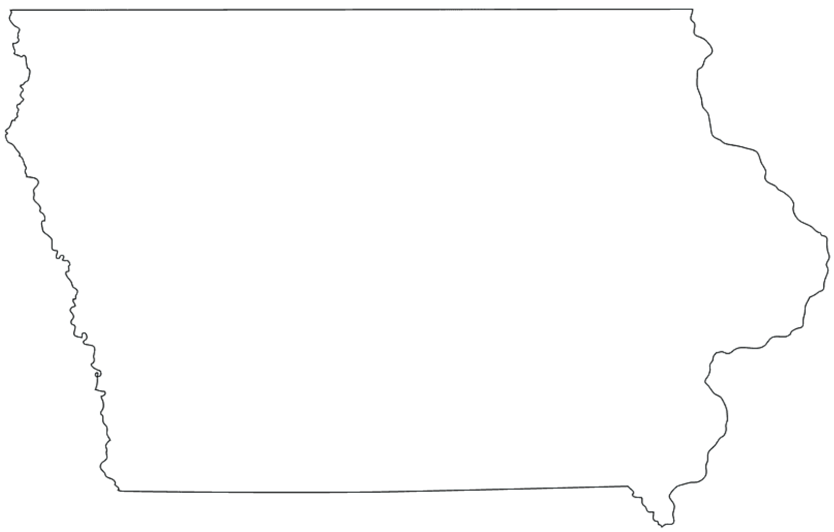
CODES

Iowa Crash Outcomes Data Evaluation System



2023
Year-End Report

Prepared by the
University of Iowa Injury Prevention Research Center
Under contract to Iowa Governor's Traffic Safety Bureau and Iowa Department of Transportation

IOWA

Injury Prevention
Research Center

GTSB
GOVERNOR'S TRAFFIC SAFETY BUREAU
Reducing Death and Injuries On Iowa's Streets and Highways
IOWA DEPARTMENT OF PUBLIC SAFETY

EXECUTIVE SUMMARY

The goal of the Iowa CODES project is to examine outcomes related to motor vehicle crash-related injuries in the State of Iowa. We performed probabilistic linkage to join the Iowa police crash report data with Iowa Emergency Department data, Iowa Hospital Inpatient data, and FARS fatal motor vehicle crashes from 2016 through 2020 to get a single and comprehensive dataset for analysis. This report gives an overview of data preparation, data quality prior to linkage, the data linkage, assessment of data linkage quality, calculation of injury severity scores, and a descriptive look at some major themes identified in the linked dataset.

About 86% of the crash records for injured persons included data for three of the variables used to link the crash data with the Emergency Department and Inpatient hospital data: month of birth, sex, and zip code. The addition of the injured person's zip code as a linkage variable was very advantageous for discriminating individuals in the crash data. Without zip code, the combination of crash month, month of birth, and sex was unique for only 39.7% of the records. With zip code, the combination of the four linkage variables was unique for 98.3% of the crash records. For Emergency Department and Inpatient records, the combination of the four linkage variables identified a unique individual for 92.2% and 95.7% of the records, respectively. This finding indicates that it should be possible to achieve a high-quality linkage using this combination of variables.

Altogether, 44% of the injured persons in the crash records were linked to a record from HCUP. Most (88%) of the linked records linked only to an Emergency Department record, 8% linked only to Inpatient record, and 4% linked to both Emergency Department and Inpatient records.

Crash records that were missing month of birth were much less likely to be linked relative to other records missing sex or zip code. Even when all linkage variables were known, less than half (48%) were linked to HCUP records. Considering only the injured persons who were reported as transported by EMS from the crash scene, about 70% were linked. Among this set of linked records, hospital information entered in the crash report corresponded with the hospital identifier in the HCUP data 87% of the time, which indicates the linkage produced high quality matches. However, there is additional room for improving the number of linked records, particularly with regard to the 30% of injured persons transported by EMS who were not linked to HCUP records.

Last but not least, while preparing, linking and analyzing the data, we discovered several significant and previously unknown crash data quality issues, including missing personal data for non-motorists across all the years of crash data and errors in the dataset containing non-motorist crash data elements. These incomplete records had a negative impact on the ability to assess the crash-related outcomes for non-motorists.

PROJECT DESCRIPTION

The goal of the Iowa CODES project is to examine outcomes (injuries, long-term disability, hospital charges, discharge status) related to motor vehicle crash-related injuries in the State of Iowa. This is accomplished by probabilistically linking identified person (patient)-level crash, hospital, and mortality data.

The hospital data sources are the Iowa State Inpatient Database (SID) and Iowa State Emergency Department Database (SEDD), which comes from AHRQ's Healthcare Cost and Utilization Project (HCUP). Crash data are obtained from the Iowa Department of Transportation in the form of "Z tables" generated by the Bureau of Traffic & Safety. Mortality data are obtained from the NHTSA's Fatality Analysis Reporting System (FARS).

CURRENT STATUS

We performed probabilistic linkage to join the Iowa police crash report data with Iowa Emergency Department data, Iowa Hospital Inpatient data, and FARS fatal motor vehicle crashes from 2016 through 2020 to get a single and comprehensive dataset for analysis. In this performance period, we improved the data linkage quality by adding zip code as a linkage variable. We acquired the Emergency Department data and the Hospital Inpatient data for 2021 from HCUP and intended to expand the analysis dataset to include this additional year of data. However, the crash data we received from Iowa DOT (Traffic & Safety Bureau) for 2021 and 2022 does not include dates of birth (DOBs) and we have been informed that they will not provide them. We are exploring a process for obtaining the DOBs from another Iowa DOT division, but the work-around is labor intensive for both our team and personnel in the other DOT division. In addition to omitting DOBs, as well as other data fields that are useful for our purposes, the formatting of the crash data files was changed, leading to incompatibility issues with our software programs. These issues were reported to Iowa DOT (T&S) during the summer and have not yet been resolved.

The following sections of the report give an overview of data preparation, crash data quality issues discovered this performance period, data quality prior to linkage, the data linkage, assessment of data linkage quality, calculation of injury severity scores, and a descriptive look at some major themes identified in the linked dataset.

DATA PREPARATION

In the previous performance period (FY22), a substantial amount of data correction had to be completed in order to use the crash data for linkage and for analyses. The primary issue was that the unit numbers (UNITNUMs) listed in the person-level data were often inconsistent with the unit numbers for the vehicles individuals were operating or riding in. This issue was partially corrected by matching the injured persons' dates of birth with the drivers' dates of birth within each crash. Injured persons who were not drivers were assigned to a vehicle based on the number of occupants recorded for each vehicle in the crash. When a crash involved more than one vehicle carrying more than one occupant, the passengers could not be assigned to a vehicle.

During this performance period (FY24), we continued to inspect and find issues with the data as we attempted to add an additional year of data (2021), worked to improve the quality of the linkage to the HCUP data, and conducted analyses.

Numerous updates and additions to the data correction script were made, including:

- Added 5-digit zip code to person-level data prior to cleaning the data. Only Dr. Hamann and Ms. Reyes, as the UI administrators of memorandum of understanding (MOU) with the Iowa DOT to use the crash data, had access to the names and addresses of the injured persons. Ms. Reyes converted the injured persons' 9-digit zip codes to 5 digits and extracted a dataset containing only the PERSONKEY and the 5-digit zip code, which was then joined to the injured person data (Z tables zinj and zuni).
- Renamed the ALCTEST, ALCRESULT, DRUGTEST, and DRUGRESULT data elements for non-motorists to NM_ALCTEST, NM_ALCRESULT, NM_DRUGTEST, and NM_DRUGRESULT to distinguish them from the corresponding data elements for drivers. We also requested that the Iowa DOT make the same change in the znmt table (i.e., the Z table containing data elements for non-motorists).
- Reordered code to clean the non-motorist data before it was merged with other person-level tables.
- Made updates to the code for correcting driver DOBs to identify and remove driver records that were duplicates.
- Filled in missing injured person data elements with null values prior to dividing the data into driver, passenger, and non-motorist datasets.
- Revised the method for matching unknown drivers to unknown injured persons within a given crash.
- Added data counts to ensure that all vehicles and drivers were accounted for in the final datasets.
- Added person-level records to data set to account for drivers who were not included in zinj or zuni (Z tables for injured and uninjured persons, respectively). This included devising a way of generating a unique PERSONKEY for the omitted individual.
- Wrote code to identify and flag crashes where a driver or an injured person was associated with an unoccupied vehicle.
- Updated the assignment of injured persons to vehicles to account for multi-vehicle crashes where only one unit was occupied.
- Wrote code to identify and correct data when a non-motorist was also listed as a driver (e.g., a bicycle was entered into the crash report as both a unit and as a non-motorist).

These updates were completed in June 2023.

CRASH DATA QUALITY ISSUES DISCOVERED

While preparing the data, several previously unknown data issues were discovered. All these issues are suspected but not confirmed to be issues with the way the Z tables are created using data from the Iowa DOT's Accident Processing System (APS).

1. Upon integrating the injured person zip code, we discovered that zip code was missing for nearly all the non-motorists across all 5 years of crash data. Since zip code is a linkage variable, this could have a significant impact on our ability to link non-motorists to the HCUP data and access the injury outcomes for this population.
2. For the 2017 data, there were 463 PERSONKEYs with zip code that did not match to a person in zinj/zuni. This was the only year of data with this pattern.

3. The znmt file (Z table containing non-motorist data elements) for 2020, contained no more than one non-motorist per crash. Upon further examination, we believe approximately 16 non-motorists across 12 different crashes were omitted from the znmt file.
4. Drivers and injured persons being recorded for units where 0 was entered for the number of occupants.
5. For crash data received from 2021 and 2022, the following fields were omitted: DRIVERDOB, INJUREDDOB, TRANSTO, TRANSBY, NARRATIVE.
6. For crash data received from 2021 and 2022, database files (.dbf) could not be imported into SAS (Statistical Analysis Software) due to format changes.

QUALITY OF DATASETS USED FOR LINKAGE

COMPLETENESS OF LINKAGE VARIABLES IN CRASH DATA

The level of completeness for the linkage variables was compiled for the person-level crash data (after it had been partially corrected by UI team). When a person-level record is missing data for one or more of the linkage variables, it could significantly impact the likelihood of linking that record to one in the other dataset, especially if only a few linkage variables are being used, which is the case here. Additionally, completeness is one of six performance attributes for traffic records quality. Tables 1, 2, and 3 detail the completeness of three data linkage variables, date of birth (subsequently converted to month of birth, MMYYYY, for compatibility with HCUP data), sex, and zip-code, respectively, by the injury status recorded in the crash report. The fourth data linkage variable, date of the crash, was known for all crash records (subsequently converted to month of crash and year of the crash for compatibility with the HCUP data).

Month of birth was known for 93% of injured persons. Within the fatal, suspected serious, suspected minor, and possible categories of injury status, month of birth was known for 98.8-99.6% of the individuals. Within the unknown injury category, month of birth was unknown for 72.8% of the individuals. Overall, sex was known for 94.3% of injured persons. Zip code was recorded for 88.0% of injured persons. Surprisingly, the proportion of injured persons with a known zip code increased as injury level decreased, going from 88.7% for fatal injury up to 95.1% for possible injury.

Table 1. Number (%) of records for injured persons in the crash data with known and unknown month of birth, by injury status.

Injury status	Known month of birth	Unknown month of birth	Total
Fatal	1,720 (99.6%)	*	*
Suspected serious/ incapacitating	6,897 (99.1%)	*	*
Suspected minor/ non-incapacitating	30,964 (98.8%)	363 (1.6%)	31,327
Possible	51,532 (98.9%)	563 (1.1%)	52,095
Unknown	2,252 (27.2%)	6,016 (72.8%)	8,268
Total	93,365 (93.0%)	7,013 (7.0%)	100,378

Table 2. Number (%) of records for injured persons in the crash data with known and unknown sex, by injury status.

Injury status	Known sex	Unknown sex	Total
Fatal	1,710 (99.0%)	17 (1.0%)	1,727
Suspected serious/ incapacitating	6,838 (98.2%)	123 (1.8%)	6,961
Suspected minor/ non-incapacitating	30,803 (98.3%)	524 (1.7%)	31,327
Possible	51,412 (98.7%)	683 (1.3%)	52,095
Unknown	3,892 (47.1%)	4,376 (52.9%)	8,268
Total	94,655 (94.3%)	5,723 (5.7%)	100,378

Table 3. Number (%) of records for injured persons in the crash data with known and unknown zip code, by injury status.

Injury status	Known sex	Unknown sex	Total
Fatal	1,532 (88.7%)	195 (11.3%)	1,727
Suspected serious/ incapacitating	6,268 (90.0%)	693 (10.0%)	6,961
Suspected minor/ non-incapacitating	28,820 (92.0%)	2,507 (8.0%)	31,327
Possible	49,522 (95.1%)	2,573 (4.9%)	52,095
Unknown	2,232 (27.0%)	6,036 (73.0%)	8,268
Total	88,374 (88.0%)	12,004 (12.0%)	100,378

Completeness across all three linkage variables was also compiled, as shown in Table 4. Across all injured person records, 86.2% included month of birth, sex, and zip code. Completeness varied by the person's role in the crash. For drivers and passengers, 93.7% and 88.0% of records, respectively, included all three of these linkage variables. However, as mentioned previously in this report, none of the non-motorist records included zip-code.

Table 4. Number (%) of records for injured persons in the crash data with known ("Yes") and unknown ("No") month of birth, sex, and zip code, by role in the crash.

Month of birth	Sex	Zip code	All injured persons	Drivers	Passengers	Non-motorists
Yes	Yes	Yes	* (86.2%)	70,432 (93.7%)	15,429 (88.8%)	*
Yes	Yes	No	5,645 (5.6%)	731 (1.0%)	1,045 (6.0%)	3,751 (96.7%)
Yes	No	Yes	*(1.1%)	981 (1.3%)	91 (0.5%)	*
Yes	No	No	84 (0.1%)	29 (0.0%)	31 (0.2%)	13 (0.3%)
No	Yes	Yes	613 (0.6%)	25 (0.0%)	535 (3.1%)	*
No	Yes	No	*(1.8%)	769 (1.0%)	111 (0.6%)	32 (0.8%)
No	No	Yes	125 (0.1%)	34 (0.0%)	80 (0.5%)	*
No	No	No	4,437 (4.4%)	2,199 (2.9%)	62 (0.4%)	85 (2.2%)
Total			100,378	75,200	17,384	3,881

QUALITY OF E-CODES IN HOSPITAL DATA

External cause-of-injury codes (a.k.a. "e-codes") within the International Classification of Diseases, Tenth Revision, Clinical Modification (i.e., ICD-10-CM) classify injury events by mechanism and intent of injury. Intent of injury categories include unintentional, homicide/assault, suicide/intentional self-harm, legal intervention or war operations, and undetermined intent.

The UI team investigated the completeness of the e-code data in the HCUP data to assess whether it would be appropriate to restrict the HCUP data to include only E-codes associated with "transport accidents." We contacted the Iowa Hospital Association who reported that the External Cause of Injury codes are required to be entered into the Inpatient and Emergency Department datasets for the State of Iowa if they are available. The Council of State and Territorial Epidemiologists advise that the data quality of e-codes can be evaluated by calculating the percent of injury records for which an e-code is entered and comparing the proportion of records over time.

Both in the Emergency Department data and Inpatient hospitalization data, the percentage of injury records that have an e-code was stable from 2017 to 2021, around 98%, which indicates good quality of e-codes (see Figure 1 and Figure 2, respectively). In 2016, the proportion of records with e-codes is 92%, lower than other years. Upon further investigation, we learned that in 2016 the e-codes variables were represented in a separate field (I10_ECAUSE) and beginning in 2017, e-codes were included in the diagnosis fields (I10_DX).

Both in Emergency Department data and Inpatient hospitalization data, the percentage of e-codes that are associated with transport accidents was stable from 2016 to 2021, around 11-12%, which also indicates good data quality.

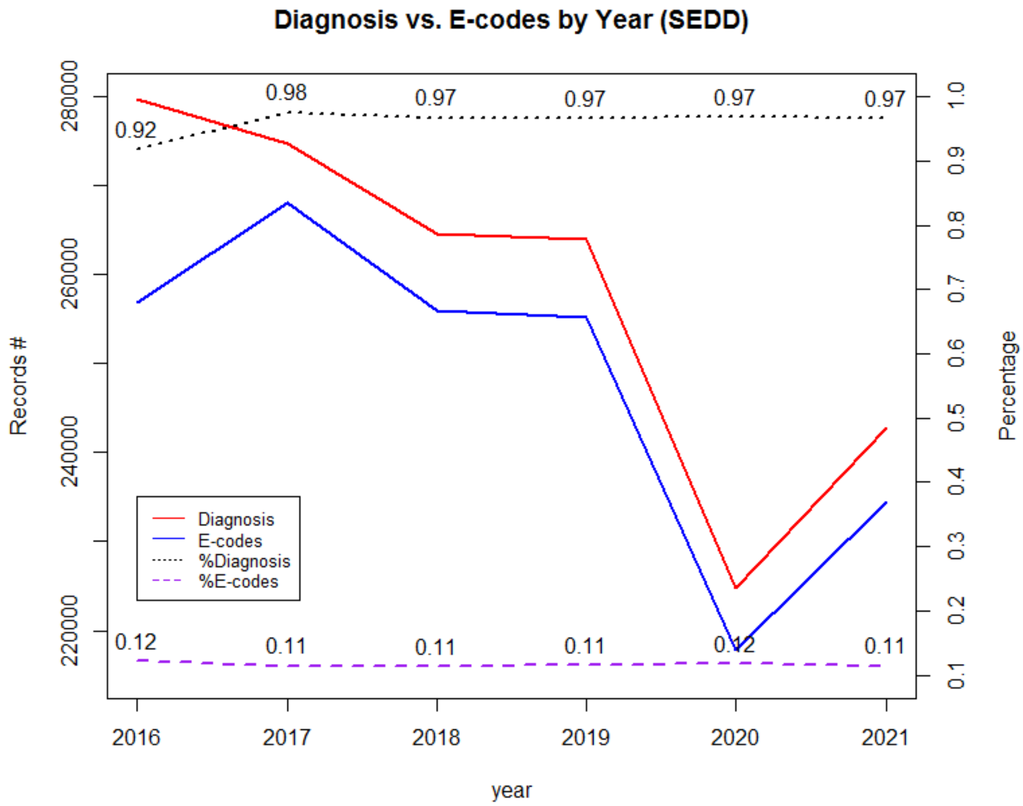


Figure 1. The number of records with an injury diagnosis code (red line) and with an external cause of injury code (e-code; blue line), the proportion of records with an injury diagnosis code that also had an e-code (black dotted line), and the proportion of e-codes associated with transport accidents (dashed purple line) in the Iowa Statewide Emergency Department Data from HCUP.

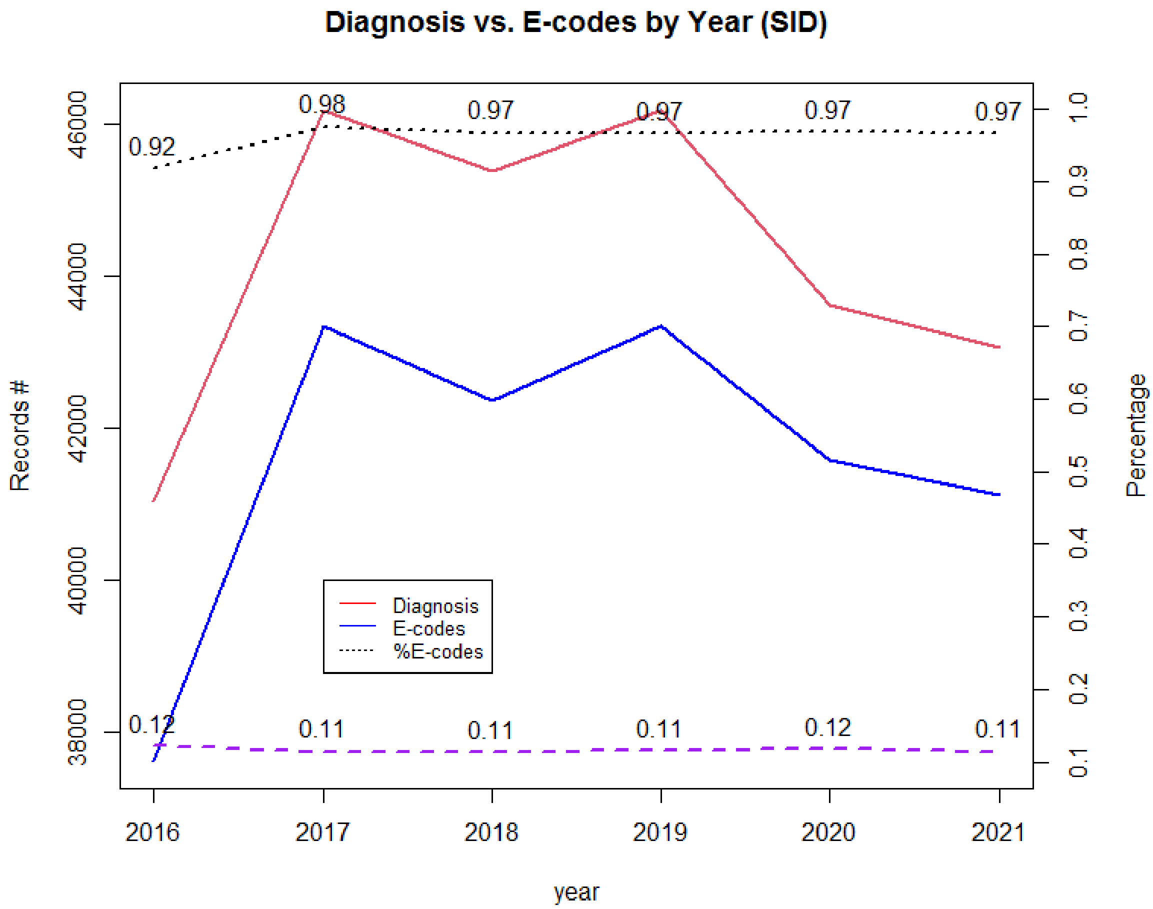


Figure 2. The number of records with an injury diagnosis code (red line) and with an external cause of injury code (e-code; blue line), the proportion of records with an injury diagnosis code that also had an e-code (black dotted line), and the proportion of e-codes associated with transport accidents (dashed purple line) in the Iowa Statewide Inpatient Data from HCUP.

UNIQUENESS OF LINKAGE VARIABLES

In the initial linkage of data for crash-injured persons to the HCUP data, the linkage variables were the injured person's month of birth, their gender, and the month of the crash. For the 93,366 crash records that included the injured person's month of birth, the combination of these three linkage variables identified a unique individual for 37,061 (39.7%) the records.

In the subsequent linkage of data, the injured person's zip code was added as a linkage variable. For the crash records that included the injured person's date of birth, the combination of the four linkage variables (i.e., month of birth, gender, month of crash, and zip code) identified a unique individual for 91,737 (98.3%) of the records.

For the 145,431 hospital Emergency Department records with external cause of injury codes associated with transport accidents, the combination of the four linkage variables identified a unique individual for 134,120 (92.2%) of the records.

For the 14,181 hospital Inpatient records with external cause of injury codes associated with transport accidents, the combination of the four linkage variables identified a unique individual for 12,149 (95.7%) of the records.

SUMMARY

About 86% of the crash records for injured persons included month of birth, sex, and zip code. The addition of the injured person's zip code as a linkage variable was very advantageous for discriminating individuals in both the crash and HCUP datasets, boosting the proportion of unique records from 39.7% without zip code to 98.3% with it. This finding indicates that it should be possible to achieve a high-quality linkage using this combination of variables. However, about 12% of the crash records did not include the injured person's zip code.

DATA LINKAGE AND DATA QUALITY CHECKS

Probabilistic linkage was performed to link the crash records for injured persons with the HCUP data that were identified as transport accidents. The software LinkSolv from Strategic Matching, a probabilistic record linkage software, was used to do the probabilistic linkage. The primary linkage variables were the month and year of the crash, the month and year of the injured person's birth, their sex, and their zip code.

Altogether, 44% of the injured persons in the crash records were linked to a record from HCUP. Most (88%) of the linked records linked only to an Emergency Department record, 8% linked only to Inpatient record, and 4% linked to both Emergency Department and Inpatient records.

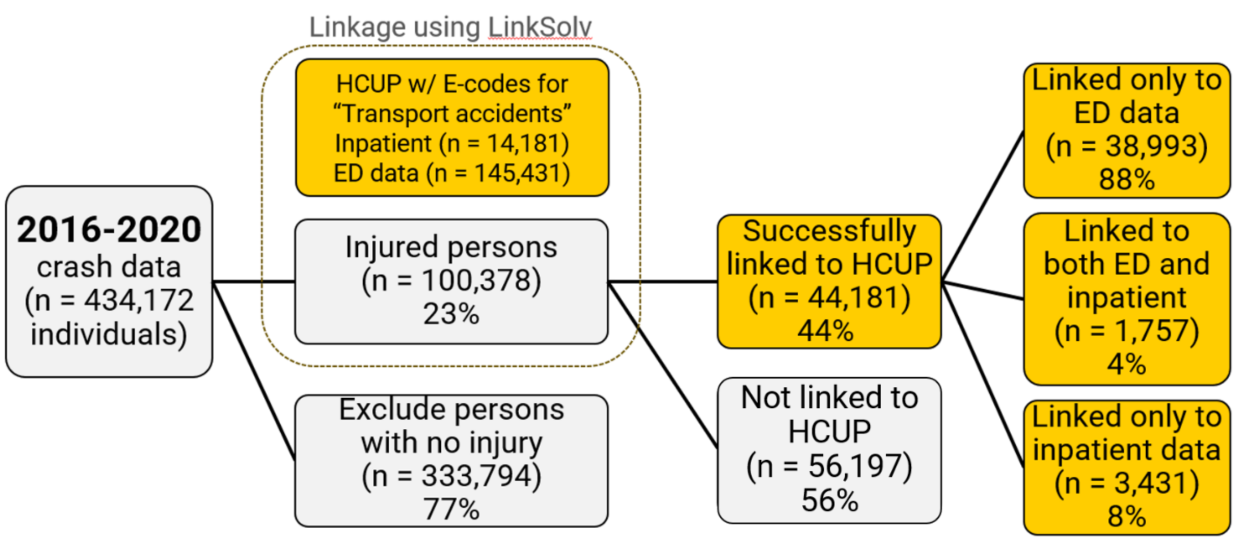


Figure 3. Diagram showing linkage of crash data to Iowa Inpatient and Emergency Department hospital data obtained from HCUP.

EFFECT OF MISSINGNESS ON LINKAGE QUALITY

The linked data were examined to evaluate the effect of data completeness on the linkage quality. Table 5 displays the missingness of linking variables among all crash records, the number crash records that were successfully linked to HCUP records despite missing data, and the number of crash records missing data that were not linked to HCUP. The distribution of missingness was very similar within each year, so only the collective missingness is shown in the table. A record that was missing month of birth for the injured person was very unlikely to be linked, with only 3.5% of those records being linked to an HCUP record. Only 8.1% of the records missing sex were linked. Just under 17% of the records missing zip code for the injured person were linked.

Table 5. Records in crash data missing linking variables, 2016-2020 and results of linkage with Iowa Emergency Department and Inpatient hospital data from HCUP.

	All records n (%) of all injured person crash records missing linkage variable	Linked records n (%) of records missing linkage variable that were linked	Unlinked records n (%) of records missing linkage variable that were not linked
Month of birth	7,013 (7.0%)	246 (3.5%)	6,767 (96.5%)
Sex	5,723 (5.7%)	464 (8.1%)	5,259 (91.9%)
Zip code	12,004 (12.0%)	2,011 (16.8%)	9,993 (83.2%)

Missingness across combinations of linkage variables was also examined (see Table 6). None of the records missing both month of birth and zip code were linked. About a third of the records missing month of birth with known zip code were linked, but such records represented a very small proportion (<1%) of all records.

Table 6. Number (%) of records for injured persons in the crash data with known ("Yes") and unknown ("No") month of birth, sex, and zip code that were and were not linked to HCUP records.

Month of birth	Sex	Zip code	All injured persons	Linked to HCUP	Not linked to HCUP
Yes	Yes	Yes	86,559 (86.2%)	41,539 (48.0%)	45,020 (52.0%)
Yes	Yes	No	5,645 (5.6%)	1,974 (35.0%)	3,671 (65.0%)
Yes	No	Yes	1,077 (1.1%)	385 (35.7%)	692 (64.3%)
Yes	No	No	84 (0.1%)	37 (44.0%)	47 (56.0%)
No	Yes	Yes	613 (0.6%)	204 (33.3%)	409 (66.7%)
No	Yes	No	1,838 (1.8%)	0 (0%)	1,838 (100%)
No	No	Yes	125 (0.1%)	42 (33.6%)	83 (66.4%)
No	No	No	4,437 (4.4%)	0 (0%)	4,437 (100%)
Totals			100,378	44,181 (44.0%)	56,197(56.0%)

Finally, missingness across combinations of linkage variables was also examined for the records associated with injured persons who were recorded as transported by air or ground EMS. As shown in Table 7, the records for about 90% of these individuals included month of birth, sex, and zip code. However, about 27% of these records were not linked to an individual in the HCUP data, despite being transferred from the crash scene by EMS. There were 2,499 records where the injured person's month of birth and sex were known but not their zip code. Only 39.4% of these records were linked, compared with a linkage rate of 73.1% when all three variables were known. Overall, 30.3% of injured persons transported by EMS were not linked to records in the HCUP data.

Table 7. Number (%) of records for injured persons in the crash data reported as transported by air or ground EMS with known ("Yes") and unknown ("No") month of birth, sex, and zip code that were and were not linked to HCUP records.

Month of birth	Sex	Zip code	All EMS transported persons	Linked to HCUP	Not linked to HCUP
Yes	Yes	Yes	30,786 (89.9%)	22,499 (73.1%)	8,287 (26.9%)
Yes	Yes	No	2,499 (7.3%)	985 (39.4%)	1,514 (60.6%)
Yes	No	Yes	448 (1.3%)	249 (55.6%)	199 (44.4%)
Yes	No	No	31 (0.1%)	19 (61.3%)	12 (38.7%)
No	Yes	Yes	299 (0.9%)	112 (37.5%)	187 (62.5%)
No	Yes	No	83 (0.2%)	0 (0%)	83 (100%)
No	No	Yes	52 (0.2%)	17 (32.7%)	35 (67.3%)
No	No	No	54 (0.2%)	0 (0%)	54 (100%)
Totals			34,252	23,881 (69.7%)	10,371 (30.3%)

CHECKING LINKAGE QUALITY USING HOSPITAL IDENTIFIER

When completing a crash report, the investigating officer can enter the location to which an injured person was transported (TRANSTO) and what agency or person transported them (TRANSBY). This information was used to assess the quality of the CODES data linkage.

The TRANSTO and TRANSBY fields are open text fields in the crash report. A SAS script was created to extensively clean the data entered by the officers into these fields and associate them with the lists of hospitals providing care in the State of Iowa each year from 2016-2020. Specifically, the script:

- Corrected observed misspellings (e.g., "CEDR RAPIDS" rather than "CEDAR RAPIDS")
- Applied rules for discriminating which location was applicable for hospital systems with multiple locations (e.g., MercyOne, UnityPoint)
- Applied rules for discriminating which location was applicable when the same name could refer to multiple hospitals (e.g., "KEOKUK" could refer to Keokuk County Hospital in Sigourney or to Blessing Health in the city of Keokuk; "GRMC" could refer to Grinnell Regional Medical Center, Great River Medical Center, or Greater Regional Medical Center)

- Applied rules to consolidate the many various names that could be used to refer to a unique hospital due to health system merger and name changes (e.g., Central Iowa Healthcare became UnityPoint Health – Marshalltown) as well as abbreviations (e.g., "UI", "UIHC", "UOFI", "UOI", "U OF I", "THE U", or entries containing both "UNIV" and "IOWA" all refer to the University of Iowa Hospitals and Clinics)
- Identified when the named hospital to which the injured person was transferred was located outside the State of Iowa
- Identified when TRANSTO may indicate transfer from first hospital to a higher level of care (e.g., "CLINTON MERCY ER / U OF I HOSP"). Note that officers may not always be aware of transfers, nor are they required to enter both hospitals.
- Identified and consolidated entries into other categories of transport locations other than specific hospitals (e.g., jail, home, medical examiner)

Next the cleaned TRANSTO data from 2016-2019 (2020 was not included because the hospital identifiers were not yet available from HCUP at the time) were joined with the CODES data (i.e., data output by the linkage procedure including zip code as a linkage variable). Table 8 summarizes the findings. Crash records that indicated the injured person was transferred to a single known in-state hospital were linked to hospital records about 70% of the time.

Table 8. Number of individual crash records with data entered for TRANSTO that were linked to Iowa Emergency Department and Inpatient hospital data (HCUP) by entry type.

Description of TRANSTO category	Number of injured person crash records	Linked to HCUP	Not linked to HCUP
Total records with entry in TRANSTO after cleaning	47,670	29,562 (62.0%)	18,108 (38.0%)
Records with clean in-state hospital name	39,064	27,423 (70.2%)	11,641 (29.8%)
OOS – out of state hospital	728	70 (9.6%)	199 (90.4%)
MULTI – entry suggests care at multiple hospitals	180	117 (65.0%)	63 (35.0%)
EMS – TRANSTO indicates name of EMS agency rather than hospital	*	23 (71.9%)	*
Funeral home, medical examiner or morgue	420	14 (3.3%)	406 (96.7%)
Unspecified hospital	*	227 (68.2%)	*
Not transported/treated at scene/refused	4,916	1,054 (21.4%)	3,862 (78.6%)
Other/unclassified	1,242	465 (37.4%)	777 (62.6%)

The sample of data was further restricted to records where the officer indicated that the injured person had been transported by air or ground EMS and the record had been linked to HCUP. There were 21,516 injured person records meeting these inclusion criteria. Next the hospital identifier from the HCUP data was compared to the cleaned TRANSTO entry from the crash report. For 18,728 records (87.0%), the name of the hospital in the crash report corresponded with the hospital ID in the linked HCUP record. For 1,969 records (9.2%), the hospital name entered on the crash report did not correspond to the hospital ID in the linked HCUP record. For 819 records (3.2%), the information entered by the officer into the TRANSTO field did not indicate the name of a known hospital.

SUMMARY

Only 44% of the crash records for injured persons were linked to HCUP records. Records that were missing month of birth were much less likely to be linked relative to other records missing sex or zip code. Even when all linkage variables were known, less than half (48%) were linked to HCUP records. Considering only the injured persons who were reported as transported by EMS from the crash scene, about 70% were linked. Among this set of linked records, hospital information entered into the crash report corresponded with the hospital identifier in the HCUP data 87% of the time, which indicates the linkage produced high quality matches. However, there is additional room for improvement, and we intend to pursue several of strategies for this in the next performance period.

CALCULATION OF INJURY SEVERITY SCORES

During this performance period, the research team acquired AIS ICD ISS Map from the Association for the Advancement of Automotive Medicine. The ICD map is a set of tables that help transform ICD-9CM or ICD-10 CM injury diagnosis codes into single patient severity scores. We used the ICD Map to calculate Injury Severity Score (ISS), New ISS, and Maximum Abbreviated Injury Scale (MAIS) scores, which are commonly used in clinical settings to understand injury severity. Injury scores were mapped from ICD-10-CM codes for 2,949 (86.0%) of the individuals linked only to the Inpatient hospital data, for 30,193 (77.4%) of the individuals linked to the Emergency Department data, and 1,433 (81.6%) of the individuals linked to both HCUP datasets. The distribution of these scores for the 2016-2020 linked dataset are included in Table 9.

Table 9. Distribution of injury severity scores in SEDD and SID, 2016-2020

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NAs
SEDD							
ISS	0	1	1	2.2	2	75	8,919
NISS	0	1	2	2.8	3	51	8,919
MAIS	0	1	1	1.2	1	6	8,919
SID							
ISS	0	5	9	10.4	13	75	709
NISS	0	6	12	13.1	17	86	709
MAIS	0	2	2	2.5	3	6	709

The ISSs calculated from Emergency Department were compared to the injury status entered into the crash report (see Figure 4). Generally, officers' assessments of minor injury in the crash reports tended to be accurate relative to the ISSs. The indication of "possible" injury typically corresponded to minor injuries in the ED data. Though suspected serious/incapacitating injuries in the crash report corresponded to higher injury scores (relative to minor or possible injury), most often in these cases the ISS score from the ED data indicated overall minor injuries.

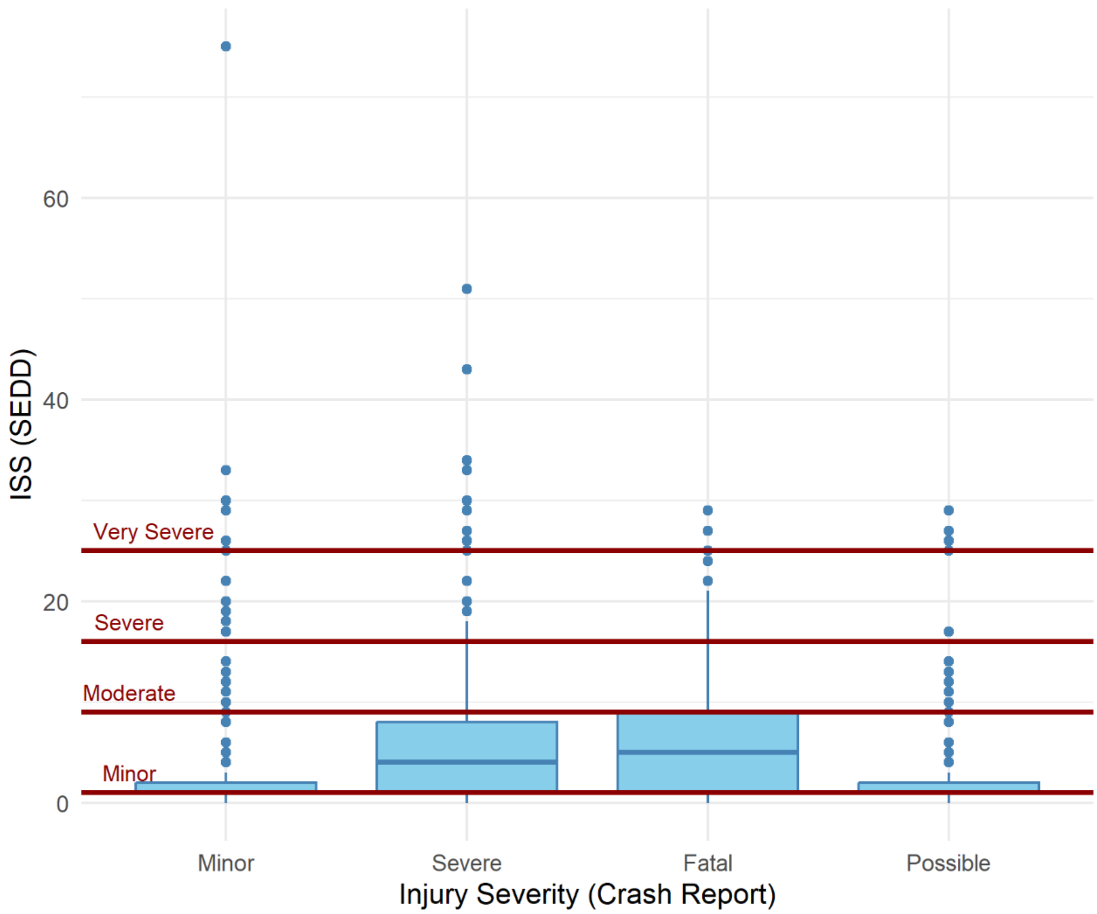


Figure 4. Boxplot of Injury Severity Scores (ISS) in the HCUP Emergency Department records compared to injury severity entered in the crash report.

DESCRIPTIVE ANALYSIS

The following sections provide results from descriptive analyses of the 2016-2020 CODES dataset including characteristics of the linked records (Table 10) and several safety emphasis areas (alcohol involvement, older drivers, occupant protection, and teen drivers).

The median age of the linked dataset was 36, with a range of 0-99 (Table 10). The dataset had slightly more males (51.4%) compared to females (47.6%) and included 1.1% unknown gender. Nearly three-quarters of the linked persons where white (74.0%), followed by Black (7.9%), and Hispanic (5.2%). Asian or Pacific Islander and Native American racial groups made up less than 2% of records combined. However, 10.4% had unknown race. Most linked persons were drivers (76.7%) or passengers (20.2%) in the crash and 3% were non-motorists.

Nearly half (48.2%) of the linked persons were indicated as 'possible injury' in the crash report, followed by minor injury (39.6%), and severe injury (10.7%). One percent of the linked persons were recorded as having died in the crash report.

Table 10. Descriptive characteristics of linked records

	N = 44,181	
	N (Median)	Percent (Range)
Age	(36)	(0-99)
Gender		
Female	22,701	51.4%
Male	21,016	47.6%
Unknown	464	1.1%
Race		
Asian or Pacific Islander	624	1.4%
Black	3,512	7.9%
Hispanic	2,316	5.2%
Native American	123	0.3%
Other	315	0.7%
Unknown	4,611	10.4%
White	32,680	74.0%
Role in Crash		
Driver	33,884	76.7%
Non-Motorist	1,325	3.0%
Passenger	8,944	20.2%
Unknown	28	0.1%
Injury status (Crash report)		
Died	438	1.0%
Severe Injury	4,728	10.7%
Minor Injury	17,501	39.6%
Possible Injury	21,307	48.2%
Unknown	207	0.5%

SAFETY EMPHASIS AREAS

This section includes a number of different descriptive tables, based on several State Highway Safety Plan (SHSP) Safety Emphasis Areas, which report hospital costs, length of stay, and discharge disposition. We created descriptive tables for these characteristics and hospital costs stratified by alcohol involvement, older driver age groups, and vehicle occupant protection.

Each selected theme is represented by two tables. The first table displays the total hospital costs and average cost per person recorded in the HCUP data for linked persons within the 2016-2020 time period. These are further broken down by whether the costs are incurred through Emergency Department (SEDD) and Inpatient hospital (SID) visits. The second table contains summary information on other variables of interest: the record type (SID, SEDD, or both), length of stay (LOS), and discharge disposition. Length of stay is categorized and capped at three days due to censoring of the data. Summary statistics for length of stay and discharge disposition are calculated separately for SID and SEDD records.

ALCOHOL-IMPAIRED DRIVING

Alcohol involvement is defined for a crash when one of the drivers has either a measured BAC over the legal limit for Iowa, or, if BAC information was not collected, was marked by law enforcement as "Under the influence of alcohol." Crashes included in the "Not alcohol-involved category" either had BAC below the statutory limit or the influence of alcohol was not detected, and thus are not guaranteed to be completely alcohol-free.

Overall, 2,231 persons injured in crashes involving alcohol were linked to the HCUP data. In the Emergency Department (SEDD) records, the average cost per person is $3,044 higher for a person who was in a crash involving alcohol versus a person in a crash not involving alcohol. In the Inpatient (SID) records, the average cost per person is $2,276 lower for a person who was in a crash involving alcohol versus a person in a crash not involving alcohol. Combining costs across all SID and SEDD records, the average hospital cost per person is $9,782 higher for a person who was in a crash involving alcohol versus a person in a crash not involving alcohol. Injured persons involved in alcohol-related crashes were proportionally more likely to be linked with Inpatient records than just Emergency Department records (21.4% vs 78.6%) compared to those individuals who were involved in a crash not known to be alcohol-related (11.2% vs 88.8%). The length of stay for injured persons linked to Inpatient (SID) records was similar for both categories of alcohol involvement.

Table 11. Hospital costs by alcohol involvement in crash

	Not alcohol-involved N = 41,930*			Alcohol-involved N = 2,231		
	N	**Total charges (USD)**	**Avg. charges per person (USD)**	**N**	**Total charges (USD)**	**Avg. charges per person (USD)**
SEDD	38,775	212,696,060	5,485	1,955	16,675,069	8,529
SID	4,712	367,060,014	77,899	476	35,996,622	75,623
Overall	41,930	579,756,074	13,827	2,231	52,671,691	23,609
*A total of 41,950 linked individuals were not known to be in alcohol-involved crashes; however, 20 individuals were missing hospital charge data and were not included in this table.						

Table 12. Record Type, length of stay (LOS), and Discharge Disposition by crash alcohol involvement

	Not alcohol-involved N = 41,950	Alcohol-involved N = 2,231
Record Type		
Both	1,557 (3.7%)	200 (9.0%)
SEDD Only	37,238 (88.8%)	1,755 (78.6%)
SID Only	3,155 (7.5%)	276 (12.4%)
Length of Stay (SEDD)		
0	34,458 (88.8%)	1,471 (75.2%)
1	3,879 (10.0%)	449 (23.0%)
2	334 (0.9%)	28 (1.4%)

3+	124 (0.3%)	7 (0.4%)
Length of Stay (SID)		
0	140 (3.0%)	23 (4.8%)
1	621 (13.2%)	83 (17.4%)
2	741 (15.7%)	78 (16.4%)
3+	3,210 (68.1%)	292 (61.3%)
Discharge Disposition (SEDD)		
AMA/Unknown	208 (0.5%)	32 (1.6%)
Died	187 (0.5%)	15 (0.8%)
Routine/Discharge Alive	35,935 (92.6%)	1,567 (80.2%)
Short-Term Hospital	2,096 (5.4%)	299 (15.3%)
Transfer Other	369 (1.0%)	42 (2.1%)
Discharge Disposition (SID)		
AMA/Unknown	21 (0.4%)	4 (0.8%)
Died	162 (3.4%)	14 (2.9%)
Routine/Discharge Alive	2,940 (62.4%)	355 (74.6%)
Short-Term Hospital	120 (2.5%)	12 (2.5%)
Transfer Other	1,469 (31.2%)	91 (19.1%)

Older Drivers

The average SEDD charge per person is quite similar among the three age groups examined ($6,176 for the referent age group of 45-54, $6,312 for the 65-74 age group, and $6,743 for the 75+ age group). In terms of SID charges, the average cost per person is $2,782 higher for the 65-74 age group than the 75+ age group. Meanwhile, the referent age group of 45-64 had a much larger mean cost per person then either of the two older age groups of interest. The lengths of stay were similar across the three age groups as well, though those age 75+ had slightly higher lengths of stay that were three or more days (76% vs 74.5%: ages 65-74 and 69.3%: ages 45-64).

Table 13. Hospital costs of older drivers by Age Group; 45-64 Age group included for reference

	Age 45-64 N = 11,094		Age 65-74 N = 3,412		Age 75+ N = 2,410	
	Total Charges (USD)	Avg. Charges per person (USD)	Total Charges (USD)	Avg. Charges per person (USD)	Total Charges (USD)	Avg. Charges per person (USD)
SEDD	62,295,719	6,176	190,421,59	6,312	13,835,845	6,743
SID	128,463,030	85,075	38,818,260	66,356	36,237,186	63,574
Overall	190,758,749	17,195	57,860,419	16,958	50,073,031	20,777
*A total of 11,097 linked individuals were in the age range of 45-094 and a total of 2,410 were in the age group of 75+; however, 3 from the 45-64 age group and 1 from the 75+ age group were missing hospital charge data and were not included in this table.						

Table 14. Record type, LOS, and Discharge Disposition for older drivers by Age Group; 45-64 Age group included for reference.

	Age 45-64 N = 11,097	Age 65-74 N = 3,412	Age 75+ N = 2,411
Record Type			
Both	502 (4.5%)	190 (5.6%)	212 (8.8%)
SEDD Only	9,587 (86.4%)	2,827 (82.8%)	1,841 (76.4%)
SID Only	1,008 (9.1%)	395 (11.6%)	358 (14.8%)
Length of Stay (SEDD)			
0	8,937 (88.6%)	2,677 (88.7%)	1,765 (86.0%)
1	990 (9.8%)	271 (9.0%)	219 (10.7%)
2	117 (1.2%)	52 (1.7%)	50 (2.4%)
3+	45 (0.4%)	17 (0.6%)	19 (0.9%)
Length of Stay (SID)			
0	43 (2.8%)	*	*
1	185 (12.3%)	*	*
2	235 (15.6%)	76 (13.0%)	81 (14.2%)
3+	1,047 (69.3%)	436 (74.5%)	433 (76.0%)
Discharge Disposition (SEDD)			
AMA/Unknown	56 (0.6%)	*	*
Died	60 (0.6%)	*	*
Routine/Discharge Alive	9,240 (91.6%)	2,695 (89.3%)	1,713 (83.4%)

Short-Term Hospital	641 (6.4%)	242 (8.0%)	230 (11.2%)
Transfer Other	92 (0.9%)	46 (1.5%)	72 (3.5%)
Discharge Disposition (SID)			
AMA/Unknown	*	*	*
Died	*	*	39 (6.8%)
Routine/Discharge Alive	923 (61.1%)	291 (49.7%)	191 (33.5%)
Short-Term Hospital	50 (3.3%)	25 (4.3%)	16 (2.8%)
Transfer Other	485 (32.1%)	246 (42.1%)	323 (56.7%)

Occupant Protection

In the table below, restraint status refers to whether the linked occupant is restrained or not with some known form of vehicle occupant protection. To directly compare known restrained individuals with known non-restrained individuals in vehicles, all vehicle occupants with unknown restraint status are excluded. Non-motorists are also excluded.

Of the vehicle occupants with known occupant protection status linked to HCUP data (total N = 34,893), 88% were known to be restrained. In the Emergency Department (SEDD) records, the average cost per person is over $3,000 higher for a vehicle occupant who was not restrained versus one who was restrained. In the Inpatient (SID) records, the average cost per person is $25,278 higher for a vehicle occupant who was not restrained versus one who was restrained. Combining costs for SID and SEDD records, the average hospital cost per person is $21,899 higher for a vehicle occupant who was not restrained versus one who was restrained. Compared to individuals who were known to be restrained, a greater proportion of individuals who were known to not be restrained were linked to an Inpatient record rather than just an Emergency Department record.

Table 15. Hospital costs by Restraint Status

	Not Restrained N = 4,161		**Restrained N = 30,713**	
	Total Charges (USD)	**Avg. Charges per person (USD)**	**Total Charges (USD)**	**Avg. Charges per person (USD)**
SEDD	28,596,800	8,251	150,877,601	5,198
SID	106,720,132	94,193	175,320,349	68,915
Overall	135,316,932	32,520	326,197,950	10,621
*A total of 30,732 linked individuals were restrained; however, 19 were missing hospital charge data and were not included in this table.				

Table 16. Record Type, LOS, and Discharge Disposition by Restraint Status

	Not Restrained N = 4,161	**Restrained N = 30,732**
Record Type		
Both	438 (10.5%)	856 (2.8%)
SEDD Only	3,028 (72.8%)	28,188 (91.7%)
SID Only	695 (16.7%)	1,688 (5.5%)
Length of Stay (SEDD)		
0	2,778 (80.2%)	26,064 (89.7%)
1	619 (17.9%)	2,680 (9.2%)
2	52 (1.5%)	224 (0.8%)
3+	17 (0.5%)	76 (0.3%)
Length of Stay (SID)		
0	43 (3.8%)	67 (2.6%)
1	145 (12.8%)	373 (14.7%)
2	156 (13.8%)	435 (17.1%)
3+	789 (69.6%)	1,669 (65.6%)

Discharge Disposition (SEDD)		
AMA/Unknown	23 (0.7%)	151 (0.5%)
Died	100 (2.9%)	70 (0.2%)
Routine/Discharge Alive	2,639 (76.1%)	27,479 (94.6%)
Short-Term Hospital	621 (17.9%)	1,122 (3.9%)
Transfer Other	83 (2.4%)	222 (0.8%)
Discharge Disposition (SID)		
AMA/Unknown	*	11 (0.4%)
Died	65 (5.7%)	78 (3.1%)
Routine/Discharge Alive	686 (60.5%)	1,603 (63.0%)
Short-Term Hospital	*	59 (2.3%)
Transfer Other	341 (30.1%)	793 (31.2%)

TEEN DRIVER CULPABILITY ANALYSES

During this performance period a paper titled “Direct medical charges of all parties in teen-involved vehicle crashes by culpability” was published using the Iowa CODES 2016-2020 dataset. The paper is available here: https://injuryprevention.bmj.com/content/29/4/334 The abstract of the paper is included below:

Background: Motor vehicle crashes among teen drivers often involve passengers in the teen’s vehicle and occupants of other vehicles, and the full cost burden for all individuals is largely unknown. This analysis estimated direct hospitalisation and emergency department charges for teen-involved crashes by teen culpability, comparing charges for the teen driver, passengers and occupants of other vehicles.

Methods: Probabilistic linkage was performed to link the Iowa police crash reports with Iowa emergency department and Iowa hospital inpatient data. Teen drivers aged 14–17 involved in a crash from 2016 through 2020 were included. Teen culpability was determined through the crash report and examined by teen and crash characteristics. Direct medical charges were estimated from charges through linkage to the Iowa hospital inpatient and the Iowa emergency department databases.

Results: Among the 28 062 teen drivers involved in vehicle crashes in Iowa between 2016 and 2020, 62.1% were culpable and 37.9% were not culpable. For all parties involved, the inpatient charges were $20.5 million in culpable crashes and $7.2 million in non-culpable crashes. The emergency department charges were $18.7 million in teen culpable crashes and $6.8 million in teen non-culpable crashes. Of the $20.5 million total inpatient charges in which a teen driver was culpable, charges of $9.5 million (46.3%) were for the injured teen driver and $11.0 million (53.7%) for other involved parties.

Conclusions: Culpable teen-involved crashes lead to higher proportions of injury and higher medical charges, with most of these charges covering other individuals in the crash.

NEXT STEPS

1. Obtain date of birth information for individuals involved in crashes during 2021, join it with the rest of the crash data, and link the crash data to the 2021 Emergency Department and Inpatient hospital data from HCUP.
2. Pursue obtaining missing zip codes for non-motorists involved in crashes from 2016 to present
3. Modify (or consider modifying) CODES linkage to
    a. Omit persons who died at the crash scene or who were transported to an out-of-state hospital prior to linkage
    b. Further restrict e-codes to motor vehicle crashes specifically
    c. Capture multiple visits to unique hospitals (i.e., transfers) associated with the same crash event
    d. Export and examine match quality metrics
    e. Examine match quality metrics and adjust criteria to relax or tighten as necessary
4. Report missingness within relevant HCUP data elements, linkage variables, and outcome measures of interest (ICD-10 CMs to calculate ISS, LOS, costs).

5. Investigate how to calculate ISS scores for individuals linked to both Emergency Department and Inpatient data.
6. Conduct adjusted analyses to examine 2016-2021 vehicle crash outcomes in relation to:
    - Benefits of safety devices (e.g., seat belts, helmets)
    - Medical and economic burden estimates
    - Trends in vehicle crashes and related injury outcomes for high-risk and vulnerable road user populations, specific injury types, and specific crash and vehicle characteristics (e.g., older drivers, motorcyclists, traumatic brain injuries)
7. Integrate justice (charge/conviction data) with CODES data

CONCLUSION

During this performance period, our team made continued progress toward our objectives of building a comprehensive CODES dataset and analyzing outcomes related to motor vehicle crash-related injuries in the State of Iowa for different populations and safety emphasis areas. We improved and examined the quality of the probabilistic linkage to join the Iowa police crash report data with Iowa Emergency Department data, Iowa Hospital Inpatient data, and FARS fatal motor vehicle crashes from 2016 through 2020. The current set of linkage variables resulted in unique combinations (i.e., only one individual in the dataset had that combination) for 98.3%, 92.2% and 95.7% of the crash, Emergency Department, and Inpatient records, respectively.

The linkage yielded probabilistic matches to the hospital data for 44% of the crashes. Considering only the injured persons who were reported as transported by EMS from the crash scene, about 70% were linked to hospital records. Within this group of linked individuals, the quality of the matches seems to be quite good, as the hospital information entered in the crash report corresponded with the hospital identifier in the hospital data nearly 90% of the time. However, there is additional room for improving the number of linked records, particularly with regard to the other 30% of injured persons transported by EMS who were not linked to hospital records.

During this course of this performance periods, as with several other past projects funded by GTSB and completed by investigators at the University of Iowa, as we cleaned the data in preparation for linkage and attempted to analyze the data, we discovered a number of new data quality issues. In this year specifically, these have hampered our progress on linking more recent crash data to Emergency Department and Inpatient hospital data and on our ability to evaluate crash outcomes for non-motorists.

Appendix A

Iowa's 2023 CODES Project Personnel

Conducted by the University of Iowa Injury Prevention Research Center
Under contract to Iowa Governor's Traffic Safety Bureau and Iowa Department of Transportation

Key Investigators
Cara J. Hamann, MPH, PhD
Associate Professor
The University of Iowa Injury Prevention Research Center

Michelle Reyes
Senior Research Associate
University of Iowa Driving Safety Research Institute

Jonathan Davis
Research Assistant Professor
The University of Iowa Injury Prevention Research Center

Data Analysts
Ling Zhang
Stephanie Jansson

Research Team
Ryan Dusil
Corinne Peek-Asa
Elizabeth O'Neal

GTSB Contact
Mick Mulhern