	2712	1317	660	542	0.034	0.200	0.501
Bourbon	1875	310	247	136	100	0.073	0.323	0.551
Boyd	4686	779	407	215	184	0.046	0.236	0.528
Boyle	3185	414	533	234	128	0.073	0.309	0.439
Bracken	435	83	113	54	29	0.124	0.349	0.478
Breathitt	902	225	9	1	1	0.001	0.004	0.111
Breckinridge	738	151	175	41	38	0.056	0.252	0.234
Bullitt	6816	1346	730	322	334	0.047	0.248	0.441
Butler	634	125	189	57	40	0.090	0.320	0.302
Caldwell	1289	171	143	53	42	0.041	0.246	0.371
Calloway	3686	416	391	186	105	0.050	0.252	0.476
Campbell	10482	1185	679	377	249	0.036	0.210	0.555
Carlisle	149	70	73	16	13	0.107	0.186	0.219
Carroll	1150	204	243	64	44	0.056	0.216	0.263
Carter	1911	226	433	122	57	0.064	0.252	0.282
Casey	544	107	169	55	34	0.101	0.318	0.325
Christian	6033	1406	622	324	302	0.054	0.215	0.521
Clark	4011	529	824	187	139	0.047	0.263	0.227
Clay	1296	383	405	100	80	0.077	0.209	0.247
Clinton	424	83	108	31	26	0.073	0.313	0.287
Crittenden	450	116	72	31	28	0.069	0.241	0.431
Cumberland	241	47	49	16	13	0.066	0.277	0.327
Daviess	14174	2031	2014	1165	792	0.082	0.390	0.578
Edmonson	509	89	120	58	51	0.114	0.573	0.483
Elliott	167	32	84	22	12	0.132	0.375	0.262
Estill	619	134	289	60	37	0.097	0.276	0.208
Fayette	50566	7844	3733	1282	1032	0.025	0.132	0.343
Fleming	838	138	128	48	31	0.057	0.225	0.375
Floyd	2607	619	777	298	231	0.114	0.373	0.384
Franklin	4837	699	528	245	150	0.051	0.215	0.464
Fulton	323	43	56	22	15	0.068	0.349	0.393
Gallatin	852	149	168	48	35	0.056	0.235	0.286
Garrard	1148	221	305	160	88	0.139	0.398	0.525
Grant	2650	341	296	164	113	0.062	0.331	0.554
Graves	3120	588	446	195	149	0.063	0.253	0.437
Grayson	2158	474	275	103	86	0.048	0.181	0.375

Green	587	104	102	43	35	0.073	0.337	0.422
Greenup	2261	388	126	62	57	0.027	0.147	0.492
Hancock	302	34	86	31	14	0.103	0.412	0.360
Hardin	10289	1693	1160	376	295	0.037	0.174	0.324
Harlan	1656	397	228	99	91	0.060	0.229	0.434
Harrison	1472	238	71	35	34	0.024	0.143	0.493
Hart	1580	307	347	111	57	0.070	0.186	0.320
Henderson	5307	894	736	373	267	0.070	0.299	0.507
Henry	1242	201	224	71	53	0.057	0.264	0.317
Hickman	150	32	15	4	3	0.027	0.094	0.267
Hopkins	4268	645	482	187	132	0.044	0.205	0.388
Jackson	380	77	126	15	10	0.039	0.130	0.119
Jefferson	100713	17766	10626	5508	4555	0.055	0.256	0.518
Jessamine	5735	930	801	423	311	0.074	0.334	0.528
Johnson	1718	332	303	134	100	0.078	0.301	0.442
Kenton	20284	2671	1530	855	661	0.042	0.247	0.559
Knott	691	179	138	62	53	0.090	0.296	0.449
Knox	2487	606	462	144	117	0.058	0.193	0.312
Larue	875	203	196	85	66	0.097	0.325	0.434
Laurel	6699	1360	1071	410	349	0.061	0.257	0.383
Lawrence	768	192	209	85	61	0.111	0.318	0.407
Lee	191	25	0	0	0	0.000	0.000	0.000
Leslie	71	21	176	9	4	0.127	0.190	0.051
Letcher	1200	300	317	111	89	0.093	0.297	0.350
Lewis	531	121	146	40	21	0.075	0.174	0.274
Lincoln	1374	257	360	96	66	0.070	0.257	0.267
Livingston	508	110	174	69	55	0.136	0.500	0.397
Logan	2069	404	282	112	87	0.054	0.215	0.397
Lyon	702	140	156	49	39	0.070	0.279	0.314
McCracken	9395	1170	1504	998	414	0.074	0.354	0.516
McCreary	680	181	325	112	65	0.147	0.359	0.422
McLean	1644	260	106	40	85	0.083	0.327	0.493
Madison	2704	566	1354	699	200	0.101	0.353	0.464
Magoffin	442	89	237	100	16	0.061	0.180	0.333
Marion	2016	252	278	137	55	0.029	0.218	0.527
Marshall	8504	1986	588	273	853	0.117	0.430	0.664
Martin	651	190	81	27	78	0.172	0.411	0.345
Mason	615	135	112	59	32	0.065	0.237	0.377
Meade	1292	433	292	172	162	0.133	0.374	0.589
Menifee	164	40	98	23	12	0.140	0.300	0.235
Mercer	1483	250	197	85	66	0.057	0.264	0.431
Metcalfe	780	156	129	72	57	0.092	0.365	0.558
Monroe	368	65	51	12	12	0.033	0.185	0.235
Montgomery	2506	488	388	179	163	0.071	0.334	0.461
Morgan	622	105	11	2	1	0.003	0.010	0.182
Muhlenberg	2661	529	426	197	163	0.074	0.308	0.462
Nelson	3358	582	465	218	175	0.065	0.301	0.469
Nicholas	406	60	83	37	25	0.091	0.417	0.446
Ohio	1969	449	425	166	132	0.084	0.294	0.391
Oldham	4010	642	437	185	159	0.046	0.248	0.423
Owen	539	120	126	57	42	0.106	0.350	0.452
Owsley	73	19	1	0	0	0.000	0.105	0.000
Pendleton	826	245	164	66	55	0.080	0.224	0.402
Perry	3275	775	1034	454	295	0.139	0.381	0.439
Pike	5153	1231	1043	483	327	0.094	0.266	0.463
Powell	828	164	218	94	61	0.114	0.372	0.431
Pulaski	6527	930	4	3	0	0.000	0.000	0.750
Robertson	87	15	11	1	2	0.011	0.133	0.091

Rockcastle	2490	402	185	58	44	0.023	0.109	0.314
Rowan	2797	341	3	0	0	0.000	0.000	0.000
Russell	1392	154	140	56	35	0.040	0.227	0.400
Scott	5853	997	996	401	291	0.069	0.292	0.403
Shelby	4914	816	625	284	222	0.058	0.272	0.454
Simpson	1832	356	354	130	106	0.071	0.298	0.367
Spencer	863	207	78	33	31	0.038	0.150	0.423
Taylor	2801	317	307	156	100	0.056	0.315	0.508
Todd	656	159	123	49	32	0.075	0.201	0.398
Trigg	969	205	173	61	47	0.063	0.229	0.353
Trimble	482	108	103	39	26	0.081	0.241	0.379
Union	976	183	150	58	41	0.059	0.224	0.387
Warren	15907	2606	1276	705	565	0.044	0.217	0.553
Washington	877	205	113	53	47	0.060	0.229	0.469
Wayne	1319	285	270	88	69	0.067	0.242	0.326
Webster	631	156	182	47	33	0.074	0.212	0.258
Whitley	3769	849	59	11	12	0.003	0.014	0.186
Wolfe	448	92	0	0	0	0.000	0.000	0.000
Woodford	2885	406	369	181	143	0.063	0.352	0.491

Table A2 Queries requested for EMS data

Dataset Grouping	Attribute	For Modeling	For Linking	Query	Attribute
DEMDataSet	dAgency				
	dConfiguration				
	dContract				
	dCustom Configuration				
	dCustomResults				
	dDevice				
	dFacility	x		Type of facility? Hospital/Urgent Care	• Hospital Designation • Facility Name • Facility Location Code
	dLocation	x	x		• EMS location type • EMS location GPS • US national Grid Coordinates
	dPersonnel				
	dState				
	dVehicle	x	x	Vehicle type? Ambulance	• Crew State Certification/Licensure Levels • Number of Each EMS Personnel Level on Normal 911 Ambulance Response
EMSDataSet	eAirway				
	eArrest	x			• Cardiac arrest
	eCrew				
	eCustomConfiguration				
	eCostumResults				
	eDevice				
	eDispatch	x	x	Complaint reported by dispatch? Traffic/Transportation incident	• Dispatch Priority (Patient Acuity)
	eDisposition	x	x		• Destination transferred to, name • Destination GPS location • Destination location US national grid coordinate • Incident/patient disposition • How patient was moved to ambulance • Position of patient during transport • How patient was transported from ambulance • Final patient acuity • Reason for choosing destination • Type of destination • Date/time pre-arrival or activation
	eExam	x			all
	eHistory	x			all
	eInjury	x	x		all
	eMedication	x			all
	eNarrative	x	x		PCR
	eOther				

	eOutcome	x	x		all
	ePatient	x	x		• Gender • Race • Age
	ePayment				
	eProcedures	x			• Date/Time Procedure Performed • Procedure Complication • Response to Procedure
	eProtocols	x			• Protocols Used
	eRecord				
	eResponse	x	x	Type of service requested? 911 response (scene)	all
	eScene	x	x		all
	eSituation	x	x		all
	eState				
	eTimes	x	x		all
	eVitals	x	x		all

Table A3 Queries requested for Trauma data

	Attribute	For Modeling	For Linking	Query	Attribute
Kentucky State Specific Data Collection Fields	Admit Service		x		TRAUMA BURN
	Cause Code		x		BIKE MC MV MVO MVU PEDESTRIAN
	Discharge Destination Code	x			
	ED Destination Code	x			
	Blood alcohol level	x			
	Injury details		x		
	Medication code	x			
	Medication location code	x			
	Medication Start Date				
	Outcome		x		
	Position in Vehicle		x		
	Referring Hospital Arrival Time		x		
	Toxicology / Drug Screen Results	x			
	Trauma Type	x			
	Transport destination		x		
	Transport Origin		x		
	AGE		x		
	RACE		x		
	ETHNICITY		x		
	SEX		x		
Injury information	INJURY INCIDENT DATE		x		
	INJURY INCIDENT TIME		x		
	ICD-9 PRIMARY E-CODE	x			
	ICD-10 PRIMARY E-CODE	x			
	ICD-9 LOCATION E-CODE	x			
	ICD-10 LOCATION E-CODE	x			
	ICD-9 ADDITIONAL E-CODE	x			
	ICD-10 ADDITIONAL E-CODE	x			
	INCIDENT LOCATION ZIP CODE		x		
	INCIDENT COUNTY		x		
	INCIDENT CITY		x		
	PROTECTIVE DEVICES	x			
	CHILD SPECIFIC RESTRAINT		x		
	AIRBAG DEPLOYMENT		x		
Pre-hospital information	EMS DISPATCH DATE	x			
	EMS DISPATCH TIME	x			
	EMS UNIT ARRIVAL DATE AT SCENE OR TRANSFERRING FACILITY	x			
	EMS UNIT ARRIVAL TIME AT SCENE OR TRANSFERRING FACILITY	x			
	EMS UNIT DEPARTURE DATE FROM SCENE OR TRANSFERRING FACILITY	x			
	TRANSPORT MODE		x		

	OTHER TRANSPORT MODE		x		
	INITIAL FIELD SYSTOLIC BLOOD PRESSURE		x		
	INITIAL FIELD PULSE RATE		x		
	INITIAL FIELD RESPIRATORY RATE		x		
	INITIAL FIELD OXYGEN SATURATION		x		
	INITIAL FIELD GCS - EYE		x		
	INITIAL FIELD GCS - VERBAL		x		
	INITIAL FIELD GCS - MOTOR		x		
	INITIAL FIELD GCS - TOTAL		x		
	INTER-FACILITY TRANSFER		x		
Emergency Department Information	ED/HOSPITAL ARRIVAL DATE		x		
	ED/HOSPITAL ARRIVAL TIME		x		
	INITIAL ED/HOSPITAL SYSTOLIC BLOOD PRESSURE	x			
	INITIAL ED/HOSPITAL PULSE RATE	x			
	INITIAL ED/HOSPITAL RESPIRATORY RATE	x			
	INITIAL ED/HOSPITAL RESPIRATORY ASSIST ANCE	x			
	INITIAL ED/HOSPITAL OXYGEN SATURATION	x			
	INITIAL ED/HOSPITAL SUPPLEMENTAL OXYGEN	x			
	INITIAL ED/HOSPITAL GCS - EYE	x			
	INITIAL ED/HOSPITAL GCS - VERBAL	x			
	INITIAL ED/HOSPITAL GCS - MOTOR	x			
	INITIAL ED/HOSPITAL GCS ASSESSMENT QUALIFIERS	x			
	INITIAL ED/HOSPITAL HEIGHT	x			
	INITIAL ED/HOSPITAL WEIGHT	x			
	ALCOHOL USE INDICATOR	x			
	DRUG USE INDICATOR	x			
	ED DISCHARGE DISPOSITION	x			
	SIGNS OF LIFE	x			
	ED DISCHARGE DATE	x			
	ED DISCHARGE TIME	x			
Hospital Procedure Information	ICD-9 HOSPITAL PROCEDURES	x			
	ICD-10 HOSPIT AL PROCEDURES	x			
	HOSPITAL PROCEDURE START DATE		x		
	HOSPITAL PROCEDURE START TIME		x		
	CO-MORBID CONDITIONS	x			

Diagnosis Information	ICD-9 INJURY DIAGNOSES	x			
	ICD-10 INJURY DIAGNOSES	x			
Injury Severity Information	AIS PREDOT CODE	x			
	AIS SEVERITY	x			
	ISS BODY REGION	x			
	AIS VERSION	x			
	LOCALLY CALCULATED ISS	x			
Outcome Information	TOTAL ICU LENGTH OF STAY	x			
	TOTAL VENTILATOR DAYS	x			
	HOSPITAL DISCHARGE DATE	x			
	HOSPITAL DISCHARGE TIME	x			
	HOSPITAL DISCHARGE DISPOSITION	x			
Financial Information	PRIM ARY METHOD OF PAYMENT	x			
Quality Assurance Information	HOSPITAL COMPLICATIONS	x			
Measures for Processes of Care	HIGHEST GCS TOTAL	x			
	GCS MOTOR COMPONENT OF HIGHEST GCS TOTAL	x			
	GCS ASSESSMENT QUALIFIER COMPONENT OF HIGHEST GCS TOTAL	x			
	CEREBRAL MONITOR	x			
	CEREBRAL MONITOR DATE	x			
	CEREBRAL MONITOR TIME	x			
	VENOUS THROMBOEMBOLISM PROPHYLAXIS TYPE	x			
	VENOUS THROMBOEMBOLISM PROPHYLAXIS DATE	x			
	VENOUS THROMBOEMBOLISM PROPHYLAXIS TIME	x			
	LOWEST ED/HOSPITAL SYSTOLIC BLOOD PRESSURE	x			
	TRANSFUSION BLOOD (4 HOURS)	x			
	TRANSFUSION PLASM A (4 HOURS)	x			
	TRANSFUSION PLATELETS (4 HOURS)	x			
	CRYOPRECIPITATE	x			
	TRANSFUSION BLOOD (24 HOURS)	x			
	TRANSFUSION PLASM A (24 HOURS)	x			
	TRANSFUSION PLATELETS (24 HOURS)	x			
	CRYOPRECIPITATE (24 HOURS)	x			
	ANGIOGRAPHY	x			
	EMBOLIZATION SITE	x			
	SURGERY FOR HEMORRHAGE CONTROL TYPE	x			

	WITHDRAWAL OF CARE	x			