



Passing Behavior on Rural Roadways: Knowledge Extraction from SHRP-2 Data

A Project of TTI Strategic Highway Research Program

Subasish Das, Ph.D. (PI)

Lisa Minjares-Kyle, M.S. (Co-PI)

September 1, 2016

Table of Contents

Page

EXECUTIVE SUMMARY	3
BACKGROUND	5
OBJECTIVES	5
DATABASE PREPARATION	5
SHRP-2 Roadway Information Database	6
SHRP-2 Naturalistic Driving Study.....	8
ANALYSIS ON SHRP-2 RID	9
ANALYSIS ON SHRP-2 NDS	9
STATUS OF THE PROJECT	11
FUTURE RESEARCH.....	11
Background.....	11
Research objectives for project extension	13
Work Plan for Project Extension	13
Schedule and Budget for Project Extension	14
APPENDIX [2017 TRB PAPER].....	16

List of Tables

TABLE 1 Passing behavior related attributes in RID.....	7
TABLE 2 Improper passing related attributes in RID	20
TABLE 3 Descriptive Statistics of the Key variables	22
TABLE 4 Two Itemset Precision Rules for Divided and Undivided Roadways.....	28
TABLE 5 Three Itemset Precision Rules for Divided and Undivided Roadways.....	30

List of Figures

FIGURE 1 SHRP-2 Rural Two-lane Routes.	7
FIGURE 2 Improper passing crashes on SHRP-2 Rural Two-lane Routes.....	8
FIGURE 3 Selected SHRP-2 Rural Two-lane NDS Routes.	8
FIGURE 4 Mosaic plot for Incident Types.....	9
FIGURE 5 Mosaic plot for length of Vehicle Ownership.	10
FIGURE 6 Mosaic plot for Age of the Drivers.....	10
FIGURE 7 Mosaic plot for Use of Hands while Driving.	11
FIGURE 8 Improper passing crashes on divided and undivided roadways in Florida.....	21
FIGURE 9 Passing and Non-passing roadways markings.....	23
FIGURE 10 Passing permitted and No passing Criteria.....	24

EXECUTIVE SUMMARY

The current research findings show that improper passing behavior was associated with higher numbers of fatal or injury crashes. The results also showed that these crashes on divided roadways are higher in inclement weather condition. In addition, the percentage of large trucks was a key association factor on undivided roadways.

What are the findings?

This study used the Second Strategic Highway Research Program (SHRP 2) Roadway Inventory Database (RID) and Naturalistic Driving Study (NDS) data for Florida to investigate the improper passing related crashes. Key findings from the RID study included:

- Improper passing crashes on divided roadways were associated with higher Annual Average Daily Traffic (AADT), wider roadways, and higher speeds. On the other hand, undivided roadways incorporated lower AADT, narrower roadways, and higher average percentage of trucks.
- Fatal or severe injuries were dominant on undivided roadways.
- Improper passing crashes on divided roadways were higher in inclement weather condition.
- The percentage of large trucks was a key contributing factor on undivided roadways.

Initial findings from the NDS study included:

- The probability of passing/overtaking incidents (crash or near crash) was higher than incidents involving changing of lanes.


Improper passing behavior causes higher number of fatal or injury crashes

Case Study: Florida

Improper passing crashes on divided roadways are higher in inclement weather.

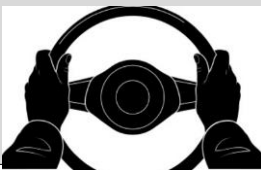
For undivided roadways, crash severity is higher.

Large truck



is a significant factor in improper passing crashes on **undivided roadways.**

While passing use both hands



- The probability of passing/overtaking incidents was significantly higher in the group of vehicle ownership of 1 year or less than 1 year.
- Single hand driving associated with more incidents while passing or overtaking.

Why is this important?

A significant number of crashes occurring on rural two-lane undivided roadways are due to the lack of effective countermeasures to separate opposing traffic flows. As a result, a major concern involves vehicles crossing the centerline resulting in either sideswiping or head-on collisions. These types of crashes result in around 10% to 15% of fatal crashes annually in the United States. From a behavioral perspective, passing behavior and driver's interactions with the surroundings are two significant contributors in these crashes. Investigation on the key association factors is thus called for.

What is unique in this study?

No prior studies have been conducted using the SHRP-2 NDS and RID to investigate the association between passing behavior related crashes and contributing factors. This study is designed to explore both geometric and human related factors to identify the key associations. This study used a recent data mining algorithm (association rules negative binomial miner) to generate the key association rules.

How was the study done?

This study used data from three states (Florida, New York, and Washington) as case studies. Due to unavailability of NDS data at the beginning of the project start date, the research team used Florida RID to investigate passing behavior related crashes. The study used association rules negative binomial miner to perform the analysis. The research team wrote a TRB paper explaining the research methodology and findings. The research team is currently exploring the NDS data to investigate the effects of human factors and driving parameters on passing behavior related crashes.

BACKGROUND

A 4-year-old was killed and two others were injured on May 27, 2015 when their school bus crossed the centerline and collided with a tractor-trailer on a rural two-lane road in Aiken County, South Carolina. This particular incident is not a distinct case. A significant numbers of crashes occur on rural roadways due to the lack of effective countermeasures. For rural undivided roadways, a major concern involves vehicles crossing the centerline and ending in either sideswiping or head-on collisions. These types of crashes result in around 10% to 15% fatal crashes annually in the U.S. From the behavioral perspective, passing behavior and driver's interactions with the surroundings are two significant contributors in these crashes. Compared with data from traditional crash databases, SHRP 2 NDS data provide an unprecedented opportunity to perform this research. The main objective of this research is to develop data mining models to predict risk associated with passing behavior by using a wide range of variables from roadway geometrics and human factors. The outcomes of this research will provide deep insights on this potential safety concern and which can aid in the prevention of another preschooler's death on rural two-lane roadways.

OBJECTIVES

The objectives of this project are to (a) explore the passing behavior of the drivers on the rural roadways, (b) examine interaction between roadway factors and behaviors of drivers of different age groups, (c) develop statistical and algorithmic predictive models to determine the risk outcomes, and (d) create a sense of urgency and action towards the improvement of rural roadway safety.

DATABASE PREPARATION

To conduct the analysis, key information were obtained from the following databases: (1) Crash, Roadway, and Vehicle data for Florida, and Washington from SHRP 2 RID, and (2) SHRP 2 NDS data for selected suitable sites in Florida, and New York. The selection of states in SHRP 2 RID was based on the availability of the passing behavior information on rural two-lane roadways. In selecting NDS sites, the research team focused on availability of the continuous rural two-lane sites from NDS routes. Florida met criteria for both databases, while New York

RID did not provide passing behavior information and Washington NDS routes did not meet required suitability criteria for NDS data request.

The research team divided the study into three parts: 1) Study on linkable dataset of RID and NDS (Florida), 2) Study on RID (Washington), and 3) Study on NDS (New York). This study design would be beneficial to explore the research scopes in the individual dataset and the linkable datasets.

SHRP-2 Roadway Information Database

The SHRP-2 Project populated RID with data from the SHRP-2 mobile data collection project (S04B); existing roadway data from government, public, and private sources; and supplemental data that further characterize traffic operations. This database provides good quality data that is linkable to the SHRP-2 NDS database utilizing geographic information system (GIS) tools. The RID is the supplementary tools for safety researchers to look at data sets of selected road characteristics and study matching NDS trips in order to explore the relationships between driver, vehicle, and roadway.

The research team selected two states (Florida, and Washington) based on the availability of the passing behavior information in the crash database on rural two-lane roadways. Florida RID maintains traffic crash data for eight years (2005-2012). The dataset contains crash, vehicle, and person information. The database is divided into two parts: 1) Local Roadways, and 2) State Highway Systems (SHS) roadways. Washington RID maintains crash data for eight years (2006-2013). The database maintains comprehensive information on crash, vehicle, and person. The database is divided into two parts: 1) State route roadways and 2) Non-state route roadways.

The primary focus of the database preparation for this study is to create a detailed database on the passing behavior related crashes. TABLE 1 lists the passing behavior related attributes in RID for Florida, and Washington.

TABLE 1 Passing behavior related attributes in RID

State	Attribute
Florida	CRSHCAUSE1 - 05 (Improper Lane Change) - 15 (Improper Passing)
Washington	CONTRIBUTING_CIRCUMSTANCE_1 CONTRIBUTING_CIRCUMSTANCE_2 CONTRIBUTING_CIRCUMSTANCE_3 - Improper Passing

FIGURE 1 and FIGURE 2 illustrate the roadway network for rural two-lane SHRP-2 roadways, and improper passing crashes on these routes respectively.

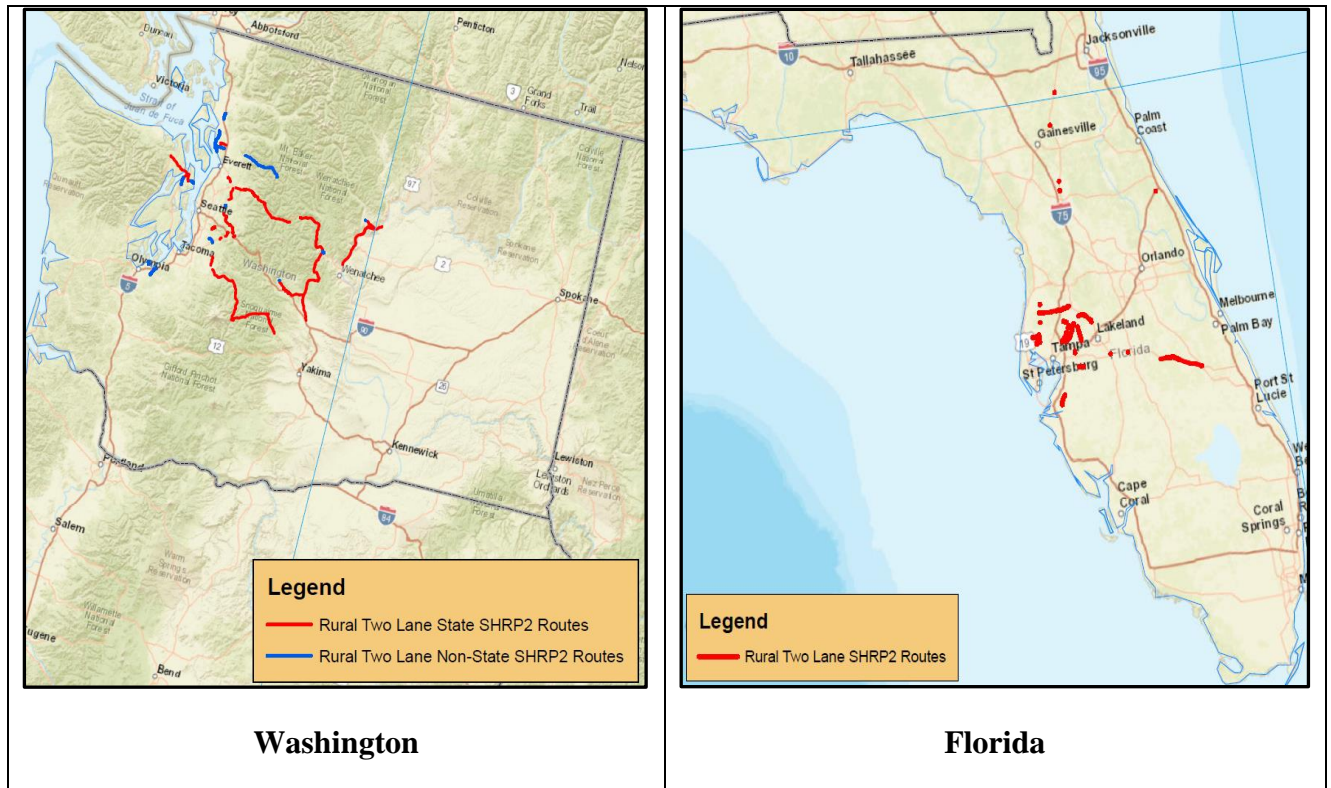


FIGURE 1 SHRP-2 Rural Two-lane Routes.

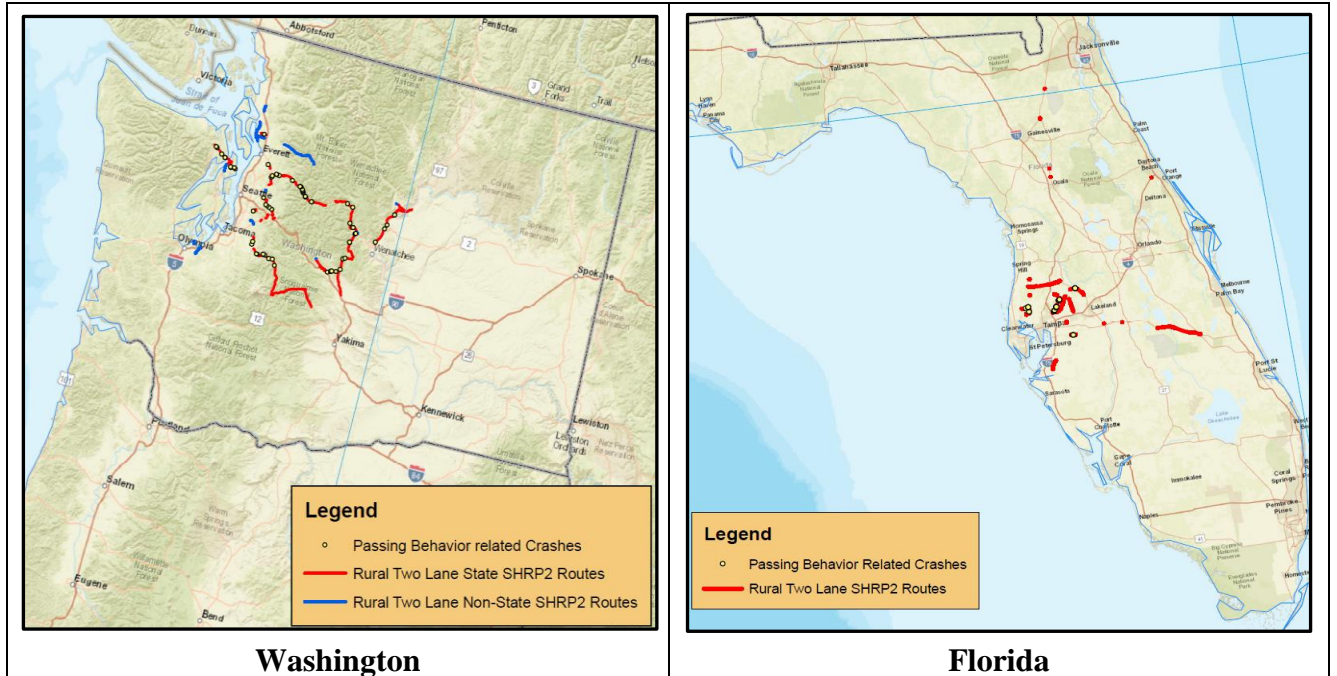


FIGURE 2 Improper passing crashes on SHRP-2 Rural Two-lane Routes.

SHRP-2 Naturalistic Driving Study

SHRP-2 NDS data provides an opportunity to the researchers to investigate on the variables that are usually not present in the conventional police reported crash data. The research team has requested NDS data for selected locations in Florida and New York. FIGURE 3 shows the selected routes for the NDS data.

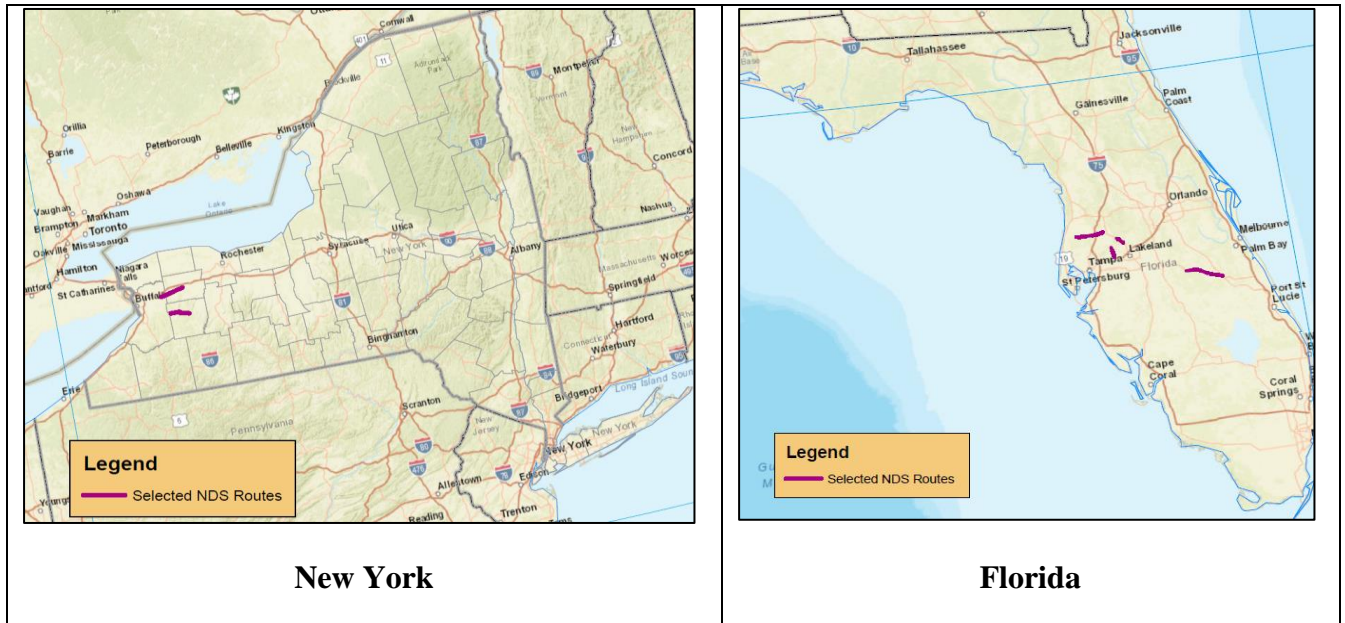


FIGURE 3 Selected SHRP-2 Rural Two-lane NDS Routes.

ANALYSIS ON SHRP-2 RID

This research prepared a finalized dataset for passing behavior related crashes from SHRP-2 RID for Florida and Washington. The research team wrote a TRB paper (documented in the Appendix) from the analysis of Florida data.

ANALYSIS ON SHRP-2 NDS

The research team collected required NDS data for analysis. A preliminary exploratory analysis is presented in this section.

Mosaic plots are helpful to depict the relationship of different variables to the probability that an incident may occur due to changing of lanes or passing/overtaking. FIGURE 4 depicts the relationship of incident type. Note that the width of each bar represents the number of observations in a mosaic plot. For example, non-crash incidents associated with more counts than crash/near-crash incidents, but the probability of passing/overtaking incidents was higher than changing lane incidents.

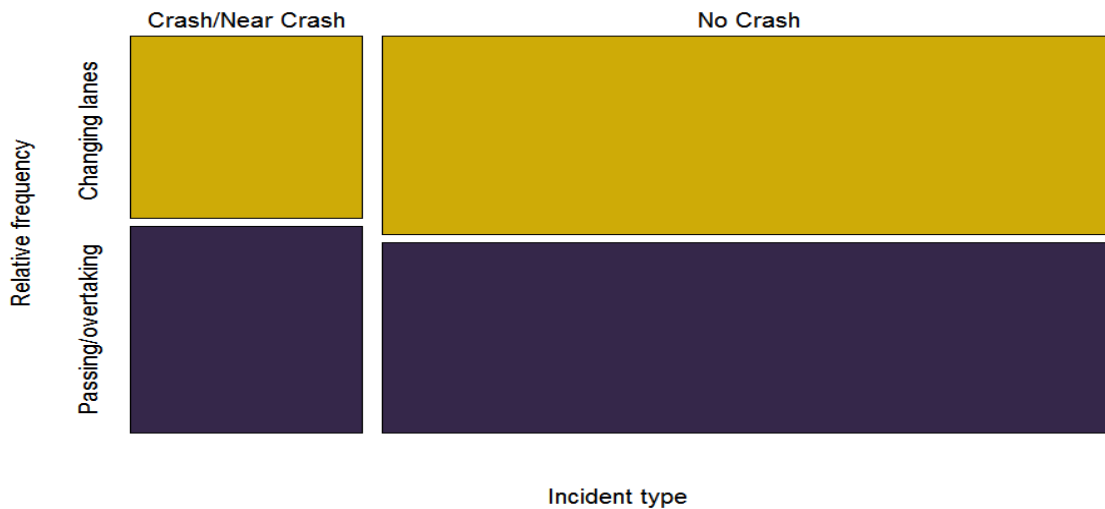


FIGURE 4 Mosaic plot for Incident Types.

FIGURE 5 illustrates the mosaic plot for the length of vehicle ownership. It is interesting to find that lower number of vehicle ownership years associated with more incidents, which may relate

with inexperienced drivers. The probability of passing/overtaking incidents was significantly higher in the group of vehicle ownership of 1 year or less than 1 year.

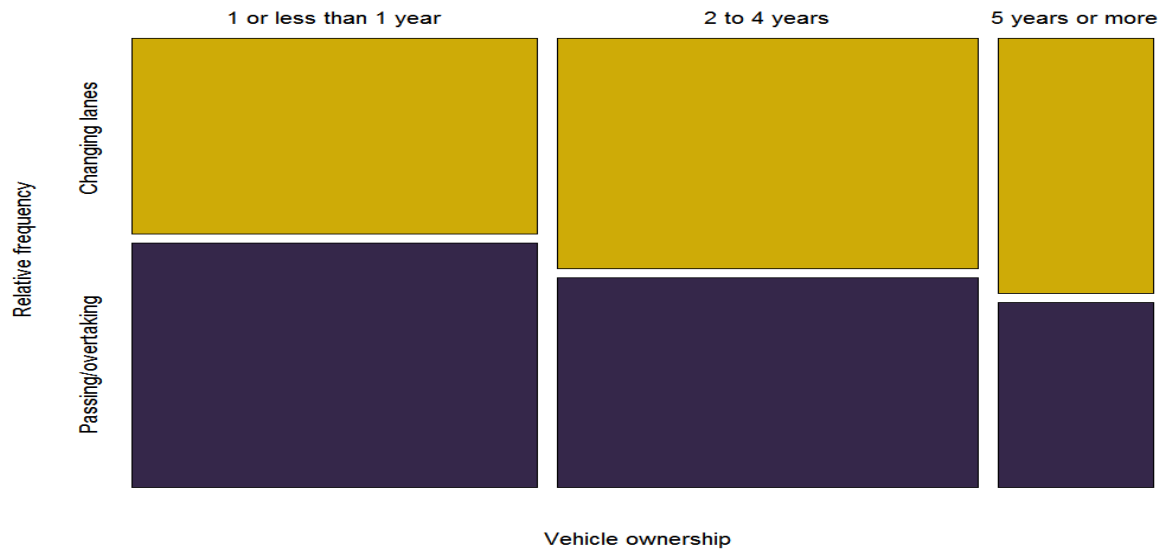


FIGURE 5 Mosaic plot for length of Vehicle Ownership.

FIGURE 6 illustrates the mosaic plot for age of the drivers. It is found that younger drivers (age less than 30 years) involved with more incidents.

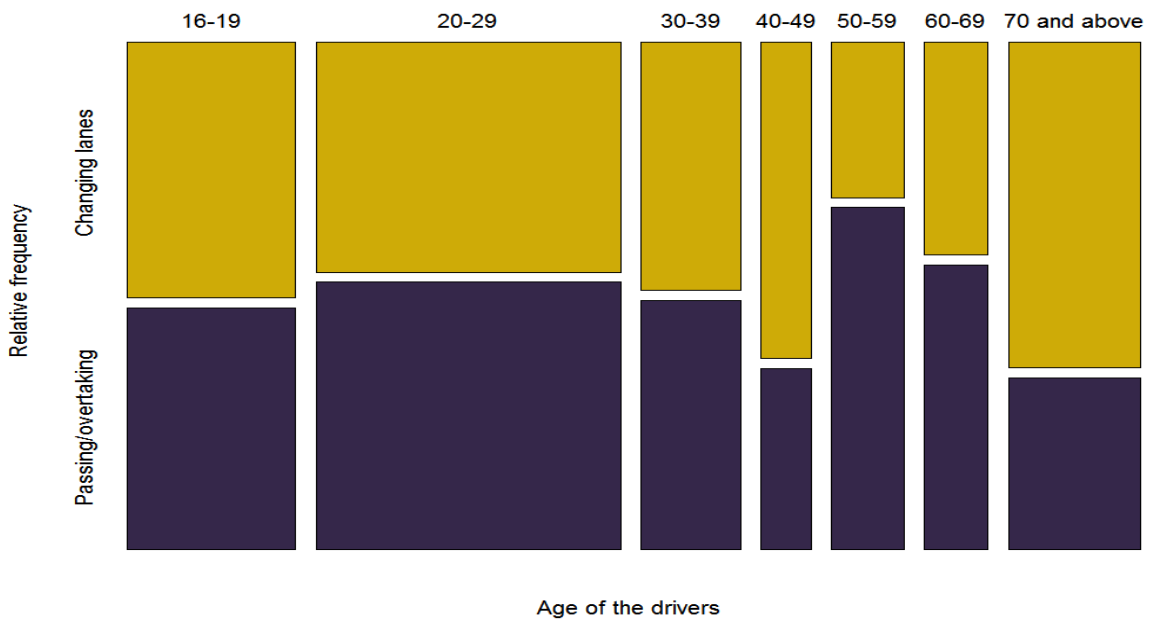


FIGURE 6 Mosaic plot for Age of the Drivers.

FIGURE 7 shows the mosaic plot for use of hands while driving. It is interesting to find that single hand driving associated with more incidents, which may relate to less control of the steering wheel while driving.



FIGURE 7 Mosaic plot for Use of Hands while Driving.

STATUS OF THE PROJECT

The status of the project is described below:

- A TRB paper was produced from the analysis of Florida data.
- The research team collected required NDS data for analysis. A preliminary exploratory analysis was conducted from the NDS data.
- The research team is currently joining NDS data with RID data to prepare a final dataset for analysis.
- A final report was written based on the tasks performed until August 31, 2016.

FUTURE RESEARCH

Background

The current project has completed analysis on passing behavior related crashes using Florida RID and while the initial findings from Florida RID data are promising, they do not offer a complete picture. This research requires an extended project to investigate the associated human

factors and driving parameters in improper passing crashes. Literature has identified various human factor variables and their impact on both crash risk and crash severity. The future direction of this project will seek a deeper review of variables including, but not limited to speeding, aggressive driving, and distractions to better understand their impact on improper passing crashes (or near crashes). The literature has shown that age and gender can also influence a driver's likelihood to speed, a driver's desire to pass and a driver's reported levels of aggressive driving tendencies, therefore, an additional avenue for future research to explore would also include the analysis of reported driver behavior for various demographic groups in relation to passing related crashes and near crashes. Research questions to be answered include: What role does driver distraction or aggressive driving have, if any, on crash severity on undivided/divided roadways? Are there significant differences in passing related crashes or severity of crashes for drivers of different ages? Are younger drivers, if involved in a passing related crash, more likely to have exhibited 'reckless' driving behaviors such as, following too closely or attempting an unsafe pass?

In addition, while the current focus of this research highlighted rural two way roadways, cursory analysis of available SHRP 2 data has highlighted several potential areas of investigation that relate to passing on other roadway types such as undivided multilane roads. Examination of these passing behavior related crashes can help researchers further understand the complexity of passing which can potentially lead to the development of risk outcome models for various roadway types. Despite the presence of an additional lane, passing can be difficult and result in crash or near crash instances due to higher levels of traffic, turning traffic, limited visibility or the presence of intersections. How do driver behaviors influence crashes on undivided and divided roadways that have multiple lanes and/or higher levels of traffic? Are there significant differences between passing crashes on multilane roads vs. two lane roadways?

The research team collected NDS data and conducted preliminary analysis. As no previous research conducted on analyzing the real-time driving parameters in passing behavior crashes; the findings from this research would be exemplary in nature.

Research objectives for project extension

The objectives of this project extension are to (a) explore the impact of human factors while passing improperly, (b) analyze driving parameters (acceleration, brake activation, headway, and other significant real-time factors) to associate with crash or near-crash events, (c) develop statistical and algorithmic predictive models to determine the risk outcomes, and (d) perform text mining on the final narratives and driver narratives for knowledge extraction.

Work Plan for Project Extension

The research team identified four tasks to accomplish the objectives of the project extension.

Task 1. Data Preparation

Objective

In this task, the research team will link collected NDS data with RID data to prepare a final dataset.

Approach

The research will use wide range of variables, but not limited to, like geometric features, environmental factors, human behavioral factors, and driving parameters. This task will:

- Prepare database for three states targeting improper passing behavior crashes
- Perform exploratory data analysis to conduct variable importance

Task 2. Data Mining

Objective

The objective of this task is to use innovative data mining algorithms to conduct the knowledge extraction.

Approach

This task will:

- Refine data mining algorithms to determine association between the factors
- Develop data mining rules from the clustered or non-clustered dataset

Task 3. Analysis of Predictive Methods

Objective

The objective of this task is to develop both statistical and machine learning models to address the passing behavior issues using the real-time contributing factors. The focus of this task is dual: methods of analysis will be assessed on their ability to predict phenomena, and on their interpretability merits.

Approach

Conventional statistical analysis tools often fail to deliver notable predications. The research team will use both statistical modeling and machine learning predictive tools to determine the best-fit model. Since the ultimate goal of the scientific endeavor is enlarge the body of knowledge, this research will also assess which approach yields a better understanding about the phenomenon that underlies the data. This task will:

- Develop statistical models
- Develop machine learning predictive models
- Perform validation on the models
- Develop two example problems to walk through the modeling approaches

Task 4. Prepare Final Report

The objective of this task is to prepare a final report documenting the whole research effort. The findings of the extended effort will be transformed into a 2018 TRB Paper. The last month of the project will be dedicated to the review and comments of the panel and preparation of the revised final paper.

Schedule and Budget for Project Extension

Budgeted Hours

Team Member	% Time	Task 1	Task 2	Task 3	Task 4	Total Hours per Researcher
Subasish Das	24	20	20	40	50	130
Lisa Minjares	9	10	10	10	20	50
Raul Avelar	6	0	0	10	20	30
PhD Graduate Student	61	80	80	80	80	320
Total Project Hours						530

Project Schedule

Task	Month			
	09/2016	10/2016	11/2016	12/2016
Task 1	DO, DA, TM1			
Task 2		DM, PA, PM1		
Task 3			PR, TM2	
Task 4				FR, PR

Legends: DA= Data Analysis, DO= Data Management, FR= Final Report, M1 = Panel Meeting 1, PA= Predictive Analysis, PR= Panel Review, TM1 = Technical Memorandum 1, TM2 = Technical Memorandum 2

Deliverables

Task	Deliverables	Month	Task	Deliverables	Month
1	Technical Memorandum 1	1	3	Technical Memorandum 2	3
2	Progress report on Task 2	2	4	Final Report	4

APPENDIX [2017 TRB PAPER]

IMPROPER PASSING RELATED CRASHES ON RURAL ROADWAYS: USING ASSOCIATION RULES NEGATIVE BINOMIAL MINER**Subasish Das, Ph.D.**

(Corresponding author)

Texas A&M Transportation Institute

3135 TAMU

College Station, TX 77843-3135

Phone: 979-845-9958, fax: 979-845-6006

Email: s-das@tti.tamu.edu

Lisa Minjares-Kyle, M.S.

Texas A&M Transportation Institute

701 N. Post Oak Rd., Suite 430

Houston, TX 77024

Phone: 713-613-9211

E-mail: l-minjares@tti.tamu.edu

Raul E. Avelar, Ph.D., P.E.

Texas A&M Transportation Institute

3135 TAMU

College Station, TX 77843

Phone: 979-862-1651

E-mail: r-avelar@tam.u.edu

Karen Dixon, Ph.D., P.E.

Senior Research Engineer

Texas A&M Transportation Institute, 3135 TAMU

College Station, TX 77843-3135

Phone: 979-845-9906, fax: 979-845-6006

Email: k-dixon@tti.tamu.edu

Bharadwaj Bommanayakanahalli

Texas A&M Transportation Institute

3135 TAMU

College Station, TX 77843-3135

Phone: 979-739-6412, fax: 979-845-6006

Email: bharadwaj472@tam.u.edu

Word Count: 7226 including 3 Figures and 4 Tables

Submitting to the 2017 TRB Annual Meeting for Presentation and Publication under Committee

ANB20 (Safety Data, Analysis and Evaluation)

ABSTRACT

Significant numbers of crashes end in rural roadways due to the lack of effective countermeasures to separate opposing traffic flows. As a result, a major concern involves vehicles crossing the centerline and ending in either sideswiping or head-on collisions on undivided roadways. At the same time, improper passing related crashes are also seen on divided roadways. Passing vehicles in either passing permitted or no passing zone requires complex interaction. For this task, a driver needs several judgments based on dynamic variables and allows little room for driver error yet can result in drastic consequences if performed improperly. By seeking to explore these types of crashes even further, researchers hope to gain knowledge that will contribute to reducing both the number of lives lost and severity of these types of crashes. This study used the second Strategic Highway Research Program's (SHRP-2) Roadway Inventory Database (RID) crash data for Florida rural roadways to investigate passing behavior related crashes. This study used an unsupervised data mining technique (known as association rules negative binomial (NB) miner) to extract the knowledge pattern of co-occurrence of the significant variables. The findings of this study would be beneficial for the safety practitioners and policy makers in decision making to reduce crashes due to improper passing.

INTRODUCTION

A 4-year-old was killed and two others were injured on May 27, 2015 when their school bus crossed the centerline and collided with a tractor-trailer on a rural two-lane road in Aiken County, South Carolina. Crash like this happening due to passing behavior of the drivers is not a distinct case. Significant numbers of crashes end in rural roadways due to the lack of effective countermeasures to separate opposing traffic flows. As a result, a major concern involves vehicles crossing the centerline and ending in either sideswiping or head-on collisions. At the same time, improper passing related crashes on divided roadways are not negligible in count. Investigating the patterns of improper passing related crashes is thus called for.

Driving is a multifaceted task, which needs perception, comprehension and projection of states of the roadway condition, as well as decision making on courses of spontaneous action and execution of driving behaviors. Passing on rural roadways is complex due to the roadway environment and opposite direction vehicles if the roadway is undivided. Many studies envisioned to explore the complex nature of passing to improve roadway standards by re-evaluating the patterns of improper passing crashes. Factors range from a wide variety in improper passing crashes: presence of curvature, curve radius, curve length, skid condition, weather, lane width, and presence of countermeasures like raised pavement markers or edge lines, size of the vehicle, posted speed, vehicle speed, lighting condition, traffic volume, and time of the day. This study has used a wider variety of geometric and environmental variables from Florida to investigate the hidden pattern of passing behavior related crashes by incorporating SHRP-2 RID data. To increase the precision of the knowledge extraction, this study has used association rules negative binomial miner algorithm.

LITERATURE REVIEW

Literature to date has yet to develop and understand the impact of various passing behaviors on crash severity but has made significant progress in identifying influences on passing behavior including roadway geometrics and human factor variables.

Historically, research has focused on refinement of roadway infrastructure guidelines, provided primarily by AASHTO and FHWA, which provides criteria for two-lane, two-way highways such as minimum passing sight distance and no passing zones (1, 2). Refinement of these guidelines included re-evaluation and classification of passing maneuvers and development of passing models that factored in higher speeds (1, 2). Benefits of passing identified included: reduction of congestion, delays and overall improvement of level of service for a roadway (3).

Current studies have explored the relationship between roadway geometry and driver behavior on rural two lane roadways through simulation, field observation and, for a limited number of studies, naturalistic driving data. For example, the presence of a guardrail, highly visible pavement markers or the addition of centerline and shoulder rumble strips greatly improved lane positioning of vehicles, which reduced departures from the roadway, i.e., encroachments (4, 5). Lower overtaking speeds were found on roadways with no centerline markings and narrower lanes, while wider lanes, that offered more space, resulted in significantly higher passing speeds (6).

Roadway curvature, specifically the size of the curve, could also impact the likelihood of lane departure, left side encroachment and the amount of time drivers took to pass a lead vehicle, i.e., critical passing gaps where larger curves, which afford larger sight distances, had smaller passing gaps compared to smaller or narrower curves (5).

Nighttime driving on rural roads, can be particularly dangerous due to limited visibility and availability of lighting which are significant factors in passing behavior (7, 8, 9). Nighttime conditions were correlated with lower traffic volumes, larger passing gaps and larger headways (5, 7, 10). The effect of speed on passing behavior has been studied through variables including: speed of subject vehicle, lead vehicle, opposing lane vehicle, effects on driver frustration, age and gender (7, 11, 12, 13). Higher roadway speeds were associated with smaller passing gaps, and posted speed limits also influenced driver's overtaking speed wherein lower posted speed limits correlated with lower overtaking speeds (6, 7, 11).

While the focus of the current research did not encompass analysis of human factor variables, it is important, nonetheless to highlight the influence of geometric variables on driver behavior in relation to passing behavior. Passing maneuvers on two lane roads require a driver to merge into an opposing lane of traffic and continually evaluate both the speed and distance of lead and oncoming vehicles (if present) accurately before choosing to execute a pass, which can be difficult (7, 8). The size of a vehicle can impact both the distance maintained from a lead vehicle and a driver's ability to accurately judge distance as larger vehicles can obscure a driver's vision and smaller vehicles can be harder to distinguish impacting passing gaps and headways (7, 9). Nighttime driver behavior studies implied higher levels of caution wherein, drivers allowed more space to accomplish a pass yet completed the pass in less time (11). Speed's effect on several variables showed: driver's motivations to pass increased when "constraints" i.e. slower moving vehicles imposed on subject's speed; females and older drivers exhibited higher levels of caution when passing at higher speeds; males and younger drivers exhibited higher acceptance of smaller passing gaps (7, 13). In addition, while the scope of this research was did not explore video data, outcomes for future research considerations will include analysis of driver glances from roadways as their role on speed has also been shown to have potential influence on driver's speed and will be an important variable for future consideration (5).

Of the available literature that has examined preventative measures, findings have shown that the addition of passing lanes and shoulders to improve traffic operations, in certain areas, be cost effective and a significant safety improvement in reducing the potential number of fatal crashes on rural roads (14, 15). Interventions that improve roadway safety have been identified as one of the top priorities for roadway users (16). Forward collision warning technology has also shown promise in its ability to track objects in front of the vehicle and provide feedback of an impending collision, a valuable system to assist with overtaking maneuvers (11).

SHRP-2 RID DATA

SHRP-2 project populated a Roadway Information Database (RID) with data from the SHRP-2 mobile data collection project (S04B); existing roadway data from government, public, and private sources; and supplemental data that further characterize traffic operations (17). This database provides good quality data that are linkable to the SHRP-2 Naturalistic Driving Study (NDS) database utilizing geographic information system (GIS) tools. The RID is the supplementary tools for safety researchers to look at data sets of selected road characteristics and study matching NDS trips to explore the relationships between driver, vehicle, and roadway.

Florida RID maintains traffic crash data for six years (2005-2010). The dataset contains crash, vehicle, and person information. The database is divided into two parts: 1) Local Roadways, and 2) State Highway Systems (SHS) roadways. To prepare the database for this study, the research team used both roadways.

The primary focus of the database preparation for this study is to create a detailed database on crashes related to improper passing. TABLE 2 lists the passing behavior related attributes in RID for Florida.

TABLE 2 Improper passing related attributes in RID

State	Attribute
Florida	CRSHCAUSE1 - 05 (Improper Lane Change) - 15 (Improper Passing)

FIGURE 8 illustrates distribution of improper passing crashes in Florida based on two major facility types: divided roadways, and undivided roadways. The primary notion of this study was to investigate the passing related crashes on rural two-lane roadways. Due to the very small sample size of the data, the research considered a broader group of facility (rural roadways) to perform the analysis. Research also directed towards diving the database into two major groups (divided and undivided roadways) to identify the safety issues associated with roadway division.

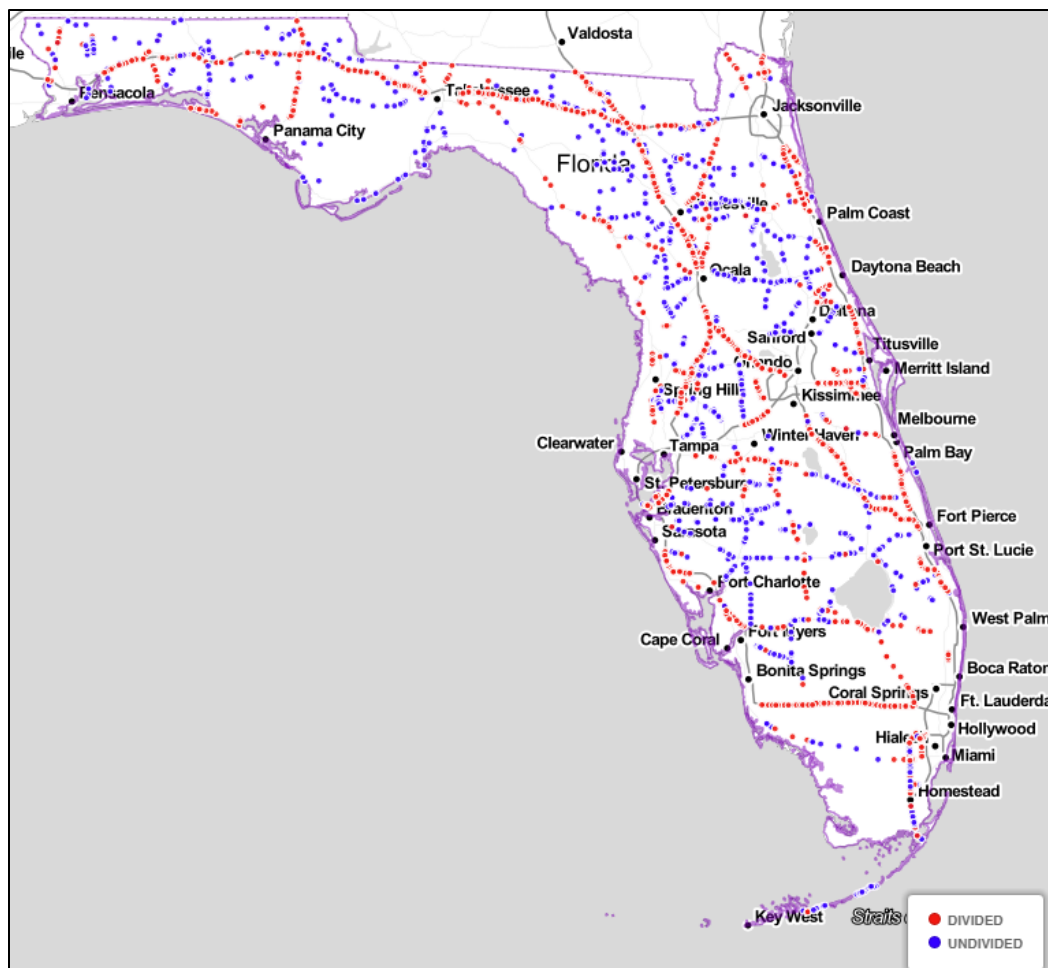


FIGURE 8 Improper passing crashes on divided and undivided roadways in Florida.

Descriptive Statistics

Florida RID database maintains a larger number of variables. The research team conducted a detailed literature review to investigate the significant factors associated with improper passing crashes. A group of major roadway and environmental variables was selected from the research synthesis and was explored in Florida RID database for availability. Due to significant number of missing information on variables like vehicle type, vehicle model, vehicle size, and driver age in the improper passing crash database, the research team omitted these variables in preparing the database. The research team also tested multi-collinearity issues on a pre-final dataset. Moreover, the research team used random forest algorithm to conduct variable importance for selecting a final group of variables. The details of the variable selection method is not described here to make the study more focused on the current scope of the research.

TABLE 3 lists descriptive statistics of the final ten variables. Significant difference is visible for different variables in the divided and undivided roadways. Skid number in between 30 to 40 is higher in percentage on divided roadways, while skid number above 40 is higher in percentage on undivided roadways. Maximum-posted speed is higher in percentage in divided roadways while comparing with undivided roadways. Divided roadways also exhibit higher percentage in high Annual Average Daily Traffic (AADT), and high average trucking

percentages. On the other hand, undivided roadways exhibit higher percentage of fatal and injury crashes.

TABLE 3 Descriptive Statistics of the Key variables

Category	Percentage		Category	Percentage	
	Divided	Undivided		Divided	Undivided
Surface Width (ft.)			AADT (vpd)		
0.00-10.00	0.03%	0.17%	0-10,000	4.97%	64.17%
10.01-20.00	0.78%	14.21%	10,000-20,000	15.17%	26.24%
20.01-30.00	62.60%	81.95%	20,000-30,000	18.82%	3.92%
30.01-40.00	36.05%	3.49%	30,000-40,000	13.21%	1.92%
> 40	0.55%	0.17%	40,000-50,000	14.33%	0.96%
Skid Number	Divided	Undivided	50,000-60,000	10.40%	0.87%
0.00-30.00	10.49%	1.66%	> 60,000	23.10%	1.92%
30.01-40.00	70.50%	33.13%	Lighting Condition		
40.01-50.00	17.93%	48.13%	Daylight	70.59%	70.70%
> 50	1.09%	17.09%	Dark(No Street light)	21.11%	21.01%
Shoulder Type			Dark(Street Light)	4.37%	3.49%
Paved	25.60%	77.77%	Dawn/Dusk	3.94%	4.80%
Paved with Warning	73.57%	11.25%	Weather Condition		
Other	0.83%	10.99%	Inclement	34.04%	28.68%
Maximum Speed (mph)			Non-inclement	65.96%	71.31%
0-50	3.36%	13.60%	Surface Condition		
50-60	10.14%	77.86%	Dry	83.65%	91.02%
> 60	86.50%	8.54%	Wet	16.35%	8.98%
Average Truck Percentage			Crash Severity		
0.00-5.00	3.48%	5.41%	Fatal	1.49%	4.80%
5.01-10.00	14.97%	24.59%	Incapacitating injury	7.93%	17.18%
10.01-15.00	23.56%	29.47%	Possible Injury	17.90%	17.96%
	26.00%	18.48%	Non-incapacitating injury	13.70%	17.44%
15.01-20.00			No Injury	58.98%	42.63%
> 20	32.00%	22.06%			

ROADWAYS MARKINGS

Passing permitted and No-passing Markings

Manual on Uniform Traffic Control Device (MUTCD) (18) provides detailed description on passing permitted and no-passing roadway markings. As per the MUTCD guidelines, the center line markings indicating the passing condition are:

1. A broken yellow line center line markings (as shown in FIGURE 9a) indicates that passing is allowed in both the directions.
2. A double yellow line centerline markings consisting of a solid line and a broken yellow line (as shown in FIGURE 9b) indicates that passing is only allowed for the traffic travelling adjacent to the broken line and passing is prohibited for the traffic travelling adjacent to the solid line.
3. A double yellow line consisting of both solid lines (as shown in FIGURE 9c) indicates that passing is prohibited in both the directions. The same markings should be used as the centerline marking for 4 or more lanes undivided highways.

4. On three lane roadways where the direction of travel in the center line transitions from one direction to the other, a no passing buffer zone marking (as shown in FIGURE 9d) should be provided. The passing is not allowed in this zone.

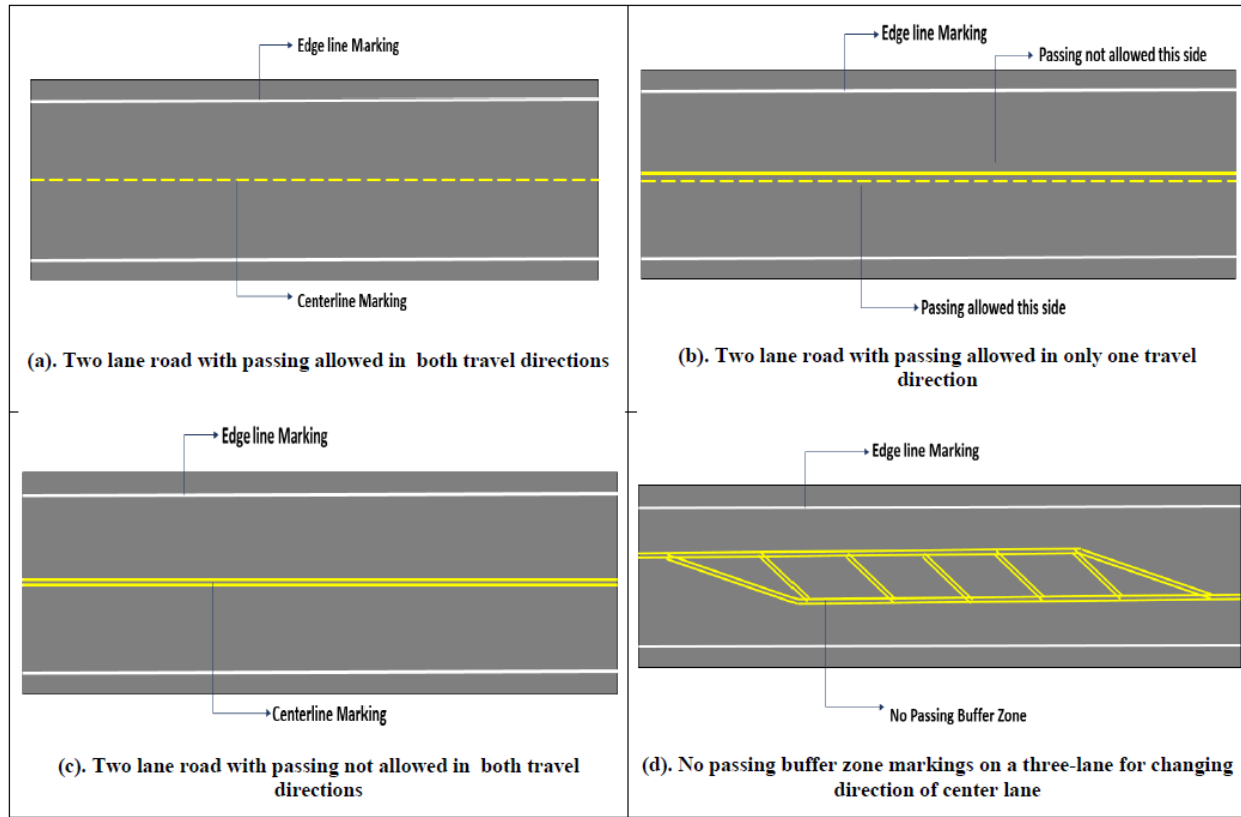


FIGURE 9 Passing and Non-passing roadway markings

Issues with Improper Passing Crashes

FIGURE 10 illustrates the criteria for passing permitted and no-passing zones. FIGURE 10a shows the permissible passing behavior on a 2 lane road with a single broken yellow centerline line. FIGURE 10b shows the permissible passing behavior on a 2 lane road with broken yellow line and solid yellow line marking. FIGURE 10c shows the improper passing behavior on a road with similar markings. FIGURE 10d shows the improper passing behavior on a road which has a solid double yellow center line marking. FIGURE 10e indicates that passing zone is not allowed near the buffer zone.

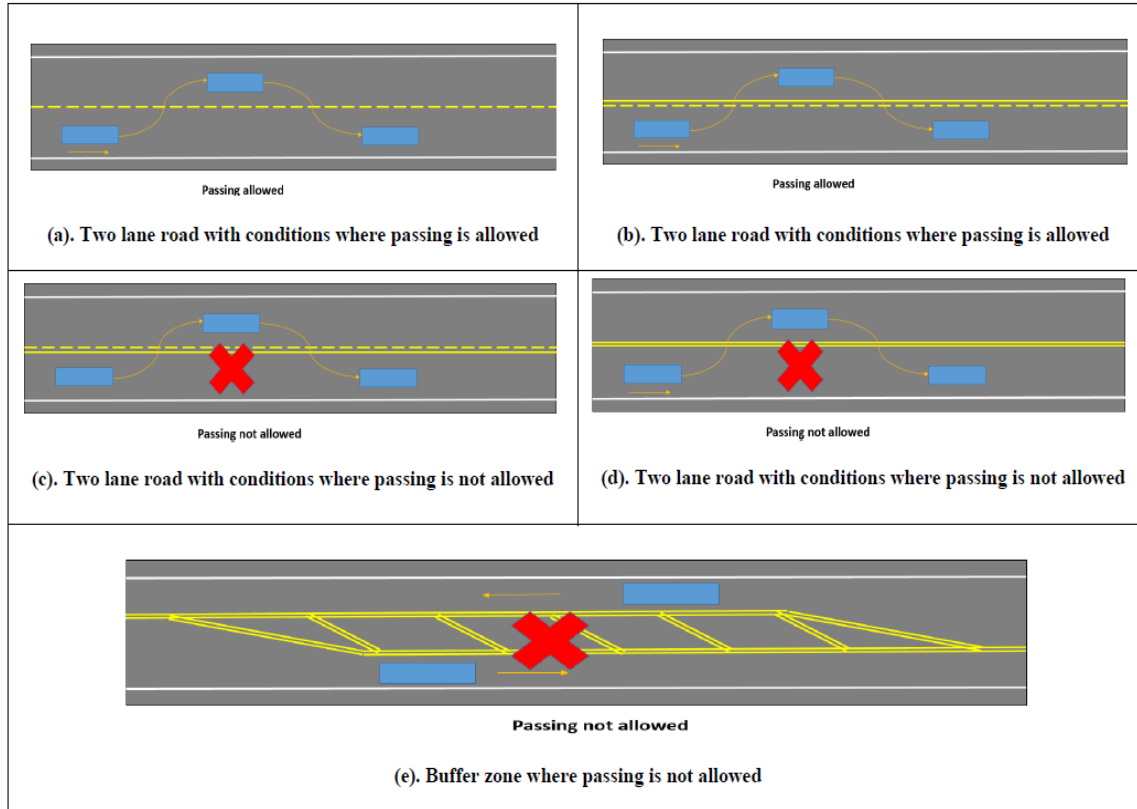


FIGURE 10 Passing permitted and No passing Criteria.

Usually no passing marking are seen in the areas with curves or low visibility ranges. Improper passing mainly indicates the behavior of passing vehicles on no-passing zones on undivided roadways. However, for divided roadways, improper passing may relate to other factors. Improper passing crashes mainly involve driver cognition and attitude while passing. Previous research show that roadway characteristics also play a key role in the likelihood of injury or fatal crashes in improper passing crashes. The current research is focused on identifying the association between the key geometric features on the zones where crashes happened due to improper passing.

ASSOCIATION RULES NEGATIVE BINOMIAL (NB) MINER

This study used association rules NB miner to perform the analysis. A short introduction on the theoretical aspects of precision rules is described here. A more detailed introduction to this algorithm can be found in Hashler study (19).

Agrawal et al. introduced the data mining on the transaction data based on the associated items using the mining association rules in 1993 (19). Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of n distinct items and $Q = \{q_1, q_2, \dots, q_m\}$ be the transactions. Each transaction in Q contains a subject of the items in I . A rule is defined as an implication of the form $\{Antecedent \rightarrow Consequent\}$ or $\{M \rightarrow N\}$ where $M, N \subseteq I$ and $M \cap N = \emptyset$. An itemset that contains d items is said to have length or size d and is called a d -itemset.

Support is defined on itemset $M \subseteq I$ as the proportion of transactions in which all items in Z :

$$support(M) = \frac{freq(M)}{|Q|} \quad (1)$$

Where:

$freq(M)$ = frequency of itemset M (number of transactions in which M occurs) in Q , and
 $|Q|$ = number of transactions in the database.

For the rule: $M \rightarrow N$, confidence can be calculated as:

$$Confidence(M \rightarrow N) = \frac{Support(M \rightarrow N)}{Support(M)} \quad (2)$$

Relevance of the itemset Z to the user depends on the constraint: $support \geq \sigma$, where: σ is a user-specified minimum support value. Itemsets that satisfy the minimum support constraint are known as frequent itemsets since their occurrence frequency surpasses a set frequency threshold.

The minimum support must be considered to ensure that a particular item in the dataset is significant. A low minimum support can increase the number of generated rules exponentially. It is very difficult to interpret such large number of rules. One of the most common approaches to the application of mining the association rules is the lift. It is a measure of the deviation from statistical independence of the relationship between M and N and is useful to identify associations that are significant deviations from the assumption of statistical independence. It can be defined as:

$$Lift(M \rightarrow N) = \frac{Support(M \rightarrow N)}{Support(M) \times Support(N)} \quad (3)$$

The methodology used in this paper, however replaces the usage of the lift for a model that can be used to evaluate the deviation for the set of all possible 1-extensions of an itemset together to find a local frequency constraint for these extensions. Under this approach, the frequency of itemsets is assumed to follow a probability distribution, negative binomial (NB) in this case. The probability mass function of NB distribution is given by:

$$Pr[P = p] = (1 + b)^{-t} \frac{\Gamma(t + p)}{\Gamma(p + 1)\Gamma(t)} \left(\frac{b}{1 + b}\right)^p, p = 0,1,2 \dots \quad (4)$$

Where:

t = dispersion parameter

b = mean parameter

p = realization of the random variable P

Let $\Sigma = \sigma f$, where f is the number of transactions in the database. Here, one can consider the frequency threshold equivalent to the minimum support σ . Then the expected number of 1-itemsets, which satisfy the frequency threshold, Σ is given by

$$xPr[P \geq \Sigma] \quad (5)$$

Where:

x = the number of variables items.

The counts for the 1-extensions of association m can be modeled by random variable P_m with the following probability mass function:

$$\Pr[P_m = p] = (1 + b_m)^{-t} \frac{\Gamma(t + p)}{\Gamma(p + 1)\Gamma(t)} \left(\frac{b_m}{1 + b_m}\right)^p, \text{ for } p = 0, 1, 2 \dots \quad (6)$$

Let M be the set of all 1-extension of a known association m which are generated by joining m with all candidate items, which co-occurrence with m in at least ρ transactions. For set M , the predicted precision is:

$$\text{predicted precision of a rule} = \begin{cases} (o - e) / o & \text{if } o \geq e \text{ and } o > 0 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Where:

o is the observed and e is the expected number of candidate items which have a co-occurrence frequency with item set m of $p \geq \rho$. The observed number is calculated as the sum of items with count p by $o = \sum_{p=\rho}^{p_{max}} o_r$, where p_{max} is the highest observed co-occurrence. The expected number is given by the baseline model as $e = (x - |m|)\Pr[P_m \geq \rho]$, where $x - |m|$ is the number of possible candidate items for pattern m .

This method overcomes the conventional drawbacks of defining a minimum threshold support. In this method, a baseline model was first developed to count co-occurrences of items in the database. For a defined precision threshold (suppose 90%), a local frequency constraint for all 1-extensions of an itemset can be found. The precision threshold is overall parameter that can be used to fine-tune the accurate rules. This method uses a faster algorithm than conventional association rules and it is very effective in identifying non-spurious (co-variation) associations.

Link to Statistical Inference

Before being able to identify underlying associations, this procedure requires a statistical model be fitted to the count of the number of transactions involving every possible pair of items in the dataset. There are two clear advantages to choosing an NB specification while doing so: (a) only the two parameters t and b are estimated from eq. (4) without parameterizing them; and (b) it is reasonable to assume that as the number of items in the database increases, the distribution of the number of pair-draws would become more smooth and could be well fitted by the geometric probability distribution. The negative binomial distribution would then emerge as a convolution of geometric random variables if their long-term mean parameter can reasonably be assumed equal.

These advantages are, naturally, contingent to a reasonable model fit to the distribution of counts of transaction involving a randomly picked pair of items. Assuming that the fit is adequate, a criterion that screens unusually large counts involving pairs of items (such as eq. 7) naturally discards cases where the observed frequency is fairly likely and flags cases where the probability is fairly unlikely. This procedure is analogous to screening minimizing type-I error, and its statistical interpretation is equivalent to the p-value. However, because of the simple parameterization of the NB model, unsupervised fit can be performed until an acceptable fit is achieved. Given an acceptable probability model for the frequency of transactions of pairs of

items, this method yields a fast algorithm, (arguably as fast as an algorithm with conventional association rules or even faster) that it is bound to be more effective in identifying non-spurious (co-variation) associations, compared with algorithms that rely on the simplistic definition of lift.

RESULTS

The research team used open source R package ‘arulesNBMiner’ to perform analysis on two broader groups of unsupervised dataset based on the roadway facility types: divided, and undivided roadways (20, 21). As no response variable was pre-selected in both of the datasets, the learning algorithm of the model development was unsupervised in nature. It is important to note that lower precision threshold usually increase significant amount of increase of the generated rules. Number of itemsets is also another issue to interpret the results. By investigating these issues, the research team used precision threshold as 0.9 and maximum number of itemset as 3. TABLE 4 lists the top twenty (based on higher precision) two itemset rules for divided and undivided roadways. For divided roadways, the significant attributes are paved shoulder with warning, higher AADT, lane widths (in between 20 to 40 ft.), and higher speed (above 50 mph). The significant attributes for undivided roadways are higher skid number, lower AADT, narrower lane widths (in between 20 to 30 ft.), average percentage of trucks, and crash severity (fatal and injury). The findings of the two itemset precision rules are consistent with the other studies that point to speed, lane widths, traffic volume to improper passing crashes (5, 6, 7, 11).

From TABLE 4, it is seen that the top twenty rules (generated from divided rural roadways) do not exhibit severity in either antecedent or consequent. On the other hand, severity is one of the dominating factor in the rules for undivided roadways. For example, rule number 8 for undivided roadways is: $\{Severity = Incapacitating Injury\} \rightarrow \{Speed = 50 - 60 mph\}$ with a precision value of 0.963. This rule indicates that the co-occurrence of incapacitating injury crashes with higher speed (50-60 mph) is over 95%.

TABLE 4 Two Itemset Precision Rules for Divided and Undivided Roadways

Divided Roadways				Undivided Roadways			
No.	Antecedent	Consequent	Precision	No.	Antecedent	Consequent	Precision
1	AADT ¹ =50,000-60,000	Shoulder=Paved with Warning	0.965	1	Skid ⁶ =50	AADT=0-10,000	0.977
2	Width ² =30.01-40.00	Shoulder=Paved with Warning	0.965	2	Skid= 50	Width=20.01-30.00	0.971
3	AADT=40,000-50,000	Shoulder=Paved with Warning	0.963	3	Skid= > 50	Speed=50-60	0.971
4	AADT= 60,000	Shoulder=Paved with Warning	0.960	4	Skid=40.01-50.00	Width=20.01-30.00	0.967
5	AADT= 60,000	Surface ⁵ =Dry	0.958	5	AADT=0-10,000	Width=20.01-30.00	0.967
6	Speed ³ =50-60	Surface=Dry	0.957	6	AADT=0-10,000	Shoulder=Paved	0.965
7	AADT=30,000-40,000	Width=20.01-30.00	0.954	7	Trucks ⁷ =5.01-10.00	Width=20.01-30.00	0.964
8	Speed=0-50	Width=20.01-30.00	0.948	8	Severity ⁸ =Incapacitating injury	Speed=50-60	0.963
9	AADT= 60,000	Skid=30.01-40.00	0.947	9	Skid=40.01-50.00	Speed=50-60	0.963
10	AADT=20,000-30,000	Surface=Dry	0.946	10	Severity=Non-incapacitating injury	Shoulder=Paved	0.962
11	Speed=50-60	Shoulder=Paved	0.946	11	Skid=40.01-50.00	Light ⁹ =Daylight	0.961
12	AADT= 60000	Width=30.01-40.00	0.946	12	Severity=Fatal	Shoulder=Paved	0.959
13	Surface=Wet	Speed= 60	0.946	13	Shoulder=Other	Speed=50-60	0.959
14	AADT=50,000-60,000	Surface=Dry	0.945	14	Trucks= 20	Speed=50-60	0.957
15	AADT=40,000-50,000	Skid=30.01-40.00	0.943	15	Skid=40.01-50.00	AADT=0-10,000	0.957
16	AADT=30,000-40,000	Surface=Dry	0.941	16	Trucks=5.01-10.00	Speed=50-60	0.955
17	AADT= 60,000	Speed= 60	0.941	17	AADT=0-10,000	Speed=50-60	0.955
18	Shoulder ⁴ =Paved with Warning	Speed= 60	0.941	18	Severity=Fatal	Width=20.01-30.00	0.955
19	AADT=0-10,000	Shoulder=Paved	0.940	19	Severity=Fatal	AADT=0-10,000	0.955
			0.939			Weather ¹⁰ =Non-	
20	Width=30.01-40.00	Surface=Dry		20	Trucks= 20	inclement	0.954

Note: ¹AADT= Annual Average Daily Traffic (vpd), ²Width= Surface Width (ft.), ³Speed=Maximum Posted Speed (mph), ⁴Shoulder=Type of Shoulder, ⁵Surface=Surface Condition, ⁶Skid=Skid Number, ⁷Trucks=Average Percentage of Trucks, ⁸Severity=Crash Severity, ⁹Light=Lighting Condition, ¹⁰Weather=Weather Condition

TABLE 5 lists the top twenty (based on higher precision) three itemset rules for divided and undivided roadways. For divided roadways, the significant attributes are paved shoulder with warning, higher AADT, weather, lane widths (in between 30 to 40 ft.), and higher speed (60 mph). The significant attributes for undivided roadways are higher skid number, lower AADT, narrower lane widths (in between 20 to 30 ft.), average percentage of trucks, and crash severity (non-incapacitating injury or no injury). The findings of the two itemset precision rules are consistent with the other studies that point to speed, lane widths, traffic volume to improper passing crashes (5, 6, 7, 11). Weather exhibits a significant factor in the group of rules generated for divided roadways. For example, rule number 2 for divided roadways is: $\{Surface\ Width = 30 - 40\ ft.,\ Surface\ Condition = Wet\} \rightarrow \{Weather = Inclement\}$ with a precision value of 0.989. This rule indicates that the co-occurrence of these three attributes in divided roadway dataset is over 98%. On the other hand, average percentage of trucks exhibits significant dominance on undivided roadways. For example, the top rule for undivided roadways is: $\{Skid\ Number = 40 - 50,\ Average\ Percentage\ of\ Trucks = 20\%\} \rightarrow \{Roadway\ Width = 220 - 30\ ft.\}$ with a precision value of 0.990. This rule indicates that the co-occurrence of these three attributes in divided roadway dataset is 99%.

The following findings below summarize the contents of knowledge extraction from the top twenty precision rules (two itemset and three itemset) for both divided and undivided roadways:

- Improper crashes on divided roadways associate with higher AADT, wider roadways, and higher speed for two itemset rules. On the other hand, undivided roadways incorporate lower AADT, narrower roadways, average percentage of trucks, and severe/fatal crashes.
- Fatal or severe injuries are dominant on undivided roadways for two itemset rules.
- For three itemset rules, inclement weather contributes significantly on improper crashes on divided roadways.
- For three itemset rules, average percentage of trucks contributes significantly on improper crashes on undivided roadways.

TABLE 5 Three Itemset Precision Rules for Divided and Undivided Roadways

Divided Roadways				Undivided Roadways			
No.	Antecedent	Consequent	Precision	No.	Antecedent	Consequent	Precision
1	Weather=Non-inclement, AADT=50,000-60000	Shoulder=Paved with Warning	0.989	1	Skid=40.01-50.00, Trucks=> 20	Width=20.01-30.00	0.990
2	Width=30.01-40.00, Surface=Wet	Weather=Inclement	0.989	2	Skid=40.01-50.00, Trucks=> 20	Speed=50-60	0.989
3	Width=30.01-40.00, Weather=Non-inclement	Shoulder=Paved with Warning	0.989	3	Skid=> 50, Shoulder=Paved	Speed=50-60	0.989
4	Weather=Non-inclement, AADT=40,000-50,000	Shoulder=Paved with Warning	0.988	4	Skid=> 50, Light=Daylight	Width=20.01-30.00	0.988
5	Weather=Non-inclement, AADT= 60,000	Shoulder=Paved with Warning	0.987	5	Skid=> 50, Light=Daylight	Speed=50-60	0.988
6	Shoulder=Paved with Warning, Weather=Non-inclement	Surface=Dry	0.986	6	Skid=40.01-50.00, AADT=0-10000	Width=20.01-30.00	0.988
7	Width=30.01-40.00, Trucks= 20	Shoulder=Paved with Warning	0.986	7	Severity=Non-incapacitating injury, Trucks=5.01-10.00	Width=20.01-30.00	0.987
8	AADT=60,000, Trucks=15.01-20.00	Shoulder=Paved with Warning	0.986	8	Skid=> 50, Severity=No Injury	AADT=0-10,000	0.987
9	Width=30.01-40.00, Trucks=15.01-20.00	Shoulder=Paved with Warning	0.986	9	Skid=> 50, Severity=No Injury	Speed=50-60	0.987
10	Weather=Non-inclement, AADT= 60,000	Surface=Dry	0.986	10	Skid=> 50, Speed=50-60	Width=20.01-30.00	0.986
11	AADT= 60,000, Trucks=10.01-15.00	Shoulder=Paved with Warning	0.986	11	Skid=> 50, Speed=50-60	AADT=0-10,000	0.986
12	AADT= 60,000, Trucks=10.01-15.00	Speed= 60	0.986	12	Skid=40.01-50.00, Trucks=> 20	AADT=0-10,000	0.986
13	Width=30.01-40.00, Weather=Non-inclement	Speed= 60	0.986	13	Skid=40.01-50.00, Trucks=> 20	Surface=Dry	0.986
14	Width=30.01-40.00, AADT=40,000-50,000	Shoulder=Paved with Warning	0.986	14	Skid=> 50, Shoulder=Paved	AADT=0-10,000	0.986
15	Width=30.01-40.00, AADT=40,000-50,000	Speed= 60	0.986	15	Skid=> 50, AADT=0-10,000	Width=20.01-30.00	0.986
16	Weather=Non-inclement, AADT=50,000-60,000	Surface=Dry	0.986	16	Skid=> 50, Weather=Non-inclement	AADT=0-10,000	0.985
17	Weather=Non-inclement, AADT=50,000-60,000	Speed= 60	0.986	17	Severity=Non-incapacitating injury, AADT=0-10,000	Shoulder=Paved	0.985
18	Width=30.01-40.00, AADT=50,000-60,000	Shoulder=Paved with Warning	0.986	18	Skid=40.01-50.00, AADT=0-10,000	Shoulder=Paved	0.985
19	Width=30.01-40.00, Light=Dark (no street light)	Shoulder=Paved with Warning	0.985	19	Skid=40.01-50.00, AADT=0-10,000	Surface=Dry	0.985
20	Shoulder=Paved with Warning, AADT=20,000-30,000	Width=20.01-30.00	0.985	20	AADT=0-10,000, Trucks=5.01-10.00	Speed=50-60	0.985

CONCLUSIONS

The current study used Florida RID crash data for six years (2005-2010) to identify the key issues associated with improper passing crashes. Facility types on rural roadways (divided and undivided) shows significant differences in the descriptive statistics of the geometric and environmental variables. The undivided roadways shows a higher likelihood of fatal and injury crashes compared to divided roadways. The research team used an unsupervised data mining method (Association Rules NB Miner) to generate two-itemset and three-itemset rules. This method is better than conventional association rules mining, as no predefined support threshold is needed to be included while generating rules. The current method is more robust because it identifies rules with higher precision (by using a single performance measure) from a large dataset by identifying key co-occurrence, while the conventional association rules use three parameters as performance measures. For two itemset precision rules in divided roadways, paved shoulder with warning, high AADT, are high speed are dominant in both antecedents and consequents. On the other hand, for undivided roadways, the dominant associated factors are crash severity, low AADT, percentage of trucks, and skid number. For three itemset rules while considering divided roadways, the dominant factors are weather, high AADT, and percentage of trucks. For divided roadways, skid, low AADT, and crash severity are dominant. The findings of this research are consistent with those of previous studies. The top twenty rules for both roadways clearly show that facility type (either divided or undivided) play a significant role in the higher likelihood of injury or fatal crashes. However, divided roadways also need attention in reducing improper passing crashes on roadways with high AADT and speed. The findings of the current study will help the safety professionals in mitigating improper passing crashes and crash severities.

One of the limitations of this study is that it considers only geometric variables to conduct analysis. Therefore, future research should incorporate at least two major issues: identify exact passing permitted and no passing locations from the spatially coded crash information to perform a more robust analysis; use NDS data for exploring behavioral perspectives.

ACKNOWLEDGEMENTS

The research reported here was sponsored by the Texas A&M Transportation Institute (TTI) Strategic Research Program. The authors also appreciate the efforts of the TTI staff Bahar Dadashova who helped in collecting RID.

REFERENCES

1. Carlson, P., J. Miles, and P. Johnson. Daytime High-Speed Passing Maneuvers Observed on Rural Two-Lane, Two-Way Highway: Findings and Implications. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1961, 2006, pp. 9–15.
2. Jenkins, J., and L. Rilett. Classifying Passing Maneuvers: A Behavioral Approach. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1937, 2005, pp. 14–21.

3. Romana, M. Passing Activity on Two-Lane Highways in Spain. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1678, 1999, pp. 90–95.
4. Gates, T., P. Savolainen, T. Datta, R. Todd, B. Russo, and J. Morena. Use of Both Centerline and Shoulder Rumble Strips on High-Speed Two-Lane Rural Roadways. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2301, 2012, pp. 36–45.
5. Hallmark, S. L., S. Tyner, N. Oneyear, C. Carney, and D. Mcgehee. Evaluation of driving behavior on rural 2-lane curves using the SHRP 2 naturalistic driving study data. *Journal of Safety Research*, Vol. 54, 2015, pp. 17-27
6. Shackel, S. C., and J. Parkin. Influence of road markings, lane widths and driver behaviour on proximity and speed of vehicles overtaking cyclists. *Accident Analysis & Prevention*, Vol. 73, 2014, pp. 100–108.
7. Farah, H., and T. Toledo. Passing behavior on two-lane highways. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 13, No. 6, 2010, pp. 355–364.
8. Papakostopoulos, V., D. Nathanael, E. Portouli, and N. Marmaras. The effects of changes in the traffic scene during overtaking. *Accident Analysis & Prevention*, Vol. 79, 2015, pp. 126–132.
9. Levulis, S. J., P. R. Delucia, and J. Jupe. Effects of oncoming vehicle size on overtaking judgments. *Accident Analysis & Prevention*, Vol. 82, 2015, pp. 163–170.
10. Llorca, C., A. Moreno, A. García, and A. Pérez-Zuriaga. Daytime and Nighttime Passing Maneuvers on a Two-Lane Rural Road in Spain. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2358, 2013, pp. 3–11.
11. Chen, R., K. D. Kusano, and H. C. Gabler. Driver Behavior During Overtaking Maneuvers from the 100-Car Naturalistic Driving Study. *Traffic Injury Prevention*, Vol. 16, No. sup2, 2015, pp. S176-S181.
12. Kinnear, N., S. Helman, C. Wallbank, and G. Grayson. An experimental study of factors associated with driver frustration and overtaking intentions. *Accident Analysis & Prevention*, Vol. 79, 2015, pp. 221–230.
13. Vlahogianni, E. I., and J. C. Golias. Bayesian modeling of the microscopic traffic characteristics of overtaking in two-lane highways. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 15, No. 3, 2012, pp. 348–357.
14. Schrock, S., R. Parsons, and H. Zeng. Estimation of Safety Effectiveness of Widening Shoulders and Adding Passing Lanes on Rural Two-Lane Roads. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2203, 2011, pp. 57–63.
15. Brewer, M., S. Venglar, K. Fitzpatrick, L. Ding, and B.-J. Park. Super 2 Highways in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2301, 2012, pp. 46–54.
16. Mutabazi, M., E. Russell, and R. Stokes. Drivers' Attitudes, Understanding, and Acceptance of Passing Lanes in Kansas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1628, 1998, pp. 25–33.
17. SHRP-2 RID. <http://www.ctre.iastate.edu/shrp2-rid/> Last Accessed: July, 2016.

18. *Manual on uniform traffic control devices: for streets and highways*. U.S. Dept. of Transportation, Federal Highway Administration, Washington, D.C., 2009.
19. Hashler, M. A Model-Based Frequency Constraint for Mining Associations from Transaction Data. *Data Mining and Knowledge Discovery*, Vol. 13, pp. 137–166, 2006.
20. 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>. Last accessed: July, 2016.
21. Hashler, M. R package ‘arulesNBMiner’. <https://cran.r-project.org/web/packages/arulesNBMiner/arulesNBMiner.pdf>. Last accessed: July, 2016.