Safety Impacts of Reduced Visibility in Inclement Weather

Final Report
ATLAS-2017-19

by

Subasish Das, Ph.D.
Associate Transportation Researcher
Texas A&M Transportation Institute

Bradford K. Brimley, Ph.D.
Traffic Engineer
Binkley and Barfield, Inc.

Tomas Lindheimer, Ph.D.
Associate Transportation Researcher
Texas A&M Transportation Institute

Ashesh Pant, M.S.
Assistant Transportation Researcher
Texas A&M Transportation Institute

ATLAS CENTER
Advancing Transportation Leadership and Safety

University of Michigan
2901 Baxter Rd., Room 124
Ann Arbor, MI 48109-2150

Texas A&M Transportation Institute
College Station, TX 77843-3135

April 2017

DISCLAIMER

ACKNOWLEDGMENTS

This research project was supported by the Center for Advancing Transportation Leadership and Safety (ATLAS Center). The ATLAS Center is supported by a grant from the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Transportation, University Transportation Centers Program (DTRT13-G-UTC54). The ATLAS Center is a collaboration between the University of Michigan Transportation Research Institute and the Texas A&M Transportation Institute.

The authors would like to thank Karen Dixon and Michael Manser for their help and guidelines. The authors would also like to thank Robert Wunderlich, David W. Eby, Barbara Lorenz, and Beth Jakubowski for their patience as the authors needed to extend the project to the end of 2016.

Technical Report Documentation Page

1. Report No. ATLAS-2017-19	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Safety Impacts of Reduced Visibility in Inclement Weather		5. Report Date April 4, 2017
		6. Performing Organization Code
7. Author(s) Subasish Das, Bradford K. Brimley, Tomas Lindheimer, and Ashesh Pant		8. Performing Organization Report No.
9. Performing Organization Name and Address Texas A&M Transportation Institute The Texas A&M University System College Station, TX 77843-3135		10. Work Unit no. (TRAIS)
		11. Contract or Grant No. DTRT13-G-UTC54
12. Sponsoring Agency Name and Address Advancing Transportation Leadership and Safety (ATLAS) Center 2901 Baxter Rd., Room 124 Ann Arbor, MI 48109-2150		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes Supported by a grant from the U.S. Department of Transportation, OST-R, University Transportation Centers Program		

16. Abstract

This research seeks to better understand the ramifications of inclement weather on safety from a perspective of visibility. Visibility conditions at the time of a crash are rarely documented at a high level of detail. While vision is a key component of how drivers acquire information, a direct relationship between quantified levels of visibility related issues and safety (in terms of crashes) should be identified. One of the factors affecting visibility is inclement weather. According to the Federal Highway Administration, inclement weather contributed to over 13 percent of injury crashes and over 10 percent of fatal crashes in 2013. To examine safety effects of inclement weather driving, the research team used three methods: 1) parametric model (ordinal logistic regression) development to quantify visibility issues by using Florida SHRP2 Roadway Inventory Database (RID), 2) non-parametric analysis (multiple correspondence analysis) to identify key associated factors for inclement weather crashes by using Washington SHRP2 RID, and 3) topic model development by analyzing inclement weather related crash narratives from SHRP2 Insight database. Data collected from the National Oceanic and Atmospheric Administration maintained airport weather stations that record visibility scores in real-time were linked with Florida RID. The findings indicate that, as expected, the likelihood of a crash increases during periods of low visibility, despite the tendency for less traffic and lower speeds to prevail during these times. The findings also identified key associating factors: younger and older drivers, low friction roadways, roadways with higher posted speed, curved roadways, undivided roadways, signalization, and no lighting at dark. The insights from this study could promote a better understanding of the safety impacts of reduced visibility related issues in inclement weather.

17. Key Words Visibility, Inclement weather, Ordinal logistic regression, Multiple correspondence analysis, Text Mining, Topic Modeling		18. Distribution Statement Unlimited	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 60	22. Price

Technical Report Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

Page

List of Figures v
List of Tables vi
Chapter 1: Introduction 1
    Problem 1
    Key Objective 1
    Research Approach 1
    Outline of the Report 2
Chapter 2: Literature Review 3
    Visibility 3
    Factors Other than Visibility 4
        Road Characteristics 4
        Environmental Characteristics 5
        Temporal Factors 6
        Other Factors 6
Chapter 3: Parametric Model on Visibility Issues 7
    Data Collection and Processing 7
    NOAA Weather Data 8
    SHRP2 RID Crash Data 9
    Multicollinearity Check 12
    Variable Importance 12
    Selected Variables 13
    Model Development with Ordinal Logistic Regression 15
    Odds Ratios 17
    Crash Probabilities by Visibility Level 17
    Findings 19
Chapter 4: Non-Parametric Model on Inclement Weather Crashes 21
    Multiple Correspondence Analysis 21
    Data Collection and Analysis 23
    Limitations 29
    Findings 29
Chapter 5: Analysis on Crash Narratives 31
    SHRP2 InSight 31
    Crash Narratives from SHRP2 InSight 32
    Comparison Word Clouds 33
    Structural Topic Modeling (STM) 35
    Findings 38
Chapter 6: Conclusions 39
References 41
Appendix 45

LIST OF FIGURES

Page

Figure 1. Flowchart of Data Compilation. 8
Figure 2. Airports, Buffer Areas, and Crashes in Florida. 10
Figure 3. Variable Importance from Random Forest Algorithm. 13
Figure 4. Histogram of Key Variables. 14
Figure 5. Probability of Occurrence for Different Variable Levels. 18
Figure 6. MCA Plot Different Categories (Das et al., 2017). 23
Figure 7. MCA Plot for Non-inclement Weather Crashes. 27
Figure 8. MCA Plot for Inclement Weather Crashes. 27
Figure 9. MCA Plot for Non-Inclement Weather Crashes (Center Zoomed). 28
Figure 10. MCA Plot for Inclement Weather Crashes (Center Zoomed). 29
Figure 11. Screenshot of SHRP2 NDS InSight Home Page. 32
Figure 12. Most Frequent Terms. 33
Figure 13. Comparison Word Cloud for Non-inclement Crash Narratives. 34
Figure 14. Comparison Word Cloud for Inclement Crash Narratives. 34
Figure 15. Structural Topic Model (Source:). 35
Figure 16. Top 10 Topic Models for Non-inclement Crash Narratives. 36
Figure 17. Top 10 Topic Models for Inclement Crash Narratives. 37
Figure 18. Network Plots from the Topic Models. 38

LIST OF TABLES

Page

Table 1. Studies Using Factors Other than Visibility. 5
Table 2. Descriptive Statistics.................... 11
Table 3. Estimates from Ordinal Cumulative Link Model (Model 1). 16
Table 4. Model Comparison. 17
Table 5. Odds Ratios for Model Variables. 17
Table 6. Descriptive Statistics of the Key Factors. 24
Table 7. Percent Variance Explained in Top 10 Dimensions. 25
Table 8. Parentage of Variations on Dimensions (Non-inclement Weather Crashes). 45
Table 9. Parentage of Variations on Dimensions (Inclement Weather Crashes). 46
Table 10. Coordinates of the Attributes on First Five Dimensions (Non-inclement Weather Crashes). 47
Table 11. Coordinates of the Attributes on First Five Dimensions (Inclement Weather Crashes).................... 49
Table 12. Sample of Crash Narratives for Non-inclement Weather Crashes. 51
Table 13. Sample of Crash Narratives for Inclement Weather Crashes. 52
Table 14. Sample SHRP2 Insight Final Narrative and Metadata. 53

CHAPTER 1: INTRODUCTION

PROBLEM

Inclement weather influences roadway safety, mobility, and productivity. It affects roadway safety through higher crash risk. Federal Highway Administration (FHWA) defines inclement crashes as “those crashes that occur in adverse weather (i.e., rain, sleet, snow, and/or fog) or on slick pavement (i.e., wet pavement, snowy/slushy pavement, or icy pavement)” (FHWA, 2016). Drivers tend to adapt their driving behavior to adjust the conditions presented by inclement weather. Depending on the surroundings and visibility issues, drivers drive more vigilantly by keeping longer headways, reducing operating speeds, or being more cautious. Approximately 24 percent of the U.S. traffic crashes happen due to inclement weather, resulting in 7,130 fatalities and 629,000 injuries (Murphy et al., 2012). According to FHWA, inclement weather contributed to over 13 percent of injury crashes and over 10 percent of fatal crashes in 2013. Moreover, there is a perception that transportation planners and managers can do little about weather induced crash prevention. Visibility conditions at the time of a crash are rarely documented at a high level of detail in the conventional police reports. While vision is a key component of how drivers acquire information, a direct relationship between quantified levels of visibility related issues and safety (in terms of crashes) needs to be identified. This research seeks to understand the ramifications of inclement weather on safety from a perspective of visibility and other key issues.

KEY OBJECTIVE

The key objective of this study is to examine safety impacts of reduced visibility and associated issues during inclement weather driving.

RESEARCH APPROACH

To perform a robust analysis on inclement weather crashes, the research team performed a systematic literature review to determine the key factors associated with inclement weather related crashes. The research team used the second Strategic Highway Research Program’s (SHRP2) Roadway Information Database (RID) database to determine the suitable states for data analysis. After performing quality check on the availability of the variables and percentage of missing information in each of the variables, the research team selected two states for analysis: Florida and Washington. To examine the safety impacts of inclement weather crashes, this study used three methods: 1) parametric model (ordinal logistic regression) development to quantify visibility issues by using Florida SHRP2 RID, 2) non-parametric analysis (multiple correspondence analysis [MCA]) to identify key associated factors for inclement weather crashes by using Washington SHRP2 RID, and 3) topic model development by analyzing inclement weather related crash narratives from SHRP2 Insight website.

OUTLINE OF THE REPORT

This report is organized as follows. Chapter 1 discusses research problem, key objective, and research approach. Chapter 2 provides the literature review on the existing body of research. Chapter 3 describes parametric statistic models, while Chapter 4 documents non-parametric analysis. Chapter 5 provides description on the text mining performed on the crash narratives. Chapter 6 describes conclusions.

CHAPTER 2: LITERATURE REVIEW

There is a substantial body of research and knowledge on inclement weather crashes. This study provides a brief summary of past studies.

VISIBILITY

Studies evaluating driver behavior in inclement weather have been oriented toward large-scale observations rather than changes at the level of the individual driver. For example, researchers identified how inclement weather affects traffic speed, flow, and density on freeways in Seattle, Baltimore, and Minneapolis-Saint Paul (Rakha et al., 2008). Light rain was found to reduce free-flow speed 2 to 3.6 percent, capacity 10 to 11 percent, and the speed at capacity 8 to 10 percent. The reductions on free-flow speed and capacity increased with rain intensity, and snow had a larger impact on free-flow speed and capacity than rain. The researchers also found that visibility affects traffic conditions. Free-flow speeds on freeways in inclement weather were also studied in Spain (Camacho et al., 2010), identifying a reduction in speed by 5.5 to 7 km/h for rain and 9 to 13.7 km/h for snow. Wind affects free-flow speed only when the wind speed is above 8 m/s. Visibility affects free-flow speed only when visibility is less than 2,000 m. In a separate study (Bartlett et al., 2013), wind speed was also shown to have a small effect on traffic compared to other weather factors. Analyzing traffic volumes, the researchers identified that volume is reduced 13 to 34 percent during inclement weather. Traffic volume and speed have also been observed to decrease during inclement weather in China (Shang et al., 2015).

In support of drivers responding to inclement weather by driving more conservatively, the research indicates that there is a change in traffic during inclement weather, not only in characteristics such as speed and capacity, but also in the number of vehicles on roads. The reduction in traffic volumes during inclement weather is a reflection of driving as a derived demand. Depending on the driver and the severity of the weather, it is worth postponing the commercial and recreational activities that would have occurred to a later time. While less driving occurs during inclement weather, there are also reductions in free-flow and operating speeds. This suggests that, while some drivers forego driving altogether, those that still choose to drive do so with some level of caution.

While many drivers naturally drive slower during periods of inclement weather and low visibility, some agencies use variable speed limits on changeable message signs to encourage all drivers to respond similarly. Hassan et al. (2011) surveyed drivers in Florida to investigate whether drivers respond to reduced speed limits in low visibility. Their models indicate that drivers 18 to 25 years and female drivers 51 years and older are more likely to reduce their speed in response to a variable speed limit when it is used during fog in low to medium-high traffic. Drivers are also more likely to reduce their speed if they are on a two-lane road.

Despite the behavioral changes triggered by inclement weather, there are safety concerns with driving in poor conditions. Several studies have focused on the effects of weather on crashes, including those related to visibility. One notable area for studying weather and visibility effects has been the state of Florida, which has several locations that frequently experience fog. Abdel-Aty et al. (2011) compared crashes occurring in Florida during periods of fog and smoke with crashes during periods of clear visibility. They identified that there are a disproportionate number of crashes in fog and smoke when the speed limit is 55 mph or higher, light conditions are dark, and there is no street lighting. An odds ratio analysis showed that the probability of a crash in fog or smoke is 3.24 times more likely to result in a severe injury and is 1.53 times more likely to be a multiple vehicle crash. Head-on collisions are also more likely. One interesting note is that likelihood of young and middle-aged drivers to be in a crash increased during fog and smoke, but not for old drivers, suggesting that old drivers tend to apply more caution in periods of low visibility. Also in Florida, Wang et al. (2015) studied crashes on expressway ramps during periods of low visibility, finding an increase in the likelihood of a crash as visibility decreases. Though not focused on visibility, Sun et al. (2011) calculated an increase in crash risk for rainy weather compared to dry weather. Das and Sun (2014) used crash data of Louisiana to investigate the pattern of crashes under rainy weather.

FACTORS OTHER THAN VISIBILITY

Although visibility is a key concern in inclement weather crashes, researchers have considered many additional variables in their studies. Table 1 enlists significant factors other than visibility and related studies. The factors are divided into four major groups: 1) road characteristics, 2) environmental characteristics, 3) temporal factors, and 4) other factors. A short review on these major groups is described in this section.

Road Characteristics

Researchers have found that median width is likely to reduce crash frequency, while curves, shoulder, and steep downgrade slopes may increase crash frequencies under adverse weather conditions (Yu et al., 2015). In addition, steep grades increase the severity of a crash (Yu and Abdel-Aty, 2014). Kopelias et al. (2007) found that driver behavior was more dominant factor in crash frequency than road characteristics. Other researchers have investigated the effects of road characteristics on crashes and found that the effects of road geometry were not as significant as other variables in the models (Wei et al., 2017; Naik et al., 2016; Shaheed et al., 2016; El-Basyouny et al., 2014b).

Table 1. Studies Using Factors Other than Visibility.

Factors other than Visibility	Study
Road characteristics	
Median width (ft)	Yu et al., 2015; Yu and Abdel-Aty, 2014
Shoulder width (ft)	Yu et al., 2015
Number of lanes	Yu et al., 2015; Yu and Abdel-Aty, 2014; Kopelias et al., 2007
Alignment	Yu et al., 2015; Yu and Abdel-Aty, 2014; Kopelias et al., 2007; Wei et al., 2017
Curvature	Yu et al., 2015; Yu and Abdel-Aty, 2014; Kopelias et al., 2007; Wei et al., 2017
Road surface	Naik et al., 2016; Shaheed et al., 2016; Wei et al., 2017
Lighting condition	Naik et al., 2016; El-Basyouny et al., 2014b; Wei et al., 2017
Posted speed (mph)	Xu et al., 2013; Kopelias et al., 2007
Environmental characteristics	
Temperature	El-Basyouny et al., 2014a; Naik et al., 2016; El-Basyouny et al., 2014b; Yu and Abdel-Aty, 2014
Wind speed	Usman et al., 2012; Naik et al., 2016
Precipitation intensity	Yu et al., 2015; Usman et al., 2012; Xu et al., 2013; Qiu et al., 2008
Adverse weather type	Chakrabartya and Gupta, 2013; El-Basyouny et al., 2014a; Wei et al., 2017; Naik et al., 2016; Qiu et al., 2008; El-Basyouny et al., 2014b
Temporal factors	
Time of the day	Naik et al., 2016; Wei et al., 2017
Day of the week	Brijs et al., 2008; Shankar et al., 1995; El-Basyouny et al., 2014a; Kopelias et al., 2007; El-Basyouny et al., 2014b
Other factors	
Season	El-Basyouny et al., 2014b
AADT (vpd)	Ahmed et al., 2014; Xu et al., 2013; Kopelias et al., 2007; McCann and Fontaine, 2016
Truck percentage	Yu et al., 2015; Yu and Abdel-Aty, 2014; Kopelias et al., 2007
Speed	Xu et al., 2013; McCann and Fontaine, 2016;
Collision type	El-Basyouny et al. 2014a
Injury severity	Wei et al., 2017; Shaheed et al., 2016
Contributing Factor	Naik et al., 2016
Major harmful event	Shaheed et al., 2016
Severity	Shaheed et al., 2016; El-Basyouny et al., 2014b; Qiu et al., 2008
Driver age	Naik et al., 2016
Driver gender	Shaheed et al., 2016
Registered vehicles	El-Basyouny et al., 2014b

Environmental Characteristics

El-Basyouny et al. (2014a) determined that wind gusts were mostly insignificant with a few exceptions. Factors such as wind speed, rain, snow, and lower temperatures had a positive correlation with increase in crash rates (Naik et al., 2016; El-Basyouny et al., 2014a; El-Basyouny et al., 2014b). Qiu et al. (2014) reviewed crash records from previous studies. The study determined that snow has a greater effect on crash occurrence than rain. According to the study results, snow increases the crash rate by 84 percent and injury rate by 75 percent.

Temporal Factors

Researchers found that duration of daylight was negatively related to crashes (El-Basyouny et al., 2014b). Other researchers have used these variables to account for fluctuations in traffic volume and visibility.

Other Factors

Researchers have investigated the effects of other factors that may be affected by inclement weather or low visibility. The effect of the season during the year was found to be significant by a recent study (El-Basyouny et al., 2014b). When looking at the effects of adverse weather on crash type, El-Basyouny et al. (2014b) found that inclement weather increases all crashes between 9 to 73.7 percent, and the highest increase was observed on run-off-the-road crashes. Ahmed et al. (2014) found that higher AADT during low-visibility periods led to a higher crash frequency. Xu et al. (2013) observed that speed difference had the largest impact during low-visibility periods. McCann and Fontaine (2016) pointed out that although motorists slow down during low visibility conditions, there is a significant difference between observed speeds during the study and safe speed calculated using stopping sight distance. Naik et al. (2016) attempted to determine the effects of contributing factors such as driver age and alcohol consumption, among other factors, and found that the effects varies across individual driver crash injury severity.

CHAPTER 3: PARAMETRIC MODEL ON VISIBILITY ISSUES

One of the primary causes of vehicular crashes is human error. This error ranges from complete negligence (e.g., distracted or impaired driving) to limitations of human abilities (e.g., slower reflexes with age, low visibility in inclement weather). One limitation that is often neglected is reduced visibility during inclement weather. The reduction of crashes in inclement weather will not be fully realized unless key association factors are thoroughly investigated. This study seeks to identify the effects of reduced visibility on the likelihood of crashes and the factors that influence crashes during periods of reduced visibility.

Inclement weather presents a safety concern for vehicular traffic from multiple perspectives. One is the moisture on the road that reduces friction. Friction is reduced even more if the temperature is near freezing. Another safety concern is impaired visibility from precipitation on the windshield. Windshield wipers, despite their utility in removing the precipitation, can be distracting and intermittently impede the driver's vision. The moisture droplets in the air also impact visibility. Visibility changes based on the size and concentration of the water droplets.

Visibility can be measured with specialized instrumentation. The National Oceanic and Atmospheric Administration (NOAA) regularly measures atmospheric and ground-level conditions, including visibility, at airports so pilots and air traffic control can make informed decisions for flying. These readings are stored in historical databases maintained by NOAA. This study used historical weather data to identify the relationships between visibility and crashes.

While low-visibility events tend to be associated with weather, there is one notable exception. Smoke is not a direct result of weather or precipitation, but usually a consequence of human dealings. The impact of smoke on crashes, in addition to the effects of fog, has been studied in previous work (Abdel-Aty et al., 2011). Smoke-related visibility events are infrequent and have a dramatically different quality than periods of reduced visibility caused by precipitation and moisture (such as fog or rain) because there is no water on the windshield nor need to use wipers. For consistency in the analysis, events with smoke are not of interest in this study.

DATA COLLECTION AND PROCESSING

The research team assembled a database of data collected in Florida from two different sources. These two sources are: 1) National Oceanic and Atmospheric Administration (NOAA) airport weather station data, and 2) SHRP2 RID data. Figure 1 illustrates a framework of the data compilation work to prepare the final dataset.

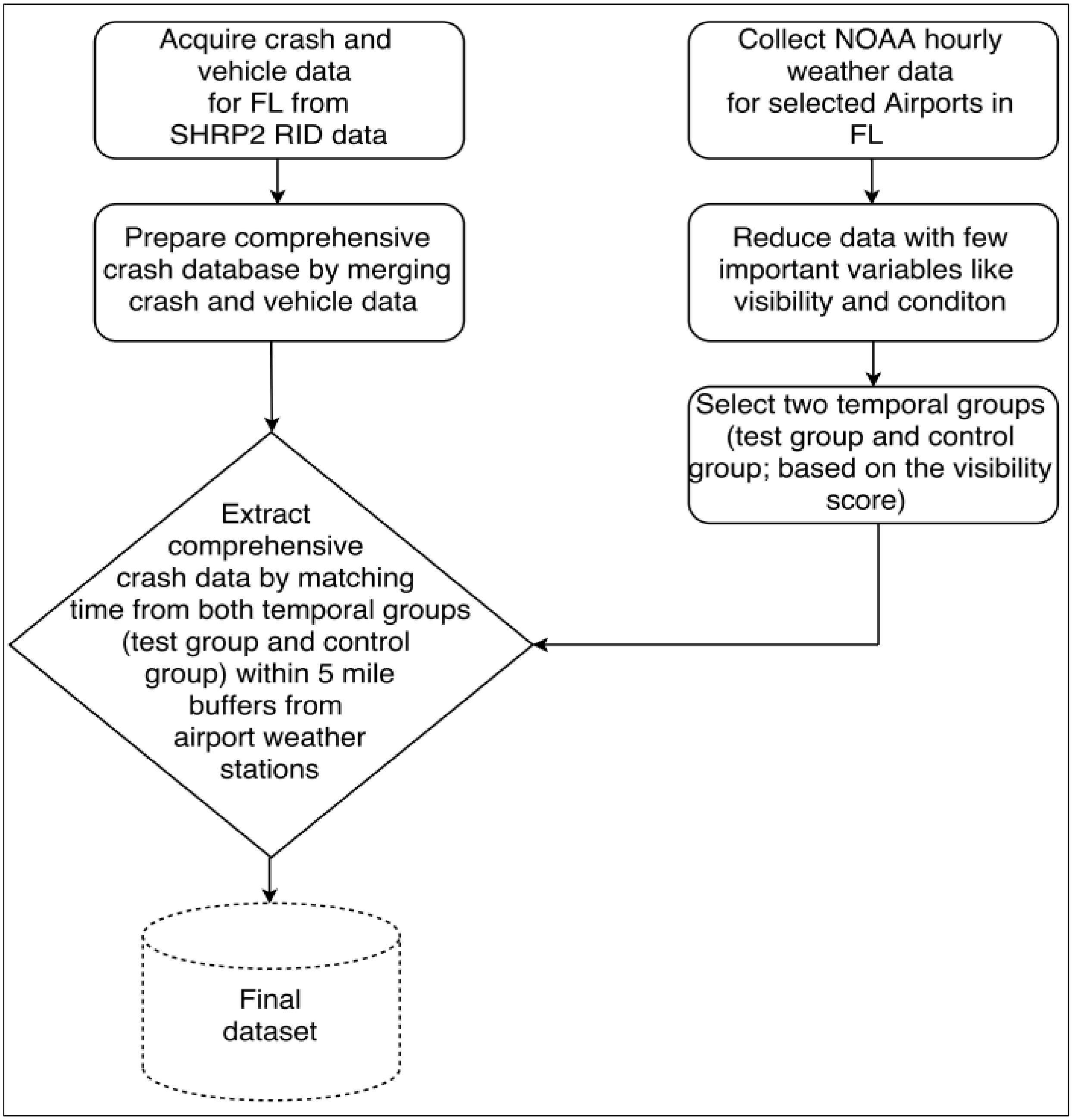


Figure 1. Flowchart of Data Compilation.

NOAA WEATHER DATA

The weather data used in the analysis was extracted from data collected by weather stations at airports in Florida. The data are now available in databases maintained by NOAA. The equipment at each weather station reported hourly measurements. Because timely visibility information is important at airports, weather measurements are made more frequently whenever there is a change in the recorded visibility distance. Having more observations increases the precision with which periods of low visibility are identified. Researchers sifted through the weather data to identify periods when the visibility was within a defined range of poor visibility (0–0.5 mi) and medium visibility (0.5–4.0 mi). Each entry in the original weather data included a visibility level (in miles), temperature, and code for type of weather.

The categorical options for type of weather include events such as rain, thunderstorm, drizzle, mist, hail, snow, and fog. Modifiers light or heavy can be applied to indicate a heavy thunderstorm or light rain. Because different types of weather events cause reduced visibility, it was critical to have consistency in the types of weather observed within the same category of visibility. Fog is the primary weather event associated with poor visibility; mist and rain are the primary weather events for medium visibility. Periods that had any record of snow, freezing temperatures, thunderstorms, and hail were removed as these incidences are very limited in numbers in Florida. Smoke was also occasionally the cause of low visibility. Periods with smoke were removed from the analysis to focus exclusively on reduced visibility associated with moisture.

Each period of reduced visibility was matched with a period of normal visibility (9–10 mi) precisely one week earlier or one week later. This matching procedure, similar to the method used by Sun et al. (2011), produced a control sample. This method addresses spatial and temporal variations that are found in crash frequencies, where crashes can be more frequent during certain times of the year or in certain locations. Time periods with reduced visibility that could not be matched with a control period of normal visibility (either one week earlier or one week later) were removed. Both matches were removed if either the control or the test period included a state or national holiday or other day near the holiday when travel would be different from that of a typical day. If visibility is assumed to have no effect on crash frequencies, the number of crashes observed during the tested periods of reduced visibility should be equivalent to the number of crashes observed during the matched control period.

SHRP2 RID CRASH DATA

SHRP2 data consist of two databases: 1) RID, and 2) Naturalistic Driving Study (NDS). The RID contains crashes, roadway characteristics, and traffic volumes in the format of GIS layers from six states: Florida, New York, Washington, Pennsylvania, Indiana, and North Carolina. Crash records from 2010–2012 were acquired from the SHRP2 Florida RID data and included both the crash-level and vehicle-level fields. These data were merged to create a comprehensive dataset of crash variables. The recorded locations of the crashes were used to isolate the crashes to those occurring 5 miles within an airport, and the recorded times of the crashes were used to further reduce the crash dataset to those crashes occurring during a period of interest, whether during a test period of reduced visibility or a control period of normal visibility. Figure 2 shows a map graphically showing the airports, the 5-mile buffers, and observed crashes.

The raw crash data, compiled from law enforcement records and supplemented with roadway information from SHRP2, contain several fields that can be tested as variables. While the focus of this study is visibility, the influence of the road features and conditions at the time of the crash should be accounted for in studying how visibility impacts crash frequencies. There were several variables in the crash data that were often not reported. These variables were excluded from the analysis so efforts could be directed at the variables that were consistently reported.

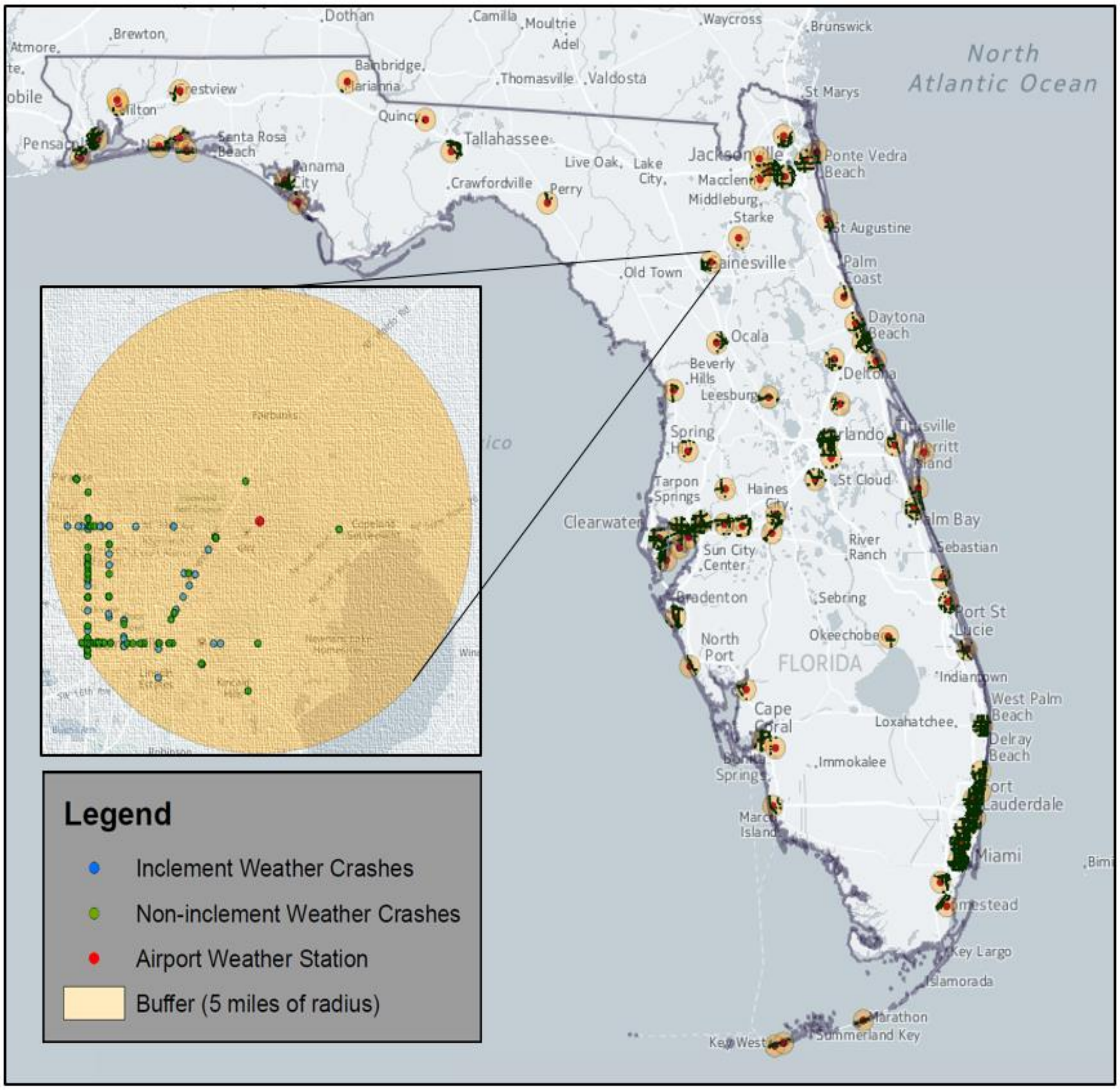


Figure 2. Airports, Buffer Areas, and Crashes in Florida.

Table 2 shows descriptive statistics of the variables that were tested. The information in Table 2 shows how often a particular variable category is observed in the crash records during a control period (excellent visibility) or one of the test periods (medium visibility or poor visibility). From the descriptive statistics, for example, 16.6 percent of crashes in excellent visibility occur when the maximum speed (speed limit) is 45 to 60 mph. Crashes increased to 19.7 percent when visibility is poor. For injury severity, there appears to be an increase in the percentage of crashes that were recorded as injury crashes, from nearly 48 percent of crashes in excellent visibility to 50 percent in poor visibility. The percentage of crashes reported as property damage only decreases from over 51 percent in excellent visibility to less than 49 percent in poor visibility. This suggests that crash severity is likely to increase as visibility decreases. The analyses that follow are able to account for other variables to test that observation.

Table 2. Descriptive Statistics.

	Percentage of Crashes				Percentage of Crashes		
Category	**Excellent Visibility**	**Medium Visibility**	**Poor Visibility**	**Category**	**Excellent Visibility**	**Medium Visibility**	**Poor Visibility**
Functional Class				**Facility Type**			
Rural major collector	0.02%	0.01%	0.00%	Divided	67.21%	69.08%	69.69%
Rural minor arterial	0.17%	0.17%	0.14%	Undivided	31.42%	29.54%	29.33%
Rural principal arterial	0.82%	0.79%	0.75%	Unknown	1.36%	1.39%	0.98%
Urban collector	0.57%	0.60%	0.23%	**Skid Number**			
Urban local	0.01%	0.01%	0.00%	20–30	6.90%	7.80%	7.81%
Urban minor arterial	21.22%	18.71%	18.62%	30–40	67.73%	67.46%	65.39%
Urban other principal	53.91%	53.48%	56.69%	40–50	20.24%	19.27%	21.05%
Urban principal arterial	19.71%	22.09%	19.93%	> 50	1.23%	0.83%	1.22%
Unknown	3.57%	4.14%	3.65%	Unknown	3.91%	4.65%	4.54%
Maximum Speed (mph)				**Lighting Condition**			
0–30	7.89%	7.55%	7.76%	Dark(No Street light)	3.42%	3.02%	5.75%
30–45	65.24%	61.73%	62.82%	Dark(Street Light)	28.29%	23.34%	38.63%
45–60	16.62%	18.52%	19.74%	Dawn/Dusk	5.01%	4.87%	6.22%
>60	10.25%	12.20%	9.68%	Daylight	63.29%	68.77%	49.39%
AADT (vpd)				**Number of Vehicles**			
0–9999	3.76%	3.66%	3.41%	Multi-Vehicle	92.31%	91.66%	91.25%
10,000–34,999	35.08%	33.46%	35.08%	Single Vehicle	7.69%	8.34%	8.75%
35,000–54,999	30.92%	30.14%	30.92%	**Severity**			
55,000–124,999	16.47%	18.03%	16.79%	Fatal	0.90%	0.56%	0.98%
>124,999	12.85%	13.81%	12.96%	Injury	47.96%	47.14%	50.09%
Unknown	0.92%	0.91%	0.84%	No Injury	51.14%	52.30%	48.92%
Percentage of Trucks				**Driver Age**			
0–5	58.91%	56.90%	60.62%	15–19	6.24%	6.58%	6.74%
5–10	35.19%	36.25%	32.32%	20–29	26.14%	26.10%	27.08%
10–20	5.75%	6.55%	6.78%	30–39	19.78%	18.70%	19.27%
>20	0.15%	0.30%	0.28%	40–49	18.63%	18.28%	18.29%
Avg. Shoulder Width (ft.)				50–59	15.56%	13.89%	12.96%
0.00–1.00	39.41%	38.73%	40.74%	60–69	8.37%	7.94%	7.39%
1.01–3.00	16.67%	15.88%	15.34%	> 70	5.19%	4.84%	4.07%
3.01–5.00	12.44%	11.22%	11.46%	Unknown	0.09%	3.68%	4.21%
5.01–10.00	23.72%	26.00%	25.12%				
> 10.00	6.19%	7.35%	6.50%				
Unknown	1.57%	0.83%	0.84%				

MULTICOLLINEARITY CHECK

The variance inflation factor (VIF) is used to detect collinearity (strong correlation between two or more predictor variables) that causes instability in parameter estimation in regression models. VIF can be defined as:

$$VIF = 1/(1 - R_i^2) \quad (1)$$

Where:

R_i^2 = co-efficient of determination of *i*th variable on all other variables.

A general rule of thumb for multicollinearity is to check whether VIF is greater than 10. No variable has a VIF value greater than 10 in the final dataset. The correlation values of Lighting, Weather, and Number of vehicles are closer to 1, suggesting that these variables have strong linear relationship with at least one other variables among these three variables.

VARIABLE IMPORTANCE

The research team used a random forest algorithm assess variance importance using a package with the R statistical software (Liaw and Wiener, 2002). With random forest algorithms, a randomly selected vector of input variables ($\boldsymbol{X} = X_1, \dots, X_n$) to a random response variable $Y \in \mathbb{y}$ is considered for analysis. The importance of a variable X_k while predicting or estimating Y is calculated by the Gini index, calculated from adding the impurity decreases in the following equation:

$$Importance(X_k) = \frac{1}{N_T} \sum_{T} \sum_{t \in T : v(s_t) = X_k} p(t).d \quad (2)$$

where:

t= node.

T= all nodes.

$p(t)$ = proportion N_t / N of sample for node t.

s_t = split for which all variables are sampled into two major nodes t_L and t_R to maximize the decrease, d.

$v(s_t)$ = variable used in split s_t.

$d = i(t) - p_L i(t_L) - p_R i(t_R)$ = decrease.

N_T = all variables.

If the Gini index is considered an impurity function, the measurement is known as Mean Decrease Gini. Variables with higher values of Mean Decrease Gini are considered more important to the model. Figure 3 shows the variable importance plot for the selected variables. Each variable is shown on the y-axis and variable importance on the x-axis. The variables on the y-axis are ordered from most to least important.

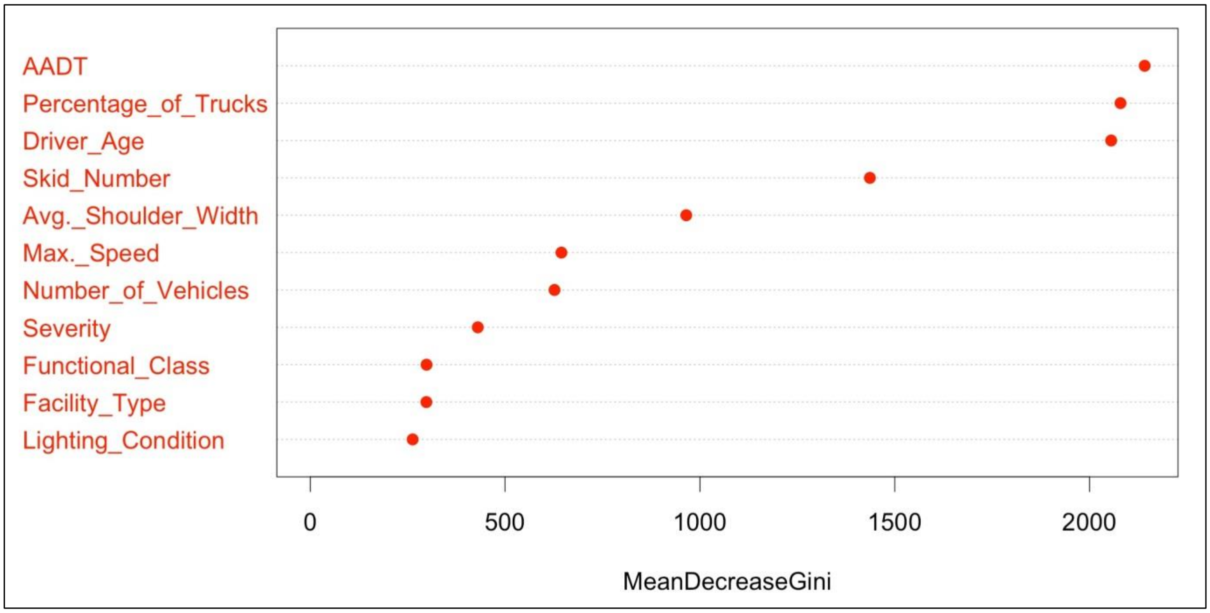


Figure 3. Variable Importance from Random Forest Algorithm.

SELECTED VARIABLES

Seven variables were selected for testing in models based on values of VIF, correlation, and Mean Decrease Gini. The selected variables are:

- AADT.
- Driver Age.
- Percentage of Trucks.
- Skid Number.
- Average Shoulder Width.
- Maximum Speed.
- Crash Severity.

Histograms of the first six of these variables are shown in Figure 4, divided by crashes occurring in the three visibility groups. The distributions appear to be quite similar for all variables. The percentage in each of the groups indicates relative percentages of the attributes in that particular group. The most notable differences are for skid number and driver age where there are peaks in the distributions for excellent visibility that are not seen in the other visibility levels.

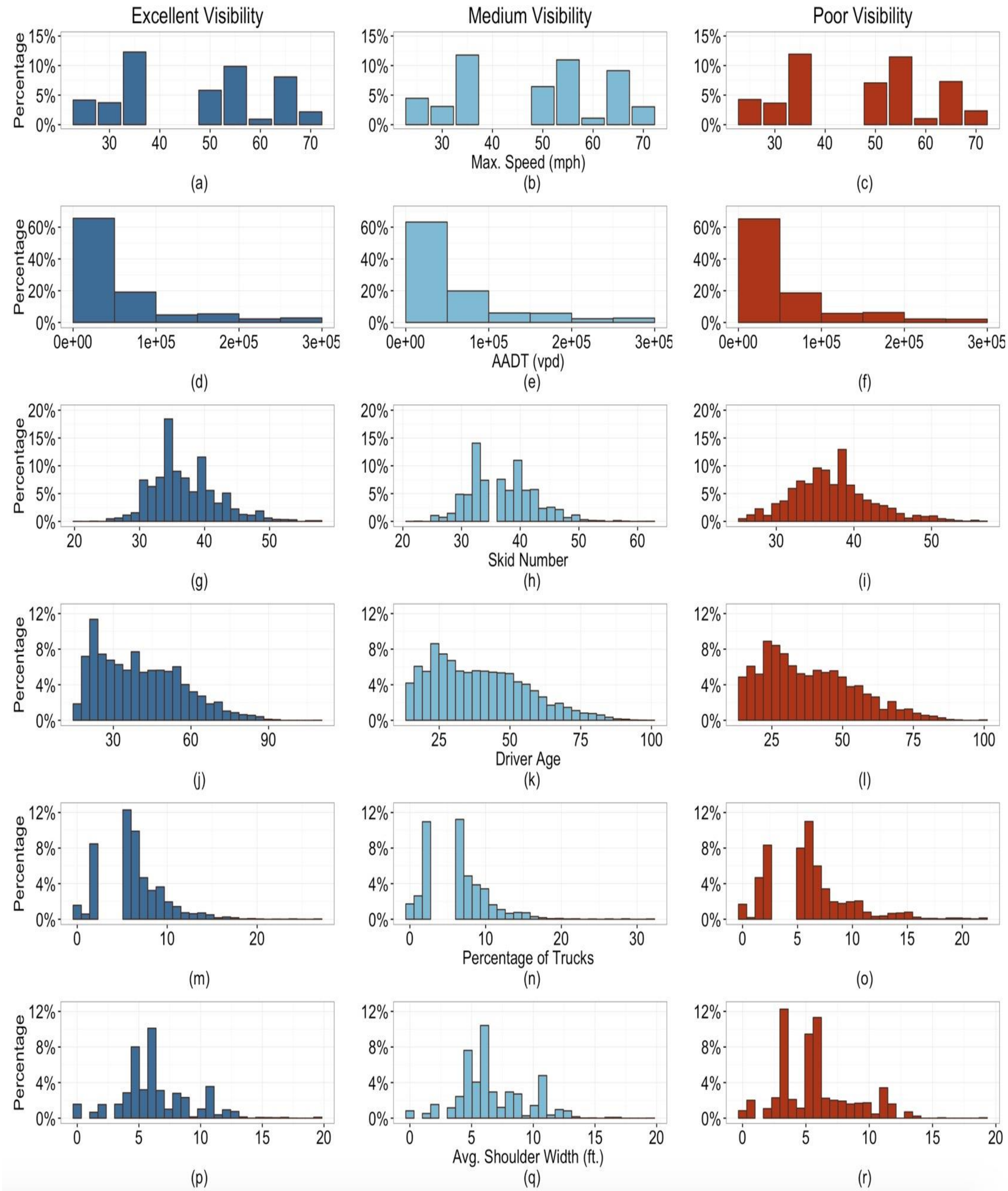


Figure 4. Histogram of Key Variables.

MODEL DEVELOPMENT WITH ORDINAL LOGISTIC REGRESSION

The research team used ordinal logistic regression to perform the analysis. The ordinal logistic regression models are also known as cumulative link models. In this analysis, the response variable is the visibility score. Three ordinal response categories are considered: excellent visibility, medium visibility, and poor visibility. A cumulative link model will be developed based on the ordinal response variable, for example, Y_i with $k = 1, \dots, K$ categories [where $K \geq 2$]. Then Y_i follows a multinomial distribution with parameter π. The cumulative probabilities can be defined as:

$$Y_{ik} = P(\,Y_i \leq k) = \pi_{i1} + \ldots + \pi_{ik} \tag{3}$$

Where:

π_{ik} = probability of ith observation for response category k.

The cumulative logistic function is:

$$logit(y_{ik}) = logit\big(P(\,Y_i \leq k)\big) = \log[\frac{P(\,Y_i \leq k)}{1 - P(\,Y_i \leq k)}]\; where\!: k = 1, \dots, k-1 \tag{4}$$

Note that the logit functions are defined as $\text{logit}(\pi) = \log[\frac{\pi}{1-\pi}]$. A cumulative link with a logistic link can be written as:

$$logit(y_{ik}) = \theta_k - X\mu \tag{5}$$

Where:

θ_k = parameters act as intercepts or horizontal displacements.
X = transpose of a vector of predictor variables for the ith observation.
μ = matching set of regression parameters.

$X\mu$ is dependent on k categories. Therefore, it is considered that μ has the same effect for each of the $K - 1$ cumulative logits. The odds ratio (OR) of the event $Y(\leq k)$ at x_1 relative to event $Y(\leq k)$ at x_2 is:

$$Odds\ Ratio\ (OR) = \frac{\dfrac{y_k(x_1)}{[1 - y_k(x_1)]}}{\dfrac{y_k(x_2)}{[1 - y_k(x_2)]}} = \frac{exp(\theta_k - X\mu)}{exp(\theta_k - X\mu)} = exp[(x_2^T - x_1^T)\mu] \tag{6}$$

Here, the odds ratio is independent of k. Thus, the cumulative odds ratio is proportional to the distance between x_1 and x_2. Therefore, the cumulative logit model is also known as proportional odds model. The analysis was performed with open source statistical software (R Development Core Team, 2013; Christensen, 2016). Table 3 lists the values of the estimates of the first model. The first part of the outputs lists the regression coefficient values, standard errors, and p- values. By observing the p-values, it is found that maximum speed, skid number, and driver age have higher significance, and severity and AADT have somewhat significance. One reason is that the impact of AADT is not highly significant in the model, as it is not considered in log scale. The 2.5 percent and 97.5 percent confidence interval values of these variables do not contain zero, which is also a good indicator of significance. The next part of Table 3 lists the estimates for the two intercepts (intercept between excellent visibility and medium visibility and intercept between medium visibility and poor visibility) to form three response categories. The intercepts indicate where the latent variable is cut to make the ordered subdivision in visibility score. The final part of the outputs provide $-2 loglikelihood$ of the model and the AIC value, used later for model selection.

Table 3. Estimates from Ordinal Cumulative Link Model (Model 1).

Variables and Statistical Measures	**Combined**						
	Estimate	**2.50%**	**97.50%**	**St. Er.**	**z**	**Pr(>|z|)**	**Significance**
Variables							
Severity: Injury	0.265	-0.032	0.562	0.152	1.750	0.080	.
Severity: NoInjury	0.263	-0.034	0.560	0.151	1.736	0.083	.
Maximum Speed	0.008	0.004	0.012	0.002	4.048	0.000	***
AADT	0.000	0.000	0.000	0.000	-2.536	0.011	*
Skid Number	-0.006	-0.010	-0.003	0.002	-4.209	0.000	***
Percentage of Trucks	0.001	-0.008	0.010	0.005	0.197	0.844	
Avg. Shoulder Width	0.003	-0.007	0.013	0.005	0.536	0.592	
Driver Age	-0.006	-0.007	-0.004	0.001	-7.273	0.000	***
Intercepts							
Visibility: Excellent | Medium	0.124			0.171	0.722		
Visibility: Medium| Poor	2.462			0.172	14.305		
Statistical Measures							
AIC	43888.17						
Log likelihood	-21934.08						
Maximum Gradient	6.38E-07						
Conditional H	4.60E+12						

*Note: p < 0.10, *: p < 0.05, **: p < 0.01, ***: p < 0.001*

The research team developed another model (model 2) by omitting percentage of trucks and average shoulder width, because both of these factors were insignificant in Model 1. Table 4 compares the two models. The p-value of the test is 0.836, meaning that the Model 2 is not significantly different from Model 1. As Model 2 is not significantly different from Model 1, the research team used Model 1 as the final model.

Table 4. Model Comparison.

Variables	AIC	loglikehood	LR Stat	Pr (> Chisq)
Model 2	43885	-21934		
Model 1	43888	-21934	0.3582	0.836

ODDS RATIOS

The interpretation of proportional odds ratios in cumulative link models is same as the odds ratios from a binary logistic regression model. Table 5 lists the odds ratio of the variables for the comparison of poor visibility to excellent and medium visibility. For injury crashes, the odds of injury severity in poor visibility versus medium visibility or excellent visibility combined is 1.304, indicating a 30 percent greater chance of a crash resulting in injury during poor visibility (assuming all other variables are held constant). For continuous variables, the odds ratio indicates the change in likelihood of a poor visibility crash for each unit change in the listed variable. For the maximum speed (odds ratio is 1.008 per 5 mph), the probability that the visibility at the time of the crash is poor (instead of medium or excellent visibility) increases 1.008 times higher for each 5 mph increase in speed. The odds ratio for skid number indicates there is a decrease in the likelihood of a poor visibility crash for increases in skid number. The odds ratio for AADT is 1, which indicates that the effect of AADT is negligible when considering the effect of visibility on crashes.

Table 5. Odds Ratios for Model Variables.

Variables	Odds Ratio
Severity: Injury	1.304
Severity: No Injury	1.301
Maximum Speed	1.008 (per 5 mph)
AADT	1.000 (per 10,000 vpd)
Skid Number	0.994 (per 5 units of skid resistance)
Percentage of Trucks	1.001
Average Shoulder Width	1.002
Driver Age	0.994 (per 10 years of age)

CRASH PROBABILITIES BY VISIBILITY LEVEL

Using the model results, the probabilities of a crash with a specific visibility level are plotted in Figure 5, divided by the injury severity and independent variable. Plots for four independent variables are shown (the variables shoulder width and percentage of trucks are not shown as they were not significant in the original model). These plots can be used to examine how the probability that a particular visibility level is represented in the crash data changes as each independent variable changes.

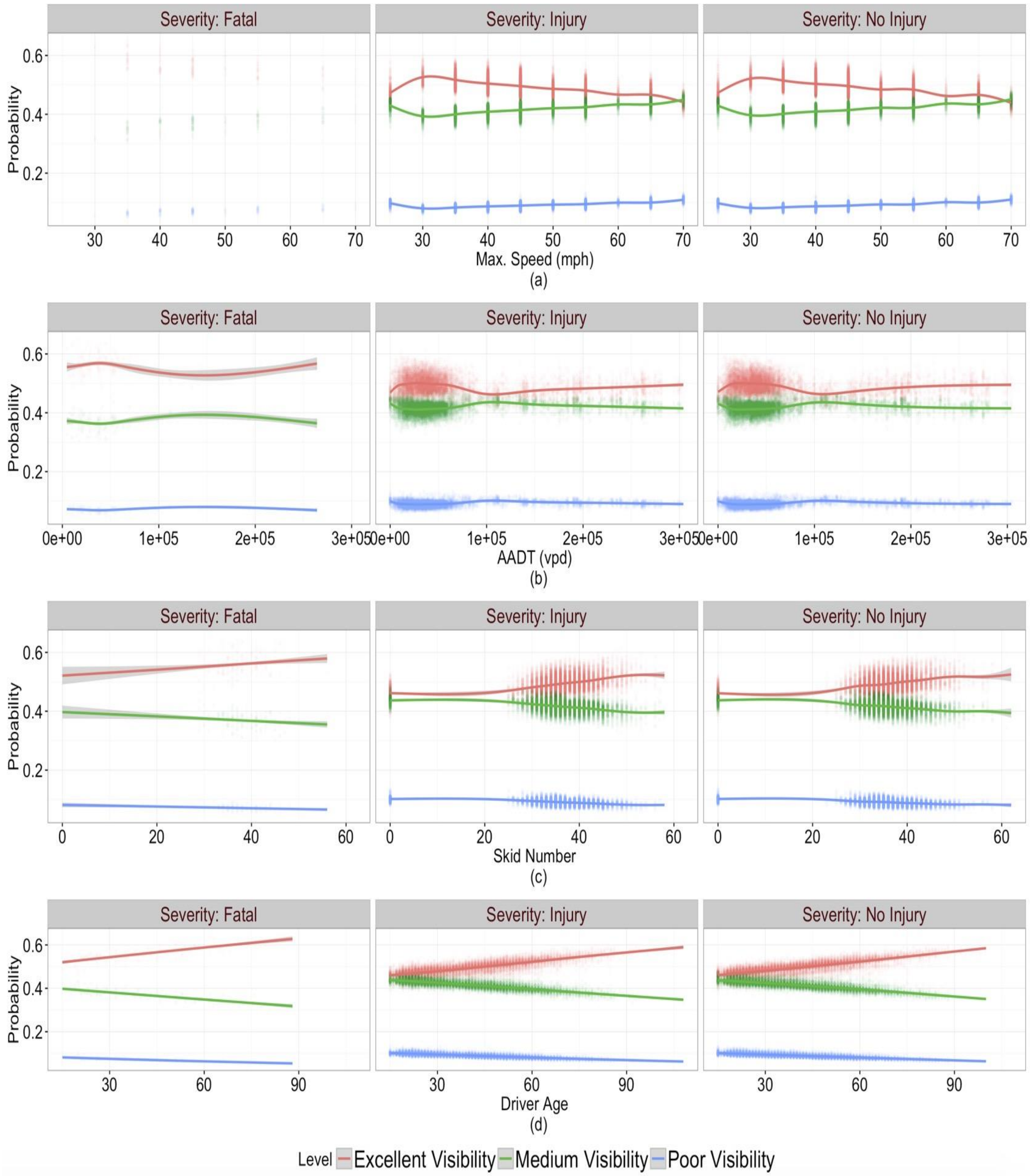


Figure 5. Probability of Occurrence for Different Variable Levels.

In Figure 5a, there were not enough crashes to identify a connection between maximum speed and visibility level for fatal crashes. However, as the maximum speed increases, there is a decrease in the likelihood that an observed crash occurred in excellent visibility compared to medium and poor visibility for both the injury and no injury crashes. There appear to be no common trends for AADT in Figure 5b, but Figure 5c suggests that increases in measured friction lead to crash reductions in medium and poor visibility. The probability that a crash from the dataset is coded as occurring in medium and poor visibility decreases while the probability for excellent visibility increases. This happens for all severity levels. Finally, driver age is shown as affecting the crash probabilities for all severity levels as well. The probability for a crash occurring in medium and poor visibility decreases compared to that of excellent visibility as age increases. This should not be surprising as drivers are less likely to drive in inclement weather as they age.

FINDINGS

This study evaluated several of the factors that influence crashes during reduced periods of visibility during inclement weather. While previous research confirms that driver behavior changes during inclement weather, it is clear that there remain safety concerns due to reduced visibility. Unless technology is able to adapt to the conditions drivers regularly drive in and overcome the limitations that human drivers have, the crash patterns are likely to continue with the same effects identified in this study. Ahmed et al. (2014) were unable to connect roadway geometry and traffic to models that identify effects of visibility conditions. The findings of this study concur that there is difficulty finding geometric or traffic effects when focusing on visibility. Shoulder width and truck percentage were not significant effects, and AADT has small consequence. The variables that appear most interesting are skid number and driver age, suggesting that higher friction reduces crashes in inclement weather and the old drivers are less likely to be in crashes in inclement weather. Moreover, higher speed is associated with fatal and injury crashes during reduced visibility conditions.

The findings suggest that drivers are less likely to be involved in a crash during poor visibility conditions, as they get older. This is not surprising as older drivers are likely to apply caution and simply not drive in inclement weather (Office of the Aging, 2015). A large part of the reduced traffic volumes in inclement weather as observed by Bartlett et al. (2013) is likely to first come from the older drivers that tend to be more cautious, as suggested by Abdel-Aty et al. (2011).

One limitation of the current study is the less availability of more comprehensive data like NDS data, curve characteristics, roadway marking, and roadway signage data. Future researchers may choose additional constructs and related variables to further examine the effect of inclement weather.

CHAPTER 4: NON-PARAMETRIC MODEL ON INCLEMENT WEATHER CRASHES

Many factors could be associated with an inclement weather crash. Since the objective of this study is to determine if there is an association between crash risk and driving in inclement weather conditions, select variables from past research were considered in the analysis. One of the major tasks in highway safety analysis is the identification of the key contributing factors for different types of crashes. MCA, a popular non-parametric tool, is a dimensionality reduction method to describe the significance of co-occurrence of groups of variables or variable categories from a big dataset. It is also referred to as the pattern recognition method that treats arbitrary datasets as combination of points in a larger dimensional space. It uniquely simplifies complex data into knowledge extraction in a completely different way than parametric estimation does (Das and Sun, 2016).

The objectives of using MCA in this study were: (1) to identify the relative closeness of the key association factors and (2) to identify significant knowledge patterns. The findings of this study could help authorities to determine effective and efficient crash countermeasures.

MULTIPLE CORRESPONDENCE ANALYSIS

French Statistician Jean-Paul Benz´ecri (Roux and Rouanet, 2010) introduced the idea of MCA in 1970. This method has similarities with Principal Component Analysis (PCA) and Factor Analysis, two well-documented multivariate statistical methods. PCA mainly deals with numerical data, and MCA is a familiar tool for multidimensional categorical data.

The concept of MCA has been improved many times under different frameworks while keeping the goals similar (De Leeuw, 1973: Hoffman and De Leeuw, 1992). Limited number of transportation engineering related studies used MCA. Hoffman and De Leeuw (1992) associated different vehicle models with crash severities by using MCA. Fontaine (1995 performed MCA on one year of pedestrian crash data to determine the statistical proximity of the key variables. Factor et al. (2010) applied MCA in determining the association between driver’s social characteristics and their involvements in crash related injuries. This study exposed new facets in the social organization of fatalities. Das and Sun (2015) used eight years (2004–2011) of pedestrian crash data in Louisiana to determine key associations between risk factors. Das and Sun (2016) applied MCA on eight years (2004–2011) of fatal run-of-road crashes in Louisiana to examine the degree of association between risk factors. The finding revealed some specific factor groups that require careful attention from the safety professionals.

MCA is a non-parametric learning algorithm. In MCA, one does not need to distinguish between explanatory variables and the response variable. It requires the construction of a matrix based on pairwise cross-tabulation of each variable. For example, the dimension of the final dataset of a study is: $A \times b$. For a table of qualitative or categorical variables with dimension $A \times b$, MCA can be explained by taking an individual record (in row), *i [i= 1 to A]*, where *b* categorical variables (represented by *b* columns) have different sizes of categories. MCA can generate the spatial distribution of the points by different dimensions based on these *b* variables. For theoretical details, readers can consult related studies (Roux and Rouanet, 2010; Das et al., 2015; Das et al., 2016a, Das et al., 2017a).

Let P be the number of variables (i.e., columns) and I is the number of transactions (i.e., rows). This will generate a matrix of *I multiplied by P*. If L_p is the number of categories for variable p, the total number of categories for all variables can be expressed as $L = \sum_{p=1}^{P} L_p$. In matrix *I multiplied by L*, each of the variables will contain several columns to show all of their possible categorical values.

The cloud of categories is considered as a weighted combination of J points. Category j is represented by a point denoted by C^j with weight of n_j. For each of the variables, the sum of the weights of category points is n. In this way, for the whole set J the sum is nP. The relative weight w_j for point C^j is $w_j = n_j/(nP) = f_j/P$. The sum of the relative weights of category points is 1/P, which makes the sum of the whole set as 1.

$$w_j = \frac{n_j}{nP} = \frac{f_j}{P} \quad \textit{with} \; \sum_{j \in J_q} w_j = \frac{1}{P} \qquad \textit{and} \; \sum_{j \in J} w_j = 1$$

Here, $n_{jj'}$ represents the number of individual records, which have both categories k and k'. The squared distance between two categories C^j and $C^{j'}$ can be represented by:

$$(C^j C^{j'})^2 = \frac{n_j + n_{j'} - 2n_{jj'}}{n_j n_{j'} / n} \tag{7}$$

The numerator of Equation (4) is the number of individual records associating with either j or j' but not both. For two different variables, p and p', the denominator is the familiar theoretical frequency for the cell (j, j') of the $J_p \times J_{p'}$ two-way table.

Figure 6 shows a representation of the cloud generation for combination of categories. In this figure, three variables are considered. Variable A has three categories (A_1, A_2, and A_3), variable B has four categories (B_1, B_2, B_3, and B_4), and variable C has six categories (C_1, C_2, C_3, C_4, C_5, and C_6). The categories are plotted in the MCA plot representing their relative proximity on the two-dimensional space. The plot shows a distinct cloud (red ellipse) associated with A_2, B_1, C_2, and C_6. This cloud is created based on the proximity of the coordinates of these four categories.

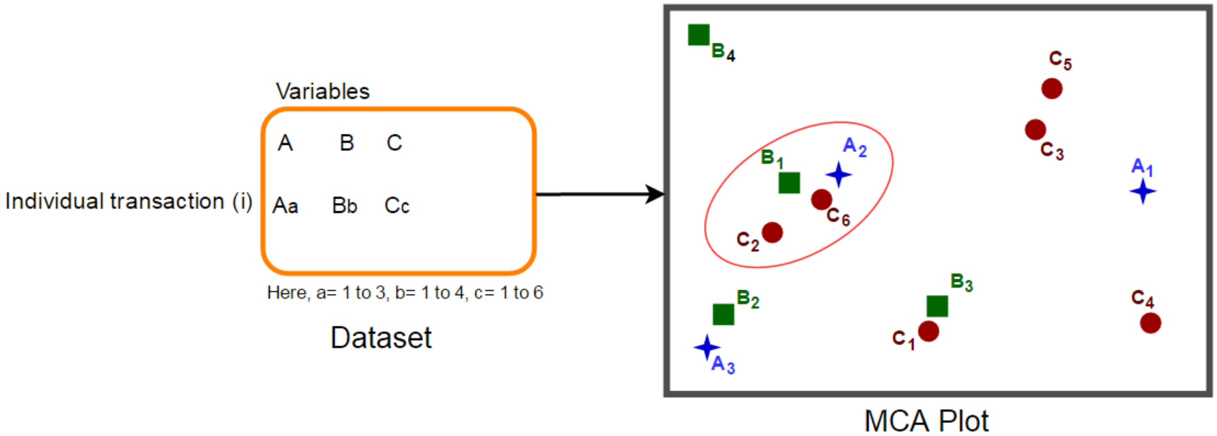


Figure 6. MCA Plot Different Categories (Das et al., 2017a).

DATA COLLECTION AND ANALYSIS

To accomplish the objectives of this study, the research team examined the influence of significant variables by conducting a systematic literature review. Florida and Washington RID do not have all of the required variables to examine. This study used MCA for two reasons:

- Conventional statistical models require assumptions. Deviation from the assumptions will make the research findings biased and trivial. As a non-parametric method, MCA does not need to consider any prior assumptions, so the findings from MCA are unbiased.
- Identification of key associations is difficult to measure in statistical modeling. This study also considers additional variables listed in Washington RID to perform a robust analysis on inclement weather crashes.

This study used five years (2009–2013) of Washington RID to perform this analysis. After conducting variable importance (as performed in Chapter 3), a final list of nine variables were selected. Table 6 provides descriptive statistics of the key variables for both non-inclement and inclement weather crash datasets.

Table 6. Descriptive Statistics of the Key Factors.

Category	Percentage of Crashes: Non-inclement Weather	Percentage of Crashes: Inclement Weather
Functional Class		
Interstate	47.89%	53.90%
Principal Arterial	37.92%	33.62%
Minor Arterial	11.28%	10.07%
Collector	2.86%	2.40%
Local	0.04%	0.01%
Lighting		
Daylight	75.96%	61.04%
Dark-Street Lights On	13.13%	22.54%
Dark-No Street Lights	6.57%	9.68%
Dawn/Dusk	3.60%	5.72%
Dark-Street Lights Off	0.42%	0.60%
Other	0.32%	0.42%
Road type		
Two Way - Undivided	44.31%	44.60%
Two Way - Divided, with Barrier	39.15%	39.69%
Center-Two Way Left Turn Lane	7.31%	6.43%
One Way	2.81%	2.25%
Two Way - Divided, no Barrier	1.96%	2.36%
Driveway	1.68%	1.44%
Interchange Ramp	1.21%	1.58%
Reversible Road	0.27%	0.36%
Alley	0.01%	0.00%
Other	1.29%	1.28%
Posted Speed (mph)		
0–40	41.25%	36.43%
41–50	12.99%	13.05%
51–60	41.12%	45.43%
61–70	4.64%	5.09%

Category	Percentage of Crashes: Non-inclement Weather	Percentage of Crashes: Inclement Weather
Location		
Urban	82.94%	82.80%
Rural	17.06%	17.20%
Alignment		
Straight & Level	68.01%	60.93%
Straight & Grade	18.99%	22.85%
Curve & Level	5.55%	6.12%
Curve & Grade	5.48%	7.66%
Straight at Hill crest	0.69%	0.78%
Straight in Sag	0.52%	0.63%
Curve in Sag	0.20%	0.29%
Curve at Hillcrest	0.15%	0.23%
Other	0.40%	0.51%
Driver Gender		
Male	53.16%	60.93%
Female	43.18%	42.65%
Not included	3.66%	3.97%
Driver Age		
15–24	25.79%	26.76%
25–34	21.17%	22.15%
35–44	16.59%	16.71%
45–54	15.97%	15.72%
55–64	11.70%	11.30%
65–75	5.66%	4.85%
> 75	3.12%	2.52%
Driver Severity		
PDO	65.72%	64.40%
Injury	41.53%	35.50%
Fatal	0.64%	0.46%

Graphical representations in MCA help to interpret data in a convenient way as they effectively summarize large, complex datasets by simplifying the structure of the associations between variable categories with a relatively simple view of the data. MCA analyzes the rows and columns of a dataset while treating them as high-dimension geometry elements. It shows the co-occurrence of the categories in a lower dimensional space where proximity in the space potentially indicates meaningful associations among the categories of different variables. A larger distance indicates a distant association. If the distance for a particular category is very far away from the centroid, it indicates that such category is different from the average profile. Each of the categories is independent in principle and a co-occurrence based on weight proximity (by associating certain categories together in a cloud) tends to form a complete picture of certain scenarios. A percentage distribution for a single category conveys little meaning in many cases, but when combined with other categories due to the proximity in the space, various implications can be potentially interpreted (Das et al., 2017a).

Table 7 shows the percentages of variance explained by the top 10 dimensions. The first principal axis explained 6.05 percent of the principal inertia; the second principal axis explained 4.64 percent for non-inclement weather crashes. For inclement weather crashes, these percentages are 6.14 percent and 4.74 percent, respectively. These percentages are calculated based on the eigenvalue (a value in between 0 and 1). Dimension with larger variance has a higher eigen-value magnitude. First two dimensions explain around 11 percent of variance, and none of the remaining major dimensions explain more than 2.88 percent. As the first plane associated dimension 1 and dimension 2 represented the largest inertia, the remaining analysis was based on this plane. A complete list of the percentage of variance is provided in the Appendix (Table 8 and Table 9).

Table 7. Percent Variance Explained in Top 10 Dimensions.

Dimension	**Non-inclement Weather**			**Inclement Weather**		
	Eigenvalue	Percent of Variance	Cumulative percent of Variance	Eigenvalue	Percent of Variance	Cumulative percent of Variance
Dimension 1	0.269	6.047	6.047	0.273	6.140	6.140
Dimension 2	0.206	4.642	10.689	0.210	4.736	10.876
Dimension 3	0.128	2.880	13.569	0.127	2.868	13.744
Dimension 4	0.124	2.795	16.364	0.124	2.795	16.538
Dimension 5	0.121	2.730	19.094	0.121	2.732	19.270
Dimension 6	0.120	2.696	21.790	0.119	2.680	21.950
Dimension 7	0.118	2.649	24.439	0.118	2.657	24.608
Dimension 8	0.116	2.615	27.055	0.117	2.626	27.234
Dimension 9	0.114	2.573	29.627	0.115	2.593	29.826
Dimension 10	0.114	2.559	32.186	0.115	2.584	32.411

The contribution of the variables depends on its number of categories, whereas the contribution of a category depends on the number of occurrences. This study used open source R software package FactoMineR to perform MCA (Le et al., 2008). This package also produces related graphics from the analysis. The research team used two other visualization packages ggplot2 and ggrepel to develop MCA plot shown in Figure 7–Figure 10 (Wickham, 2009; Slowikowski, 2016). The flexibility of functions in these two packages makes the graphics neat and easy to interpret. The research team used similar combination number (for example, Cloud 1a for non-inclement weather crashes, and Cloud 1b for inclement weather crashes) to compare the results effectively. Descriptions of these combination clouds are:

- *Cloud 1 (Cloud 1a and Cloud 1b):* Cloud 1a consists of three attributes: *Rural-Fatal-Curve at Hill Crest* and Cloud 1b consists of *Rural-Fatal-Dark No Lighting at Night.* Due to high speed, rural roadway crashes are usually more severe in nature. This cloud indicates that dark with no lighting was a significant factor for inclement weather crashes. Other studies (Naik et al., 2016; El-Basyouny et al., 2014b; Wei et al., 2017) showed similar findings.
- *Cloud 2 (Cloud 2a and Cloud 2b):* Cloud 2a has four attributes. The attributes in this cloud are: *Curve and level-Curve* and *grade- Curve in sag- Interchange ramp*. Cloud 2b has all alignment related attributes. These are *Curve and level-Curve* and *grade-Curve in Hillcrest- Curve in sag- Interchange ramp*. Cloud 2 indicates that curve roads were risky irrespective of weather condition. Moreover, ramps are found associated with non-inclement weather crashes. Similar findings were found in other studies (Yu et al., 2015; Yu and Abdel-Aty, 2014; Kopelias et al., 2007; Wei et al., 2017).
- *Cloud 3 (Cloud 3a and Cloud 3b):* Both Cloud 3a and Cloud 3b consist of four attributes: *Minor Arterial-Posted speed (41–50 mph)-Two way undivided roadways-Older drivers.* This combination clearly indicates particular age groups faced difficulty for certain roadway properties.
- *Cloud 4 (Cloud 4a and Cloud 4b):* The attributes in cloud 4a are: *Principal Arterial-Alley-Driveway-Two way undivided with no Barrier- Lower posted speed*. Cloud 4b has attributes like *Principal Arterial-Local-Driveway - Lower posted speed*. This cloud indicates that driving during inclement weather was risker in the presence of higher number of driveways.

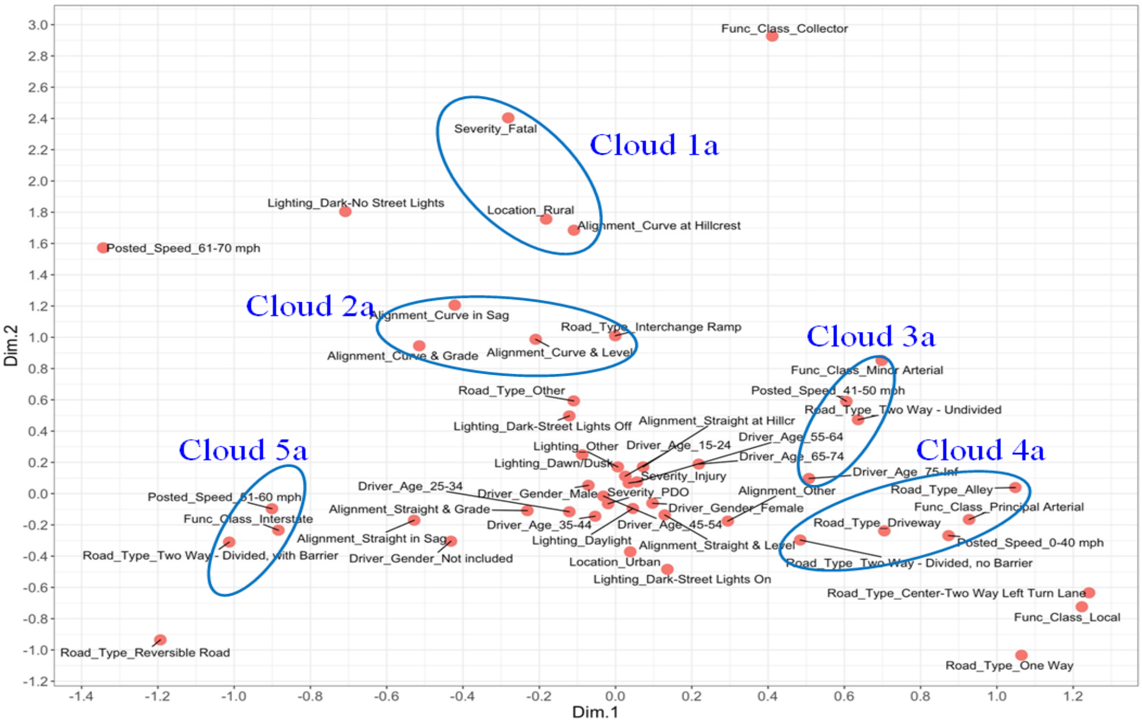


Figure 7. MCA Plot for Non-inclement Weather Crashes.

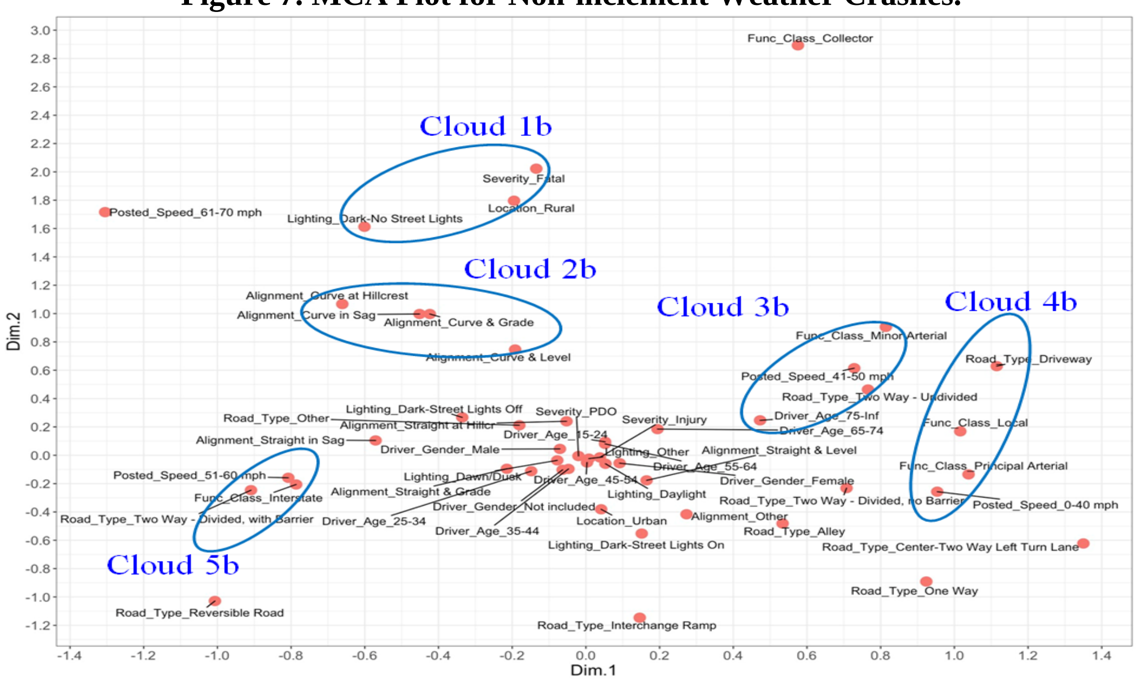


Figure 8. MCA Plot for Inclement Weather Crashes.

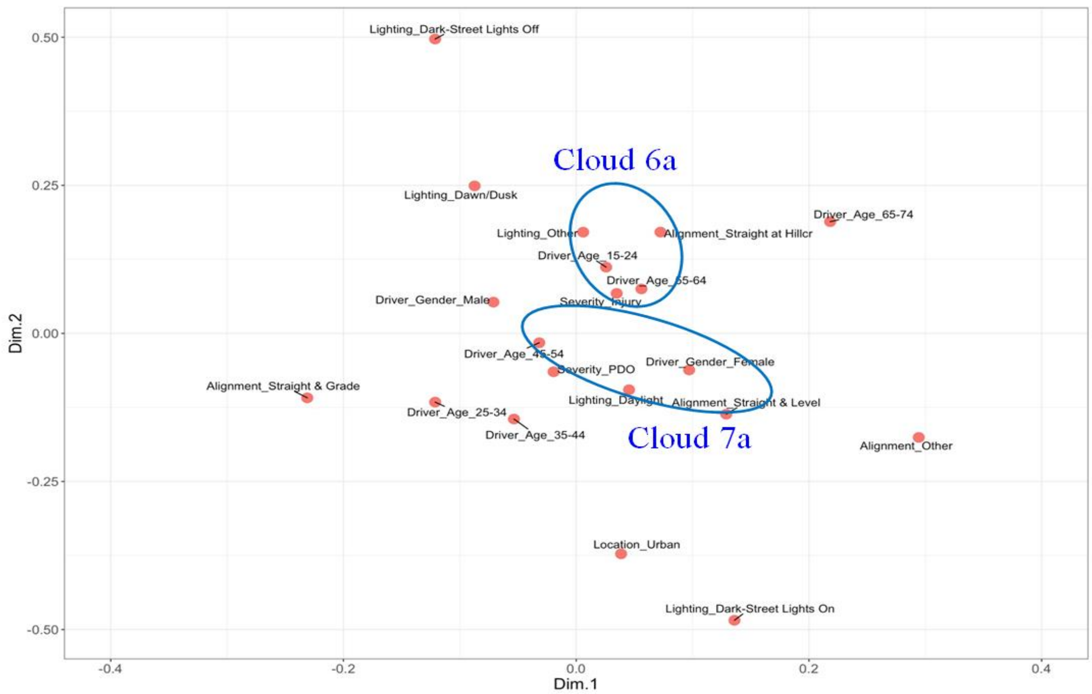


Figure 9. MCA Plot for Non-Inclement Weather Crashes (Center Zoomed).

- *Cloud 5 (Cloud 5a and Cloud 5b):* Both Cloud 5a and Cloud 5b consist of four attributes: *Posted speed (51–60 mph)-Interstate-Two way divided roadway with Barrier.* This combination indicates that two lane instate roadways separated with barrier were associated with higher number of crashes.
- *Cloud 6 (Cloud 6a and Cloud 6b):* Cloud 6a associates driver age and severity with lighting condition. This cloud is consisted of four attributes: *Driver age (15–24, 55–64 years)-Injury-Other types of Lighting*. On the other hand, Cloud 6b for inclement weather crashes contains only age groups in the combination group. It shows that younger and older drivers were disproportionately involved in inclement weather crashes. Naik et al. (2016) also found that young driver indicator was statistically significant in inclement weather crashes.
- *Cloud 7 (Cloud 7a and Cloud 7b):* Cloud 7a combines five attributes: *Driver age (45–54 years)-Female-PDO-Straight aligned-Daylight*. Cloud 7b has six attributes: *Driver age (45–54 years)-Female-PDO and Injury -Daylight* and other types of Lighting. This cloud indicates that a certain age group of female drivers was disproportionately involved in crashes.

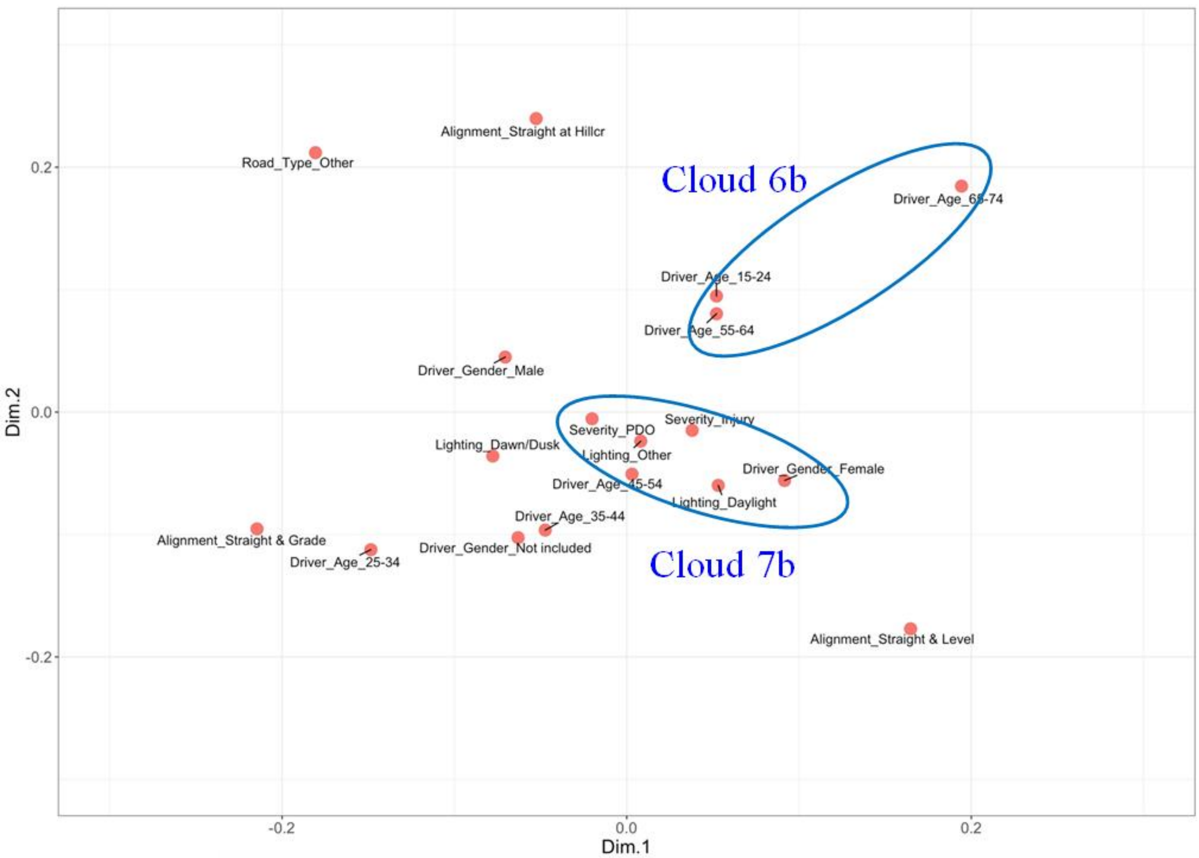


Figure 10. MCA Plot for Inclement Weather Crashes (Center Zoomed).

LIMITATIONS

A graphical display like MCA plot is useful as it shows multifaceted associations involving many factors into a lower dimensional space. One disadvantage of MCA is that it fails to quantify any significance test for the cloud groups. Other cluster techniques like K-means clustering and principle component analysis (PCA) have similar strengths and disadvantages as the MCA (Das et al., 2016a , Das et al., 2017a). Future research can improve the findings of MCA by connecting it statistical models like log-linear and risk models. Another limitation is that the study findings are based on the first plane that explained only 11 percent of the data inertia. Inclusion of more dimensions can potentially increase the number of association patterns underlying the dataset.

FINDINGS

MCA is useful in presenting proximity map of the variable categories in a low dimensional plane by revealing the main features from a multidimensional dataset. This study used MCA on five years (2009–2013) of crash data in Washington. Systematic literature review of the past studies was used to identify the key variables. The key groups of confluence of factors in inclement weather crashes found in this research are the following:

- Older drivers face difficulty in certain roadway characteristics (two way undivided arterial roadways with posted speed 41–50 mph) during inclement weather.
- Rural roadways are more involved in fatal crashes. For inclement weather, rural roadways with no lighting at dark are more risky.
- Roadways with curves are risky irrespective of weather condition.
- Younger and older drivers seem to appear as a meaningful factor by itself for inclement weather crashes.
- Interstate two lane roadways with barriers are always more likely to be involved with crashes.

Results from this study could help transportation agencies improve conventional warning systems and countermeasures for inclement weather driving.

CHAPTER 5: ANALYSIS ON CRASH NARRATIVES

Police reported crash report contains a final narrative of a crash occurrence. It explains how the crash happened. These narrative descriptions usually elaborate on the coded variables. The research team envisioned to use the crash narratives of inclement and non-inclement weather crashes to clarify the range and nature of events under each condition. To obtain further insight into the inclement weather crashes, this study extracted and analyzed crash narrative data (SHRP2 InSight, 2016).

TEXT MINING

Text data usually lack any formal structure. Moreover, text data contain a significant amount of redundant words, prepositions, conjunctions, and punctuations. Sophisticated data mining tools are required to perform text analysis. Text mining is an applied method originated from data mining or knowledge discovery. This method is used to identify effective, original, potentially useful, and ultimately comprehensible patterns in unstructured textual data. Knowledge Discovery in Text can be viewed as a multistage process that comprises all activities from document collection to knowledge discovery. It uses approaches like data mining, information retrieval, supervised and unsupervised machine learning, and computational semantics. Extraction of useful information from data resources through pattern recognition helps identifying contributing factors in associated task. In text mining, corpus represents a collection of text documents. After developing a corpus, it is important to clean the data by removing redundant words, numbers, and particular parts of speech.

SHRP2 INSIGHT

SHRP2 InSight (screenshot of InSight data access website is shown in Figure 11) contains the following data components:

- Information describing the 3,400+ drivers and vehicles that participated in the NDS.
- More than 5,400,000 trip summary records that describe individual trips recorded during the study.
- SHRP2 NDS status information including data collection and processing progress.
- More than 36,000 crash, near-crash, and baseline driving events.
- Background information about the project and the data being collected.
- Discussion forums for questions about the project and available data.

Figure 11. Screenshot of SHRP2 NDS InSight Home Page.

CRASH NARRATIVES FROM SHRP2 INSIGHT

There is limited amount of research conducted on crash narrative analysis. Stutts et al. (2001) used two years of NASS Crashworthiness Data System crash narratives to identify additional information on distracted driving. Fitzpatrick et al. (2017) investigated speeding-related crash designation through crash narrative reviews sampled via logistic regression.

To perform text mining, final narratives from SHRP2 InSight crashes were manually extracted. Samples of these crash narratives are listed in the Appendix (Table 12, Table 13, and Table 14). The objective of crash narrative analysis in this study was to perform text mining and develop topic models. The research team used structural topic modeling (STM) to develop the topic models. This method requires both narrative and supportive metadata to develop relevant topic models. A database of 100 crash narratives was manually extracted from SHRP2 Insight. Around 20 percent of the data contain final narratives of inclement weather. Figure 12 illustrates the most frequently cited terms found in the combined corpus from each group of crash narratives. The terms ‘driver,’ ‘lane,’ ‘intersection,’ ‘brake,’ and ‘light’ are found as the most frequent terms in both of the datasets. For inclement weather crash narratives, ‘rain/snow’ and ‘wet’ are found in the top 10 most frequent terms.

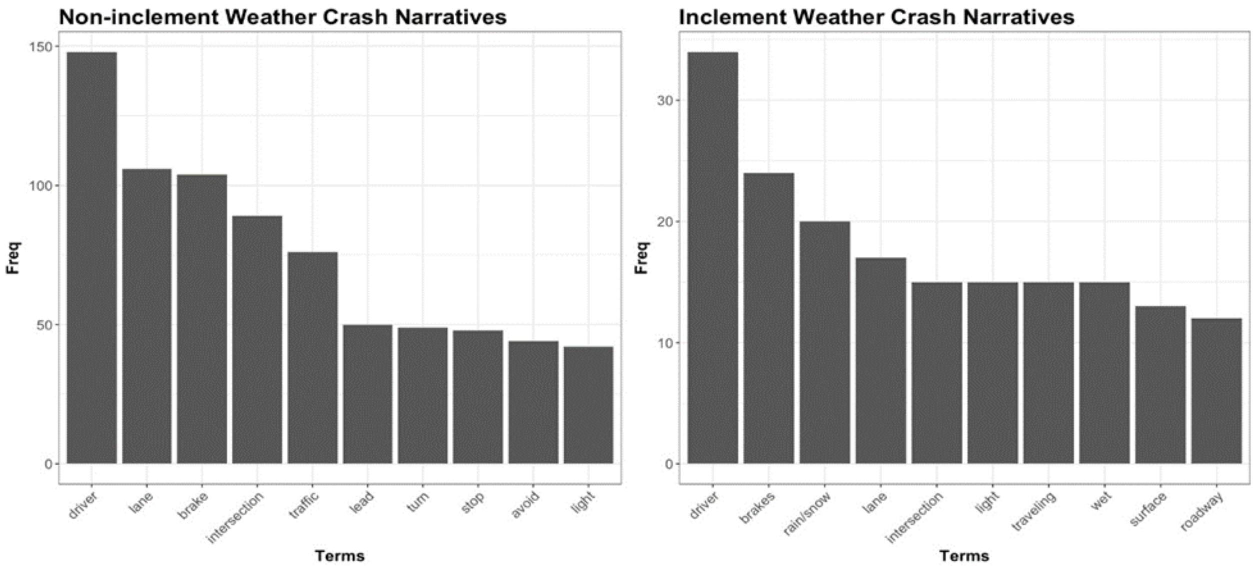


Figure 12. Most Frequent Terms.

COMPARISON WORD CLOUDS

A word cloud is the most convenient way of visualizing the most frequent terms in unstructured documents. The more frequently a word appears, the larger would be the text size (Ignatow and Mihalcea, 2016). This study developed comparison word clouds to visualize the research trends over time. If $p_{x,y}$ is the rate at which word x occurs in document y, and p_y is the average rate across documents $\left(\frac{\sum_y p_{x,y}}{n}\right)$, where n is the number of documents. In comparison clouds, the size of each word is mapped to its maximum deviation ($max_y(p_{x,y} - p_y)$), and its angular position is determined by the document in which that maximum occurs (Das et al., 2016b). Figure 13 shows comparison word cloud for non-inclement crash narratives. Different color denotes different states, for example, words in green color indicate that these words are extracted from Florida non-inclement crash narratives. Texts with larger fonts (for example, ‘stop,’ ‘driver,’ ‘oncoming’) indicate that these words appeared in the crash narratives more frequently. Figure 14 illustrates comparison word clouds for inclement weather crashes. Relevant terms like ‘snow,’ ‘rain,’ and ‘surface’ are noticed in this word cloud.

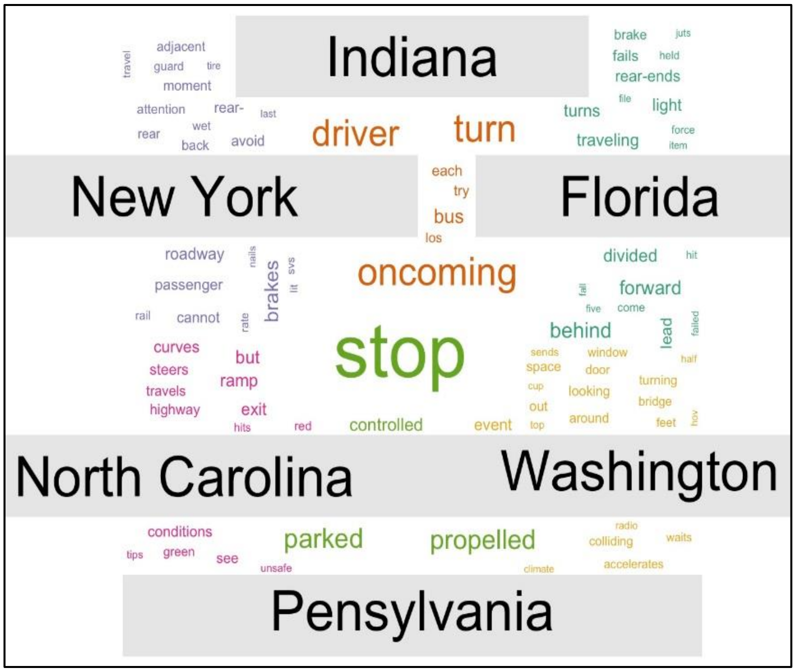


Figure 13. Comparison Word Cloud for Non-inclement Crash Narratives.

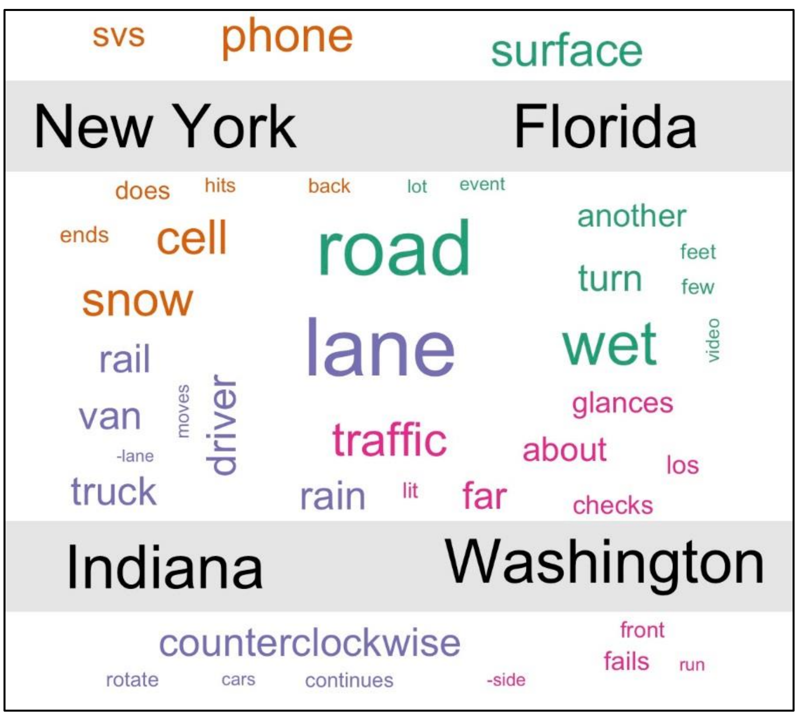


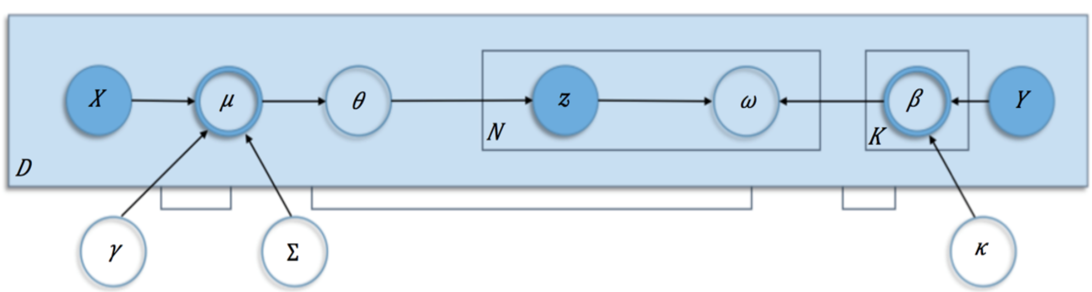
Figure 14. Comparison Word Cloud for Inclement Crash Narratives.

STRUCTURAL TOPIC MODELING

Topic models are extensively used in many branches of science to identify trends, patterns, and knowledge extraction. In transportation engineering, there is limited number of studies that used topic modeling. Das et al. (2016b) used Latent Dirichlet Allocation (LDA) to identify research trends by using several years of abstracts from Transportation Research Board Annual Meeting compendium papers. Das et al. (2017b) used STM on the same dataset to develop topic models by using document-level covariates.

LDA lacks additional document-level information on which variation can be seen in different theoretical interests. Applying LDA and then performing a post-hoc evaluation of variation with a certain covariate of interest can be considered as a plausible solution. STM accommodates corpus structure through document-level covariates. The key idea behind STM is to specify the priors as generalized linear models through which one can weigh on arbitrary observed data. This directly allows an estimation of the quantities of interest in the unstructured textual contents (Roberts et al., 2016; Das et al., 2017b).

Figure 15 shows three existing models used in the STM structure: the correlated topic model, the Dirichlet-Multinomial Regression topic model, and the Sparse Additive Generative topic model.



Topic Prevalence:

$$\mu_{d,k} = X_{d\gamma_k}$$
$$\gamma_k \sim \mathcal{N}\left(0, \sigma_k^2\right)$$
$$\sigma_k^2 \sim Gamma(s^\gamma, r^\gamma)$$

Language Model:

$$\theta_d \sim LogisticNormal(\mu_d, \Sigma)$$
$$z_{d,n} \sim Mult(\theta_d)$$
$$\omega_{d,n} \sim Mult(\beta_d^{k=z_{d,n}})$$

Topical Content:

$$\beta_{d,v}^k \propto \exp(m_v + \kappa_v^{.,k} + \kappa_v^{y,.} + \kappa_v^{y,k})$$
$$\kappa_v^{y,k} \sim Laplace(0, \tau_v^{y,k})$$
$$\tau_v^{y,k} \sim Gamma(s^\kappa, r^\kappa)$$

Figure 15. Structural Topic Model (Source: Das et al., 2017b).

The notations used for the theoretical part are:

$d \in \{1 \ldots . D\}$= index of the documents.

$n \in \{1 \ldots . N\}$= index of the tokens in the documents.

$v \in \{1 \dots . V\}$= index of a vocabulary of words.

$\omega_{d,n}$ = observed token (a conditionally independent drawn from $v \in \{1 \dots . V\}$).

$k \in \{1 \dots . K\}$= index of topics.

X= topic prevalence matrix with dimension D by P.

Y= topic content matrix with dimension D by A.

m_v = baseline log-frequency of each word in the vocabulary.

s, r, σ^2, ρ= hyper parameters.

Customization of a topic model to a particular dataset involves specifying a model for the linear predictors of topic prevalence (permits the expected document-topic proportions to vary by covariates) and/or topical content (parameterizes the distribution over words as deviations in log-space from a corpus-wide baseline). STM uses a semi-collapsed variational Expectation-Maximization (EM) algorithm to fit the model. Regularizing prior distributions are used to enhance interpretation and prevent overfitting. In the E-step, the joint optimum of the document's topic proportions and the token-level assignments are solved. In the M-step, the global parameters are inferred that control the priors on topical prevalence and content (Roberts et al., 2016). This study used 'maneuver information,' 'driver behavior,' 'event information' as metadata to develop topic models. Figure 16 lists the corpus level visualization of the top topics from a 10-topic model from non-inclement crash narratives and the frequency of words in each of these topics. Topic 1 has high expected topic proportions. The theme of each topic is expressed by the three top words in each topic. For example, Topic 9 is related to signalized intersection. Similarly, Topic 3 indicates content related to rear end crashes at the intersections.

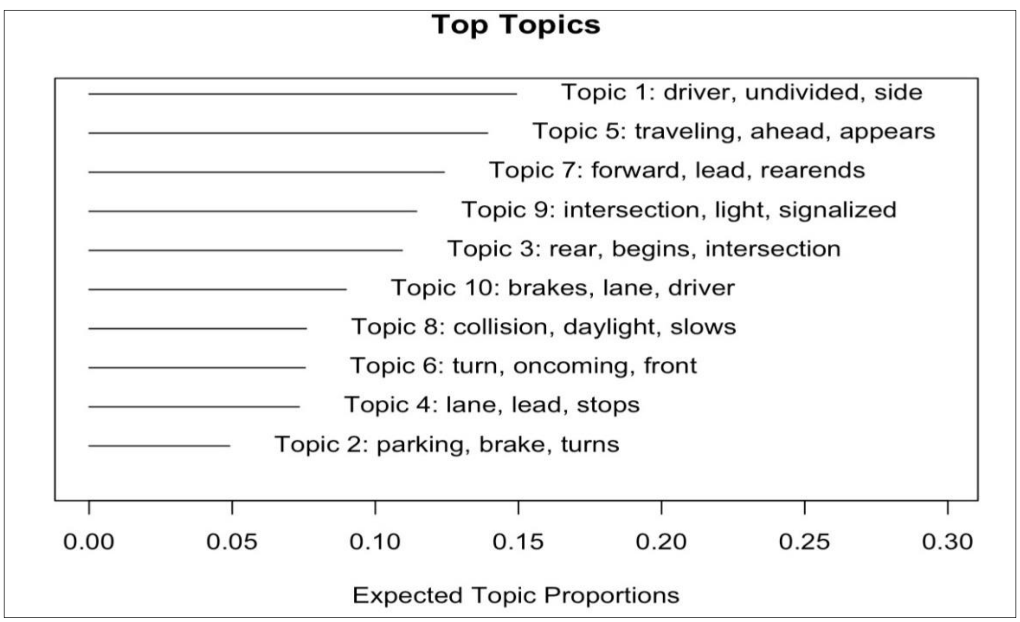


Figure 16. Top 10 Topic Models for Non-inclement Crash Narratives.

Figure 17 lists the corpus level visualization of the top topics from a 10- topic model from inclement crash narratives and the frequency of words in each of these topics. Topic 3 shows higher expected topic proportions. The top 10 topic models contain ‘surface,’ ‘wet,’ and ‘rain.’ Higher presence of ‘wet’ and ‘surface’ indicate that these crashes are associated with lower surface friction due to the wet surface. For key themes emerge in these topic models: 1) friction, 2) friction and lighting, 3) intersection, 4) signalized intersection, and 5) undivided roadways.

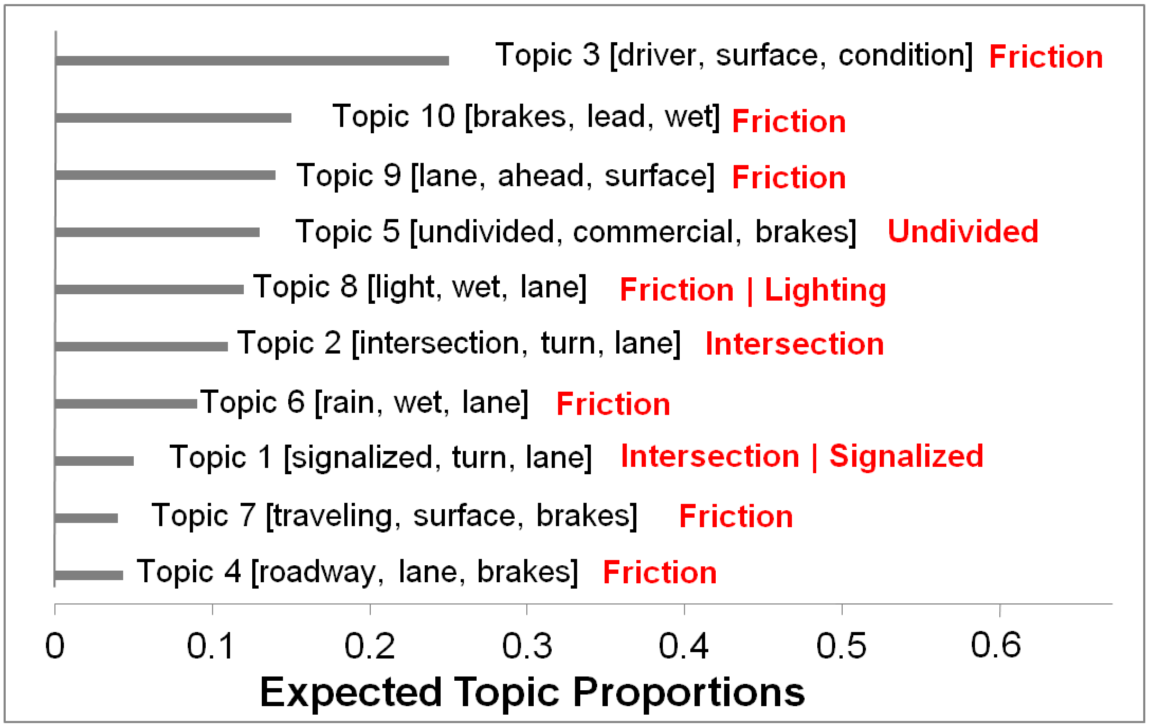
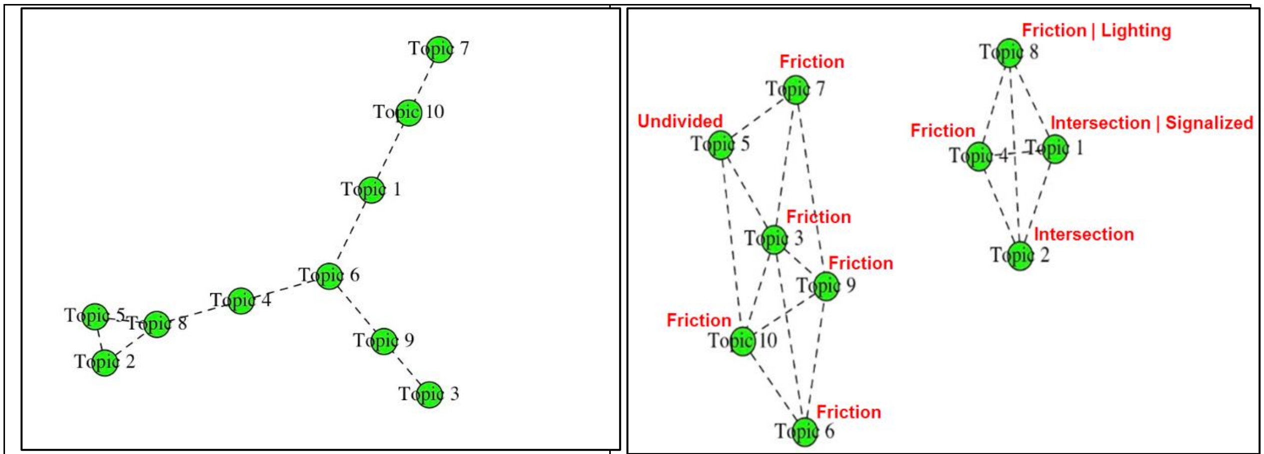


Figure 17. Top 10 Topic Models for Inclement Crash Narratives.

In addition, STM permits correlations between topics. Positive correlations between topics indicate that both topics are likely to be discussed within a document. Only one cluster is visible (cluster associated with Topic 8, Topic 5, and Topic 2) from the network plot generated from non-inclement crashes. Two separate clusters are found from inclement weather crash narratives. The first cluster is associated with six topic models (Topic 3, Topic 5, Topic 6, Topic 7, Topic 9, and Topic 10). The second cluster contains four topics (Topic 1, Topic 2, Topic 4, and Topic 8). The first cluster shows correlation between all of the friction related topics. The second cluster associates a relationship of intersection related topical contents.



Non-inclement Weather Crash Narratives | Inclement Weather Crash Narratives

Figure 18. Network Plots from the Topic Models.

FINDINGS

Identification of trends from text corpus is fundamental in most of the topic mining models. Mining the potential knowledge hidden inside unstructured textual content has been a popular research topic around the world, but limited research was conducted in the field of transportation engineering. In this study, the research team performed text mining on crash narratives collected from SHRP2 Insight. To develop document specific STM, this study used ‘maneuver information,’ ‘driver behavior,’ ‘event information’ as metadata. The developed top topic models for inclement weather crashes emerge five specific attention areas: friction, friction and lighting, intersection, signalization of intersection, and undivided roadways. The study provides a unique method to explore topical prevalence and topical content to identify additional information in the crash narratives.

CHAPTER 6: CONCLUSIONS

The main objective of this study was to examine safety impacts of reduced visibility and associated factors during inclement weather driving. The research team performed a systematic literature review to determine the key factors associated with inclement weather related crashes. Two SHRP2 RID states, Florida and Washington, were selected for final analysis. This study used three methods: 1) parametric model (ordinal logistic regression) development to quantify visibility issues by using Florida SHRP2 RID, 2) non-parametric analysis (multiple correspondence analysis) to identify key associated factors for inclement weather crashes by using Washington SHRP2 RID, and 3) topic model development by analyzing inclement weather related crash narratives from SHRP2 Insight database. The key findings from this study are:

- Higher friction reduces crashes in inclement weather, and old drivers are less likely to be in crashes in inclement weather.
- Higher speed is associated with fatal and injury crashes during reduced visibility conditions.
- Younger and older drivers seem to appear as a meaningful factor by itself for inclement weather crashes.
- Older drivers face difficulty in certain roadway characteristics (two way undivided arterial roadways with posted speed 41–50 mph) during inclement weather.
- Rural roadways are more involved in fatal crashes. For inclement weather, rural roadways with no lighting at dark are more risky.
- Interstate two lane roadways with barriers are always more likely to be involved with crashes.
- Five specific areas require attention: friction, friction and lighting, intersection, signalization of intersection, and undivided roadways.

The insights from this study could promote a better understanding of the safety impacts of reduced visibility related issues in inclement weather.

REFERENCES

Abdel-Aty, M., Ekram, A., Huang, H. and Choi, K. “A study on crashes related to visibility obstruction due to fog and smoke.” *Accident Analysis and Prevention*, Vol. 43, 2011, pp. 1730–1737.

Ahmed, M., Abdel-Aty, M., Lee, J., and Yu, R. “Real-time assessment of fog-related crashes using airport weather data: A feasibility analysis.” *Accident Analysis and Prevention*, No. 72, 2014, pp. 309–317.

Bartlett, A., Lao, W., Zhao, Y., and Sadek, A. *Impact of Inclement Weather on Hourly Traffic Volumes in Buffalo, New York*. Transportation Research Board 92nd Annual Meeting, No. 13-3240, 2013.

Brijs, T., Karlis, D., and Wets, G. “Studying the effect of weather conditions on daily crash counts using a discrete time-series model.” *Accident Analysis and Prevention*, No. 40, 2008, pp. 1180–1190.

Camacho F., Garcia, A., and Belda, E. “Analysis of Impact of Adverse Weather on Freeway Free-Flow Speed in Spain. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2169, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 150–159.

Chakrabartya, N., and Gupta, K. “Analysis of Driver Behavior and Crash Characteristics during Adverse Weather Conditions.” 2nd Conference of Transportation Research Group of India procedia, *Social and Behavioral Sciences*, No. 104, 2013, pp. 1048–1057.

Christensen, R. *Analysis of ordinal data with cumulative link models- estimation with the R-package ordinal. Vignette of ‘ordinal’ package.* https://cran.r-project.org/web/packages/ordinal/vignettes/clm_intro.pdf Accessed: December 2016.

Das, S. and Sun, X. *Investigating the Pattern of Traffic Crashes under Rainy Weather by Association Rules in Data Mining.* In the proceedings of the 93rd TRB Annual Meetings, Washington D.C., 2014.

Das, S. and Sun, X. “Factor Association with Multiple Correspondence Analysis in Vehicle–Pedestrian Crashes.” *Transportation Research Record: Journal of the Transportation Research Board*, No. 2519, Transportation Research Board, Washington, D.C., 2015, pp. 95–103.

Das, S. and Sun, X. “Factor Association knowledge for fatal run-off-road crashes by Multiple Correspondence Analysis.” *IATSS Research*. Vol. 39, Iss. 2, 2016a, pp. 146–155.

Das, S., X. Sun, and A. Dutta. Text Mining and Topic Modeling on Compendium Papers from Transportation Research Board Annual Meetings. Transportation Research Record: Journal of the Transportation Research Board, Volume 2552, 2016b, pp. 48-56.

Das, S. Avelar, R., Dixon, K., and Sun, X. *Identifying Patterns in Wrong Way Driving Crashes using Multiple Correspondence Analysis.* Prepared manuscript, 2017a.

Das, S., K. Dixon, X. Sun, A. Dutta, and M. Zupancich. Trends in Transportation Research: Exploring Content Analysis in Topics. Transportation Research Record: Journal of the Transportation Research Board, 2017b (in press).

De Leeuw, J. *Canonical analysis of Categorical Data*. PhD Thesis, University of Leiden, 1973. Republished in 1985 by DSWO- Press, Leiden.

El-Basyouny, K., Barua, S., and Islam, M. "Investigation of time and weather effects on crash types using full Bayesian multivariate Poisson lognormal models." *Accident Analysis and Prevention*, No. 73, 2014a, pp. 91–99.

El-Basyouny, K., Barua, S., Islam, M., and Li, R. "Assessing the Effect of Weather States on Crash Severity and Type by Use of Full Bayesian Multivariate Safety Models." *Transportation Research Record: Journal of the Transportation Research Board*, No. 2432, Washington D.C., 2014b, pp. 65–73.

Factor, R., G. Yair, and D. Mahalel. Who by Accident? The Social Morphology of Car Accidents. Risk Analysis, Vol. 30, No. 9, 2010, pp. 1411–1423

FHWA. *How Do Weather Events Impact Roads? Road Weather Management Program*. http://www.ops.fhwa.dot.gov/weather/q1_roadimpact. htm. Accessed December 2016.

Fitzpatrick, C., Rakasi, S., and Knodler, M. "An investigation of the speeding-related crash designation through crash narrative reviews sampled via logistic regression." *Accident Analysis and Prevention*, No. 98, 2017, pp. 57–63.

Fontaine, H. *A Typological Analysis of Pedestrian Accidents.* Presented at the 7th workshop of ICTCT, Paris, 26-27 October 1995.

Hassan, H., Abdel-Aty, M., and Oloufa, A. *The Effect of Warning Messages and Variable Speeds in Different Visibility Conditions.* In Transportation Research Board 90th Annual Meeting, No. 11-0354, 2011.

Hoffman, D. and De Leeuw, J. "Interpreting Multiple Correspondence Analysis as a Multidimensional Scaling Method." *Marketing Letter*, Vol. 3, pp. 259–272, 1992.

Ignatow, G. and Mihalcea, R. *Text Mining: A Guidebook for the Social Sciences*. Sage Publications, 2016.

Kopelias, P., Papadimitriou, F., Papandreou, K., and Prevedouros, P. “Urban Freeway Crash Analysis: Geometric, Operatinal, and Weather Effects on Crash Number and Severity.” *Transportation Research Record: Journal of the Transportation Research Board*, No. 2015, Washington, D.C., 2007, pp. 123–131.

Le, S., Josse, J., and Husson, F. “FactoMineR: An R Package for Multivariate Analysis.” *Journal of Statistical Software*, Vol 25(1), 2008.

Liaw, A., and Wiener, M. Classification and Regression by randomForest. R News, Vol. 2, No. 3, 2002, pp. 18-22.

McCann, K. and Fontaine, M. “Assessing Driver Speed Choice in Fog with the Use of Visibility Data from Road Weather Information Systems.” *Transportation Research Record: Journal of the Transportation Research Board*, No. 2551, Washington D.C., 2016, pp. 90–99.

Murphy, R., Swick, R., and Guevara, G. *Best Practices for Road Weather Management*, Version 3.0. Report No. FHWA-HOP-12-046, 2012.

Naik, B., Tung, L., Zhao, S., and Khattak, A. “Weather impacts on single-vehicle truck crash injury severity.” *Journal of Safety Research*, No. 58, 2016, pp. 57–65.

Office of the Aging, New York. *Are you concerned about an Older Driver?* 2015.

Qiu, L., and Nixon, W. “Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis.” *Transportation Research Record: Journal of the Transportation Research Board*, No. 2055, Washington D.C., 2008, pp. 139–146.

R Development Core Team. *R: A Language and Environment for Statistical*. Computing. Version 2.10.1. R Foundation for Statistical Computing, Vienna, Austria, 2013. http://www.R-project.org.

Rakha, H., Farzaneh, M., Arafeh, M., and Sterzin, E. “Inclement Weather Impacts on Freeway Traffic Stream Behavior.” In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2071, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 8–18.

Roberts M., B. Stewart, and D. Tingley. *stm: R Package for Structural Topic Models.* 2016. http://www.structuraltopicmodel.com, Accessed: December 2016.

Roux B., and Rouanet H. *Multiple Correspondence Analysis*. Sage Publications, Washington D.C. 2010.

Shaheed, M., Gkritza, K., Carriquiry, A., and Hallmark, S. “Analysis of occupant injury severity in winter weather crashes: A fully Bayesian multivariate approach.” *Analytic Methods in Accident Research*, No. 11, 2016, pp. 33–47.

Shang, P., Li, R., Liu, Z., and Li, X. *Inclement Weather Impacts on Urban Traffic Conditions*. In 15th COTA International Conference of Transportation Professionals, 2015.

Shankar, V., Mannering, F., and Barfield, W. "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies." *Accident Analysis and Prevention*, No. 27 (3), 1995, pp. 371–389.

SHRP2 *Insight Webpage*. https://insight.shrp2nds.us/ Accessed: December 2016.

Slowikowski, K. *ggrepel: Repulsive Text and Label Geoms for 'ggplot2'*. R package version 0.6.3. Accessed in: November, 2016. https://CRAN.R-project.org/package=ggrepel.

Stutts, J., Reinfurt, D., Staplin, L., and Rodgman, E. *The Role of Driver Distraction in Traffic Crashes*. AAA Foundation for Traffic Safety, 2001.

Sun, X., Hu, H., Habib, E., and Magri, D. "Quantifying Crash Risk under Inclement Weather with Radar Rainfall Data and Matched-Pair Method." *Journal of Transportation Safety & Security*, Vol. 3, 2011, pp. 1–14.

Usman, T., Fu, L., and Miranda-Moreno, L.F. "A disaggregate model for quantifying the safety effects of winter road maintenance activities at an operational level." *Accident Analysis and Prevention*, No. 48, 2012, pp. 368–378.

Wang, L., Shi, Q., Abdel-Aty, M., and Kuo, P. "Predicting Crashes on Expressway Ramps With Real-Time Traffic and Weather Data." In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2514, Transportation Research Board of the National Academies, Washington, D.C., 2015, pp. 32–38.

Wei, Z., Xiaokun, W., and Dapeng, Z. "Truck crash severity in New York city: An investigation of the spatial and the time of day effects." *Accident Analysis and Prevention*, No. 99, 2017, pp. 249–261.

Wickham, H. *ggplot2: elegant graphics for data analysis.* Springer New York, 2009.

Xu, C., Wang, W., and Liu, P. "Identifying crash-prone traffic conditions under different weather on freeways." *Journal of Safety Research*, No. 46, 2013, pp. 135–144.

Yu, R., and Abdel-Aty, M. "Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data." *Safety Science*, No. 63, 2014, pp. 50–56.

Yu, R., Xiong, Y., and Abdel-Aty, M. "A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway." *Transportation Research Part C*, No. 50, 2015, pp. 68-77.

APPENDIX

Table 8. Parentage of Variations on Dimensions (Non-inclement Weather Crashes).

Dimension	Eigenvalue	Percent of Variance	Cumulative Percent of Variance
Dimension 1	0.2688	6.05	6.05
Dimension 2	0.2063	4.64	10.69
Dimension 3	0.1280	2.88	13.57
Dimension 4	0.1242	2.79	16.36
Dimension 5	0.1213	2.73	19.09
Dimension 6	0.1198	2.70	21.79
Dimension 7	0.1177	2.65	24.44
Dimension 8	0.1162	2.62	27.05
Dimension 9	0.1143	2.57	29.63
Dimension 10	0.1137	2.56	32.19
Dimension 11	0.1134	2.55	34.74
Dimension 12	0.1133	2.55	37.29
Dimension 13	0.1128	2.54	39.83
Dimension 14	0.1124	2.53	42.35
Dimension 15	0.1121	2.52	44.88
Dimension 16	0.1118	2.52	47.39
Dimension 17	0.1114	2.51	49.90
Dimension 18	0.1112	2.50	52.40
Dimension 19	0.1110	2.50	54.90
Dimension 20	0.1107	2.49	57.39
Dimension 21	0.1105	2.49	59.88
Dimension 22	0.1101	2.48	62.36
Dimension 23	0.1101	2.48	64.83
Dimension 24	0.1097	2.47	67.30
Dimension 25	0.1094	2.46	69.76
Dimension 26	0.1092	2.46	72.22
Dimension 27	0.1084	2.44	74.66
Dimension 28	0.1077	2.42	77.08
Dimension 29	0.1074	2.42	79.50
Dimension 30	0.1071	2.41	81.91
Dimension 31	0.1051	2.36	84.27
Dimension 32	0.1021	2.30	86.57
Dimension 33	0.1007	2.27	88.83
Dimension 34	0.0982	2.21	91.04
Dimension 35	0.0956	2.15	93.19
Dimension 36	0.0933	2.10	95.30

Dimension 37	0.0882	1.98	97.28
Dimension 38	0.0668	1.50	98.78
Dimension 39	0.0305	0.69	99.47
Dimension 40	0.0236	0.53	100.00

Table 9. Parentage of Variations on Dimensions (Inclement Weather Crashes).

Dimension	**Eigenvalue**	**Percent of Variance**	**Cumulative Percent of Variance**
Dimension 1	0.2729	6.14	6.14
Dimension 2	0.2105	4.74	10.88
Dimension 3	0.1274	2.87	13.74
Dimension 4	0.1242	2.79	16.54
Dimension 5	0.1214	2.73	19.27
Dimension 6	0.1191	2.68	21.95
Dimension 7	0.1181	2.66	24.61
Dimension 8	0.1167	2.63	27.23
Dimension 9	0.1152	2.59	29.83
Dimension 10	0.1149	2.58	32.41
Dimension 11	0.1144	2.57	34.98
Dimension 12	0.1135	2.55	37.54
Dimension 13	0.1131	2.55	40.08
Dimension 14	0.1122	2.52	42.61
Dimension 15	0.1119	2.52	45.13
Dimension 16	0.1116	2.51	47.64
Dimension 17	0.1115	2.51	50.15
Dimension 18	0.1114	2.51	52.65
Dimension 19	0.1110	2.50	55.15
Dimension 20	0.1110	2.50	57.65
Dimension 21	0.1106	2.49	60.14
Dimension 22	0.1103	2.48	62.62
Dimension 23	0.1101	2.48	65.10
Dimension 24	0.1098	2.47	67.57
Dimension 25	0.1093	2.46	70.02
Dimension 26	0.1087	2.45	72.47
Dimension 27	0.1084	2.44	74.91
Dimension 28	0.1076	2.42	77.33
Dimension 29	0.1069	2.41	79.73
Dimension 30	0.1061	2.39	82.12
Dimension 31	0.1051	2.36	84.49
Dimension 32	0.1035	2.33	86.81
Dimension 33	0.1010	2.27	89.09
Dimension 34	0.0984	2.21	91.30
Dimension 35	0.0955	2.15	93.45
Dimension 36	0.0908	2.04	95.49
Dimension 37	0.0873	1.96	97.46
Dimension 38	0.0604	1.36	98.82
Dimension 39	0.0286	0.64	99.46
Dimension 40	0.0239	0.54	100.00

Table 10. Coordinates of the Attributes on First Five Dimensions (Non-inclement Weather Crashes).

Attributes	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Location_Rural	-0.1820	1.7558	0.3121	-0.1140	0.1128
Location_Urban	0.0386	-0.3722	-0.0662	0.0242	-0.0239
Func_Class_Collector	0.4110	2.9265	-1.2280	-0.4465	0.2291
Func_Class_Interstate	-0.8839	-0.2344	0.0070	-0.0020	0.0235
Func_Class_Local	1.2227	-0.7238	3.5348	3.1509	8.3281
Func_Class_Minor Arterial	0.6980	0.8516	-0.7294	0.4844	0.2279
Func_Class_Principal Arterial	0.9271	-0.1657	0.2967	-0.1096	-0.1244
Severity_Fatal	-0.2813	2.4035	-0.4928	2.1041	0.2252
Severity_Injury	0.0348	0.0675	-0.1770	-0.3736	-0.2475
Severity_PDO	-0.0194	-0.0647	0.1166	0.2173	0.1546
Lighting_Dark-No Street Lights	-0.7084	1.8040	0.8505	0.2764	-0.6172
Lighting_Dark-Street Lights Off	-0.1212	0.4970	0.7767	1.0287	-0.0078
Lighting_Dark-Street Lights On	0.1360	-0.4844	-0.3239	1.2909	-0.8822
Lighting_Dawn/Dusk	-0.0873	0.2490	-0.1146	0.9438	-0.0570
Lighting_Daylight	0.0454	-0.0951	-0.0177	-0.2973	0.2078
Lighting_Other	0.0059	0.1709	-0.5542	0.3504	0.2241
Alignment_Curve & Grade	-0.5146	0.9454	0.0904	0.2840	0.2593
Alignment_Curve & Level	-0.2093	0.9871	-0.7745	0.6107	0.3302
Alignment_Curve at Hillcrest	-0.1090	1.6850	1.0777	2.8258	-1.4682
Alignment_Curve in Sag	-0.4216	1.2056	0.6771	0.9558	0.5340
Alignment_Other	0.2944	-0.1756	0.1351	4.6674	-1.4645
Alignment_Straight & Grade	-0.2311	-0.1088	-0.3334	-0.0480	-0.1728
Alignment_Straight & Level	0.1289	-0.1364	0.1533	-0.1025	0.0006
Alignment_Straight at Hillcr	0.0724	0.1710	-0.5153	1.2747	1.0448
Alignment_Straight in Sag	-0.5280	-0.1704	-0.1813	-0.7227	-0.1731
Driver_Age_15–24	0.0257	0.1118	0.2144	0.3788	-0.5884
Driver_Age_25–34	-0.1212	-0.1160	-0.0833	0.2302	-0.3192
Driver_Age_35–44	-0.0535	-0.1446	-0.2489	-0.1043	0.0605
Driver_Age_45–54	-0.0318	-0.0158	-0.1128	-0.2091	0.3716
Driver_Age_55–64	0.0560	0.0751	0.0455	-0.3704	0.5318
Driver_Age_65–74	0.2182	0.1886	0.1112	-0.6139	0.9403
Driver_Age_75–Inf	0.5072	0.0969	0.2918	-0.8714	1.5797
Driver_Gender_Female	0.0970	-0.0617	0.1306	-0.1408	-0.0899
Driver_Gender_Male	-0.0712	0.0528	-0.1039	0.1098	0.0820
Driver_Gender_Not included	-0.4316	-0.3046	0.0409	0.1228	-0.8325

fdas

Table 10. Coordinates of the Attributes on First Five Dimensions (Non-inclement Weather Crashes).

Attributes	**Dim.1**	**Dim.2**	**Dim.3**	**Dim.4**	**Dim.5**
Posted_Speed_0–40 mph	0.8733	-0.2686	0.2162	0.2866	0.2396
Posted_Speed_41–50 mph	0.6052	0.5906	-0.7170	-0.8522	-1.0115
Posted_Speed_51–60 mph	-0.9009	-0.0965	-0.3106	0.0253	0.1405
Posted_Speed_61–70 mph	-1.3432	1.5722	2.9720	-0.3279	-0.4964
Road_Type_Alley	1.0484	0.0388	2.9987	-1.6063	-0.4680
Road_Type_Center-Two Way Left Turn Lane	1.2416	-0.6344	0.5217	-1.1465	-1.4109
Road_Type_Driveway	0.7048	-0.2385	-0.7769	-0.3590	-1.9033
Road_Type_Interchange Ramp	-0.0007	1.0097	-3.1577	1.3701	7.9086
Road_Type_One Way	1.0645	-1.0339	2.3721	0.9641	1.8295
Road_Type_Other	-0.1097	0.5931	0.8909	9.4512	-2.9168
Road_Type_Reversible Road	-1.1935	-0.9357	-3.1990	-1.0975	4.0205
Road_Type_Two Way - Divided, no Barrier	0.4843	-0.2967	-0.3541	0.4812	-0.3065
Road_Type_Two Way - Divided, with Barrier	-1.0129	-0.3101	0.0745	-0.0913	-0.0989
Road_Type_Two Way - Undivided	0.6361	0.4726	-0.2681	0.1811	0.2003

Table 11. Coordinates of the Attributes on First Five Dimensions (Inclement Weather Crashes).

Attributes	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Location_Rural	-0.1950	1.7964	0.2415	-0.0114	0.0894
Location_Urban	0.0414	-0.3812	-0.0512	0.0024	-0.0190
Func_Class_Collector	0.5755	2.8929	-1.4902	-1.1391	-0.0288
Func_Class_Interstate	-0.7863	-0.2062	0.0103	0.0057	0.0245
Func_Class_Local	1.0160	0.1689	5.8173	20.2843	20.8022
Func_Class_Minor Arterial	0.8141	0.9057	-1.1039	0.4516	0.2515
Func_Class_Principal Arterial	1.0384	-0.1351	0.4193	-0.0687	-0.1220
Severity_Fatal	-0.1346	2.0225	-0.5712	2.0687	-1.2839
Severity_Injury	0.0380	-0.0150	-0.2194	-0.2705	-0.0424
Severity_PDO	-0.0201	-0.0055	0.1254	0.1357	0.0322
Lighting_Dark-No Street Lights	-0.6009	1.6127	0.4412	0.5584	-0.4403
Lighting_Dark-Street Lights Off	-0.3355	0.2672	-0.8667	0.1925	-1.3851
Lighting_Dark-Street Lights On	0.1515	-0.5526	-0.3698	0.8628	-0.5275
Lighting_Dawn/Dusk	-0.0777	-0.0359	-0.1917	-0.1065	-0.2140
Lighting_Daylight	0.0531	-0.0599	0.0931	-0.3975	0.2915
Lighting_Other	0.0082	-0.0237	-0.4852	0.0988	0.9589
Alignment_Curve & Grade	-0.4236	0.9969	-0.1643	-0.0648	0.1989
Alignment_Curve & Level	-0.1920	0.7463	-0.7629	0.8533	0.1731
Alignment_Curve at Hillcrest	-0.6611	1.0671	1.4288	-1.3756	0.6305
Alignment_Curve in Sag	-0.4514	0.9969	0.1185	1.0941	-0.3691
Alignment_Other	0.2732	-0.4164	-0.3773	2.5938	-0.4746
Alignment_Straight & Grade	-0.2146	-0.0955	0.0556	-0.0877	-0.1268
Alignment_Straight & Level	0.1648	-0.1771	0.0924	-0.0798	-0.0077
Alignment_Straight at Hillcr	-0.0525	0.2397	-0.6103	1.2548	1.5823
Alignment_Straight in Sag	-0.5707	0.1045	-0.9921	-0.0615	-0.3106
Driver_Age_15–24	0.0521	0.0948	-0.0455	0.5904	-0.4358
Driver_Age_25–34	-0.1485	-0.1124	-0.0060	0.2330	-0.2769
Driver_Age_35–44	-0.0473	-0.0964	0.0060	-0.2037	0.1069
Driver_Age_45–54	0.0031	-0.0507	-0.0831	-0.3652	0.0882
Driver_Age_55–64	0.0521	0.0803	-0.0542	-0.4564	0.5673
Driver_Age_65–74	0.1943	0.1846	0.4169	-0.8640	1.1140
Driver_Age_75–Inf	0.4727	0.2465	0.5302	-1.2299	1.3900
Driver_Gender_Female	0.0916	-0.0559	0.0255	0.0090	-0.0173
Driver_Gender_Male	-0.0705	0.0449	-0.0109	-0.0262	0.0251
Driver_Gender_Not included	-0.0631	-0.1026	-0.6676	1.4287	-0.8702

Table 11. Coordinates of the Attributes on First Five Dimensions (Inclement Weather Crashes).

Attributes	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Posted_Speed_0–40 mph	0.9536	-0.2572	0.1238	0.3496	0.3313
Posted_Speed_41–50 mph	0.7283	0.6146	-0.1898	-0.9017	-1.1483
Posted_Speed_51–60 mph	-0.8075	-0.1591	-0.3398	-0.0930	0.0661
Posted_Speed_61–70 mph	-1.3047	1.7160	2.6410	0.6030	-0.0812
Road_Type_Alley	0.5338	-0.4815	2.7419	1.9837	0.7224
Road_Type_Center-Two Way Left Turn Lane	1.3501	-0.6225	1.3026	-0.8342	-1.6813
Road_Type_Driveway	1.1149	0.6289	-2.8659	0.8473	3.2391
Road_Type_Interchange Ramp	0.1461	-1.1464	0.3947	3.1762	2.4702
Road_Type_One Way	0.9239	-0.8913	1.9106	1.4032	2.7922
Road_Type_Other	-0.1806	0.2119	1.6783	2.3442	1.0020
Road_Type_Reversible Road	-1.0067	-1.0285	-3.4006	-2.2687	2.1120
Road_Type_Two Way - Divided, no Barrier	0.7073	-0.2317	-0.1806	0.9937	-0.4132
Road_Type_Two Way - Divided, with Barrier	-0.9089	-0.2463	0.0861	-0.0514	-0.0852
Road_Type_Two Way - Undivided	0.7656	0.4639	-0.3823	0.0690	0.2143

Table 12. Sample of Crash Narratives for Non-inclement Weather Crashes.

Site	Final Narrative
New York	Subject vehicle (SV) is stopped at a stop sign controlled intersection with a lead vehicle (V2) starting to enter the intersecting road as soon as traffic is clear. Traffic clears and V2 begins to accelerate off. The subject driver is talking on a cell phone. SV checks for traffic and begins to accelerate into the intersecting road but becomes distracted by the rear passenger for a moment while entering the intersection. At the same moment, V2 brakes in the middle of the intersection. (There was no traffic coming.) The subject turns her attention back to the roadway and brakes immediately but still rear end strikes V2. Both drivers exit pull over and exit there vehicles to discuss the incident.
Florida	SV is traveling in a business area on a divided road in the left lane. V2 is behind SV and V3 is ahead of SV, all of which are in the same lane. V3 begins decelerating for traffic ahead and then SV begins decelerating as well. V2 behind SV appears to be decelerating as well, but fails to notice that SV has come to a complete stop and rear-ending SV. SV skids forward and subject tries to brake in time, but ends up rear-ending V3 ahead.
New York	SV is traveling through an intermittently lit residential neighborhood at night. While yawning, the subject seems to stop paying attention and does not notice V2 parallel parked ahead. Subject approaches V2 at about 21 mph. Upon noticing V2, the subject brakes severely and steers to the left, but she is unable to avoid a rear-end collision. V2 rolls up and over the curb and keeps rolling until it strikes a tree.
Washington	SV is stopped in a traffic lane preparing to turn left into an alley from an undivided four lane roadway. Just prior to the turn, the subject is distracted by talking with the front seat passenger. Vehicles in the left of two oncoming lanes stop to allow the subject to turn. Subject begins to turn, but his view of any oncoming traffic in the right lane is likely obscured by stopped vehicles in the left oncoming lane. V2, traveling in the right oncoming lane at constant speed, appears to brake in reaction to the subject turning into his path, but he is unable to avoid a collision with the front right corner of the subject vehicle at approximately −2.2g.
Washington	The subject driver and following V2 approach a signalized intersection in stop-and-go traffic in the left lane of an undivided commercial road. The subject driver is texting throughout the event. When lead traffic allows, the subject driver accelerates, then comes to a stop again. V2 accelerates with the subject vehicle, but does not use brake when the subject driver does. V2 seems to attempt no evasive maneuvers and rear-ends the subject vehicle.
Indiana	SV is traveling at a constant speed on a one-lane undivided road through a business area when V2, ahead, begins to execute a U-turn. SV begins to brake to try to avoid impact. V2 does not show any signs of evasive maneuvers. SV collides with the rear driver's side of V2.
North Carolina	The subject driver is traveling in the right lane of a divided highway at night, in icy conditions. The subject driver attempts to merge onto an exit ramp at roughly 50 miles per hour and loses control of the vehicle. He brakes and steers to the left, and SV rotates clockwise, sliding off the road and into a ditch.

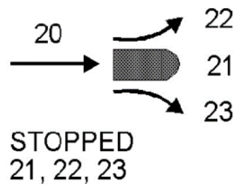
Table 13. Sample of Crash Narratives for Inclement Weather Crashes.

Site	Final Narrative
Indiana	The subject driver is traveling at night on an interstate in heavy rain. The subject driver is moving at a legal but unsafe (for the conditions) speed (60–65 mph) as he moves into the right lane. Immediately after the lane change, the subject vehicle begins to hydroplane. The subject driver steers left, causing the subject vehicle to rotate 180° counterclockwise, across four lanes of traffic, off the left side of the road, and into the guard rail. The subject vehicle continues to travel backward, knocking against the guard rail, until it runs out of momentum.
Washington	SV and V2 travel on a right-curving entrance ramp approaching a highway. It is raining, and the road surface is wet. It is night, and the area is lit by streetlights at regular intervals. The subject driver glances to his left to check traffic conditions. V2 comes to a complete stop at the point where the entrance ramp and highway intersect. The subject driver glances back ahead of the subject vehicle and brakes, but cannot avoid rear-ending V2.
Florida	The subject driver travels in the right lane of a commercial highway, approaching a signalized intersection. It is raining, and the road surface is wet. V2 is ahead, stopped at the end of a queue of vehicles at a red light. The subject driver brakes, and the subject vehicle loses traction on the wet road surface and skids forward, rear-ending V2 at roughly 20 mph.
Washington	SV is traveling at constant speed along an undivided roadway approaching a signalized intersection. The light is red, and SV is too distracted with making external glances and talking to an unknown audience. Once SV realizes she is about to run the light the subject looks very surprised and applies the brakes. Meanwhile V2 is making a left turn through the intersection at the same moment the subject enters. SV brakes and steers right before colliding with V2. SV then exits the vehicle to discuss the incident with driver 2.
Washington	The subject driver negotiates a curve to the right on an undivided commercial road in rainy conditions, traveling at roughly 45 mph. V2 drives straight across traffic from the left toward a parking lot entrance on the subject driver's right. The subject driver brakes and steers right. The subject vehicle loses traction and skids forward. V2 appears to not notice the subject driver and attempts no evasive maneuvers. The subject vehicle hits the front right side of V2.
New York	SV travels along an undivided road at night with no lead traffic. It is night, but the roadway is lit. It is also raining with lightning, and the roadway is wet. SV approaches a signalized intersection with a yellow light and begins to decelerate slightly. The light turns red before SV enters the intersection. SV drives through the red light. V2 in the dedicated left-turn lane enters the intersection and crosses SV's path from the left. V2 turns as SV's light turns red, so it appears V2 illegally enters the intersection as well. SV brakes as an evasive maneuver, but ends up skidding forward due to the wet roadway. SV strikes the right side of V2 going 23 mph. Both cars rotate clockwise slightly before coming to complete stops.
New York	SV is approaching a signalized intersection and comes to a stop. The subject looks in her rearview mirror and notices V2 closing at an unsafe speed and braces herself for impact. V2 following SV continues the same rate of speed and rear end strikes SV above 2.0g.
Florida	The subject is traveling on a one-lane, non-divided roadway through a business area. The roadway is visibly wet and it is lightly raining. The subject is approaching a signalized intersection that has two dedicated turn lanes.

Table 14. Sample SHRP2 Insight Final Narrative and Metadata.

Participant ID	XXX[1]
Vehicle ID	XXX
Vehicle Class	CAR
Trip ID	XXX
Site	New York
Crash Severity 1	I - Most Severe
Crash Severity 2	Not Applicable
Event ID	116168906
Event Severity 1	Crash
Event Severity 2	Not Applicable
Event Nature 1	Conflict with a lead vehicle
Event Nature 2	None
Vehicle 1 Configuration	20
Vehicle 2 Configuration	21
Vehicle 3 Configuration	9999
Final Narrative	SV is stopped at a stop sign controlled intersection with V2 starting to enter the intersecting road as soon as traffic is clear. Traffic clears and V2 begins to accelerate off. The subject driver is talking on a cell phone. SV checks for traffic and begins to accelerate into the intersecting road but becomes distracted by the rear passenger for a moment while entering the intersection. At the same moment V2 brakes in the middle of the intersection. (There was no traffic coming.) The subject turns her attention back to the roadway and brakes immediately but still rear end strikes V2. Both drivers exit pull over and exit there vehicles to discuss the incident.
Precipitating Event	Other vehicle ahead - slowed and stopped 2 seconds or less
Event Start	XXX
Event End	XXX
(Non)Motorist 2 Evasive Maneuver	No reaction
(Non)Motorist 2 Pre-Incident Maneuver	Starting in traffic lane
(Non)Motorist 3 Evasive Maneuver	Not applicable
(Non)Motorist 3 Pre-Incident Maneuver	Not applicable
Airbag Deployment	No
Construction Zone	Not construction zone-related
Contiguous Trvl Lanes	2
Driver Seatbelt	Lap/shoulder belt properly worn
Driving Behavior 1	Distracted
Driving Behavior 2	No Additional Driver Behaviors
Driving Behavior 3	No Additional Driver Behaviors

Fault	Subject driver
Front Seat Passengers	1
Grade	Grade Down
Hands on Wheel	Left hand only
Impact Time	XXX
Impairments	None apparent;
Incident Type 1	Rear-end, striking
Incident Type 2	None
Infrastructure	None
Intersection Influence	Yes, Stop Sign
Light	Daylight
Locality	Open Residential
Maneuver Judgment	Safe and legal
Object 2 Location	A = In front of the subject vehicle
Object 2 Type	Van (minivan or standard van)
Object 3 Location	Not applicable
Object 3 Type	Not applicable
Objects/Animals	0
Others Involved	1
Pre-Incident Maneuver	Starting in traffic lane
Rd Alignment	Straight
Reaction Start	XXX
Rear Seat Passengers	2
Relation to Junction	Intersection
Sec Task 1 End Time	Cell phone, Talking/listening, hand-held
Sec Task 1 Outcome	XXX
Sec Task 1 Start Time	Yes
Sec Task 1	XXX
Sec Task 2 End Time	Passenger in rear seat - interaction
Sec Task 2 Outcome	XXX
Sec Task 2 Start Time	Yes
Sec Task 2	XXX
Sec Task 3 End Time	No Additional Secondary Tasks
Sec Task 3 Outcome	0
Sec Task 3 Start Time	Not applicable
Sec Task 3	0
Surface Condition	Dry
Thru Travel Lanes	1
Traffic Control	Stop sign
Traffic Density	Level-of-service B

Traffic Flow	Not divided - simple 2-way traffic way
V1 Evasive Maneuver 1	Braked (no lockup)
V1 Evasive Maneuver 2	Not Applicable
V1 Lane Occupied	1
V1 Post-Maneuver Control 1	Control maintained
V1 Post-Maneuver Control 2	Not Applicable
Vehicle Factors	None
Vehicle Rollover	No
Vis Obstructions	No obstruction
Weather	No Adverse Conditions
Diagram	

Note: [1] Dummy id.