CENTER FOR TRANSPORTATION STUDIES

UNIVERSITY OF MINNESOTA

Minnesota Crash Records Audit

Final Report

Nichole Morris

HumanFIRST Laboratory
Department of Mechanical Engineering
University of Minnesota

CTS 20-11

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.
CTS 20-11		

4. Title and Subtitle	5. Report Date
Minnesota Crash Records Audit	August 2020
	6.

7. Author(s)	8. Performing Organization Report No.
Nichole L. Morris, David Libby, Colleen Peterson, Andrew Ryan, McKenzie Sheppard	

9. Performing Organization Name and Address	10. Project/Task/Work Unit No.
HumanFIRST Laboratory Department of Mechanical Engineering 111 Mechanical Engineering, 111 Church St. SE Minneapolis, MN 55455	CTS #2017064
	11. Contract (C) or Grant (G) No. (c) 1003325 (wo) 18

12. Sponsoring Organization Name and Address		13. Type of Report and Period Covered
MnDPS 445 Minnesota Street St. Paul, MN 55101-5155	MnDOT 395 John Ireland Blvd. St. Paul, MN 55155-1800	Final Report
		14. Sponsoring Agency Code

15. Supplementary Notes

http://www.cts.umn.edu/Publications/ResearchReports/

16. Abstract (Limit: 250 words)

In 2016, Minnesota deployed a state-of-the art electronic crash reporting system to all law enforcement officers across the state. The updated crash report was created with the officer in mind through an extensive usability and design project led by HumanFIRST. Preliminary testing of the new system suggested high user satisfaction but could not determine the extent of the crash data improvement. This study employed a mixed-methods approach to conduct a series of qualitative analyses of the crash data records collected by the legacy crash reporting system and the new MNCrash reporting system. The first analysis compared 360 serious injury and fatal crash reports from 2015 and 2016 and found a year-to-year decrease in mismatching data but a slight increase in missing data. However, overall, the MNCrash system increased the number of data queries, so data access was increased in 2016. The second analysis of serious and fatal injury crashes compared 996 reports sampled from 2015 and 1,572 reports sampled from 2016. The analysis showed a decrease in missing data and a decrease in erroneous data entry in 2016. Finally, the converted 2015 legacy data was analyzed to reveal few errors in the MNCrash conversion process.

17. Document Analysis/Descriptors	18. Availability Statement
Crash reports, Testing, Iterative methods, Human factors	No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312

19. Security Class (this report)	20. Security Class (this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	101	

MINNESOTA CRASH RECORDS AUDIT

FINAL REPORT

Prepared by:

Nichole L. Morris
David Libby
Colleen Peterson
Andrew Ryan
McKenzie Sheppard
Department of Mechanical Engineering
University of Minnesota

AUGUST 2020

Published by:

Center for Transportation Studies
University of Minnesota
440 University Office Plaza
2221 University Avenue SE
Minneapolis, MN 55414

ACKNOWLEDGMENTS

The authors would like to acknowledge the Minnesota Department of Public Safety for funding and supporting this project. We would like to give a special thank you to Kathleen Haney, VHB, for her support and encouragement of this work. Thank you to staff of the Minnesota Department of Public Safety Office of Traffic Safety: Alex Gohar, Chelsea Palmateer, Shannon Ryder, Karen Aldridge, Allison Hawley, Bob Hoemke, Michael Hanson, and Craig Flynn. Thank you to our partners at the Minnesota Department of Transportation: Katie Fleming, Nathan Drews, Loren Hill, and Eric DeVoe. Also, thank you to the HumanFIRST research team members for their support in this work: Curtis Craig, Katelyn Schwieters, and Savindie Liyanagamage.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A: Serious Injury (2015) or Suspected Serious Injury (2016)

ACCN: Crash Number

CMV: Commercial Motor Vehicle

K: Fatality

MMUCC: Model Minimum Uniform Crash Criterion

MNCrash: Minnesota Crash Reporting System

MnDPS: Minnesota Department of Public Safety

MSP: Minnesota State Patrol

MV: Motor Vehicle

NA: Not Applicable

RPN: Relative Person Number

RVN: Relative Vehicle Number

EXECUTIVE SUMMARY

On January 1, 2016, Minnesota replaced its legacy accident reporting system, in use since 2004, with the updated MNCrash reporting system, with the goal of improving crash data quality, accuracy, and ease of entry for the responding officers. This new front-end interface corresponded with a new MNCrash database. To test the efficacy of this new crash reporting system, the HumanFIRST research team conducted an audit of the crash data from 2015, which was entered using the legacy system, and compared it to crash data from 2016, which was entered utilizing the new MNCrash reporting system. The goal of this crash audit was to determine the extent to which the MNCrash reporting system has improved data quality characteristics, compared to the legacy crash reporting system.

The audit included three main activities to measure potential quality improvements in the crash data from the MNCrash system. This included a qualitative analysis and quantitative analysis of 2015 and 2016 crash data and an examination of the conversion process of the data from 2015 and earlier into the new MNCrash data formatting structure.

The first qualitative analysis of the data examined the data's validity, consistency, and verifiability in a comparative analysis with the crash narratives. The method used for this assessment involved a detailed examination of 360 crash reports from 2015 and 2016 and compared the data entry fields to a 'gold standard' account by leveraging the officer-recorded narratives. In general, there was a year-to-year improvement in data accuracy, as measured by a decrease in mismatching information in 2016. Minnesota State Patrol tended to provide the most complete data relative to the high- and low-volume reporting agencies in 2015. In 2016, there was a significant increase for the high- and low-volume reporting agencies resulting in no differences in data completeness across agencies in 2016. There was a small, but statistically significant increase in missing information in 2016 through this analysis; however, the 2016 MNCrash reporting system has nearly twice as many form items compared to the 2015 system. The slight decrease in data completeness, as measured by the percent of missing form information, is surpassed by the overall gain of crash information added by the MNCrash system. The qualitative analyses suggest that further design modifications should be considered in an effort to continue to support officers as they complete the report to reduce remaining errors.

The second qualitative analysis of the data examined the data's completeness and accuracy. The method used for this assessment involved a larger examination of serious injury and fatal crash data across years with 996 reports sampled from 2015 and 1,572 reports sampled from 2016. The analysis of the MNCrash and legacy datasets compared the frequencies of missing, unknown, other, and not applicable responses along with system logic errors and user logic errors. The results showed that when the backend conditional skip logic issues for missing data were accounted for, the overall completeness of the crash data significantly increased in 2016. Agencies that report greater than 500 crashes a year experienced the greatest improvement in reducing the amount of missing data in their crash reports with the use of the MNCrash system. The targeted design components of the MNCrash interface, particularly conditional skip logic features, successfully reduced erroneous data entry for single unit crashes and motorcycle crashes. The nested structure of the sequence of events also showed significant improvements in reducing erroneous data entry for single unit crashes, improving accuracy and

agreement in most harmful event data capture, and increasing non-collision event documentation for first events and prior to collision with fixed object crashes. The results suggest opportunities to further improve data completeness through more comprehensive conditional logic functions to auto-populate fields as 'NOT APPLICABLE' when they are skipped by the system.

Third, the audit concluded with an examination of the conversion process of the legacy data into the new MNCrash data format. The compliance of the new MNCrash system to updated Model Minimum Uniform Crash Criteria (MMUCC data standard meant that several data variables or attributes would need to be added, removed, or modified, making the legacy data format incompatible with the new MNCrash database. A detailed data dictionary and conversion process were created to transform the legacy data into the new MNCrash format. The final activity of this project sampled converted data and the conversion notes to determine if any data quality was lost within the process. Few issues were discovered involving incomplete translation notes, erroneously inserting data where missing fields had been, and losing data for 'NOT APPLICABLE' data. Overall, the analysis of the MNCrash data conversion process revealed that the integrity of the data was maintained in a vast majority of the variables examined.

The outcome of the analysis indicated that the MNCrash system not only improved user satisfaction, as found in the original design study, but it also increased the quality of the data from 2015 to 2016 through completeness, verifiability, and accuracy. This is an important finding to support user-centered design as a standard practice for all crash reporting systems across the United States to not only ease officer frustrations but also improve our understanding of crashes and the factors that contribute to them. This work will be continued through additional research focused on iterative user-centered design and user-centered training modules expected to further improve the data quality gains seen with the first implementation of the MNCrash system.

CHAPTER 1: INTRODUCTION

On January 1, 2016, Minnesota replaced its legacy accident reporting system, in use since 2004, with the updated MNCrash reporting system, with the goal of improving crash data quality, accuracy, and ease of entry for the responding officers. This new front-end interface was created based on an in-depth, human factors, user-centered design project (Morris, Achtemeier, Ton, Plummer, & Sykes, 2016) and corresponded with a new MNCrash database. To test the efficacy of this new crash reporting system, the HumanFIRST research team conducted an audit of the crash data from 2015, which was entered using the old legacy system, and compared it to crash data from 2016, which was entered using the new MNCrash reporting system. The goal of this crash audit was to determine the extent to which the MNCrash reporting system had improved data quality characteristics, compared to the legacy crash reporting system. The crash audit compared the validity, completeness, and accuracy of the two reporting systems using a mixed-methods analysis of the crash report data. The research team created and used a coding rubric to determine if certain business rules had been met, documented instances of incomplete or inconclusive data entry, and cross-referenced each officer's narrative account of a crash against select elements within the report. In addition, the team analyzed a larger sample of data to determine the extent of missing data, incomplete/inconclusive data, or violation of certain business rules. The goal of the crash audit was to determine the extent to which the MNCrash reporting system had improved data quality characteristics, compared to the legacy crash reporting system.

1.1 MNCRASH BACKGROUND

Crash report data helps to provide objective information about the conditions and factors that lead to their occurrences. Law enforcement personnel enter this data based on their observations, witness statements, and driver admissions, which are subject to inaccuracies, especially if there is urgency in drafting the report at the scene of the crash site (Wickens & Hollands, 1999). The responsibility placed on law enforcement officers to report accurate crash details is immense because the data are used in implementing traffic safety policy and costly infrastructure changes (Porter, 2011). These impactful, data-driven, societal and financial decisions may be influenced by inconsistencies in reporting practices across different law enforcement agencies (Farmer, 2003). Furthermore, the data entry errors that users of the crash report may make can "ruin statistical results and conclusions" that may not be detected through examinations of frequencies or histograms (Barchard & Pace, 2011). Providing a user-centered interface to help account for the capabilities and limitations of users is an effective approach to reduce data entry errors and better support the demands of users.

Minnesota embarked on a new crash record database and reporting system in 2014 to improve data accuracy through integrated technology features and increased data standardization with federal guidance (Model Minimum Uniform Crash Criterion Fourth Edition or MMUCC, 2012). The process involved an innovated, iterative design approach to create a user-centered interface for Minnesota law enforcement officers (Morris et al., 2016). The goal of the user-centered design project was to use human factors techniques to provide an interface to support officers in timely, accurate, reliable, and meaningful crash report data.

In total, 82 Minnesota law enforcement agencies (including Minnesota State Patrol, 66 local police departments, and 15 county sheriff's departments) participated in the user-centered design research effort (Morris et al., 2016). Officers' feedback and suggestions were integrated into the design of the report interface from the early stages of the prototype all the way through to the finished product. The results of the series of interviews, card sorting tasks, surveys, and usability tests led to the creation of two interfaces (i.e., a form-based interface and a wizard-based interface), with the same features and functionality, to accommodate users of varying technical skill levels and preferences. The interfaces were built by a separate technology vendor, Appriss LLC, and the research team oversaw beta testing, quality assurance, and final usability of the software release.

The new MNCrash reporting system launched on January 1, 2016, was well-received by local law enforcement agencies and the Minnesota State Patrol. The project demonstrated the importance of user-centered design for supporting human-computer interactions and user satisfaction. This project was the first known of its kind in the United States to apply an in-depth usability practice in a statewide interface. The average submission time of crash reports statewide fell to less than one day, meeting the project goals of improving data timeliness. However, it has been harder to ascertain whether the goal of improving the interface to result in more accurate, reliable, and meaningful data has been met. This project aims to compare the 2016 MNCrash data to the 2015 legacy system crash data through a systematic audit to determine the extent to which the original study goals have been met.

1.2 MNCRASH AUDIT OVERVIEW

The audit included four main activities to measure potential quality improvements in the crash data from the MNCrash system. This included assessing the current state of usability for the MNCrash system and examining the conversion process of the data from 2015 and earlier into the new MNCrash data formatting structure.

The first qualitative analysis of the data examined the data's validity, consistency, and verifiability in a comparative analysis with the crash narratives. The method used for this assessment involved a detailed examination of each crash report and compared the data entry fields to a 'gold standard' account by leveraging the officer-recorded narratives. This comparison allowed the research team to consider the data's validity, consistency, and the team's confidence in the data through its verifiability against the officer's plain language description of the crash. The analysis examined select samples of 360 crash reports from 2015 and 2016. Identical sampling strata were utilized for 2015 and 2016 to ensure the comparability of the validation tests between years. Coding results were analyzed to determine year-to-year differences in mismatching, missing, and matching accounts between form items and narrative descriptions.

The second qualitative analysis of the data examined the data's completeness and accuracy. Similar to the first qualitative analysis, a sample was defined to ensure comparable examination of crash factors across reports and years and resulted in 996 reports sampled from 2015 and 1,572 reports sampled from 2016. The analysis of the MNCrash and legacy datasets compared the frequencies of missing, unknown, other, and not applicable responses. In addition, the quantitative analysis focused on several

design features of the MNCrash interface and compared the number of system logic errors and user logic errors in the crash reports between the MNCrash and legacy reporting systems.

The qualitative comparative and data analysis activities were followed by a series of usability tests to directly observe officers interacting with the MNCrash interface and determine design solutions to address errors. These tests helped to better identify some of the remaining confusion points within the interface, common misconceptions, or frustrations creating barriers to supporting users in entering timely, accurate, and complete data. The results were integrated into the initial design recommendations and training guides to assist users of the MNCrash system.

Finally, the project concluded with an examination of the conversion process of the legacy data into the new MNCrash data format. The compliance of the new MNCrash system to the updated Model Minimum Uniform Crash Criterion Fourth Edition (MMUCC) data standard meant that several data variables or attributes would need to be added, removed, or modified, making the legacy data format incompatible with the new MNCrash database. A detailed data dictionary and a conversion process were created to transform the legacy data into the new MNCrash format. The final activity of this project sampled converted data and the conversion notes to determine if any data quality was lost within the process.

CHAPTER 2: QUALITATIVE COMPARATIVE ANALYSIS

2.1 INTRODUCTION

The goal of this qualitative comparative analysis was to support the crash data audit by determining the extent to which the MNCrash reporting system has improved data quality characteristics, compared to the legacy crash report system. The officer-recorded narratives for each report were used as the 'gold standard' to validate selected form items in the reports in a side-by-side analysis. The detailed coding required for this activity limited the total number of reports that could be feasibly analyzed, and this required a controlled and representative sample of the crash reports to ensure adequate statistical power and generalizability. Additionally, this process required the research team to create a structured process to consistently assess and document potential errors of each report examined. The research team created and used a rubric to determine if certain business rules had been violated, documented instances of incomplete or inconclusive data entry, and cross-referenced the officer's narrative account of the crash against select elements within the report.

The statistical analysis was first aimed at determining if any differences in accuracy or consistency existed in the detailed data coding of the sample crash reports. Additionally, the analysis helps to determine if any gains in crash understanding have been accomplished through the implementation of the new report with updated MMUCC guidelines and intelligent data capture.

2.2 CRASH SAMPLE STRATIFICATION

A target of 720 crash reports was selected for the analysis with 360 crash reports for each of the years, 2015 and 2016. Since reporting agencies started utilizing the MNCrash reporting system on January 1st, 2016, crash reports for the first three months (January to March) were excluded. This intended to reduce the effect of the initial 'learning period' by decreasing the frequency of officer errors in the crash data that were caused by unfamiliarity with the new system. Similarly, to reduce the potential for seasonal differences in the data, crashes for the first three months of 2015 were also removed. The sampling plan was further stratified with the purpose of determining if any differences in data validity, consistency, or verifiability were affected by crash severity, reporting agency, or crash type. Target sampling within each stratum was identified to be representative of historic reporting frequencies. However, contingency over-sampling was planned in order to make adjustments in the event that one or more strata did not have the required number of crash records with useable narratives in the sampling pool. Crash reports documented through the form-based interface and the wizard-based interface were not considered through this analysis. This was because the wizard-based interface was used too infrequently (i.e., less than 10% of users) for an adequate statistical comparison and no functional or major feature differences were present between the two interfaces.

2.2.1 Crash Severity

After dividing the crash sample by year, the sample was further separated by crash severity. Notably, previous definition provided to officers for an (A) injuries was "Incapacitating Injury" which corresponded with MMUCC definitions at the time of the crash report's last update. The MMUCC 4th edition, which was the latest manual version at the time of the MNCrash update, maintained the same "KABCO" acronym as previous editions; however, injury definitions were changed so that (A) injuries indicate "Suspected Serious Injury" and (B) injuries were changed from "Non-incapacitating Injury" to "Suspected Minor Injury", while (C) injuries remained "Possible Injury". While this definition change may have shifted officer decision making or injury classifications from 2015 to 2016, this analysis treated (A) injuries from each year as equivalent and for the purposes of this analysis will describe them as "Serious Injury" Crashes.

Only fatal crashes (K) and serious injury crashes (A) were utilized for this sample. There were several reasons for this decision: (1) serious crashes tended to have more nuances regarding what led up to the crash and thus would potentially provide more data points on the crash report; (2) crashes that have a high injury or fatality rate tended to carry more weight in informing legislative or infrastructure changes, so accurate and complete data for these crashes is of great importance; and (3) crashes that involve an injury or fatality were more likely to have a longer and more detailed narrative from the reporting officer regarding what transpired when the vehicle(s) crashed. For both 2015 and 2016, there was approximately a 4:1 ratio of serious injury crashes (A) to fatal crashes (K), which was translated into a similar sampling ratio for the crash reports. This resulted in a total of 150 fatal crash reports and 570 serious injury crash reports, divided evenly between each year. Table 2.1 shows the breakdown of crash reports based on year, crash type, reporting agency, and crash severity.

Table 2.1 Stratified Crash Report Sample Sizes for Qualitative Coding

	2015 (April-Dec)		2016 (April-Dec)	
	Fatal Crash (K)	Serious Injury (A)	Fatal Crash (K)	Serious Injury (A)
Single-Unit Crash				
State Patrol	10	41	11	40
High Volume Reporting Agency	10	38	9	40
Low Volume Reporting Agency	10	41	10	40
Two-Unit Crash				
State Patrol	10	40	10	40
High Volume Reporting Agency	10	40	10	40
Low Volume Reporting Agency	10	40	10	40
Pedestrian/Cyclist Crash				
State Patrol	5	6	5	11
High Volume Reporting Agency	5	20	5	17
Low Volume Reporting Agency	5	19	5	17

2.2.2 Crash Type

The sample crash reports were further divided into three separate crash types: (1) single-unit crashes; (2) two-unit crashes; and (3) crashes involving one motor vehicle in transit and a pedestrian or bicyclist. Crashes involving more than two motor vehicles in transit were excluded from the sample due to their overall infrequency. For both 2015 and 2016, there were far fewer crashes involving cyclists or pedestrians compared to single-unit and two-unit crashes. This resulted in a total of 120 crashes involving pedestrians/cyclists, compared to 300 crash reports for both the single and two-unit crash types, see Table 2.1.

2.2.3 Reporting Agency Type

Agency type was also a focus of interest to determine if any differences between years were observable based on the agency that reported the crash. The crash reports were divided into three separate categories of agencies: (1) crashes reported by Minnesota State Patrol (MSP); (2) crashes reported by agencies/cities that reported more than 500 crashes each year; and (3) agencies/cities that reported fewer than 500 crashes each year. This was done specifically to see if familiarity with the system or experience in crash reporting played a role in the validity and consistency of the data. Officers from MSP and agencies from larger metropolitan areas responded to more crashes and tended to compile more crash reports than smaller agencies, which may result in differences in data quality due to a greater familiarity with the reporting system. Sampling was similar for each of the three agency types with 229 sampled for MSP, 244 samples for high volume agencies (i.e., 500 or greater reports per year), and 247 for low volume agencies (i.e., fewer than 500 reports per year), see Table 2.1

2.3 DATA CODING

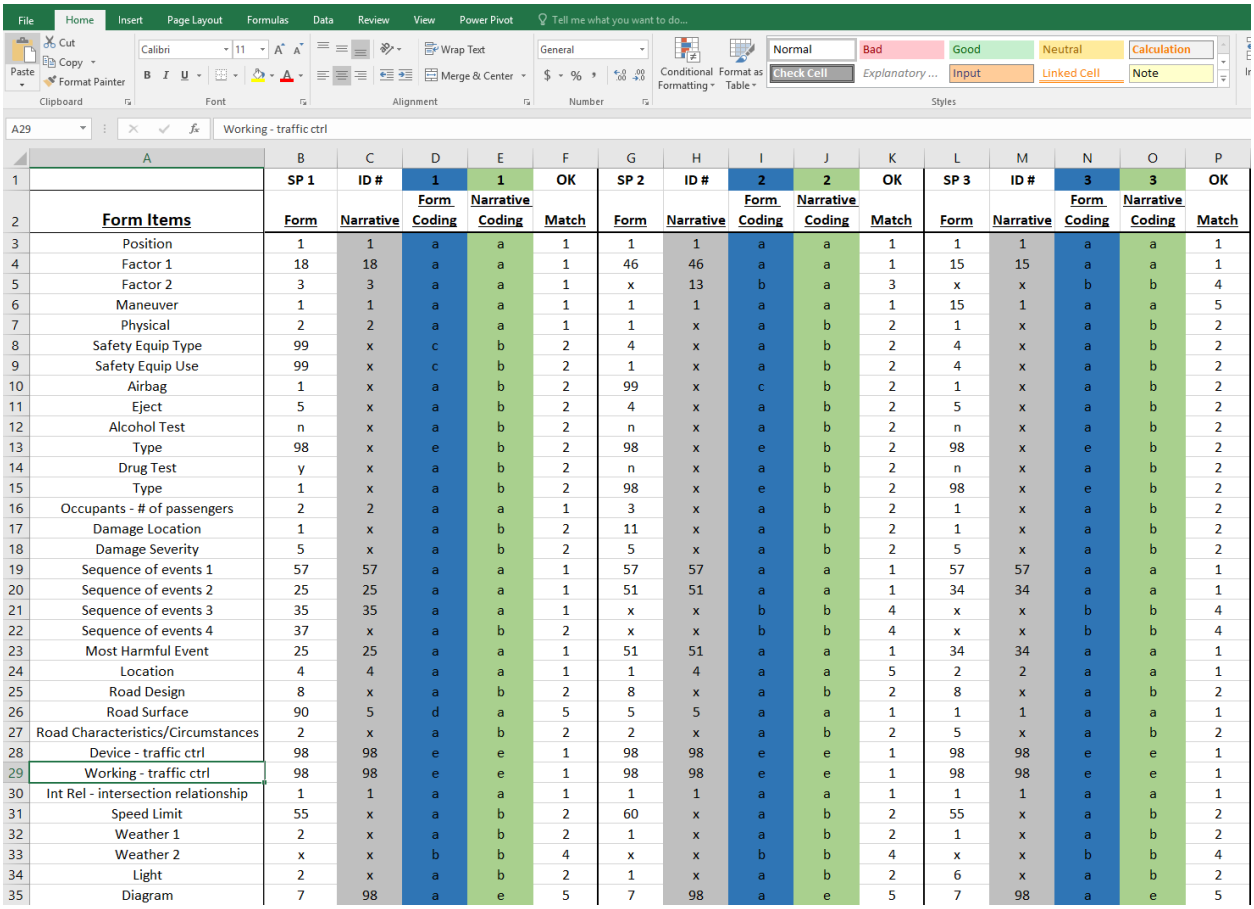
To determine the sampled reports' validity, consistency, and verifiability, it was necessary to examine and code the narratives that were written by the officer for each crash report. The narratives were then cross-compared with the crash report form (i.e., data entry fields) to determine if there were any conflicting or missing pieces of information between the form and the narrative. This was done for all 720 reports and the results were then cross-compared based on year, reporting agency type, crash severity, and crash type. Due to many of the items not being pertinent to the crash itself, not every entry item from the crash reports were coded. Only the entry items that were in some way related to the actual crash event or what precipitated the crash were coded. For the 2015 crash reports, there were 32 items from the form that were used for the comparison and 64 items from the 2016 reports (see Appendix A for complete list).

Details provided in the narrative section of the report were compared with individual elements in the reports. The coding focused on elements that were related to crash occurrence or severity and omitted all personal and vehicle descriptors. Elements were coded as missing, matching, or not matching the narrative. For example, the attending officer may indicate in the narrative that the driver of the vehicle was not wearing a safety belt. The coding protocol determined if the corresponding report element, related to safety equipment use, matched the narrative account, provided information different from

the narrative account, or was missing information altogether. The percentage of missing information within and across elements was calculated for the MNCrash reports and the legacy reports, as well as the percentage agreement and sensitivity for non-missing report elements. Statistical differences among years, agencies, or crash types would be tested using chi-square or Fisher's exact tests.

2.3.1 Building the Codebooks

Researchers used Microsoft Excel for the data entry and comparison of the crash reports. Figure 2.1 shows a depiction of how the codebooks were setup. After listing all of the pertinent crash report entry items to be coded, columns were added for the information entered on the crash form fields (Column B in Figure 2.1) and the crash narrative (Column C in Figure 2.1). Since all the entry items within the reporting system have a corresponding number, the researchers, who conducted the data entry, used the same numeric entry style in the codebooks as the officers did in the crash report system. Next, two more columns were added which would code what was entered for both the form side and narrative side of the report (Column D & E in Figure 2.1). This coding was based on whether information for each entry item was marked as present, absent, other (90), not applicable (98), or unknown (99). Then, a fifth column (Column F in Figure 2.1) was used to determine the relationship between the information entered for each entry item between the form and the narrative. If the information entered for both was the same, it was coded as a match (1). If the information for both was entered, but the information was conflicting, it was coded as a mismatch (5). The codebooks were also set up to code if the information was present for the form field (2), but absent in the narrative, as well as if there was information absent in the form field, but present in the narrative (3). Lastly, it also coded if the information was absent for both the form and the narrative (4).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		SP 1	ID #	1	1	OK	SP 2	ID #	2	2	OK	SP 3	ID #	3	3	OK
2	Form Items	Form	Narrative	Form Coding	Narrative Coding	Match	Form	Narrative	Form Coding	Narrative Coding	Match	Form	Narrative	Form Coding	Narrative Coding	Match
3	Position	1	1	a	a	1	1	1	a	a	1	1	1	a	a	1
4	Factor 1	18	18	a	a	1	46	46	a	a	1	15	15	a	a	1
5	Factor 2	3	3	a	a	1	x	13	b	a	3	x	x	b	b	4
6	Maneuver	1	1	a	a	1	1	1	a	a	1	15	1	a	a	5
7	Physical	2	2	a	a	1	1	x	a	b	2	1	x	a	b	2
8	Safety Equip Type	99	x	c	b	2	4	x	a	b	2	4	x	a	b	2
9	Safety Equip Use	99	x	c	b	2	1	x	a	b	2	4	x	a	b	2
10	Airbag	1	x	a	b	2	99	x	c	b	2	1	x	a	b	2
11	Eject	5	x	a	b	2	4	x	a	b	2	5	x	a	b	2
12	Alcohol Test	n	x	a	b	2	n	x	a	b	2	n	x	a	b	2
13	Type	98	x	e	b	2	98	x	e	b	2	98	x	e	b	2
14	Drug Test	y	x	a	b	2	n	x	a	b	2	n	x	a	b	2
15	Type	1	x	a	b	2	98	x	e	b	2	98	x	e	b	2
16	Occupants - # of passengers	2	2	a	a	1	3	x	a	b	2	1	x	a	b	2
17	Damage Location	1	x	a	b	2	11	x	a	b	2	1	x	a	b	2
18	Damage Severity	5	x	a	b	2	5	x	a	b	2	5	x	a	b	2
19	Sequence of events 1	57	57	a	a	1	57	57	a	a	1	57	57	a	a	1
20	Sequence of events 2	25	25	a	a	1	51	51	a	a	1	34	34	a	a	1
21	Sequence of events 3	35	35	a	a	1	x	x	b	b	4	x	x	b	b	4
22	Sequence of events 4	37	x	a	b	2	x	x	b	b	4	x	x	b	b	4
23	Most Harmful Event	25	25	a	a	1	51	51	a	a	1	34	34	a	a	1
24	Location	4	4	a	a	1	1	4	a	a	5	2	2	a	a	1
25	Road Design	8	x	a	b	2	8	x	a	b	2	8	x	a	b	2
26	Road Surface	90	5	d	a	5	5	5	a	a	1	1	1	a	a	1
27	Road Characteristics/Circumstances	2	x	a	b	2	2	x	a	b	2	5	x	a	b	2
28	Device - traffic ctrl	98	98	e	e	1	98	98	e	e	1	98	98	e	e	1
29	Working - traffic ctrl	98	98	e	e	1	98	98	e	e	1	98	98	e	e	1
30	Int Rel - intersection relationship	1	1	a	a	1	1	1	a	a	1	1	1	a	a	1
31	Speed Limit	55	x	a	b	2	60	x	a	b	2	55	x	a	b	2
32	Weather 1	2	x	a	b	2	1	x	a	b	2	1	x	a	b	2
33	Weather 2	x	x	b	b	4	x	x	b	b	4	x	x	b	b	4
34	Light	2	x	a	b	2	1	x	a	b	2	6	x	a	b	2
35	Diagram	7	98	a	e	5	7	98	a	e	5	7	98	a	e	5

Figure 2.1 Screenshot of a Single-Unit Crash Codebook

Each codebook was then set up to tally the total number of matches, mismatches, and missing data for each particular form item. Figure 2.2 depicts how this looked for each individual codebook. There were 12 codebooks created for this study. Six were used for 2015 and six for 2016. The three codebooks for each year were then broken down based on crash type (i.e., single-unit, two-unit, and pedestrian/cyclist). Within each codebook, there were three identical tabs for the three different agency types that the crashes were sampled from (i.e., MSP, high volume-reporting agencies, and low-volume reporting agencies). This structure allowed eased data entry, coding, and classification based on agency type, crash type, crash severity, and year. This then allowed researchers to analyze the accuracy and completeness of the data broken down by each particular form item to determine which items were more likely to have missing or mismatching information between the form and narrative.

	A	GT	GU	GV	GW	GX	GY	GZ	HA	HB	HC	HD	HE	HF	HG	HH
1		SP 41	ID #	41	41	OK	Form & Narrative Match					Form Frequency				
2	Form Items	Form	Narrative	Form Coding	Narrative Coding	Match	Match	Narrative Blank	Form Blank	Both Blank	Mismatch	Present	Absent	Unknown	Other	N/A
3	Position	1	1	a	a	1	41	0	0	0	0	41	0	0	0	0
4	Factor 1	3	3	a	a	1	27	7	0	0	7	38	0	1	2	0
5	Factor 2	x	13	b	a	3	9	5	7	19	1	15	26	0	0	0
6	Maneuver	1	1	a	a	1	31	3	0	0	7	39	0	1	1	0
7	Physical	1	x	a	b	2	3	37	0	0	1	36	0	3	2	0
8	Safety Equip Type	4	4	a	a	1	6	35	0	0	0	21	0	4	4	12
9	Safety Equip Use	1	1	a	a	1	9	31	0	0	1	34	0	6	0	1
10	Airbag	6	x	a	b	2	2	39	0	0	0	22	0	2	1	16
11	Eject	1	3	a	a	5	18	20	0	0	3	38	0	2	0	1
12	Alcohol Test	n	x	a	b	2	5	36	0	0	0	41	0	0	0	0
13	Type	98	x	e	b	2	5	36	0	0	0	11	0	0	0	30
14	Drug Test	n	x	a	b	2	0	41	0	0	0	41	0	0	0	0
15	Type	98	x	e	b	2	0	41	0	0	0	8	0	0	0	33
16	Occupants - # of passengers	1	1	a	a	1	29	11	0	1	0	40	1	0	0	0
17	Damage Location	7	x	a	b	2	0	41	0	0	0	41	0	0	0	0
18	Damage Severity	5	x	a	b	2	0	41	0	0	0	41	0	0	0	0
19	Sequence of events 1	58	58	a	a	1	31	0	0	0	10	41	0	0	0	0
20	Sequence of events 2	57	57	a	a	1	20	0	9	4	8	27	13	0	1	0
21	Sequence of events 3	30	30	a	a	1	8	0	7	22	4	12	29	0	0	0
22	Sequence of events 4	x	x	b	b	4	2	3	4	32	0	5	36	0	0	0
23	Most Harmful Event	30	30	a	a	1	37	2	0	0	2	41	0	0	0	0
24	Location	4	4	a	a	1	32	2	0	0	7	41	0	0	0	0
25	Road Design	3	x	a	b	2	0	40	0	0	1	41	0	0	0	0
26	Road Surface	1	x	a	b	2	3	37	0	0	1	40	0	0	1	0
27	Road Characteristics/Circumstances	5	x	a	b	2	1	40	0	0	0	41	0	0	0	0
28	Device - traffic ctrl	98	98	e	e	1	38	2	0	0	1	2	0	0	1	38
29	Working - traffic ctrl	98	98	e	e	1	38	3	0	0	0	3	0	0	0	38
30	Int Rel - intersection relationship	1	1	a	a	1	38	2	0	0	1	41	0	0	0	0
31	Speed Limit	65	x	a	b	2	0	41	0	0	0	41	0	0	0	0
32	Weather 1	1	x	a	b	2	1	40	0	0	0	41	0	0	0	0
33	Weather 2	x	x	b	b	4	0	6	0	35	0	6	35	0	0	0
34	Light	1	x	a	b	2	0	41	0	0	0	41	0	0	0	0
35	Diagram	7	98	a	e	5	1	0	0	0	40	36	0	0	4	1

Figure 2.2 Screenshot of a Single-Unit Codebook Results

2.3.2 Standards and Strategies for Coding Narratives

Researchers wanted to control for bias in entering information into the codebooks. To minimize bias, researchers first copied and entered in all of the form items from the corresponding crash reports into the codebooks for all strata and years, ignoring any narratives. All fields were double-checked after the first data entry to limit data entry errors. After all the data from the crash form fields were entered, the narratives for each crash report were then blindly coded and double-checked by researchers, without reference to the form fields. In order to code the narratives, researchers took the information written in the narrative and, based solely on what was explicitly written, filled in any entry items using what was written in the narrative by the officers. This information was entered in a separate excel sheet and then copied and pasted over. This was done to ensure that while coding the narratives, the researchers could not see, and potentially be influenced by, what they had already entered for each individual entry item from the form side.

2.3.3 Inter-Rater Reliability

To further ensure that there was as minimal bias and as few entry mistakes as possible, a second coder was used for a portion of both the 2015 and 2016 data. The second coder went through 1/3 of all crash reports from both 2015 and 2016. Their results were then compared to the results of the primary coder. If there were differences in results, the two coders went through the specific report and came to a mutually agreed-upon result.

2.4 RESULTS

For these qualitative analyses, the research team's goal was to validate the completeness and accuracy of the data entered into the form fields against what was written in the narrative section of each report. Six focus areas of the data were selected for the analysis:

1) Events prior to the crash;
2) Person factors contributed to the crash;
3) Roadway and environmental factors associated with the crash;
4) Crash relative to the roadway;
5) Unit condition after the crash; and,
6) Condition of vehicle(s) occupant(s) after the crash.

For this purpose, only form items that were directly associated with these focus areas were included in the analyses. It is important to note when interpreting the results of the analyses that the new MNCrash report contained far more crash-related form items, in compliance with MMUCC 4, compared to fewer items in the 2015 legacy system. This is pertinent when examining the results for the percent of missing form items per year. In some cases, there was a slight increase in the amount of missing data from 2015 to 2016; however, on a per crash report basis, there was still an increase in the number of reported form items, which may provide a more complete picture of what took place in each reported crash. In the 360 crash reports that the research team examined for each year, there were 914 more crash-related form items reported in 2016 relative to 2015.

2.4.1 Crash Severity Comparison

In the first year-to-year comparison, data validity and completeness by injury severity were examined. The two primary factors examined for each analysis were mismatching data and missing data.

2.4.1.1 Overall Missing Information Based on Crash Severity

The amount of missing form information by reporting year for the serious injury crashes (A) was examined. There was a 2.4% increase in missing form information for 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, N= 7939) = 18.92, $p < 0.0001$. However, as shown in Table 2.2, despite the increase in overall missing data, there were 607 more crash-related form items reported in 2016 (285 reports examined) relative to 2015 (285 reports examined) due to the increased number of crash-related form items in 2016.

Table 2.2 Serious Injury Missing Form Data for All Agencies

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 285 reports	3,413 (94.94%)	182 (5.06%)	3,595
2016 – 285 reports	4,020 (92.54%)	324 (7.46%)	4,344

Similarly, the amount of missing form information by reporting year for fatal injury crashes (K) was examined. There was a 4.06% increase in missing form information for 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, *N*= 2171) = 21.22, $p < 0.0001$. Again, as shown in Table 2.3, despite the increase in overall missing data, there were 114 more crash-related form items reported in 2016 (75 reports examined) relative to 2015 (75 reports examined) due to the increased number of crash-related form items in 2016.

Table 2.3 Fatal Injury Missing Form Data for All Agencies

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 75 reports	981 (97.81%)	22 (2.19%)	1,003
2016 – 75 reports	1,095 (93.75%)	73 (6.25%)	1,168

2.4.1.2 Overall Mismatching Information Based on Crash Severity

The amount of mismatching form information by reporting year for serious injury crashes (A) was examined. There was a 4.72% decrease in the amount of mismatching form information for 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, *N*= 7421) = 27.06, $p < 0.0001$. As shown in Table 2.4, the decrease in mismatching form items resulted in 57 fewer crash-related form items being inaccurately reported in 2016 (285 reports examined) relative to 2015 (285 reports examined).

Table 2.4 Serious Injury Mismatching Data for All Agencies

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 285 reports	2,689 (78.79%)	724 (21.21%)	3,413
2016 – 285 reports	3,353 (83.41%)	667 (16.59%)	4,020

Similarly, the amount of mismatching form information by reporting year for fatal injury crashes (K) was examined. There was a slight, 2.02%, decrease from 2015 to 2016 in mismatching information for fatal injury accidents. A chi-square test of independence was performed which found the relationship between these variables to be not statistically significant, χ^2 (1, *N*= 2076) = 1.67, $p = 0.19$. As shown in Table 2.5, the slight decrease in mismatching form items resulted in 4 fewer crash-related form items being inaccurately reported in 2016 (75 reports examined) relative to 2015 (75 reports examined), despite the pronounced increase in required reportable crash-related form items in 2016.

Table 2.5 Fatal Injury Mismatching Data for All Agencies

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 75 reports	825 (84.10%)	156 (15.90%)	981
2016 – 75 reports	943 (86.12%)	152 (13.88%)	1,095

2.4.2 Comparison of Reporting Agencies

The next data strata examined was year-to-year differences among the three reporting agency types: 1) Minnesota State Patrol (MSP), 2) agencies that reported 500 crashes or more per year, and 3) agencies that reported fewer than 500 crashes per year.

Missing Information

The amount of missing form information inputted by the Minnesota State Patrol (MSP) was examined by reporting year for serious injury (A) crashes. There was a 3.28% increase in missing form information for State Patrol in 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, N= 2244) = 9.24, $p < 0.01$. However, as shown in Table 2.6, despite the increase in overall missing data, there was a greater capture of data per crash report in 2016 (i.e., 14.4 items per report) compared to 2015 (i.e., 9.2 items per report) due to the increased number of required crash-related form items in 2016.

Table 2.6 State Patrol Serious Injury Missing Form Data

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 87 reports	785 (95.62%)	36 (4.38%)	821
2016 – 91 reports	1,314 (92.34%)	109 (7.66%)	1,423

Similarly, the amount of missing form information inputted by MSP was examined by reporting year for fatal injury (K) crashes. There was a 6.42% increase in missing form information for MSP in 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, N= 798) = 18.44, $p < 0.0001$. However, as shown in Table 2.7, despite the increase in overall missing data, there were 54 more crash-related form items reported in 2016 (25 reports examined) relative to 2015 (26 reports examined) due to the increased number of crash-related form items in 2016.

Table 2.7 State Patrol Fatal Injury Missing Form Data

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 25 reports	356 (98.89%)	4 (1.11%)	360
2016 – 26 reports	405 (92.47%)	33 (7.66%)	438

Mismatching Information

The amount of mismatching form information inputted by the Minnesota State Patrol (MSP) was examined by reporting year for serious injury (A) crashes. There was a 1.62% decrease in mismatching form information for State Patrol in 2016 compared to 2015. As shown in Table 2.8, the decrease in mismatching form items resulted in 65 fewer crash-related form items being inaccurately reported in 2016 (91 reports examined) relative to 2015 (87 reports examined). A chi-square test of independence

was performed and found the relationship between these variables to not be statistically significant, χ^2 (1, N= 2095) = 0.99, p = 0.31.

Table 2.8 State Patrol Serious Injury Mismatching Form Data

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 87 reports	656 (83.57%)	129 (16.43%)	785
2016 – 91 reports	1,120 (85.24%)	194 (14.76%)	1,314

Similarly, the amount of mismatching form information inputted by the Minnesota State Patrol (MSP) was examined by reporting year for fatal injury (K) crashes. There was a 0.57% decrease in mismatching form information for State Patrol in 2016 compared to 2015. As shown in Table 2.9, the decrease in mismatching form items resulted in only 4 more crash-related form items being accurately reported in 2016 (26 reports examined) relative to 2015 (25 reports examined), despite the increase in reportable crash-related form items in 2016. A chi-square test of independence was performed and found the relationship between these variables to not be statistically significant, χ^2 (1, N= 761) = 0.05, p = 0.81.

Table 2.9 State Patrol Fatal Injury Mismatching Data

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 25 reports	310 (87.08%)	46 (12.92%)	356
2016 – 26 reports	355 (87.65%)	50 (12.35%)	405

2.4.2.1 High-Volume Reporting Agency

Missing Information

The amount of missing form information inputted by high-volume reporting agencies was examined by reporting year for serious injury (A) crashes. There was a 2.63% increase in missing form information for high-volume reporting agencies in 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, N= 2845) = 8.59, p < 0.01. However, as shown in Table 2.10, despite the increase in overall missing data, there were 108 more form items reported in 2016 (98 reports examined) relative to 2015 (97 reports examined) due to the increased number of crash-related form items in 2016.

Table 2.10 High-Volume Reporting Agency Serious Injury Missing Form Data

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 98 reports	1,275 (94.87%)	77 (5.13%)	1,344
2016 – 97 reports	1,383 (92.14%)	118 (7.86%)	1,501

The amount of missing form information inputted by high-volume reporting agencies was examined by reporting year for fatal injury (K) crashes. There was a 1.27% increase in missing form information for high-volume reporting agencies in 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to not be statistically significant, χ^2 (1,

N= 661) = 0.59, *p* = .43. However, as shown in Table 2.11, despite the increase in overall missing data, there were 12 more crash-related form items reported in 2016 (25 reports examined) relative to 2015 (24 reports examined) due to the increased number of crash-related form items in 2016.

Table 2.11 High -Volume Reporting Agency Fatal Injury Missing Data

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 25 reports	309 (95.96%)	13 (4.04%)	322
2016 – 24 reports	321 (94.69%)	18 (5.31%)	339

Mismatching Information

The amount of mismatching form information inputted by high-volume reporting agencies was examined by reporting year for serious injury (A) crashes. There was a 5.41% decrease in mismatching form information for high-volume reporting agencies in 2016 compared to 2015. As shown in Table 2.12, the decrease in mismatching form items resulted in 48 fewer crash-related form items being inaccurately reported in 2016 (97 reports examined) relative to 2015 (98 reports examined). A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, *N*= 2654) = 12.17, $p < 0.001$.

Table 2.12 High-Volume Reporting Agency Serious Injury Mismatching Data

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 98 reports	986 (77.33%)	289 (22.67%)	1,275
2016 – 97 reports	1,142 (82.57%)	241 (17.43%)	1,383

The amount of mismatching form information inputted by high-volume reporting agencies was examined by reporting year for fatal injury (K) crashes. There was a 5.95% decrease in mismatching form information for high-volume reporting agencies in 2016 compared to 2015. As shown in Table 2.13, the decrease in mismatching form items resulted in 17 fewer crash-related form items being inaccurately reported in 2016 (285 reports examined) relative to 2015 (285 reports examined). A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, *N*= 630) = 4.51, $p < 0.05$.

Table 2.13 High-Volume Reporting Agency Fatal Injury Mismatching Data

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 25 reports	255 (82.52%)	54 (17.48%)	309
2016 – 24 reports	284 (88.47%)	37 (11.53%)	321

2.4.2.2 Low-Volume Reporting Agency

Missing Information

The amount of missing form information inputted by low-volume reporting agencies was examined by reporting year for serious injury (A) crashes. There was a 1.45% increase in missing form information for low-volume reporting agencies in 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to not be statistically significant, χ^2 (1, *N*= 2850) = 2.60, *p* = 0.11. However, as shown in Table 2.14, despite nearly double the reportable form items in 2016 (100 reports examined) relative to 2015 (97 reports examined), the low-volume reporting agencies still reported 30 fewer total form items.

Table 2.14 Low-Volume Reporting Agency Serious Injury Missing Form Data

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 100 reports	1,353 (94.62%)	77 (5.38%)	1,430
2016 – 97 reports	1,323 (93.17%)	97 (6.83%)	1,420

The amount of missing form information inputted by low-volume reporting agencies was examined by reporting year for fatal injury (K) crashes. There was a 4.07% increase in missing form information for low-volume reporting agencies in 2016 compared to 2015. A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, *N*= 712) = 7.99, *p* < 0.01. However, as shown in Table 2.15, despite the increase in overall missing data, there were 53 more crash-related form items reported in 2016 (25 reports examined) relative to 2015 (24 reports examined) due to the increased number of crash-related form items in 2016.

Table 2.15 Low-Volume Reporting Agency Fatal Injury Missing Data

	Form Items Present	Missing Form Items	Total Reportable Form Items
2015 – 25 reports	316 (98.44%)	5 (1.56%)	321
2016 – 25 reports	369 (94.37%)	22 (5.63%)	391

Mismatching Information

The amount of mismatching form information inputted by low-volume reporting agencies was examined by reporting year for serious injury (A) crashes. There was a 5.26% decrease in mismatching form information for low-volume reporting agencies in 2016 compared to 2015. As shown in Table 2.16, the decrease in mismatching form items resulted in 74 fewer crash-related form items being inaccurately reported in 2016 (97 reports examined) relative to 2015 (100 reports examined). A chi-square test of independence was performed and found the relationship between these variables to be statistically significant, χ^2 (1, *N*= 2672) = 11.51, *p* < 0.001.

Table 2.16 Low-Volume Reporting Agency Serious Injury Mismatching Data

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 100 reports	1,047 (77.38%)	306 (22.62%)	1,353
2016 – 97 reports	1,091 (82.46%)	232 (17.54%)	1,323

The amount of mismatching form information inputted by low-volume reporting agencies was examined by reporting year for fatal injury (K) crashes. There was a 0.10% decrease in mismatching form information for low-volume reporting agencies in 2016 compared to 2015. As shown in Table 2.17, the decrease in mismatching form items resulted in 9 more crash-related form items being inaccurately reported in 2016 (100 reports examined) relative to 2015 (97 reports examined), despite the significant increase in reportable crash-related form items in 2016. A chi-square test of independence was performed and found the relationship between these variables to not be statistically significant, χ^2 (1, N= 685) = 0.01, p = .97.

Table 2.17 Low-Volume Reporting Agency Fatal Injury Mismatching Data

	Matching Form Items	Mismatching Form Items	Total Comparable Form Items
2015 – 25 reports	260 (82.28%)	56 (17.72%)	316
2016 – 25 reports	304 (82.38%)	65 (17.62%)	369

2.4.3 Crash-Type Comparisons

Researchers also stratified the comparison based on crash-type, by examining three separate types of vehicular crashes: 1) Single-unit crashes, 2) Two-unit crashes, and 3) Crashes involving a motor vehicle in transit and a pedestrian/cyclist. The purpose of this was to determine if there were any differences in data accuracy and completeness relative to crash-type. In other words, did some of the inherent differences in filling out a crash report based on the type of crash that occurred have an impact on the quality of the data provided in the report by the officers?

2.4.3.1 Single-Unit Crashes

Single-Unit crashes are those involving only one motor vehicle in transit. Fatal injury single-unit crashes typically did not have a witness or other passengers to provide officers with feedback on what happened prior to the crash. Similarly, serious injury single-unit crashes also often involve a driver who is too incapacitated to provide officers with much information at the scene. Due to these constraints, officers must often assess the scene with little input from those involved to determine what happened. For this comparison, researchers were interested in the differences between agencies for reporting crash types.

Single-Unit Serious Injury Crashes 2015

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) single-unit crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, N= 1534) = 4.81, p = 0.08.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) single-unit crashes in 2015. The relationship between these variables was statistically significant, χ^2 (2, *N*= 1467) = 32.91, $p <$ 0.0001, with State Patrol having significantly fewer mismatching form items relative to the high and low-volume reporting agencies.

Single-Unit Fatal Injury Crashes 2015

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) single-unit crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 372) = 5.04, p = 0.08.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) single-unit crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 363) = 3.77, p = 0.15.

Single-Unit Serious Injury Crashes 2016

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) single-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 2468) = 2.70, p = 0.25.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) single-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 2286) = 2.65, p = 0.26.

Single-Unit Fatal Injury Crashes 2016

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) single-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 708) = 1.99, p = 0.36.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for fatal injury (K) single-unit crashes in 2016. The relationship between these variables was statistically significant, χ^2 (2, *N*= 675) = 6.84, $p <$ 0.03, with the high-volume agencies having significantly fewer mismatching form items in their single-unit fatal crash reports in 2016.

2.4.3.2 Two-Unit Crashes

Two-unit crashes are crashes involving two motor vehicles in transit. Any crash report involving a motor vehicle in transit that hit a parked vehicle were not included in this analysis. These reports were of

interest to the research team as they require officers to fill out twice as many fields as the single-unit crash reports, yet there are often fewer form items listed for several different categories, such as Sequence of Events. Since there are two drivers and two vehicles, it requires officers to pay close attention when filling out the form and narrative as to which driver was in which vehicle and what each of them was doing prior to the crash. For this comparison, researchers were interested in the differences between agencies for reporting crash types.

Two-Unit Serious Injury Crashes 2015

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) two-unit crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 1597) = 0.14, *p* = 0.92.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) two-unit crashes in 2015. The relationship between these variables was statistically significant, χ^2 (2, *N*= 1546) = 18.12, *p* < 0.0001, with State Patrol having significantly fewer mismatching form items relative to the high and low-volume reporting agencies.

Two-Unit Fatal Injury Crashes 2015

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) two-unit crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 512) = 0.98, *p* = 0.61.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) two-unit crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 504) = 1.09, *p* = 0.15.

Two-Unit Serious Injury Crashes 2016

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) two-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 3541) = 1.37, *p* = 0.50.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) two-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 3385) = 4.82, *p* = 0.08.

Two-Unit Fatal Injury Crashes 2016

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) two-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 1048) = 2.67, *p* = 0.26.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for fatal injury (K) two-unit crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 995) = 1.97, *p* = 0.37.

2.4.3.3 Pedestrian/Cyclist Crashes

This category focused on crashes involving one motor vehicle in transit colliding with either a pedestrian or cyclist. Similar to the two-unit crash reports, this requires officers to fill out a second form and keep track of more information when writing the narrative. This also requires officers to fill out a section of the form that is specific to pedestrians or cyclists; however, these types of crashes are far less common and thus officers are often not likely to have had as much experience composing reports of this nature. For this comparison, researchers were interested in the differences between agencies for reporting crash types.

Pedestrian/Cyclist Serious Injury Crashes 2015

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2015. The relationship between these variables was statistically significant, χ^2 (2, *N*= 756) = 15.51, *p* < 0.001, with State Patrol having significantly fewer mismatching form items relative to the high and low-volume reporting agencies; however, this may be due to State Patrol responding to relatively fewer pedestrian/cyclist incidents than other agencies.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2015. The relationship between these variables was statistically significant, χ^2 (2, *N*= 692) = 6.12, *p* < 0.05, with State Patrol having significantly fewer mismatching form items relative to the high and low-volume reporting agencies, however, this may be due to State Patrol responding to relatively fewer pedestrian/cyclist incidents than other agencies.

Pedestrian/Cyclist Fatal Injury Crashes 2015

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2015. The relationship between these variables was statistically significant, χ^2 (2, *N*= 212) = 9.54, *p* < 0.01, with the low-volume reporting agencies having significantly fewer mismatching form items relative to the high and low-volume reporting agencies.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2015. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 207) = 0.97, *p* = 0.61.

Pedestrian/Cyclist Serious Injury Crashes 2016

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 956) = 5.18, *p* = 0.07.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 930) = 0.05, *p* = 0.97.

Pedestrian/Cyclist Fatal Injury Crashes 2016

A chi-square test of independence was performed to examine the relationship between the amount of missing form information and the reporting agency for serious injury (A) pedestrian/cyclist crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 312) = 0.45, *p* = 0.79.

A chi-square test of independence was performed to examine the relationship between the amount of mismatching form information and the reporting agency for fatal injury (K) pedestrian/cyclist crashes in 2016. The relationship between these variables was not statistically significant, χ^2 (2, *N*= 305) = 1.25, *p* = 0.53.

2.5 CONCLUSION

In general, there was a year-to-year improvement in data accuracy, as measured by the percent of mismatching information, but a decrease in data completeness, as measured by the percent of missing form information. However, the missing form data that is presented here is a percent of missing data relative to all available form items. The MNCrash reporting system that was implemented in 2016 to replace the legacy system used in 2015 has nearly twice as many form items for officers to potentially fill out for each crash report. This means that while there were slight percentage increases in the amount of missing data in 2016, proportionate to the number of form items, the 2016 data analysis saw officers filling out and providing much more form information relative to the previous year. The two form items that had the greatest amount of missing and mismatching information for both years were the Contributing Factors and Sequence of Events. While Sequence of Events did not gain more slots on the form, Contributing Factors did and was representative of a significant percentage of the missing and mismatching data for the 2016 analysis. This increase in the number of possible form items that officers

could have filled out in 2016, along with the four slots provided for Sequence of Events, but often opted not to fill out entirely, represented the majority of the missing and mismatching data for 2016.

State Patrol tended to provide the most complete data relative to the high and low-volume reporting agencies in 2015, χ^2 (2, *N*= 3887) = 15.26, p < .001 or fatal injury (K) crashes χ^2 (2, *N*= 1096) = 6.12, p < 0.05. This would seem to indicate that familiarity with reporting crashes may play a role in data accuracy and completeness. However, in 2016, there were no significant differences between reporting agencies for serious injury (A) χ^2 (2, *N*= 6965) = 5.24, p = .07, or fatal injury (K) crashes χ^2 (2, *N*= 2068) = 0.95, p = .62. In general, while State Patrol did provide the most complete data, there was a significant increase for the high and low-volume reporting agencies from year-to-year.

Overall, despite a higher percentage of missing form data, the 2016 MNCrash reporting system was found to provide more data overall per crash relative to the 2015 legacy system and the data provided was also shown to be significantly more accurate based on the narratives that the officers composed within the report. Table 2.18 and Table 2.19 show the results of all comparable form items from both the serious injury (A) and fatal injury (K) crash reports that were analyzed for this comparative analysis. The 2016 crash reports, from all agencies, had a higher overall amount of data accuracy and completeness relative to their 2015 counterparts, which resulted in an average of 2.45 more crash-related form items being accurately reported per crash report.

Table 2.18 Summary of all Missing Form Data Analyzed by Reporting Year

	Total Comparable Form Items	Total Form Items Present	Total Missing Form Items	Percent of Overall Missing Data
2015 - 360 reports	4,598	4,394	204	4.44%
2016 - 360 reports	5,512	5,115	397	7.20%

Table 2.19 Summary of all Mismatching Form Data by Reporting Year

	Total Matching Form Items	Total Mismatching Form Items	Percent of Mismatching Form Items
2015 - 360 reports	3,514	880	20.02%
2016 - 360 reports	4,296	819	16.01%

This is an important finding to contrast the data's missing elements against the overall informational gains of the additional form items added in the MNCrash redesign and update process. Overall, the analysis results suggest that the user-centered design process that included an intuitive data entry interface and multiple automated components to reduce unnecessary data entry helped to capture a more complete and accurate picture of crashes in Minnesota while supporting law enforcement officers. This suggests that adding additional information queries does not necessarily increase the burden on the users nor result in reduce data entry effort; however, system designers should always be mindful of any disproportionate demands on users as new queries are created. All data queries should be fully inspected to ensure that the information will actually be analyzed and that it cannot be captured through an automated process.

CHAPTER 3: QUALITATIVE DATA ANALYSIS

3.1 INTRODUCTION

The first goal of the qualitative data analysis was to examine the completeness of the crash data by comparing the frequencies of missing, unknown, other, and not applicable responses between the legacy and MNCrash reporting systems. The second goal of the qualitative data analysis was to compare the number of system logic errors and user logic errors in the crash reports between the MNCrash and legacy reporting systems. System logic errors occur when the reporting system does not automatically fill in a specified value when appropriate. For example, the reporting system does not enter "not applicable" for a follow-up question, when the user entered "no" for the lead-in question. User logic errors occur when the reporting officer enters inconsistent values for form item pairs, and these errors are not flagged by the reporting system or are not subsequently changed by the reporting officer, despite the error warning.

The datasets were analyzed for differences in the proportions of missing data values between the MNCrash reports (2016 data) and the legacy reports (2015 data). In addition, cross-tabulations (two-dimensional frequency tables) were used to identify internal inconsistencies between pairs of variables. Proportions of inconsistent values were then compared between 2015 and 2016. Chi-square analyses were tested for statistical differences. Analysis of variance and means separation tests were employed to determine differences in the distributions of continuous variables between years. All analyses were stratified by agency type and crash type (K versus A crashes) to help identify agencies and crashes where the MNCrash reporting system had resulted in the greatest improvements, if any, in data recording.

3.2 DATA PREPARATION

Several considerations were taken to ensure the data was handled and selected to meaningfully and accurately indicate any possible changes to the completeness of the data with the implementation of the MNCrash system. The Minnesota Department of Public Safety provided the research team data for each year related to crash, vehicle, and person level data in the form of three text files. The researchers converted Minnesota Department of Public Safety (MnDPS) fixed-width files to readable Stata files using data dictionary conversion syntax, and the data was cleaned for use in Stata. For example, converted 2015 'Z' codes and double zeros to missing. (These had been coded as "left blank" in the 2015 codebook.) Variable labels and values were encoded based on the codebooks provided by MnDPS.

3.2.1 Data Reduction

Analyses were limited to crashes with a maximum of two units that involved serious injury (A) and fatalities (K) with a reporting date from April to December. Crashes with three or more units were excluded due to their rarity and increased complexity for analysis. A and K injury level crashes were selected because they necessitate accuracy due to the influence they have on driving-related policy and are expected to receive greater officer attention compared to lower injury or property damage only

crashes. This sample also eliminated a time of expected initial learning errors from January to March 2016 in the new MNCrash system to allow an equitable comparison to the data collected in the long-standing legacy system, of which few learning errors would be expected. Data from 2015 contained some errors that needed further attention: There were 16 crashes that were reported to have only one or two motor vehicles (per the 'nummv' variable), but had three or more entries in the vehicle dataset indicating more vehicles involved in the crash than initially reported. These observations were deleted from the analytical sample. A single crash could have fit more than one exclusion criteria (see Table 3.1). In total, 73,776 crashes were eliminated from the 2015 data and 86,271 crashes were eliminated from the 2016 data that did not meet all three inclusion criteria. Below, Figure 3.1 shows the number of crashes that were eliminated based on each criterion.

Table 3.1 Crashes Independently Meeting the Audit Eligibility Criteria (with Overlap)

	2015	2016
Starting Number Crashes	74,772	84,873
Crashes matching criteria:		
1 or 2 motor vehicles	68,637	80,607
January - March report month	19,191	21,271
Serious injury or fatality	1,307	2,056

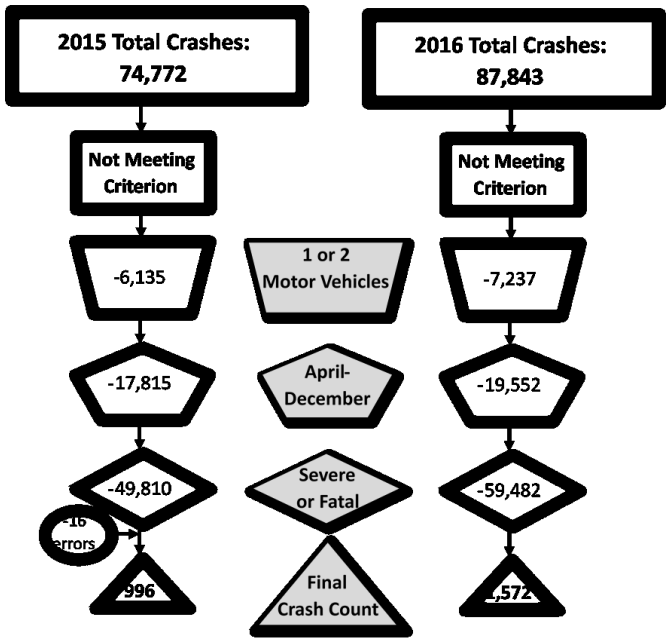


Figure 3.1 Total Crash Observations Eliminated by Audit Eligibility Criteria

Each year had a variable number of vehicles and persons involved in each crash. Overall crashes and persons documented in 2016 was higher than in 2015. The cause of this increase was not the focus of this study but may have resulted from several factors relating to an increase in electronic reporting

statewide with the MNCrash system and/or increases in crash trends in general. Table 3.2 shows the resulting number of crashes meeting study eligibility criteria.

Table 3.2 Final Total Eligible Observed Crashes, Persons, and Vehicles

Total Observed	2015	2016
Crashes	996	1,572
Persons	2,260	3,356
Vehicles	1,571	2,198

3.2.2 Data Stratification

Given that agencies reporting fewer crashes interact with the new reporting system less frequently, the research team wanted to be able to examine whether this resulted in different reporting proficiency. As such, the research team created an agency size variable corresponding to the number of crashes agencies reported in 2015 and 2016, independent of the number of workers or municipal size, see Table 3.3. The Minnesota State Patrol (MSP) used a different crash report interface in 2015 from any other agency and thus was given its own category for analysis. All other agencies were split by whether they reported 500 or more or fewer than 500 crashes in the year. Very few agencies switched agency size category from 2015 to 2016 (Seven agencies [5008, 5043, 5098, 5135, 5249, 5366, and 5403] increased crash reporting from 2015 sufficiently to be reclassified in the >500 category in 2016.) Those that changed agency size category from 2015 were put into the appropriate category for 2016 analyses.

Table 3.3 Observed Number of Eligible Crashes, Persons, and Vehicles by Agency Type

	2015			2016		
Agency	**Crashes**	**Persons**	**Vehicles**	**Crashes**	**Persons**	**Vehicles**
MSP	330	823	523	374	905	573
500+	208	507	367	414	901	577
< 500	458	930	681	784	1,550	1,048
Total	**996**	**2,260**	**1,571**	**1,572**	**3,356**	**2,198**

Each year's separate data files were merged in a many-to-one process by crash ID to examine each year's crash reporting level data with the corresponding vehicle and person-level data. This resulted in three main analytical files each year – crash level only, crash and vehicle data, and crash and person data. These combined datasets ensured that the research team only analyzed the crashes that fit the criteria outlined above and that they could evaluate how crash level data corresponded to the other types of data. Maintaining three datasets instead of combining into a single, all-merged dataset ensured that no duplicate persons or vehicles were included. Three key variables (shown in Table 3.4) were used in linking the crash files together.

Table 3.4 Key Variables Used from Crash Data to Link Files Together

Variable	Description
1. Crash Number (ACCN)	This is present in the data file for every crash, every vehicle, and every person, and is the primary way to match up the records across the three files.
2. Relative Vehicle Number (RVN)	This variable uniquely identifies each unit (vehicle or non-motorist) involved in the crash.
3. Relative Person Number (RPN)	This variable identifies each person within each unit in the crash.

An example of such linkage goes as follows: The ACCN is 123456789. The crash involved two units. The first unit, a passenger vehicle, is identified as part of ACCN 123456789 and is assigned the RVN 1. The passenger vehicle has a driver and two passengers, each of whom is identified as part of ACCN 123456789 and RVN 1. They are numbered RPN 1, 2, and 3 respectively. The second unit, a motorcycle, is also identified as part of ACCN and is assigned the RVN 2. The motorcyclist who was riding alone is identified as part of ACCN and is RVN 2 and is assigned RPN 1 as the first and only person in the vehicle.

3.3 DATA ANALYSES

Two of the most important qualities for crash data are completeness and accuracy. The research team selected a series of analyses to determine if there were any significant and meaningful changes in these two qualities with the implementation of the MNCrash system in 2016. Completeness analyses focused on missing data from year to year. It also focused on data that lacked specificity, that is, data marked as 'unknown' or 'other' or when the circumstances did not merit the inclusion of data, i.e., 'not applicable', see Table 3.5.

Table 3.5 Data Attribute Definitions for Completeness Analysis

Attribute	Definition
Missing	A missing observation indicates that the reporting officer did not select or input anything for a particular data point. In this case, it is unknown whether it was an unintentional omission or if the circumstances would have best fit one of the other data attributes provided.
Unknown	Indicates that the officer is aware of what information is requested but does not have the information necessary to supply it (e.g. driver fled scene and vehicle make unknown).
Other	Selected by officer when the options supplied for a given data point type do not apply. In some items, officers were provided an open text entry field to input their own attribute to specify which 'other' had occurred
Not Applicable	Inputted due to conditional logic for data points that are not applicable to the crash (e.g., no work zone-related data if the crash did not occur in a work zone).

Data accuracy was analyzed in terms of the frequency of data coded that was measured as violating known business rules or committing logic errors. This analysis supplemented the qualitative analysis of comparing fields to narrative reports for agreement by analyzing data that may be subject to errors that could be determined based on logic and crash reporting rules.

3.3.1 Analysis Methods

Counts (number of observations meeting the criteria) and proportions of missing data and errors were analyzed in total for 2015 and 2016 and then separately for each variable of interest with chi-square analyses. Missing data was calculated by counting the frequency of non-responses for each variable and then as a total for the year's dataset. Errors were identified in two ways: 1) the code entered did not have a corresponding codebook value and 2) logical discrepancies (e.g., crash reported as not in a work zone, but reported work zone related information other than 'not applicable'). The number of errors were counted and then compiled into one "error" code per variable of interest and for the data overall by year.

Proportions of reported frequencies for like-codes were compared across years to examine whether offering different/more dropdown values resulted in different/better reporting. Most often, "not applicable" items were compared to tell whether there was less utilization of it in 2016 compared to 2015 (i.e., a new dropdown value now existed where before 'other' was necessary). Sometimes language was not the same, so the research team came to a consensus on whether a variable value should be counted as the same across years and therefore was appropriate to compare. For example, "illegal/unsafe speed" was compared as a contributing factor in 2015 to "driver speeding" in 2016. As another example, the 2015 proportional frequency of selecting "overcorrecting" as a contributing factor was compared to the proportional frequency in 2016 of selecting "over-correcting /over-steering" factor.

3.3.2 Overall Data Trends

Without limiting the analysis to only matching variables, there was significantly more missing (Table 3.6Table 3.6) and 'unknown' observations (Table 3.7) in the 2016 data, but significantly fewer 'not applicable' (Table 3.8) and 'other' observations noted (Table 3.9) compared to the 2015 data. However, upon further examination, the significant increase in missing data often does not appear to be due to an excess of omissions by reporting officers, but due to the conditional and skip logic of the MNCrash interface design in 2016. In circumstances where the MNCrash report could eliminate unnecessary questions that were not applicable based on the logic of upstream data entry, variables were skipped or deactivated to reduce the burden on users from having to frequently and unnecessarily select 'not applicable'. This conditional logic by the system should have auto-populated the skipped or deactivated field with 'not applicable', but this analysis indicated that this system function was frequently omitted (see more on this later in section 3.3.3.2 and in Table 3.16-Table 3.18).

Table 3.6 Overall Number of Missing Observations across Data Type and Year

	2015		2016			
Data Level	**Count**	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
All	15257	9.98%	122236	33.72%	30963.89	< 0.001
Crash	4282	7.82%	12595	16.35%	2088.50	< 0.001
Person	5435	9.25%	78847	46.07%	25536.54	< 0.001
Vehicle	5540	14.11%	30794	26.94%	2666.75	< 0.001

Table 3.7 Overall Number of 'Unknown' Observations across Data Type and Year

	2015		2016			
Data Level	**Count**	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
All	1793	1.17%	5109	1.41%	45.36	< 0.001
Crash	80	0.15%	799	1.04%	383.87	< 0.001
Person	1453	2.47%	1988	1.16%	510.17	< 0.001
Vehicle	260	0.66%	2322	2.03%	331.68	< 0.001

Table 3.8 Overall Number of 'Not Applicable' Observations across Data Type and Year

	2015		2016			
Data Level	**Count**	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
All	19757	12.93%	5632	1.55%	29692.56	< 0.001
Crash	4081	7.45%	1543	2.00%	2324.98	< 0.001
Person	9459	16.10%	798	0.47%	25076.79	< 0.001
Vehicle	6217	15.83%	3291	2.88%	8440.29	< 0.001

Table 3.9 Overall Number of 'Other' Observations across Data Type and Year

	2015		2016			
Data Level	**Count**	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
All	763	0.50%	1616	0.45%	6.69	0.010
Crash	286	0.52%	224	0.29%	44.43	< 0.001
Person	279	0.47%	629	0.37%	12.81	< 0.001
Vehicle	198	0.50%	763	0.67%	12.55	< 0.001

3.3.3 Overall Completeness Using Matching Variables

The addition of new form variables in 2016 with the MNCrash system and the discovered issue of conditional skip logic questions, which should have been auto-populated with 'not applicable', but were left blank, required the team to reduce the variables to allow for a fair comparison of data entry completeness by officers in 2015 and 2016. The research team compared counts of 'missing', 'not applicable', 'unknown', and 'other' between 2015 and 2016 for a selected subset of closely equivalent variables (e.g., diagram). Matching variables were limited to the first factor if there was more than one

opportunity to report on the same issue (e.g., event1, weather1) because there may not be a need to report a second factor (e.g., no second weather type if the first is 'clear'). Variables had to be in the same reporting level to reduce comparing one-to-many data sets across years (e.g., a traffic control device was coded in the crash level for 2015, but multiple were coded for each vehicle in 2016). Crash report variables regarding very specific and limited cases were also omitted (e.g., 'waived' for commercial motor vehicle or CMV). In total, there were 60 comparable matched variables at the crash (26), person (17), and vehicle (17) reporting level. Still, different years had differing numbers of crashes, with different numbers of persons and vehicles involved. Hence, total observation opportunities for each year (across crash, person, and vehicle levels) were summed and used as the denominator for the comparison analyses (at the crash, person, and vehicle level). To improve the uniformity of these analyses across years, person-level variables were restricted to only include drivers. Passenger data was analyzed separately and is discussed later (see section 3.3.4).

Crash reports in 2016 had statistically significantly fewer 'missing' ($p < 0.001$), 'not applicable' ($p < 0.001$), and 'other' ($p = 0.002$) selected, see Table 3.1. 'Not applicable' had the largest drop, from 4.65% to 1.61% of observations. There was a slight increase in the number of 'unknown' from 2015 (0.42%) to 2016 (0.64%), which was also statistically significant, likely due to the large numbers involved (χ^2 value 39.56, $p < 0.001$).

Table 3.10 Comparison of Overall Completeness by Year for Matched Variables

	2015		2016			
	Count	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
Missing	3471	4.54%	4134	3.60%	109.63	< 0.001
NA	3554	4.65%	1852	1.61%	1549.72	< 0.001
Unknown	323	0.42%	737	0.64%	39.56	< 0.001
Other	767	1.00%	1009	0.88%	9.30	0.002

3.3.3.1 Overall Completeness Using Matching Variables by Agency Size

The research team compared counts of 'missing', 'not applicable', 'unknown', and 'other' options selected by the MSP, agencies reporting 500 or more crashes, and agencies reporting fewer than 500 crashes within 2015 and 2016 for a selected subset of matching variables. Table 3.11 outlines the total number of crashes for each agency type in each year and indicates the additional number of vehicles and persons documented (note, as stated above, persons are limited to drivers only). The total number of variables documented within each level have been listed in Table 3.12 by agency type and year.

Table 3.11 Observations by Reporting Level and Agency Size by Year for Matched Variables

	2015				2016			
Report Level	**MSP**	**500+**	**< 500**	**Total**	**MSP**	**500+**	**< 500**	**Total**
Crash	330	208	458	**996**	374	414	784	**1572**
Person*	506	288	606	**1400**	577	550	1034	**2161**
Vehicle	523	367	681	**1571**	573	577	1048	**2198**

Table 3.12 Total Matched Observations Possible (Variables*Observations) Per Report Level

	2015				2016			
Report Level	**MSP**	**500+**	**< 500**	**Total**	**MSP**	**500+**	**< 500**	**Total**
Crash	8580	5408	11908	**25896**	9724	10764	20384	**40872**
Person*	8602	4896	10302	**23800**	9809	9350	17578	**36737**
Vehicle	8891	6239	11577	**26707**	9741	9809	17816	**37366**
Total	**26073**	**16543**	**33787**	**76403**	**29274**	**29889**	**55727**	**114890**

*Person variables restricted to drivers only.

In general, there was year-over-year improvement in terms of completeness in the crash reports. MSP typically does better compared to the other agency groups in terms of completeness - they have fewer missing, unknown, and other observations in both 2015 and 2016, see Table 3.2. However, they have slightly more not-applicable observations in 2016 compared to the other collective agencies. MSP had a statistically significant increase in missing data from 2015 to 2016. It is important to note that the 2015 interface used by MSP was more rigid than what other agencies used in 2015 and likely far more restrictive for submission of forms containing missing values. While the increase in MSP missing is significant (see Table 3.13), it is likely not a meaningful increase (i.e., an increase of less than 1%). Still, those agencies reporting 500+ and less than 500 (see Table 3.15) have, in most cases, substantially improved from 2015 to 2016 in terms of reducing the number missing, not-applicable, and unknown selections. Agencies that reported greater than 500 crashes a year had the largest decrease from 2015 to 2016, 7.95% to 3.75%, respectively (see Table 3.14). 'Other' selections have slightly increased for all agencies between 2015 and 2016.

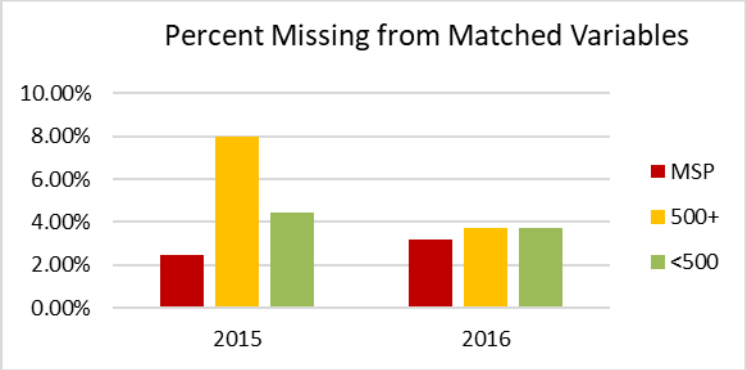


Figure 3.2 Average percent of missing data from matched variable by year and agency type. All year-to-year differences are statistically significant ($p < 0.001$).

Table 3.13 Minnesota State Patrol, All Matched Variables/Observations Combined*

	2015		2016			
	Count	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
Missing	644	2.47%	934	3.19%	25.85	< 0.001
NA	749	2.87%	548	1.87%	60.35	< 0.001
Unknown	203	0.78%	186	0.64%	4.05	0.044
Other	62	0.24%	97	0.33%	4.21	0.040

Table 3.14 Agencies reporting 500+ Crashes, All Matched Variables/Observations Combined*

	2015		2016			
	Count	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
Missing	1315	7.95%	1122	3.75%	378.03	< 0.001
NA	1151	6.96%	450	1.50%	952.46	< 0.001
Unknown	234	1.41%	317	1.06%	11.46	0.001
Other	80	0.48%	205	0.69%	7.10	0.008

Table 3.15 Agencies reporting < 500 Crashes, All Matched Variables/Observations Combined*

	2015		2016			
	Count	**%**	**Count**	**%**	**χ^2 Value**	**p-value**
Missing	1512	4.48%	2078	3.73%	30.73	< 0.001
NA	1654	4.90%	854	1.53%	875.00	< 0.001
Unknown	330	0.98%	506	0.91%	1.10	0.294
Other	181	0.54%	435	0.78%	18.37	< 0.001

*Person variables restricted to drivers only

3.3.3.2 Not Applicable System Logic Issues in 2016

There was a pronounced increase in missingness in 2016 that was not due to officer reporting errors, but backend system logic that was not adequately implemented. Specifically, certain upstream variable selection items were designed to trigger an automatic selection of 'not applicable' (NA) for the appropriate downstream corollary items. For example, in 2016 the 'diagram' item, manner of collision, was supposed to be auto-filled with NA for crashes involving a single unit or motor vehicle (MV). Instead, this item was only automatically hidden from user view, but the field remained empty or with 'missing' recorded for the crash. Because of this error type, 2016 appeared to look worse in terms of performance for certain items in comparison to 2015. Other examples that were not auto-populated with 'NA' correctly were work zone-related variables and 'eject' (i.e., conditional skip logic based on motorcycles). Discussion of the other errors in this output is detailed in Design Example 3. Configuration of Collision Impact (see section 3.3.3). In 2015, the onus would have been on officers to select 'NA' when variables were not applicable to the crash, increasing their time and effort, and provided an opportunity for the officer to erroneously provide a crash configuration on a single unit crash (see Table 3.16). This analysis indicated that officers entered a crash configuration for a single unit crash 1,398 times in 2015. The conditional skip logic implemented in the 2016 MNCrash system alleviated these unnecessary steps by reducing the potential for error. Only 36 instances of crash configuration were observed in single unit crashes in 2016 (see Table 3.17)

However, the full process on the backend was not completed to ensure the field indicated 'NA' since the officer did not have the opportunity to select it, so the variable was left as missing (see Table 3.17).

Table 3.16 2015 Crash Configuration (diagram) for Crashes with 1 MV

Crash Configuration	MSP	500+ reported	< 500 reported	Total
Rear-end	2	4	8	14
Sideswipe - same direction	2	5	2	9
Left turn	0	10	15	25
Ran off road - left	50	10	62	122
Right angle	2	43	35	80
Right turn	2	4	6	12
Ran off road - right	60	17	85	162
Head-on	16	38	41	95
Sideswipe - opposing direction	4	0	0	4
Other	29	56	91	176
Not applicable	4	6	20	30
Unknown	0	9	2	11
Missing	0	1	6	7
Total	171	203	373	747

Table 3.17 2016 Manner of Collision (diagram) for Crashes with 1 MV

Manner of Collision	MSP	500+ reported	< 500 reported	Total
Angle	1	0	2	3
Sideswipe – opposing direction	0	1	0	1
Front to rear	0	0	2	2
Other	3	3	3	9
Unknown	0	0	3	3
Missing	161	235	499	895
Total	**165**	**239**	**509**	**913**

A chi-square test of independence was performed to examine the relationship between missing errors in 2016 to missing errors from 2015 for manner of collision. The relationship between these variables was statistically significant, χ^2 (1, N=1,660) = 1560.94, p < 0.0011 (see Table 3.18). Due to this backend logic issue, there were fewer missing errors in 2015 compared to 2016; however, the number of other errors in 2015 were quite high in comparison to 2016.

Table 3.18 Average Missing for Manner of Collision (diagram) for Crashes with 1 MV

	2015	2016
Missing	7 (0.9%)	895 (98.0%)
Selected	740 (99.1%)	18 (2.0%)
Total	**747**	**913**

3.3.4 Person Factors by Driver and Non-Drivers in 2015 and 2016

The 2015 'position' variable was used to indicate whether an observed person was a driver, passenger, or other (e.g., outside a vehicle or driver was not indicated in their vehicle). The 2016 data included a 'person type' variable that readily distinguished the involved persons by driver, passenger, non-motorist (akin to 2015's 'other'), and owner status. Owners were excluded from these analyses to ensure only persons present were included and no persons were included twice. The research team compared the completeness of matching person factors in 2015 and 2016 by position status. Matching person variables examined were physical condition, alcohol test taken and type, ejection status, airbag deployment, and safety equipment (e.g., seatbelt). The new MNCrash system had two variables (options) for safety equipment usage, but only the first was shown here for the comparison. Ejection status had 20.8% of 2016 observations marked as missing compared to 1.8% missing in 2015 (Table 3.19-Table 3.20), while 2016 airbag deployment had 8.4% missing in totality compared to 2.9% in 2015 (see Table 3.21-Table 3.22). In both cases, a large proportion of the missing data is from the 'other' category (i.e., not persons in the vehicle). Still, nearly 17% of drivers in 2016 were missing an ejection status. In contrast, safety equipment usage in 2016 improved, with missing data falling to 0% for all person types in 2016 compared to 1.4% in 2015 (see Table 3.23 – Table 3.24)

Note, there were many more persons reported in 2016 (i.e., 3,356) compared to 2015 (i.e., 2,260), with a bulk of the increase stemming from more passengers denoted overall. This increase is likely due to how the MNCrash system queried passenger information compared to 2015. The MNCrash passenger section was designed to first ask the total number of passengers' present, then it presented field sections to complete for each passenger (see Figure 3.3). This contrasted with the 2015 interface that allowed officers to add each new passenger, but did not ask officers to indicate the overall number of passengers (see Figure 3.4). It was theorized that officers only "added passenger" in past years (i.e., 2015 and prior) when they had compelling information or were motivated to document information for said passenger and frequently omitted this process for passengers by which they felt they had no information to report. Given the passenger count prompt of the MNCrash system, officers may feel compelled to be exact about the total number of passengers present, whereas the previous system does not appear to have presented the information query in this same detailed manner. This suggests that historically Minnesota may have had an underrepresentation of total passengers reported in crashes, even in cases of serious or fatal injury, and is still faced with inadequate documentation regarding driver ejection status (see Table 3.2).

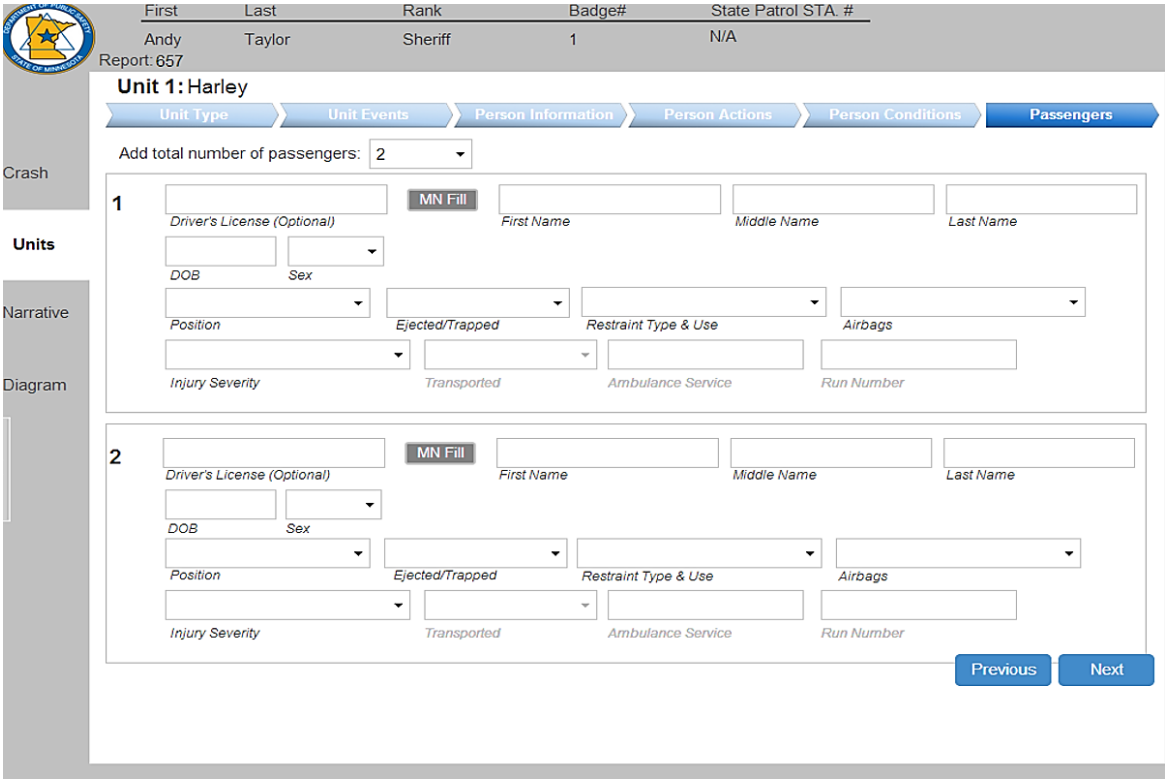


Figure 3.3 Screen image of 2016 MNCrash passenger query design

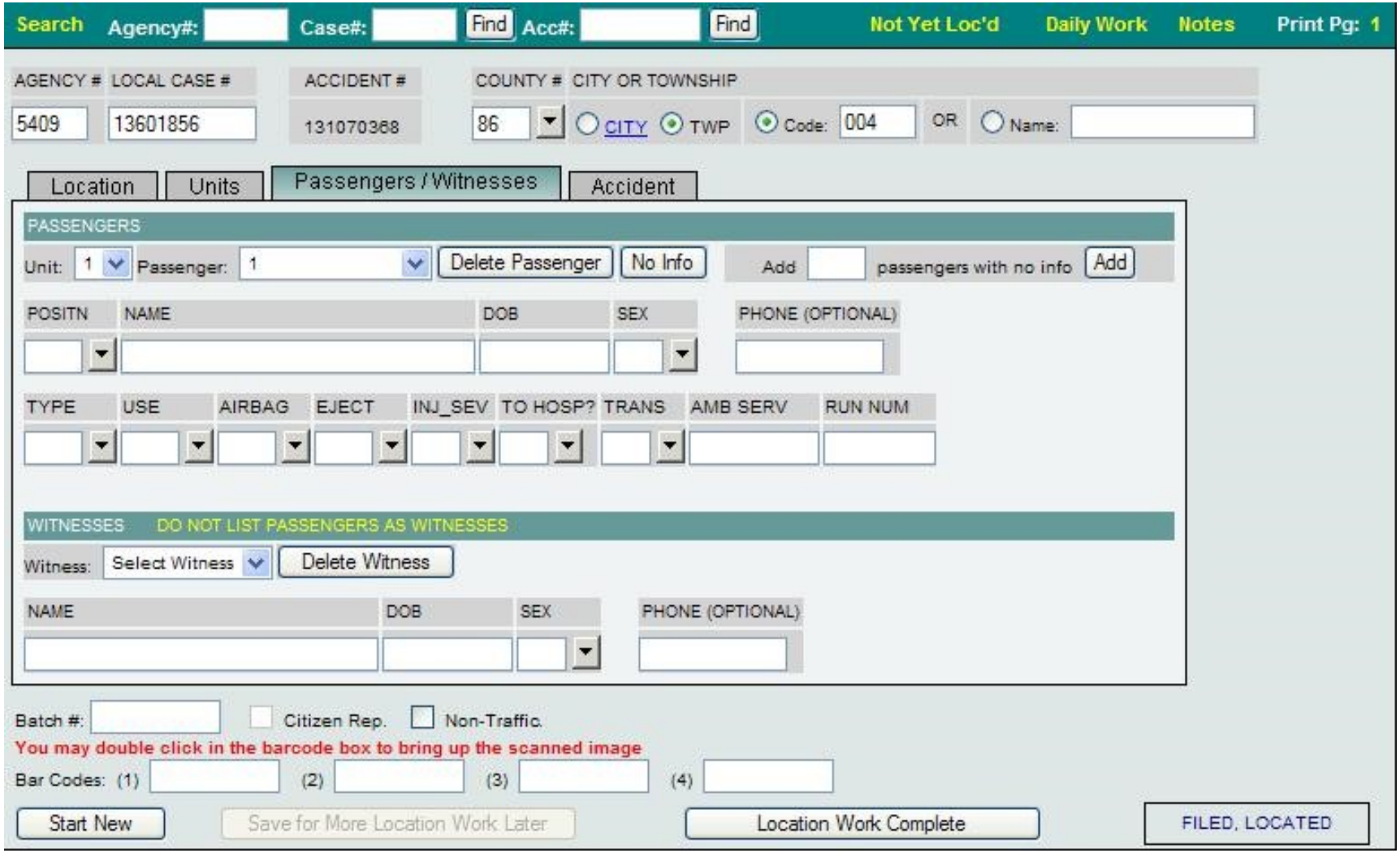


Figure 3.4 Screen image of 2015 passenger query design

Table 3.19 2015 Ejection Status of all Persons by All Agencies

Ejection status	Driver	Column %	Passenger	Column %	Other	Column %	Total	Column %
Trapped - extracted mechanically	101	7.2%	41	6.0%	0	0.0%	142	6.3%
Trapped - extracted manually	41	2.9%	10	1.5%	1	0.6%	52	2.3%
Partially ejected	19	1.4%	4	0.6%	1	0.6%	24	1.1%
Ejected	233	16.6%	69	10.0%	24	14.0%	326	14.4%
Not ejected	897	64.1%	512	74.4%	5	2.9%	1,414	62.6%
Other	12	0.9%	2	0.3%	4	2.3%	18	0.8%
Not applicable	55	3.9%	20	2.9%	135	78.5%	210	9.3%
Unknown	26	1.9%	6	0.9%	1	0.6%	33	1.5%
Missing	16	1.1%	24	3.5%	1	0.6%	41	1.8%
Total	**1,400**		**688**		**172**		**2,260**	

Table 3.20 2016 Ejection Status of all Persons by All Agencies

Ejection status	**Driver**	**Column %**	**Passenger**	**Column %**	**Other**	**Column %**	**Total**	**Column %**
Trapped - extracted mechanically	213	9.9%	61	6.7%	0	0.0%	274	8.2%
Trapped - extracted manually	114	5.3%	53	5.8%	0	0.0%	167	5.0%
Partially ejected	22	1.0%	6	0.7%	0	0.0%	28	0.8%
Ejected	77	3.6%	27	2.9%	0	0.0%	104	3.1%
Not ejected or trapped	1,304	60.3%	673	73.4%	0	0.0%	1,977	58.9%
Not applicable	33	1.5%	17	1.9%	0	0.0%	50	1.5%
Unknown	36	1.7%	22	2.4%	0	0.0%	58	1.7%
Missing	362	16.8%	58	6.3%	278	100.0%	698	20.8%
Total	**2,161**		**917**		**278**		**3,356**	

Table 3.21 Airbag Deployment of all Persons by All Agencies

Airbag deployment	Driver	Column %	Passenger	Column %	Other	Column %	Total	Column %
Deployed - front	373	26.6%	131	19.0%	0	0.0%	504	22.3%
Deployed - side	24	1.7%	29	4.2%	0	0.0%	53	2.3%
Deployed - front side	128	9.1%	60	8.7%	0	0.0%	188	8.3%
Not deployed - switch on	139	9.9%	65	9.4%	0	0.0%	204	9.0%
Not deployed- switch off	8	0.6%	2	0.3%	0	0.0%	10	0.4%
Not deployed - switch unknown	373	26.6%	170	24.7%	6	3.5%	549	24.3%
Other	7	0.5%	4	0.6%	0	0.0%	11	0.5%
Not applicable	273	19.5%	176	25.6%	142	82.6%	591	26.2%
Unknown	54	3.9%	27	3.9%	3	1.7%	84	3.7%
Missing	21	1.5%	24	3.5%	21	12.2%	66	2.9%
Total	**1,400**		**688**		**172**		**2,260**	

Table 3.22 2016 Airbag Deployment of all Persons by All Agencies

Airbag deployment	Driver	Column %	Passenger	Column %	Other	Column %	Total	Column %
Deployed - front	543	25.1%	186	20.3%	0	0.0%	729	21.7%
Deployed - side	28	1.3%	22	2.4%	0	0.0%	50	1.5%
Deployed - curtain	17	0.8%	20	2.2%	0	0.0%	37	1.1%
Deployed - other (air belt, etc.)	3	0.1%	0	0.0%	0	0.0%	3	0.1%
Deployed combination	226	10.5%	95	10.4%	0	0.0%	321	9.6%
Not deployed	805	37.3%	366	39.9%	0	0.0%	1,171	34.9%
Not applicable	439	20.3%	181	19.7%	0	0.0%	620	18.5%
Unknown	100	4.6%	43	4.7%	0	0.0%	143	4.3%
Missing	0	0.0%	4	0.4%	278	100.0%	282	8.4%
Total	**2,161**		**917**		**278**		**3,356**	

Table 3.23 2015 Safety Equipment Use by Driver Status

Safety equipment	Driver	Column %	Passenger	Column %	Other	Column %	Total	Column %
Belts not used	156	11.1%	84	12.2%	4	2.3%	244	10.8%
Lap belt only used	5	0.4%	8	1.2%	0	0.0%	13	0.6%
Shoulder belt only used	8	0.6%	4	0.6%	0	0.0%	12	0.5%
Lap & shoulder belt used	787	56.2%	339	49.3%	0	0.0%	1,126	49.8%
Child seat not used	1	0.1%	2	0.3%	0	0.0%	3	0.1%
Child seat used properly	0	0.0%	63	9.2%	0	0.0%	63	2.8%
Booster seat not used	0	0.0%	5	0.7%	0	0.0%	5	0.2%
Booster seat used improperly	0	0.0%	2	0.3%	0	0.0%	2	0.1%
Booster seat used properly	0	0.0%	11	1.6%	0	0.0%	11	0.5%
Helmet not used	115	8.2%	23	3.3%	12	7.0%	150	6.6%
Helmet used	72	5.1%	13	1.9%	8	4.7%	93	4.1%
Dark clothing	0	0.0%	0	0.0%	5	2.9%	5	0.2%
No protective pads	0	0.0%	0	0.0%	1	0.6%	1	0.0%
Other equipment used	1	0.1%	2	0.3%	0	0.0%	3	0.1%
Not applicable	21	1.5%	27	3.9%	109	63.4%	157	6.9%
Unknown	218	15.6%	91	13.2%	32	18.6%	341	15.1%
Missing	16	1.1%	14	2.0%	1	0.6%	31	1.4%
Total	**1,400**		**688**		**172**		**2,260**	

Table 3.24 2016 Safety Equipment (1) Use by Driver Status

Safety equipment (1)	Driver	Column %	Passenger	Column %	Other	Column %	Total	Column %
None used (MV occupant)	186	8.6%	101	11.0%	0	0.0%	287	8.6%
Lap & shoulder belt used	1,261	58.4%	482	52.6%	0	0.0%	1,743	51.9%
Lap belt only used	8	0.4%	2	0.2%	0	0.0%	10	0.3%
Shoulder belt only used	2	0.1%	0	0.0%	0	0.0%	2	0.1%
Restraint used - unknown	3	0.1%	6	0.7%	0	0.0%	9	0.3%
Child restraint not used	0	0.0%	2	0.2%	0	0.0%	2	0.1%
Child restraint - improper	0	0.0%	1	0.1%	0	0.0%	1	0.0%
Child restraint - rear facing	0	0.0%	14	1.5%	0	0.0%	14	0.4%
Child restraint - forward	0	0.0%	45	4.9%	0	0.0%	45	1.3%
Booster seat used properly	0	0.0%	18	2.0%	0	0.0%	18	0.5%
Child restraint - unknown	0	0.0%	8	0.9%	0	0.0%	8	0.2%
Helmet - DOT-compliance unknown	54	2.5%	5	0.5%	14	5.0%	73	2.2%
No helmet	219	10.1%	44	4.8%	14	5.0%	277	8.3%
Helmet DOT-compliant	93	4.3%	8	0.9%	8	2.9%	109	3.2%
Helmet use unknown	4	0.2%	1	0.1%	2	0.7%	7	0.2%
None	0	0.0%	1	0.1%	187	67.3%	188	5.6%
No protective pads	0	0.0%	0	0.0%	2	0.7%	2	0.1%
Protective pads used (elbow, etc.)	0	0.0%	0	0.0%	1	0.4%	1	0.0%
Dark clothing	0	0.0%	0	0.0%	15	5.4%	15	0.4%
Light clothing	0	0.0%	0	0.0%	3	1.1%	3	0.1%
Reflective clothing	0	0.0%	0	0.0%	2	0.7%	2	0.1%
Lighting	0	0.0%	0	0.0%	1	0.4%	1	0.0%
Other	4	0.2%	2	0.2%	1	0.4%	7	0.2%

Not applicable	17	0.8%	30	3.3%	19	6.8%	66	2.0%
Unknown	310	14.3%	147	16.0%	9	3.2%	466	13.9%
Missing	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Total	**2,161**		**917**		**278**		**3,356**	

3.4 DATA DESIGN ELEMENTS OF INTEREST AND ANALYSIS

The remaining sections of this report are devoted to analyzing possible changes in accuracy and completeness for individual elements within the crash report. These particular elements were known to be problematic in the legacy reporting systems used prior to 2016 or are of interest because of new Model Minimum Uniform Crash Criteria (MMUCC) requirements from the National Highway Traffic Safety Administration. The proportion of errors to non-errors for these elements were compared across years using chi-square analyses. The error tabulation is described in the text and highlighted in the tables with red text and gray background.

3.4.1 Design Example 1. Collision with MV in Transport Errors

The sequence of events within the crash report provides a critical time-series account of the crash events in data elements. The original usability testing of the MNCrash iterative design process revealed that often officers would select "motor vehicle in transit" in instances of a single unit crash, failing to realize that this was an abbreviation of "collision with a motor vehicle in transit". A collision with a motor vehicle (MV) in transit is impossible for single unit collisions because it means hitting a second traveling MV. Thus, any single unit crash reporting of this is an error. This was addressed in a new 2016 design feature of using nested menus, dividing the attributes into three sections: 1) lane departure or non-collision, 2) collision with a non-fixed object, and 3) collision with a fixed object. This helped to emphasize lane departure events to increase their documentation and emphasize the "collision with" component of "collision with a motor vehicle in transit". Instances in which this attribute, "collision with a motor vehicle in transit" was erroneously coded for single unit crashes were analyzed between 2015 and 2016, see Table 3.25 and Table 3.26.

Table 3.25 2015 Collision with MV in Transit in Sequence of Events

	Number of MVs		
Event	**1**	**2**	**Total**
1	249	757	1,006
2	24	50	74
3	11	25	36
4	9	21	30
Total	**293**	**853**	**1,146**

Table 3.26 2016 Collision with MV in Transit in Sequence of Events

	Number of MVs		
Event	**1**	**2**	**Total**
1	60	1,178	1,238
2	5	94	99
3	1	25	26
4	0	17	17
Total	**66**	**1,314**	**1,380**

A chi-square test of independence was performed to examine the relationship between collision with MV in transit reporting errors in the sequence of events and reporting year. The relationship between these variables was statistically significant, χ^2 (1, *N*= 2,526) = 221.83, $p < 0.0011$. There were fewer 'collision with MV in transit' reporting errors in 2016 compared to 2015, see Table 3.27.

Table 3.27 2015 & 2016 Collision with MV in Transit Errors Compared, n (column %)

	2015	2016
Error	293 (25.6%)	66 (4.8%)
Correct	853 (74.4%)	1,314 (95.2%)
Total	**1,146**	**1,380**

3.4.2 Design Example 2. Motorcycle Reporting Errors

Ejection status is not a reportable variable for motorcycles under MMUCC guidance. However, this rule may be unknown by officers and may result in erroneous coding when motorcycle drivers are thrown or fall from their motorcycles in a crash. Even if officers are aware of the rule, it is an unnecessary step to require them to select 'NA' when unit type of 'motorcycle' is known by the system. The MNCrash system in 2016 implemented this skip logic to eliminate the variable from being presented to officers when they indicated the unit type 'motorcycle'. The frequency of officers coding ejection status for motorcycles was examined and found to have fewer instances erroneous ejection status for motorcycles in 2016 compared to 2015, see Table 3.28 and

Table 3.29. Notably, 'missing' was overly represented in 2016 since the conditional skip logic did not auto-populate the field with 'NA' once the variable was hidden.

Table 3.28 2015 Ejection Status for Motorcyclists

	Count	Percent
Trapped, extricated (manual, non-mechanical means)	1	0.4%
Partially ejected	7	2.7%
Ejected	181	69.3%
Not ejected	17	6.5%
Other	10	3.8%
Not applicable	38	14.6%
Unknown	7	2.7%
Total	**261**	**100**

Table 3.29 2016 Ejection Status for Motorcyclists

	Count	Percent
Totally ejected	5	1.3
Not ejected or trapped	11	2.9
Not applicable	1	0.3
Missing	366	95.6
Total	**383**	**100**

The attributes were examined for accuracy. Any non-ejected or partially ejected status was considered correct. A chi-square test of independence was performed to examine the relationship between collision with motorcycle ejection reporting errors and reporting year. The relationship between these variables was statistically significant, χ^2 (1, N= 644) = 369.95, p < 0.0011. There were fewer ejection reporting errors in 2016 compared to 2015, see Table 3.30.

Table 3.30 2015 & 2016 Motorcycle Ejection Errors, n (column %)

	2015	2016
Error	188 (72.0%)	5 (1.31%)
Correct	73 (28.0%)	378 (99.3%)
Total	**261**	**383**

3.4.3 Design Example 3. Configuration of Collision Impact

As discussed in the data completeness analysis and shown in Table 3.15 and Table 3.16, the MNCrash system newly implemented conditional skip logic to prevent officers from erroneously coding crash configuration (diagram) for single-unit crashes. In addition, the 2015 report included non-collision crash configurations in the same field (e.g. ran off road to the right). On the other hand, the 2016 reporting system was designed with logic in place to only show the 'manner of collision' item, if there were more than one motor vehicle involved. As such, the 2016 report had non-collision in a separate category to avoid the confusion seen in 2015, see

Table 3.31 and Table 3.32.

Table 3.31 2015 Collision Impact Errors

	Number of MVs		
Crash Configuration (diagram)	**1**	**2**	**Total**
Rear-end	14	108	122
Sideswipe - same direction	9	34	43
Left turn	25	52	77
Ran off road - left	122	6	128
Right angle	80	342	422

Right turn	12	6	18
Ran off road - right	162	8	170
Head-on	95	160	255
Sideswipe - opposing direction	4	28	32
Other	176	70	246
Not applicable	30	6	36
Unknown	11	2	13
Missing	7	2	9
Total	**747**	**824**	**1,571**

Table 3.32 2016 Collision Impact Errors

	Number of MVs		
Manner of Collision (diagram)	**1**	**2**	**Total**
Angle	3	578	581
Sideswipe - same direction	0	46	46
Sideswipe - opposing direction	1	56	57
Front to rear	2	196	198
Front to front	0	232	232
Rear to side	0	8	8
Other	9	98	107
Unknown	3	4	7
Missing	886	76	962
Total	**904**	**1,294**	**2,198**

A chi-square test of independence was performed to examine the relationship between collision impact/angle of impact and reporting year. The relationship between these variables was statistically significant, χ^2 (1, N=1,651) = 317.71, $p < 0.0011$. There were fewer collision impact reporting errors in 2016 compared to 2015, see Table 3.33.

Table 3.33 2015 & 2016 Collision Impact Errors Compared, n (column %)

	2015	**2016**
Error	144 (32.0%)	6 (0.7%)
Correct	603 (68.0%)	898 (99.3%)
Total	**747**	**904**

3.4.4 Design Example 4. Most Harmful Event Match Errors

In the legacy system, officers had to pick the 'most harmful event' out all of the possible events (all the ones available for each of the four sequence of events), but it had to match one of the four events they already selected. In the MNCrash report, the drop-down menu is only populated with the items that

were selected from the 1-4 sequence of crash events. Any 'most harmful event' selections that do not match one of the four previously selected sequence of events are done erroneously. An analysis of 'most harmful event' selections found that the limited presentation feature in the MNCrash system resulted in fewer errors (i.e., 2 in 2016) compared to 2015 (i.e., 21 in 2015), see Table 3.34.

Table 3.34 2015 & 2016 Fatal (K) Most Harmful Event Match Errors, n (column %)

	2015	2016
Error	21 (1.3%)	2 (0.09%)
Correct	1,550 (98.7%)	2,196 (99.9%)
Total	**1,571**	**2,198**

A chi-square test of independence was performed to examine the relationship between the number of errors made in reporting the most harmful event and reporting year. The relationship between these variables was statistically significant, χ^2 (1, N= 3,769) = 23.44, p < 0.0011. There were significantly fewer 'most harmful event' reporting errors in 2016 compared to 2015.

3.4.5 Design Example 5. Report of Non-Collision Events

The 2015 report had three important non-collision sequence of event options that could be selected: ran off road – right, ran off road – left, and a combined cross median/center lane. These were all placed far down in a single long list of options that could be selected for sequence of events. In 2016, the major rework of the sequence of events that used nested menus that brought non-collision events to the top was implemented, in part, to encourage greater use of the non-collision event options. The research team compared whether all non-collision events were proportionally coded more often as the first event and whether the non-collision options for all events were coded for them more often. Chi-square analysis of proportions demonstrated that reporting of non-collision events (vs. collision events) for the first event was statistically significantly different between 2015 and 2016, χ^2 (1, N= 3,769) = 24.27, p < 0.0011, with non-collision events coded as the first event more frequently, see Table 3.35. The analysis for non-collision coding for all events, however, showed fewer non-collision events for all events in 2016 compared to 2015, χ^2 (1, N= 11,934) = 77.54, p < 0.0011, see Table 3.36.

Table 3.35 Comparison of 1st Event Reporting Non-Collision Events, n (column %)

1st Event	2015	2016	Total
Non-Collision	137 (8.7%)	307 (14.0%)	**444**
Collision	1434 (91.3%)	1891 (86.0%)	**3325**
Total	**1,571**	**2,198**	**3,769**

Table 3.36 Comparison of All Sequence of Events Reporting Non-Collision, n (column %)

All Events	2015	2016	Total
Non-Collision	313 (10.0%)	476 (5.4%)	**789**
Collision	2829 (90.0%)	8316 (94.6%)	**11,145**

Total	**3,142**	**8,792**	**11,934**

3.4.6 Design Example 6. Run Off Road before Hitting Fixed Object

The nested menu structure of sequence of events was also structured in the hopes of encouraging more accurate and complete documentation of run-off-road (ROR) crashes, which comprised a significant amount of rural fatal crashes. In cases of ROR crashes with fixed objects, the sequence of events should indicate that the vehicle crossed the centerline or median and/or go off the road to the right or left before hitting a fixed object. Analytical syntax was developed that noted whether or not any one of these non-collision events (e.g. crossed the median) was inputted in the sequence before a fixed object event was reported.

A chi-square test of independence was performed to examine the relationship between the number of errors made in the order of events and reporting year, see Table 3.37. Although, there is a smaller percentage of errors in 2016, this is not a statistically significant difference, χ^2 (1, N= 831) = 2.61, p = 0.14.

Table 3.37 2015 & 2016 Fixed Object Event Order Errors, n (column %)

	2015	2016
Error	138 (42.5%)	189 (37.4%)
Correct	187 (57.5%)	317 (62.6%)
Total	**325**	**506**

3.5 CONCLUSION

The results of this analysis show a positive outcome for the MNCrash system's efficacy to improve Minnesota crash data completeness and accuracy. The select examination of one and two-unit, serious injury or fatal crashes provides an important analysis of the roadway incidents that are most likely to lead to policy, engineering, and educational efforts both locally and nationally. It is of the utmost importance that this data provides the most complete and accurate picture of the state of crashes and their possible causes to further the pursuit of toward zero deaths.

The results showed that when the backend conditional skip logic issues for missing data were accounted for, the overall completeness of the crash data increased significantly in 2016. Agencies that reported greater than 500 crashes a year experienced the greatest improvement in reducing the amount of missing data in their crash reports with the use of the MNCrash system. Future improvements to auto-population 'NA' to fields involved in conditional skip logic will help to improve overall data completion in the future. The design of the MNCrash system also helped to identify some potential losses in data that the state may be experiencing, relating to information capture regarding passengers and others. The MNCrash system better captured a greater number of passengers present in 2016 compared to 2015.

The targeted design of many of the components of the MNCrash interface showed success in improving data accuracy. Reducing erroneous data entry for single unit crashes and motorcycle crashes were found

through some of the conditional skip logic design of the report. The nested structure of the sequence of events also showed significant improvements in reducing erroneous data entry for single unit crashes, improving accuracy and agreement in most harmful event data capture, and increasing non-collision event documentation for first events and prior to collision with fixed object crashes. The analysis identified some areas for greater improvement to the MNCrash report to further data completeness and accuracy in future design iterations.

CHAPTER 4: CONVERTED DATA ANALYSIS

4.1 DATA PREPARATION

The goal of this analysis was to determine the accuracy of the conversion process by comparing the converted data values to the legacy values. A total of 1,307 crash records in 2015 were selected for conversion and included 2,274 and 3,309 records at the vehicle- and person-level, respectively. The converted data was provided by the Minnesota Department of Public Safety, in addition to a data dictionary, which provided conversion translation notes and labels for output values. To aid in the analyses, the translation notes were used to identify the 'target value' for each of the corresponding legacy values. Variables were selected for analysis across all three report levels (i.e., crash, vehicle, and person) and are described in Table 4.1. These variables were selected by the research team based on frequency of use/presence in data analyses shared with the public in Minnesota Crash Facts.

Table 4.1 Description of legacy variables and corresponding converted variables

Variable Name			Report Level	
Legacy	**Converted**	**Variable Description**	**Legacy**	**Converted**
Acctype	CrashTypeCde	Crash type	Crash	Crash
Event1	EventCde	First crash-related event	Vehicle	Vehicle
Cfct1	FactorCde1	First contributing factor	Vehicle	Person
	PrimaryContributorCde	Primary contributing factor		Vehicle
Vehtype	VehicleTypeCde	Type of vehicle	Vehicle	Vehicle
Sex	GenderCde	Gender	Person	Person
Age	BirthDte	Age on date of crash	Person	Person
Physcnd	PhysicalConditionCde	Person's physical condition	Person	Person
Trfcntl	TrafficControlCde	Traffic control device	Crash	Vehicle
Weather	WeatherCde	Weather Conditions	Crash	Crash
Light	LightConditionCde	Light conditions	Crash	Crash
Rdsurf	SurfaceConditionCde	Road surface condition	Crash	Crash
Rddes	RoadwayAlignmentCde	Road design	Crash	Vehicle
Alctest	AlcoholTestStatusCde	Alcohol tested	Person	Person
Alctype	AlcoholTestTypeCde	Type of alcohol test	Person	Person
Eject	EjectionCde	Person ejected from vehicle	Person	Person
Safeqp	SafetyEquipmentUseCde	Type of safety equipment	Person	Person
Action	VehicleManeuverCde	Pre-crash maneuver	Vehicle	Vehicle
	ManeuverCde	Maneuver for non-motorist		Person
Diagram	MannerOfCollisionCde	Manner of collision	Crash	Crash

4.2 DATA ANALYSIS

Descriptive analyses calculated percentages of records for all combinations of legacy and converted values. Problem combinations were flagged where:

1. The converted value was non-missing and did not match the target value;
2. The converted value was missing, and the legacy value was non-missing;
3. The converted value was non-missing and was not included in the conversion translation notes.

The total number of observations for each type of problem combination was calculated, in addition to the percent among all observations.

Several variables were converted from one report level to another. These included Trfcntl, a crash-level variable that was converted to a vehicle-level variable (TrafficControlCde). In this example, the converted value was replicated to fill all available observations at the lower level (a one-to-many conversion). For descriptive analyses of one-to-many conversions, all replicated values were removed prior to calculating percentages.

For two of the selected variables (Cfct1 and Action), the legacy value was converted to one of two new variables, depending on the original value range. For example, values of Cfct1 less than 21 were converted to FactorCde1 values, and values equal to or greater than 21 were converted to PrimaryContributorCde values. For these split variable conversions, in addition to identifying incorrect or missing values in the converted values, observations where the wrong variable was populated with data, based on the data range of the original variable, were also identified.

4.3 RESULTS

The analysis of the conversion revealed that a vast majority of the legacy data was properly converted based on the translation notes within the data dictionary. Several variables examined in this conversion process were found to have no errors with their target values appropriately matching the transcription notes in the data dictionary. These variables include vehtype to VehicleTypeCde, weather to WeatherCde, light to LightConditionCde, rdsurf to SurfaceConditionCde, action (vehicle-level) to VehicleManeuverCde, and action (person-level) to ManeuverCde. The remaining variable examined was found to have a small percentage of errors within the converted process. A detailed account of the frequency and percent of compliance and errors observed can be found in Appendix B. A summary of the errors is listed below.

4.3.1 Crash type: acctype to CrashTypeCde

One attribute relating to crash type was found to have an erroneous value in the conversion process. ‘CONSTRUCT EQUIP’ (21) was improperly converted to ‘Light pole’ (26) on a single instance instead of its target value of (36). This error accounts for 0.08% of the total data analyzed for this variable.

4.3.2 Sequence of Crash-Related Events: event1 to EventCde

Multiple issues were observed for event1 in the Sequence of Events variable conversion process. First, a single instance of a missing variable was assigned a target value of ‘45’ with no known assigned attribute label. Additionally, four attributes, i.e., ‘RAN OFF RD-RIGHT’ (57), ‘RAN OFF RD-LEFT’ (58), ‘CROSS

MED/CENTRL' (62), and 'NOT APPLICABLE' (98), were not assigned a target value in the transcription notes and left missing in a total of 176 instances. This resulted in 0.04% of the data from this variable erroneously converted and 7.74% of the data missing from the converted dataset.

4.3.3 Contributing Factor: cfct1 to FactorCde1 - Person-Level

Two main issues were observed in the Contributing Factor 1 converted data for cfct1 attributes (1-29 and 99) assigned to the person-level FactorCde1. First, the attributes 'FLNG TOO CLOSELY' (4) and 'DSRGRD TRF CNTRL' (5) were assigned reversed target values, i.e., 'Failure to obey traffic signs, signals, or officer' (5) and 'Following too closely' (4), respectively. These targets were applied in the conversion as the translation notes indicated. It is unknown if the label on the translation notes for the target values are misapplied or if the target values are reverse. Since the exact issue is unknown, this conversion has not been labeled as an error but should be addressed in the translation notes. Second, 104 instances of 'Unknown' (99) were missing from the converted data, which amounts to 5.17% of the data missing for this variable.

4.3.4 Contributing factor: cfct1 to FactorCde1

In correspondence to the new MNCrash data layout, attributes from cfct1 (1-29 and 99) were the only data points that should have been attributed to FactorCde1. Attributes in cfct1 ranging from 31-61, 90, and 98 should have been converted to values at the unit level in PrimaryContributorCde. However, the research team examined the converted data to determine if any cfct1 attributes destined for unit level coding (i.e., 31-42, 46, 50, 61, 90, and 98) were erroneously and duplicitously coded in FactorCde1. The analysis revealed that eight of the cfct1 attributes (i.e., 31, 32, 33, 41, 42, 46, 50, 61, and 98) were present in one of the FactorCde1 target values. The target value for each of the cfct1 attributes was often different. For example, 'SKIDDING' (46) was found to be recoded as 'No clear contributing action' (1), 'Illegal/unsafe speed' (3), 'Driver inexperience' (16), 'Chemical impairment' (18), and 'Over-correcting/over steering' (72). In total, cfct1 vehicle-level attributes were found to be duplicated in FactorCde1 26 times.

4.3.5 Contributing factor: cfct1 to PrimaryContributorCde

The research team examined the converted data of cfct1 attributes designated for unit-level coding (i.e., 31-42, 46, 50, 61, 90, and 98) as they were converted into PrimaryContributorCde. The analysis flagged five instances in which 'OTHER VEHIC FCTR' (50) was missing from the converted dataset. These missing data accounted for 2.99% of the total data examined in this variable conversion process.

4.3.6 Gender: sex to GenderCde

Multiple abnormalities were observed in the gender variable conversion process. Missing data was primarily converted properly; however, nine instances of missing data were applied a value of (9) which was not accounted for in the translation notes. Additionally, 'F=Female' and 'M=Male' were applied to

missing data eight and ten times, respectively. There were two original entries of 'U' in the dataset that were missing from the converted dataset; however, this may be a proper conversion if the entries are deemed erroneous and should be discarded. Finally, 'X=Unknown' was converted to values of (9) in seven instances rather than the target value of (99). In total, 0.82% of the gender data examined contained errors and 0.06% were missing in the converted dataset.

4.3.7 Age: age to BirthDte

The converted age data, as calculated from the difference between the person's date of birth and incident data, was examined. While the vast majority of the ages matched the original dataset or was appropriately missing, limited number of data points were found to have issues. Three instances of the converted data did not match the original dataset. These three incidents occurred on the person's birthday. In original data, the age change, due to the crash occurring on their birthday, was not accounted for, while in the converted data, it was calculated up to and including the birthday. This conversion may not be erroneous, but suggests an instance where a limited sample of the data was previously incorrectly coded (e.g., coding a teen crashing on their 16th birthday as 15 years old) and may shift such driver ages correctly to a year older with the conversion process. Second, 18 instances were observed in which the age data was previously missing but was found to have a birthday present in the converted dataset. The two issues raised account for less than 1% of the total age data examined.

4.3.8 Physical Condition: physcnd to PhysicalConditionCde

Missing data within either dataset, i.e. original or converted, accounted for all issues flagged in the analysis of Physical Condition. Missing data within the original dataset were erroneously assigned 'Apparently normal (including no drugs/alcohol)' (5) or 'Not applicable' (98) in 6 and 20 instances, respectively. Additionally, a majority, of the 'NOT APPLICABLE' (98) original data was found to be missing from the converted data. In total, 798 instances of the (98) values were missing while 56 of them were correctly assigned the target (98) value in the converted dataset. The errors observed in this examination accounted for 0.79% of the total data examined. However, the missing data observed accounted for 24% of the total data examined under this variable.

4.3.9 Traffic Control Device: trfcntrl to TrafficControlCde

The traffic control device data was originally documented at the crash-level, but the conversion process reassigned this data to the unit-level in conformity with the new MNCrash data structure. While the vast majority of this data was appropriately converted, seven instances of data originally coded as 'UNKNOWN' (99) were missing from the converted data. These missing values accounted for 0.54% of the data examined in this analysis.

4.3.10 Road Design: rddes to RoadAlignmentCde

Data pertaining to the road design were similarly re-assigned to the unit-level in the conversion process. Two attributes were observed to have issues with missing data. 'OTHER ROAD DESGN' (90) and 'UNKNOWN RD DESGN' (99) were found to be missing from the converted dataset 32 and 3 times, respectively. Both of these attributes should have been assigned the same target values as their original code value. These missing data points accounted for 2.68% of the total data examine for this variable.

4.3.11 Alcohol Tested: alctest to AlcoholTestStatusCde

Three attributes were flagged in the analysis of data related to alcohol testing. Missing data were erroneously assigned 'No, test not given' (2) and 'Not applicable' (98) on 8 and 17 instances, respectively. The 'I=INAPPLICABLE' attributes coded in the original dataset were left missing from the converted data 591 times, while 61 instances this attribute were properly assigned the target value of (98). The observed errors accounted for 0.76% of the total data analyzed, while the missing data accounted for 17.86% of the data analyzed for this variable.

4.3.12 Type of Alcohol Test: alctype to AlcoholTestTypeCde

Three types of flags were raised in the analysis for type of alcohol test. First, missing data received an erroneous value of 'Not applicable' (98) in 25 instances. Second, 'NOT APPLICABLE' (98) was missing from the converted dataset 981 times. The third issue raised was that several attributes' conversion was not denoted in the translation notes. These attributes included 'OTHER' (90), 'NOT APPLICABLE' (98), and 'UNKNOWN' (99). The original code for these attributes was mirrored in the converted data, presumably correctly; however, the target value is not listed in the translation notes to ensure that this conversion is as intended. The erroneous data accounted for 0.8% of the total data and the missing data accounted for 29.65% of the total data analyzed for this variable. While there is no reasonable suspicion that the conversion is erroneous for the attributes missing translation notes, this data accounted for 50.23% of the total data included in this analysis.

4.3.13 Ejection: eject to EjectionCde

A limited set of data was found to have issues in the analysis of the ejection data. Missing data from the original dataset was found to have new values erroneously assigned to the crash. 'Not ejected or trapped' (5) and 'Not applicable' (98) were found to be assigned to the missing data on six and 26 instances, respectively. Additionally, 'NOT APPLICABLE' (98) was found to be missing from the converted dataset 20 times. The erroneous data accounted for 0.97% of the total data analyzed, while the missing data accounted for 0.6% of the data analyzed for this variable.

4.3.14 Safety Equipment: safeqp to SafetyEquipmentUseCde

The data for safety equipment was analyzed to reveal a limited number of errors and missing values. Missing data from the original dataset was erroneously assigned with codes for 'Lap and shoulder belt used' (6), 'Not applicable' (98), and 'Unknown' (99) in 1, 27, and 4 instances, respectively. 'NOT APPLICABLE' (98) data from the original dataset was found to be missing from the converted dataset 31 times. The erroneous data accounted for 0.97% of the total data, while the missing data accounted for 0.94% of the total data analyzed for this variable.

4.3.15 Manner of Collision: diagram to MannerOfCollisionCde

Only one issue was raised for manner of collision. 'UNKNOWN' (99) was missing from the converted dataset in nine instances. This accounted for 0.69% of the total data analyzed for this variable.

4.4 CONCLUSION

Overall, the analysis of the MNCrash data conversion process revealed that the integrity of the data was maintained in a vast majority of the variables examined. The recurring issues that were flagged in the analysis were erroneously inserting data codes where missing fields had been previously and losing data for variables coded as 'NOT APPLICABLE'. The final recurring issue was incomplete translation notes. This ranged from converted values not present in the translation notes to improper or missing labels for target values.

The conversion process did reveal a potential error in previous data regarding the age of persons involved in crashes on their birthdays. This may have shifted the average age of drivers down and could have implications for teen drivers who may have been involved on their birthday but were documented as one year younger than their legal age at the time of the crash.

CHAPTER 5: DISCUSSION

The user-centered design project, leading up to the creation and deployment of the MNCrash system involved considerable time and investment and provided a useful, user-friendly crash report interface for statewide use in Minnesota. The innovative project was one-of-a-kind to support an in-depth human factors user-centered design of a state-owned data entry system. The early user testing of the MNCrash system indicated that officers found the new system to have good usability and high satisfaction. The results of the initial MNCrash prototype testing found that a simple crash report could be filled out in approximately 15 minutes. This project sought to determine whether these user experience improvements were accompanied by improvements in data quality.

Notably, the outcome of the analysis indicated that the new MNCrash system increased the quality of data from 2015 to 2016 through completeness, verifiability, and accuracy. This is of great significance because of the profound importance of crash data quality on data analyses outcomes and how they influence our ability to make sound decisions on traffic policy, infrastructure, emergency services, and educational initiatives.

The mixed methods used in this study provide useful tools to measure crash reporting data quality. The analysis using officer reported narratives demonstrates one method to verify the data entry field data against a plain-language account of the crash details. This method not only helps to determine if details have been omitted from the data entry fields or if there is a dispute between the two accounts but also provides a measure of verifiability and trust in the data's accuracy and completeness. The second, qualitative analysis demonstrates methods to review crash data for accuracy and completeness by examining missing or non-descriptive data (i.e., not applicable or other) and leveraging knowledge of crash data business rules without the time-intensive process of reviewing individual crash reports. These two analyses allowed for a broad and controlled examination of the data quality between 2015 and 2016. These analyses not only help to determine the return on investment from the original human factors design of the MNCrash system but also help to create a repeatable process to measure enhancements to the interface in years to come.

Overall, the results of this analysis indicate an improvement in data quality, as defined through data completeness and accuracy, as a result of the implementation of the MNCrash system in 2016. There are still opportunities, however, to improve data completeness through more comprehensive conditional logic functions to auto-populate fields as 'NOT APPLICABLE' when they are skipped by the system. The qualitative analyses and usability test results indicate that there are further design modifications needed to further support officers as they complete the report. This work will be continued through additional research focused on iterative user-centered design and user-centered training modules expected to further improve the data quality gains seen with the first implementation of the MNCrash system.

5.1 FUTURE WORK

The results of this study demonstrate how user-centered design can be employed to help support accurate and complete data entry, even when data entry demands are increased for the user. The

updating of the 2016 MNCrash report nearly doubles the number of form items compared to what was asked of officers in the legacy report, yet the time requirements are reduced and the accuracy, as compared to the narratives is improved. However, the extent to which user-centered design can compensate for increasing data entry requirements should be further examined. Users may give greater care to data requests if the effort demands of entry are reduced, i.e., through supportive user design, and overall time requirements are limited, such as through conditional skip logic and auto-population. Yet, there may be a point at which the number of data entry requests surpass sufficient opportunities for automation within the process given that users may begin to reject the process and revert to shortcuts and/or workarounds. Future research should examine the limits of data entry requests, in terms of both time and volume and accuracy or effort of users as they relate to crash reporting. Particularly, the demands and prioritization for other tasks by law enforcement officers may supersede that of crash reporting, and this imbalance may become more problematic in the future given that automated vehicles may increase the complexity and time demands of the crash reporting process.

This work could also be extended to examine underserved areas to determine if inaccurate or incomplete crash data have been made available for certain classes or groupings of crashes. Specifically, the crash report auditing process could be replicated with other strata selected, e.g., racial demographics of a region, rural and urban regions, or departmental funding. Given that there are multiple socioeconomic factors that have historically been related to infrastructure investments, roadway safety, and law enforcement interventions, it may be useful to examine how crash data validity, consistency, and verifiability may correspond to other factors that could help identify disproportionate responses or treatments to a transportation system.

REFERENCES

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy statistical results. *Computers in Human Behavior, 27,* 1834–1839.

Farmer, C. M. (2003). Reliability of police-reported information for determining crash and injury severity. *Traffic Injury Prevention, 4*(1), 38–44.

MMUCC Guideline. (2012). *Model Minimum Uniform Crash Criteria*, Fourth Edition (Report DOT HS 811 631). NHTSA, U.S. Department of Transportation, Washington, D.C.

Morris, N. L., Achtemeier, J., Ton, A., Plummer, J. P., & Sykes, J. (2016). *Computerized Crash Reports Usability and Design Investigation.* Center for Transportation Studies, University of Minnesota, Minneapolis. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/182433

Porter, B. E. (2011). *Handbook of Traffic Psychology*. Amsterdam: Academic Press, Elsevier Ed.

Still, J. D., & Dark, V. J. (2013). Cognitively describing and designing affordances. *Design Studies*, *34*(3), 285–301.

Sanders, M. S., & McCormick, E. J. (1998). *Human Factors in Engineering and Design* (p. 22). New York: McGraw-Hill.

Wickens, C. D., & Hollands, J. G. (1999). *Engineering Psychology and Human Performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Lee, J., Liu, Y., & Gordon, S. E. (2004). *An Introduction to Human Factors Engineering*. Upper Saddle River, NJ: Pearson Prentice Hall.

APPENDIX A
QUALITATIVE ANALYSIS NARRATIVE CODING VARIABLES

Example individual coding sheet: 2016 Agency Reporting 500+ Crashes per Year, Single Unit Crash

Form Items	500+ 20 Form	ID # Narrative	22##### Form Coding	22##### Narrative Coding	OK Match
Driver Physical Conditions	medical	medical	a	a	1
DPC2	x	x	b	b	4
DPC3	x	x	b	b	4
DPC4	x	x	b	b	4
# of Passengers	x	x	b	b	4
Point of Initial Contact	top	x	a	b	2
Most Harmful Event	ditch	tree	a	a	5
Vehicle Maneuver	mf	curve	a	a	5
Sequence of Events	ditch	ran right	a	a	5
SE2	x	sign	b	a	3
SE3	x	tree	b	a	3
SE4	x	roll	b	a	3
Vehicle Factors	no	x	a	b	2
VF2	x	x	b	b	4
Contributing Factors	99	ran off	c	a	5
CF2	x	x	b	b	4
CF3	x	x	b	b	4
CF4	x	x	b	b	4
Distracted Driving	99	x	c	b	2
Speeding	99	x	c	b	2
Alcohol Suspected	x	x	b	b	4
Tested	no	x	a	b	2
Type	x	x	b	b	4
Results	x	x	b	b	4
Drug Suspected	x	x	b	b	4
Tested	no	x	a	b	2
Type	x	x	b	b	4
Results	x	x	b	b	4
Direction of Travel	x	s	b	a	3
Alignment	x	left	b	a	3
Posted Speed	x	x	b	b	4
Traffic Ctrl Status	x	x	b	b	4
Traffic Ctrl	x	no	b	a	3
Number of Lanes	x	2	b	a	3
Vehicle Damage	x	x	b	b	4

Top	x	x	b	b	4
Undercarriage	x	x	b	b	4
All Areas	y	y	a	a	1
Unknown	x	x	b	b	4
Extent of Damage	sev	sev	a	a	1
Crash Type	ditch	sign	a	a	5
Weather	clear	x	a	b	2
Light Cond.	day	x	a	b	2
Road Surface	dry	x	a	b	2
Road Circumstances	none	x	a	b	2
On/Off Trafficway	non	not	a	a	5
Location Relative to Trafficway	90	roadside	d	a	5
Location Relative to Intersection	not	not	a	a	1
Eject	trap ext	trap ext	a	a	1
Airbag Use	f	x	a	b	2
Injury	a	a	a	a	1
Restraint Type and Use	las	x	a	b	2
Witnesses Listed?	y	y	a	a	1

Example coding totals: 2015 Agencies Reporting Less than 500 Crashes per Year, Single Unit Crash

	Form & Narrative Match				
Form Items	Match	Narrative Blank	Form Blank	Both Blank	Mismatch
Position	41	0	0	0	0
Factor 1	20	9	1	0	11
Factor 2	9	12	3	14	3
Maneuver	25	5	2	0	9
Physical	4	36	1	0	0
Safety Equip Type	1	38	0	1	1
Safety Equip Use	4	36	0	1	0
Airbag	2	38	0	1	0
Eject	11	26	0	1	3
Alcohol Test	5	35	0	1	0
Type	5	35	0	1	0
Drug Test	1	38	0	2	0
Type	0	39	0	2	0
Occupants - # of passengers	18	22	0	0	1
Damage Location	0	41	0	0	0
Damage Severity	2	39	0	0	0
Sequence of events	18	1	0	0	22
2	16	2	2	5	16
3	11	6	4	12	8
4	5	11	0	25	0
Most Harmful Event	31	1	1	0	8
Location	30	1	0	0	10
Road Design	0	41	0	0	0
Road Surface	1	40	0	0	0
Road Characteristics/Circumstances	0	41	0	0	0
Device - traffic ctrl	34	6	0	0	1
Working - traffic ctrl	27	14	0	0	0
Int Rel - intersection relationship	27	5	0	0	9
Speed Limit	0	41	0	0	0
Weather 1	0	41	0	0	0
Weather 2	0	20	0	21	0
Light	0	41	0	0	0
Diagram	3	1	0	0	37

APPENDIX B
CONVERSION ANALYSIS RESULTS

Crash type: acctype to CrashTypeCde combination counts

Obs	acctype	CrashTypeCde	N	Percent	target_value	flag
1	1=COL-SAME RDWAY	10=Motor Vehicle in Transport	564	43.2	10	
2	2=COL-PARKED VEH	11=Parked Motor Vehicle	22	1.7	11	
3	3=COL-RDEQP-SNOWP	1=COL-RDEQP-SNOWP	2	0.2	1	
4	4=COL-RDEQP-OTHER	2=COL-RDEQP-OTHER	5	0.4	2	
5	5=COL WITH TRAIN	3=COL WITH TRAIN	2	0.2	3	
6	6=COL W PEDALCYCL	9=Pedalcycle (Bike)	55	4.2	9	
7	7=COL WITH PEDEST	8=Pedestrian	149	11.4	8	
8	8=COL WITH DEER	16=Deer	26	2.0	16	
9	9=COL ANML NOT DER	4=COL ANML NOT DER	6	0.5	4	
10	12=OTHER NONFIX COL	25=Other - Non Fixed Object	1	0.1	25	
11	13=OTHER COLISN TYP	93=OTHER COLISN TYP	23	1.8	93	
12	21=CONSTRUCT EQUIP	26=LIGHT POLE	1	0.1	36	Error
13	22=TRAFFIC SIGNAL	30=Traffic Signal or Signal Structure	2	0.2	30	
14	24=LIGHT POLE	26=LIGHT POLE	6	0.5	26	
15	25=UTILITY POLE	27=UTILITY POLE	21	1.6	27	
16	26=SIGN STRUC/POST	32=Roadway Sign or Signal Structure	11	0.8	32	
17	27=MAILBOXES/POSTS	67=Mailboxes/Posts	3	0.2	67	
18	28=OTHER POLES	35=Other Post, Pole or Support	1	0.1	35	
19	30=TREE/SHRUBBERY	69=Standing Tree/Shrubbery	85	6.5	69	
20	31=BRIDGE PIERS	41=Bridge Pier or Support	4	0.3	41	
21	32=MEDIAN SAFTY BAR	53=MEDIAN SAFTY BAR	13	1.0	53	
22	34=GUARDRAIL	54=GUARDRAIL	9	0.7	54	
23	35=FNCE (NOT MD BR)	70=Fence, Not Barrier	2	0.2	70	
24	36=CULVERT/HEADWALL	40=CULVERT/HEADWALL	5	0.4	40	
25	37=EMBANK/DITCH/CRB	45=EMBANK/DITCH/CRB	64	4.9	45	
26	38=BUILDING/WALL	65=BUILDING/WALL	7	0.5	65	
27	39=ROCK OUTCROPS	66=ROCK OUTCROPS	1	0.1	66	
28	41=OTHER FIXED OBJ	75=Other - Fixed Object	6	0.5	75	

Obs	acctype	CrashTypeCde	N	Percent	target_value	flag
29	51=OVRTURN/ROLLOVER	83=Overturn/Rollover	172	13.2	83	
30	52=SUBMERSION	84=Immersion (Partial or Full)	1	0.1	84	
31	53=FIRE/EXPLOSION	85=Fire/Explosion	1	0.1	85	
32	64=OTHR TYP NONCOLL	83=Overturn/Rollover	16	1.2	83	
33	90=OTHER ACC TYPE	92=OTHER ACC TYPE	20	1.5	92	
34	99=UNKWN ACC TYPE	99=Unknown	1	0.1	99	

Count of errors and missing entries in conversion of variable: acctype to CrashTypeCde

The MEANS Procedure

Variable	Sum
error_count	1.00
error_percent	0.08
miss_count	.
miss_percent	.

Sequence of crash-related events: event1 to EventCde combination counts

Obs	event1	EventCde1	N	Percent	target_value	flag
1	.	.	166	7.3	.	
2	.	45	1	0.0	.	Error
3	1=COL UNIT-SAME RD	10=Motor Vehicle In Transport	1493	65.7	10	
4	2=COL PARKED VEHIC	11=Parked Motor Vehicle	57	2.5	11	
5	5=COLLISN W TRAIN	3	1	0.0	3	
6	6=COL W PEDALCYCL	9=Pedalcyclist (Bicyclist)	48	2.1	9	
7	7=COL W PEDESTRN	8=Pedestrian	122	5.4	8	
8	8=COLISN WITH DEER	16=Deer	20	0.9	16	
9	9=COL ANM-NOT DEER	4	5	0.2	4	
10	11=UNDERRIDE--SIDE	7	2	0.1	7	
11	12=OTHER NONFIX OBJ	25=Other - Non Fixed Object	4	0.2	25	
12	13=OTHER COLSN TYPE	93	12	0.5	93	
13	22=TRAFFIC SIGNAL	30=Traffic Signal or Signal Structure	3	0.1	30	
14	24=LIGHT POLE	26	1	0.0	26	
15	25=UTILITY POLE	27	2	0.1	27	
16	26=SIGN STRUC/POST	32=Roadway Sign or Sign Structure	9	0.4	32	
17	27=MAILBOXES/POSTS	67=Mailboxes/Posts	1	0.0	67	
18	30=TREE/SHRUBBERY	69=Standing Tree/Shrubbery	22	1.0	69	
19	32=MEDIAN SAFTY BAR	53	11	0.5	53	
20	33=CRASH CUSHION	60=Impact attenuator/ Crash Cushion	1	0.0	60	
21	34=GUARDRAIL	54	9	0.4	54	
22	35=FENCE(NONMEDBAR)	70=Fence (Non-Median Barrier)	1	0.0	70	
23	36=CULVERT/HDWALL	40	3	0.1	40	
24	37=EMBANKMT/DTCH/CB	45	28	1.2	45	
25	38=BUILDING/WALL	65	6	0.3	65	
26	39=ROCK OUTCROPS	66	1	0.0	66	
27	41=OTHER FXD OBJ	75=Other - Fixed Object	1	0.0	75	
28	42=UNKN FIXED OBJ	76	1	0.0	76	

Obs	event1	EventCde1	N	Percent	target_value	flag
29	51=OVERTURN/ROLLOVR	83=Overturn/Rollover	37	1.6	83	
30	57=RAN OFF RD-RIGHT	.	88	3.9	.	Missing
31	58=RAN OFF RD--LEFT	.	71	3.1	.	Missing
32	62=CROSS MED/CENTRL	.	11	0.5	.	Missing
33	64=NONCOL--OTH TYPE	83=Overturn/Rollover	13	0.6	83	
34	90=OTHER EVENT	92	13	0.6	92	
35	98=NOT APPLICABLE	.	6	0.3	.	Missing
36	99=UNKNOWN EVENT	99=Unknown	4	0.2	99	

Count of errors and missing entries in conversion of variable: event1 to EventCde1

The MEANS Procedure

Variable	Sum
error_count	1.00
error_percent	0.04
miss_count	176.00
miss_percent	7.74

Contributing Factor: cfct1 to FactorCde1 combination counts

FactorCde1 is person level variable, converted from a vehicle-level variable. Kept only the first vehicle observation after verifying no new information for additional persons

Obs	cfct1	FactorCde1	N	Percent	Target value	flag
1	1=NO CLR CNTR FCTR	1=No Clear Contributing Action	892	44.4	1	
2	2=FAIL TO YLD ROW	2=Failure to Yield Right-of-Way	233	11.6	2	
3	3=ILLEGAL/UNSAF SP	3=ILLEGAL/UNSAF SP	175	8.7	3	
4	4=FLNG TOO CLOSELY	5=Failure to Obey Traffic Signs, Signals, or Officer	28	1.4	5	
5	5=DSRGRD TRF CNTRL	4=Following Too Closely	83	4.1	4	
6	6=DRVNG LFT OF CTR	6=DRVNG LFT OF CTR	61	3.0	6	
7	7=IMPR PASG/OVRTKG	7=Improper Passing	13	0.6	7	
8	8=IMPRP/UNSF LN US	8=IMPRP/UNSF LN US	42	2.1	8	
9	9=IMP PKG/STR/STPG	9=IMP PKG/STR/STPG	3	0.1	9	
10	10=IMPROPER TURN	10=Improper Turn/Merge	6	0.3	10	
11	11=UNSAFE BACKING	11=Improper Backing	5	0.2	11	
12	13=OVERCORRECTING	72=Over-correcting / Over Steering	25	1.2	72	
13	14=IMPEDING TRAFFIC	14=IMPEDING TRAFFIC	4	0.2	14	
14	15=DRV INATNTN/DSTR	15=Inattentive/Distraction	133	6.6	15	
15	16=DRIVR INXPERENCE	16=DRIVR INXPERENCE	17	0.8	16	
16	17=NON-MOTRST ERROR	17=NON-MOTRST ERROR	26	1.3	17	
17	18=CHEMICAL IMPRMNT	18=CHEMICAL IMPRMNT	111	5.5	18	
18	19=FAIL TO USE LITS	19=FAIL TO USE LITS	4	0.2	19	
19	20=DRVR ON PH/CB/RA	20=DRVR ON PH/CB/RA	1	0.0	20	
20	21=OTH HMN CNTR FCT	21=OTH HMN CNTR FCT	45	2.2	21	
21	99=UNKNOWN	.	104	5.2	99	Missing

Count of errors and missing entries in conversion of variable: cfct1 to FactorCde1 (person level)

The MEANS Procedure

Variable	Sum
error_count	.
error_percent	.
miss_count	104.00
miss_percent	5.17

Observations where value is given to factorcde1 in addition to PrimaryContributorCde, when cfct1 gt 21

Obs	cfct1	FactorCde1	N	Percent
1	31=VSN OBSCRD-WNDSH	2=Failure to Yield Right-of-Way	2	7.6923
2	32=VSN OBSCD-SUN/LT	1=No Clear Contributing Action	1	3.8462
3	32=VSN OBSCD-SUN/LT	15=Inattentive/Distraction	1	3.8462
4	33=OTH VSN RLTD FCT	2=Failure to Yield Right-of-Way	2	7.6923
5	33=OTH VSN RLTD FCT	7=Improper Passing	1	3.8462
6	41=DEFECTIVE BRAKES	4=Following Too Closely	1	3.8462
7	42=DEFECT TIRE/FAIL	3=ILLEGAL/UNSAF SP	1	3.8462
8	42=DEFECT TIRE/FAIL	72=Over-correcting / Over Steering	2	7.6923
9	46=SKIDDING	1=No Clear Contributing Action	2	7.6923
10	46=SKIDDING	3=ILLEGAL/UNSAF SP	2	7.6923
11	46=SKIDDING	16=DRIVR INXPERENCE	1	3.8462
12	46=SKIDDING	18=CHEMICAL IMPRMNT	1	3.8462
13	46=SKIDDING	72=Over-correcting / Over Steering	3	11.5385
14	50=OTHER VEHIC FCTR	16=DRIVR INXPERENCE	1	3.8462
15	61=WEATHER	1=No Clear Contributing Action	1	3.8462
16	61=WEATHER	3=ILLEGAL/UNSAF SP	2	7.6923
17	61=WEATHER	21=OTH HMN CNTR FCT	1	3.8462
18	98	98=Not Applicable	1	3.8462

Primary apparent contributing factor: cfct1 to PrimaryContributorCde combination counts

Obs	cfct1	PrimaryContributorCde	COUNT	PERCENT	target_value	flag
1	31=VSN OBSCRD-WNDSH	1=VSN OBSCRD-WNDSH	3	1.8	1	
2	32=VSN OBSCD-SUN/LT	2=VSN OBSCD-SUN/LTl	6	3.6	2	
3	33=OTH VSN RLTD FCT	3=OTH VSN RLTD FCT	8	4.8	3	
4	41=DEFECTIVE BRAKES	11=Defective Brakes	2	1.2	11	
5	42=DEFECT TIRE/FAIL	12=Defective Tire or Tire Failure	4	2.4	12	
6	46=SKIDDING	4=SKIDDING	24	14.4	4	
7	50=OTHER VEHIC FCTR	.	5	3.0	.	Missing
8	61=WEATHER	5=WEATHER	27	16.2	5	
9	90=OTHR CONTRB FCTR	90=Other	87	52.1	90	
10	98	98=Not Applicable	1	0.6	98	

Count of errors and missing entries in conversion of variable: cfct1 to PrimaryContributorCde (vehicle-level)

The MEANS Procedure

Variable	Sum
error_count	.
error_percent	.
miss_count	5.00
miss_percent	2.99

Observations where value is given to PrimaryContributorCde in addition to factorcde1, when cfct1 lt 21

Obs	cfct1	PrimaryContributorCde	N	Percent
1	1=NO CLR CNTR FCTR	10=No Clear Contributing Factor	892	100

Vehicle Type: vehtype to VehicleTypeCde combination counts

Obs	vehtype	VehicleTypeCde	N	Percent	target_value	flag
1	1=PASSENGER CAR	2=Passenger Car	865	38.0	2	
2	2=PICKUP	3=Pickup	317	13.9	3	
3	3=SUV	4=Sport Utility Vehicle	328	14.4	4	
4	4=VAN OR MINIVAN	1=VAN OR MINIVAN	131	5.8	1	
5	5=MOTORHM/CAMPR/RV	20=Motor Home/Camper/RV	1	0.0	20	
6	7=BUS(7-15 INC DRV)	11=BUS(7-15 INC DRV)	3	0.1	11	
7	8=BUS(8+ INCL DRV)	12=BUS(8+ INCL DRV)	9	0.4	12	
8	9=SNOWMOBILE	20=Snowmobile	2	0.1	21	
9	10=ATV	22=ATV	18	0.8	22	
10	11=MOTORCYCLE	31=Motorcycle	245	10.8	31	
11	12=MOTORSCOOTER/BIKE	32=Moped or Motor Scooter	2	0.1	32	
12	13=MOPED/MOTORZ BIKE	30=MOPED/MOTORZ BIKE	6	0.3	30	
13	14=FARM EQUIPMENT	50=Farm Vehicle (Tractor, Combine, etc.)	5	0.2	50	
14	31=2-AX 6-T SG TR/S	40=2-AX 6-T SG TR/S	26	1.1	40	
15	32=3+AXL SGL UNT TR	41=3+AXL SGL UNT TR	20	0.9	41	
16	33=SG UNT TR W TRLR	42=SG UNT TR W TRLR	10	0.4	42	
17	34=TRK TRC W NO TRL	43=TRK TRC W NO TRL	2	0.1	43	
18	35=TRK TRAC W SEMI	44=TRK TRAC W SEMI	59	2.6	44	
19	38=HEAV TRK-TYP UNK	47=HEAV TRK-TYP UNK	3	0.1	47	
20	51=PEDESTRIAN	60=PEDESTRIAN	149	6.6	60	
21	52=SKATER	61=SKATER	1	0.0	61	

Obs	vehtype	VehicleTypeCde	N	Percent	target_value	flag
22	53=BICYCLIST	62=BICYCLIST	56	2.5	62	
23	54=OTHER NONMOTORST	63=OTHER NONMOTORST	6	0.3	63	
24	90=OTH VEHICLE TYPE	90=Other	3	0.1	90	
25	99=UNKN VEHIC TYPE	99=Unknown	7	0.3	99	

Gender: sex to GenderCde combination counts

Obs	sex	GenderCde	N	Percent	target_value	flag
1			151	4.6		
2		9	2	0.1		Error
3		F=Female	8	0.2		Error
4		M=Male	10	0.3		Error
5	F=Female	F=Female	1150	34.8	F	
6	M=Male	M=Male	1979	59.8	M	
7	U		2	0.1		Missing
8	X=Unknown	9	7	0.2	99	Error

Count of errors and missing entries in conversion of variable: sex to GenderCde

The MEANS Procedure

Variable	Sum
error_count	27.00
error_percent	0.82
miss_count	2.00
miss_percent	0.06

Age: age to BirthDte combination counts (converted age calculated from difference between DOB and incident date)

agematch	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1=Age matched	3106	93.87	3106	93.87
2=Age not matched*	3	0.09	3109	93.96
3=Both missing	182	5.50	3291	99.46
4=old missing, new not missing	18	0.54	3309	100.00

*these three incidents occurred on person's birthday. In original data, age change on birthday was not accounted for, while in the converted data it was calculated up to and including the birthday.

Physical Condition: physcnd to PhysicalConditionCde combination counts

Obs	physcnd	PhysicalConditionCde	N	Percent	target_value	flag
1	.	.	282	8.5	.	
2	.	5=Apparently Normal (Including No Drugs/Alcohol)	6	0.2	.	Error
3	.	98=NOT APPLICABLE	20	0.6	.	Error
4	1=NORML-NO ALC/DRG	5=Apparently Normal (Including No Drugs/Alcohol)	1432	43.3	5	
5	2=UNDER THE INFLU	1=UNDER THE INFLU	147	4.4	1	
6	3=HAD BEEN DRINKNG	10=Has Been Drinking Alcohol	104	3.1	10	
7	5=HAD BN TKNG DRGS	2=HAD BN TKNG DRGS	19	0.6	2	
8	6=AGGRESSIVE	8=Emotional (Depression, Angry, Disturbed, etc.)	3	0.1	8	
9	7=FATIGUED/ASLEEP	9=Asleep or Fatigued	11	0.3	9	
10	8=PHYSICAL DISABIL	6=Physical Disability (Short Term or Long Term)	7	0.2	6	
11	9=ILL	7=Medical Issue (Ill, Sick or Fainted)	5	0.2	7	
12	90=OTHR PHYS CONDTN	90=Other	20	0.6	90	
13	98=NOT APPLICABLE	.	798	24.1	98	Missing
14	98=NOT APPLICABLE	98=NOT APPLICABLE	56	1.7	98	
15	99=UNKN PHYS CONDTN	99=Unknown	399	12.1	99	

Count of errors and missing entries in conversion of variable: physcnd to PhysicalConditionCde

The MEANS Procedure

Variable	Sum
error_count	26.00
error_percent	0.79
miss_count	798.00
miss_percent	24.12

Traffic Control Device: trfcntl to TrafficControlCde combination counts

TrafficControlCde is vehicle-level variable, converted from a crash-level variable. Kept only the first crash observation after verifying no new information for additional vehicles

Obs	trfcntl	TrafficControlCde	N	Percent	target_value	flag
1	.	.	6	0.5	.	
2	1=TRAFFIC SIGNAL	20=Traffic Control Signal	176	13.5	20	
3	2=OVRHEAD FLASHERS	21=Flashing Overhead Traffic Control Signal	1	0.1	21	
4	3=STPSGN-ALL APPRS	1=STPSGN-ALL APPRS	23	1.8	1	
5	4=STPSN-NOT ALL AP	2=STPSN-NOT ALL AP	220	16.8	2	
6	5=YIELD SIGN	22=Yield Sign	10	0.8	22	
7	6=OFCR/FLGPRSN/SCH	25=Person (Including Flagger, Law Enforcement, Crossing Guard, etc.)	1	0.1	25	
8	9=NO PASSING ZONE	5=NO PASSING ZONE	12	0.9	5	
9	10=RR CROSSING GATE	11=RR or LRT CROSSING GATE	1	0.1	11	
10	11=RRXING FLSHNG LT	7=RR or LRT XING FLSHNG LT	1	0.1	7	
11	90=OTHER DEVICE	90=Other	11	0.8	90	
12	98=NOT APPLICABLE	98=Not Applicable	838	64.1	98	
13	99=UNKNOWN	.	7	0.5	99	Missing

Count of errors and missing entries in conversion of variable: trfcntl to TrafficControlCde

The MEANS Procedure

Variable	Sum
error_count	.
error_percent	.
miss_count	7.00
miss_percent	0.54

Weather Conditions: weather to WeatherCde combination counts (Same values used for converted weather variable)

Obs	weather	WeatherCde	N	Percent	target_value	flag
1	.	.	4	0.3	.	
2	1=CLEAR	1=Clear	913	69.9	1	
3	2=CLOUDY	2=Cloudy	244	18.7	2	
4	3=RAIN	3=Rain	73	5.6	3	
5	4=SNOW	4=Snow	35	2.7	4	
6	5=SLT/HAIL/FRZ RN	5=Sleet, Hail (Freezing Rain or Freezing Drizzle)	9	0.7	5	
7	6=FOG/SMOG/SMOKE	6=Fog/Smog/Smoke	7	0.5	6	
8	7=BLWNG SND/DS/SN	7=Blowing Sand/Soil/Dirt/Snow	7	0.5	7	
9	90=OTHR WEATHER	90=Other	2	0.2	90	
10	99=UNKNWN WEATHER	99=Unknown	13	1.0	99	

Light Conditions: light to LightConditionCde combination counts

Obs	light	LightConditionCde	N	Percent	target_value	flag
1	.	.	4	0.3	.	
2	1=DAYLIGHT	1=Daylight	786	60.1	1	
3	2=SUNRISE	2=Sunrise	33	2.5	2	
4	3=SUNSET	3=Sunset	47	3.6	3	
5	4=DARK-STRT LTS ON	4=Dark (Street Lights On)	212	16.2	4	
6	5=DARK-ST LTS OFF	5=Dark (Street Lights Off	8	0.6	5	
7	6=DARK-NO ST LITES	6=Dark (No Street Lights)	191	14.6	6	
8	7=DARK-UNKN LITING	7=Dark (Unknown Lighting)	18	1.4	7	
9	90=OTHER	90=Other	2	0.2	90	
10	99=UNKN LGHT CONDTN	99=Unknown	6	0.5	99	

Road Surface Condition: rdsurf to SurfaceConditionCde combination counts
(Same values used for converted Surface Condition variable)

Obs	rdsurf	SurfaceConditionCde	N	Percent	target_value	flag
1	.	.	3	0.2	9	
2	1=DRY	1=Dry	1028	78.7	1	
3	2=WET	2=Wet	146	11.2	2	
4	3=SNOW	3=Snow	27	2.1	3	
5	4=SLUSH	4=Slush	10	0.8	4	
6	5=ICE/PACKED SNOW	5=Ice/Frost	65	5.0	5	
7	7=MUDDY	7=Mud/Dirt/Gravel	4	0.3	7	
8	8=DEBRIS	8=Debris	3	0.2	8	
9	90=OTHER	90=Other	14	1.1	90	
10	99=UNKNOWN	99=Unknown	7	0.5	99	

Road Design: rddes to RoadwayAlignmentCde combination counts

RoadwayAlignmentCde is a vehicle-level variable, converted from a crash-level variable. Kept only the first crash observation after verifying no new information for additional vehicles

Obs	rddes	RoadwayAlignmentCde	N	Percent	target_value	flag
1	.	.	23	1.8	.	
2	1=FREEWAY-MAINLINE	1=FREEWAY-MAINLINE	104	8.0	1	
3	2=FREEWAY-RAMPS	2=FREEWAY-RAMPS	19	1.5	2	
4	3=OTHER DIVIDD HWY	3=OTHER DIVIDD HWY	161	12.3	3	
5	4=ONE-WAY STREET	11=One Way Trafficway	15	1.1	11	
6	5=4-6 LNS 2-3-EACH	4=4-6 LNS 2-3-EACH	170	13.0	4	
7	6=3 LANES UNDIVIDD	5=3 LANES UNDIVIDD	16	1.2	5	
8	7=5 LN UNDV W L TURN	6=5 LN UNDV W L TURN	1	0.1	6	
9	8=2-LANES 1-ECH-WY	7=2-LANES 1-ECH-WY	749	57.3	7	
10	9=ALLEY OR DRIVEWY	8=ALLEY OR DRIVEWY	7	0.5	8	
11	10=PRIVATE PROPERTY	10=PRIVATE PROPERTY	7	0.5	10	
12	90=OTHER ROAD DESGN	.	32	2.4	90	Missing
13	99=UNKNOWN RD DESGN	.	3	0.2	99	Missing

Count of errors and missing entries in conversion of variable: rddes to RoadwayAlignmentCde

The MEANS Procedure

Variable	Sum
error_count	.
error_percent	.
miss_count	35.00
miss_percent	2.68

Alcohol tested: alctest to AlcoholTestStatusCde combination counts

Obs	alctest	AlcoholTestStatusCde	N	Percent	target_value	flag
1		.	573	17.3	.	
2		2=No, Test Not Given	8	0.2	.	Error
3		98=Not applicable	17	0.5	.	Error
4	I=Inapplicable	.	591	17.9	98	Missing
5	I=Inapplicable	98=Not applicable	61	1.8	98	
6	N=No	2=No, Test Not Given	1613	48.7	2	
7	X=Unknown	99=Unknown	69	2.1	99	
8	Y=Yes	1=Yes, Test Given	377	11.4	1	

Count of errors and missing entries in conversion of variable: alctest to AlcoholTestStatusCde

The MEANS Procedure

Variable	Sum
error_count	25.00
error_percent	0.76
miss_count	591.00
miss_percent	17.86

Type of Alcohol Test: alctype to AlcoholTestTypeCde combination counts

Obs	alctype	AlcoholTestTypeCde	N	Percent	target_value	flag
1	.	.	293	8.9	.	
2	.	98=NOT APPLICABLE	25	0.8	.	Error
3	1=BLOOD	1=Blood	273	8.3	1	
4	2=SERUM	2=SERUM	2	0.1	2	
5	3=BREATH	3=PBT (Breath)	67	2.0	3	
6	4=URINE	4=Urine	6	0.2	4	
7	90=OTHER	90=Other	17	0.5	.	No translation note
8	98=NOT APPLICABLE	.	981	29.6	.	Missing
9	98=NOT APPLICABLE	98=NOT APPLICABLE	1571	47.5	.	No translation note
10	99=UNKNOWN	99=UNKNOWN	74	2.2	.	No translation note

Count of errors and missing entries in conversion of variable: alctype to AlcoholTestTypeCde

The MEANS Procedure

Variable	Sum
error_count	25.00
error_percent	0.8
miss_count	981.00
miss_percent	29.65
notrnsl_count	1662.00
notrnsl_percent	50.23

Ejection: eject to EjectionCde combination counts

Obs	eject	EjectionCde	N	Percent	target_value	flag
1	.	.	91	2.8	.	
2	.	5=Not Ejected or Trapped	6	0.2	.	Error
3	.	98=Not Applicable	26	0.8	.	Error
4	1=TRAPPED EXTRC-MEC	1=Trapped, Extricated (by Mechanical Means)	202	6.1	1	
5	2=TRAPPED EXTRC-MAN	2=Trapped, Freed by Non-Mechanical Means	77	2.3	2	
6	3=PARTIALLY EJCTED	3=Partially Ejected	29	0.9	3	
7	4=EJECTED	4=Totally Ejected	381	11.5	4	
8	5=NOT EJECTED	5=Not Ejected or Trapped	2148	64.9	5	
9	90=OTHER	90=OTHER	25	0.8	90	
10	98=NOT APPLICABLE	.	20	0.6	98	Missing
11	98=NOT APPLICABLE	98=Not Applicable	257	7.8	98	
12	99=UNKNOWN	99=Unknown	47	1.4	99	

Count of errors and missing entries in conversion of variable: eject to EjectionCde

The MEANS Procedure

Variable	Sum
error_count	32.00
error_percent	0.97
miss_count	20.00
miss_percent	0.60

Safety Equipment: safeqp to SafetyEquipmentUseCde combination counts

Obs	safeqp	SafetyEquipmentUseCde	N	Percent	target_value	flag
1	.	.	82	2.5	.	
2	.	6=Lap and Shoulder Belt Used	1	0.0	.	Error
3	.	98=Not Applicable	27	0.8	.	Error
4	.	99=Unknown	4	0.1	.	Error
5	1=BELTS NOT USED	5=None Used, Motor Vehicle Occupant	317	9.6	5	
6	2=LAPBELT ONLY USED	7=Lap Belt Only Used	18	0.5	7	
7	3=SHLDBELTONLY USED	8=Shoulder Belt Only Used	18	0.5	8	
8	4=LAP+SHLD BELT USD	6=Lap and Shoulder Belt Used	1698	51.3	6	
9	5=CHILDSEAT NOT USD	10=Child restraint system Not Used	3	0.1	10	
10	6=CHLDSEAT USD IMPR	11=Child Restraint System Seat Used Improperly	1	0.0	11	
11	7=CHLDSEAT USD PROP	2=CHLDSEAT USD PROP	82	2.5	2	
12	8=BSTRSEAT NOT USED	3=BSTRSEAT NOT USED	5	0.2	3	
13	9=BSTRSEAT USD IMPR	4=BSTRSEAT USD IMPR	3	0.1	4	
14	10=BSTRSEAT USED PRO	14=Booster Seat Properly	20	0.6	14	
15	11=HELMET NOT USED	17=No Helmet	170	5.1	17	
16	12=HELMET USED	16=Helmet Used, Unknown If DOT-Compliant	106	3.2	16	
17	13=DARK CLOTHING	21=Dark Clothing	5	0.2	21	
18	15=NO PROTECTV PADS	19=No Protective Pads	1	0.0	19	
19	90=OTHR EQUIP USE	90=Other	29	0.9	90	
20	98=NOT APPLICABLE	.	31	0.9	98	Missing
21	98=NOT APPLICABLE	98=Not Applicable	199	6.0	98	
22	99=UNKNOWN	99=Unknown	489	14.8	99	

Count of errors and missing entries in conversion of variable: safeqp to SafetyEquipmentUseCde

The MEANS Procedure

Variable	Sum
error_count	32.00
error_percent	0.97
miss_count	31.00
miss_percent	0.94

Maneuver (action) was split into two variables on conversion: 1) VehicleManeuverCde in vehicle dataset (action = 1-23, 90, 99) and2) ManeuverCde in person dataset (action = 31-57)

Vehicle Maneuver: action to VehicleManeuverCde combination counts

Obs	action	VehicleManeuverCde	N	Percent	target_value	flag
1	1=VEH STRT FLWG RD	21=Moving Forward	1424	69.8	21	
2	2=VEH WRG WY-OP TR	22=Wrong Way into Opposing Traffic	67	3.3	22	
3	3=VEH RT TN ON RED	4=VEH RT TN ON RED	5	0.2	4	
4	4=VEH LFT TN ON RD	5=VEH LFT TN ON RD	8	0.4	5	
5	5=VEH MKNG RGHT TN	23=Turning Right	25	1.2	23	
6	6=VEH MKNG LFT TRN	24=Turning Left	163	8.0	24	
7	7=VEH MAKING U TRN	25=Making a U Turn	13	0.6	25	
8	8=VEH STRG FR PRKD	6=VEH STRG FR PRKD	4	0.2	6	
9	9=VEH STRTNG N TRC	7=VEH STRTNG N TRC	26	1.3	7	
10	10=VEH SLNG IN TRFC	26=Slowing	25	1.2	26	
11	11=VEH STOPD IN TRF	8=VEH STOPD IN TRF	67	3.3	8	
12	12=VEH ENTG PRK PTN	9=VEH ENTG PRK PTN	2	0.1	9	
13	13=VEH AVDG UNT/OBJ	27=Swerved or Attempt to Avoid Object in Roadway (Due to Wind, Slippery Surface, Motor Vehicle, Object, Non-Motorist in Roadway, etc.)	34	1.7	27	
14	14=VEH CHNGING LANS	28=Changing Lanes	18	0.9	28	
15	15=VEH OVRTKNG/PSNG	29=Overtaking/Passing	26	1.3	29	
16	16=VEHICLE MERGING	10=VEHICLE MERGING	3	0.1	10	
17	17=VEHICLE BACKING	33=Backing	8	0.4	33	
18	18=VEHICLE STALLED	34=Vehicle Stopped or Stalled in Roadway	1	0.0	34	
19	21=VEH PRKD LEGALLY	1=VEH PRKD LEGALLY	54	2.6	1	
20	23=VEH STPD OFF RDW	3=VEH STPD OFF RDW	3	0.1	3	
21	90=OT ACTN-VH/PD/BK	90=Other	31	1.5	90	
22	99=UNKN PRIOR ACTN	99=Unknown	32	1.6	99	

Maneuver: action to ManeuverCde combination counts

ManeuverCde is person level variable, converted from a vehicle-level variable. Value applied to one person per vehicle ID

Obs	action	ManeuverCde	N	Percent	target_value	flag
1	31=PED XNG W SIGNAL	1=PED XNG W SIGNAL	15	5.0	1	
2	32=PED XNG AGNT SIG	2=PED XNG AGNT SIG	9	3.0	2	
3	33=PED DARTNG N2 TR	3=PED DARTNG N2 TR	28	9.2	3	
4	34=PED--OTHR IMPROP	4=PED--OTHR IMPROP	15	5.0	4	
5	35=PED XNG N MK XWK	5=PED XNG N MK XWK	23	7.6	5	
6	36=PED XNG-NO SIG/X	6=PED XNG-NO SIG/X	28	9.2	6	
7	37=PED FAIL YLD ROW	7=PED FAIL YLD ROW	24	7.9	7	
8	38=PED INATEN/DISTR	8=PED INATEN/DISTR	18	5.9	8	
9	39=PED WLKRUN W TRF	9=PED WLKRUN W TRF	17	5.6	9	
10	40=PED WLKRN AGN TR	10=PED WLKRN AGN TR	3	1.0	10	
11	41=PED STD/LY IN RD	11=PED STD/LY IN RD	7	2.3	11	
12	44=PERSN ON/OFF VEH	14=PERSN ON/OFF VEH	1	0.3	14	
13	48=NOT IN ROADWAY	18=NOT IN ROADWAY	10	3.3	18	
14	51=BIK RIDNG W TRFC	19=BIK RIDNG W TRFC	6	2.0	19	
15	52=BIK RDNG AGNT TR	20=BIK RDNG AGNT TR	1	0.3	20	
16	53=BIK MAKNG RIT TN	22=BIK MAKNG RIT TN	3	1.0	22	
17	54=BIK MKNG LFT TRN	23=BIK MKNG LFT TRN	4	1.3	23	
18	56=BIK RDNG ACRS RD	25=BIK RDNG ACRS RD	4	1.3	25	
19	57=BK SLG/STPG/STNG	26=BK SLG/STPG/STNG	7	2.3	26	
20	90=OT ACTN-VH/PD/BK	90=Other	36	11.9	90	
21	99=UNKN PRIOR ACTN	99=Unknown	44	14.5	99	

Manner of collision: diagram to MannerOfCollisionCde combination counts

Obs	diagram	MannerOfCollisionCde	N	Percent	target_value	flag
1	.	.	10	0.8	.	
2	1=REAR-END	1=REAR-END	109	8.3	1	
3	2=SIDESWP SAME DIR	10=Sideswipe - Same Direction	26	2.0	10	
4	3=LEFT TURN	3=LEFT TURN	51	3.9	3	
5	4=RAN OFF RD-LEFT	4=RAN OFF RD-LEFT	158	12.1	4	
6	5=RIGHT ANGLE	5=Angle	268	20.5	5	
7	6=RGHT TRN	6=RGHT TRN	12	0.9	6	
8	7=RAN OFF RD-RIGHT	7=RAN OFF RD-RIGHT	203	15.5	7	
9	8=HEAD-ON	8=HEAD-ON	202	15.5	8	
10	9=SIDESWPE OPPOSNG	11=Sideswipe – Opposing	24	1.8	11	
11	90=OTHER	90=Other	201	15.4	90	
12	98=NOT APPLICABLE	98=NOT APPLICABLE	34	2.6	98	
13	99=UNKNOWN	.	9	0.7	99	Missing

Count of errors and missing entries in conversion of variable: diagram to MannerOfCollisionCde

Variable	Sum
error_count	.
error_percent	.
miss_count	9.00
miss_percent	0.69