TRANSPORTATION RESEARCH
CIRCULAR

Number E-C221 May 2017

SHRP 2 Safety Data Student Paper Competition, 2015–2016

TRB TRANSPORTATION RESEARCH BOARD
The National Academies of
SCIENCES · ENGINEERING · MEDICINE

TRANSPORTATION RESEARCH CIRCULAR E-C221

SHRP 2 Safety Data Student Paper Competition, 2015–2016

Sponsored by
TRB Oversight Committee for Use and Oversight
of SHRP 2 Safety Data, Phase I

May 2017

Transportation Research Board
Washington, D.C.
www.TRB.org

TRANSPORTATION RESEARCH CIRCULAR E-C221
ISSN 0097-8515

The **Transportation Research Board** is one of seven programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal.

The **Transportation Research Board** is distributing this E-Circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this circular was taken directly from the submission of the authors. This document is not a report of the National Academies of Sciences, Engineering, and Medicine.

Foreword

Alyssa Hernandez
Program Officer, TRB

The second Strategic Highway Research Program (SHRP 2) Safety Study, completed in 2015, collected an unprecedented amount of objective data on driver behavior and the driving context. The SHRP 2 Naturalistic Driving Study (NDS) captured detailed data on 3,500+ volunteer passenger-vehicle drivers, including continuous driving data and video of the road and the driver over 35 million vehicle miles and more than 4,200 crashes and near-crashes across sites in six states. The SHRP 2 Roadway Information Database (RID) collected detailed roadway data on 12,500 centerline miles, compiled existing driving context data on another 200,000 centerline miles, and made it possible to link the roadway data to the driving data. Together, the NDS and the RID are the "SHRP 2 Safety Data." Phase 1 of SHRP 2 Safety Data Implementation and Oversight is the initial post-data collection phase, begun in 2014. The objectives of this phase are to make the Safety Data widely available to qualified researchers and to gain experience and data to support decisions about the implementation and oversight of the data beyond the first five years.

This e-Circular contains papers submitted to the first Student Paper Competition: SHRP 2 Safety Data Bonanza. The SHRP 2 Safety Data Program and the TRB Oversight Committee for Use and Oversight of SHRP 2 Safety Data, Phase 1 sponsored this competition to promote use of the SHRP 2 Safety Data, to extract new insights and applications of the data, and to foster the next generation of leaders in surface transportation. A call for abstracts was issued on July 31, 2015 to undergraduate and graduate students across the country, soliciting innovative ideas for using the data. The Review Panel for the SHRP 2 Safety Data Student Paper Competition selected six students to conduct their research proposals; they received a data export, conducted their analysis, and were sponsored to attend the TRB Annual Meeting in January 2016 to present their results at a poster session. The students went on to develop research papers from their analyses.

Three exemplary papers were selected by the review panel to be published in this e-Circular. These student papers examine such topics as driver distraction, fault status in vehicle conflicts, and turning behavior. They are among the first research papers to be published that utilize the SHRP 2 Safety Data. They represent just a sample of the diverse research projects underway, including crash analyses, driver distraction, roadway design, vehicle safety systems, vehicle automation, data modeling and machine vision.

ACKNOWLEDGMENTS

The SHRP 2 Safety Data Student Paper Competition and this e-Circular are the products of the work of the many individuals, committees, and task groups who volunteered their valuable time and insight. Thanks go to the authors for participating in the first SHRP 2 Safety Data Student Paper Competition. Their faculty sponsors, Hao Xu, Gary Davis, and Peter Savolainen, also deserve special acknowledgement. Other participants in the Student Paper Competition should also be recognized: Amirhossein Khezerzadeh, Daniel Rodriguez, Yuan Sun, Xinli Geng, Boris Claros, and Carlos Sun.

The TRB Oversight Committee for Use and Oversight of SHRP 2 Safety Data, Phase 1 (FA014) provided overall leadership of the SHRP 2 Safety Data Student Paper Competition. The

Expert Task Group for User Community Development for Safety Data, Phase 1 (FA016) provided guidance on development and management of the Student Paper Competition. From these two groups was assembled the Review Panel for the SHRP 2 Safety Data Student Paper Competition, who provided valuable reviews and feedback on the abstracts and papers. Special acknowledgment goes to Forrest Council, Charles Fay, A. Stewart Fotheringham, Timothy McDowell, Nicole Oneyear, John Shutko, Bruce Simons-Morton, and Zongwei Tao. Teams at Virginia Tech Transportation Institute and the Institute for Transportation at Iowa State University also provided data request support.

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in Phase 1 of SHRP 2 Safety Data Implementation and Oversight, which is administered by the Transportation Research Board of the National Academies of Science, Engineering, and Medicine.

Contents

Using Naturalistic Driving Study Data to Investigate the Impact of Driver Distraction on Drivers' Reaction Time in Freeway Rear-Ending Events

Jingru Gao
Department of Civil, Environment, and Geo-Engineering
University of Minnesota

Driver distraction represents a potential threat in roadway safety. Although previous research indicates that driver distraction could have negative effects on driving performance, the specific association between driver distraction and crash risk still remains largely unknown. This study aimed to understand the mechanism by which driver distraction could influence crash risk in freeway rear-ending events, and reaction time was chosen as the indicator of crash risk. Statistical analysis was conducted and both linear and causal models were established based on the 108 freeway rear-ending events extracted from the second Strategic Highway Research Program Naturalistic Driving Study database. The analysis suggests that there is an association between driver distraction and reaction time, and distraction duration and the distracted status when a leader braked is positively related to reaction time.

1. INTRODUCTION

Rear-ending crashes are one of the most common crash types in freeway crashes. To understand the mechanism of freeway rear-ending crashes, one needs to study the associated driving behavior thoroughly.

The focus of the analysis was on driver distraction as an important contributing factor in freeway rear-ending events, as it appears in nearly 60% of freeway rear-ending crashes and near-crashes selected from the second Strategic Highway Research Program (SHRP 2) Naturalistic Driving Study (NDS) dataset (*1*).

Driver distraction has become an outstanding problem in roadway safety and receives increasing attention. According to the National Highway Traffic Safety Administration (NHTSA) statistics (*2*), based on data from NHTSA's Fatality Analysis Reporting System (FARS) and National Automotive Sampling System (NASS) General Estimates System, distraction-affected crashes took up about 15 to 20 percent of total crashes every year between 2010 to 2014 in the United States. Knipling (*3*) found that about 25 to 30 percent of the crashes could be attributed to distraction based on data in the NASS Crashworthiness Data System. "The 100-Car Naturalistic Driving Study" (*4*) showed that driver distraction presented in approximately 50 percent of crashes. Among all types of distraction, visual distraction, especially glance behavior, has become the area that researchers focused on most (*4–6*), and the analysis indicated that distraction behavior has a negative effect on driver performance based on both naturalistic and simulation driving data.

This study aims to understand the influence of driver distraction on crash risk in freeway rear-ending events. One most intuitive driving feature that may be directly impacted by driver distraction is "reaction time". Association between driver distraction and reaction time was tested and quantitative models were established based on 108 sample events from the NDS database. This study provides a way to better understand distraction's effect on reaction time as a basis of understanding the mechanism of freeway rear-ending crashes.

2. METHOD

2.1 SHRP 2 Naturalistic Driving Study (NDS) Database

The study analyzed data collected from the SHRP 2 NDS database. The SHRP2 NDS conducted a 3-year data collection from 6 data collection sites. Second-by-second data on what happened in vehicle from 3,542 drivers have been recorded, and 1,600 crashes and 2,900 near crashes are included (*1*). For each event, event detail, driver, and vehicle time-series data are available, providing both qualitative and quantitative description of the event.

2.2 Included Cases

The subject events in this study are freeway rear-ending events, including crash, near-crash, and safety-related incidents. To extract the subject events, a filter based on the event variables in Table 1 was built.

2.3 Data Coding

The response variable of model structures established in this study is reaction time. Both endogenous and exogenous predictors have been taken into consideration. Endogenous predictors

TABLE 1 Event Extraction Filter

Variable *(7)*	Definition *(7)*	Conditional Statement
Event nature	The nature of the other object(s) of conflict the subject vehicle encountered for the event.	(i) "Conflict with a lead vehicle"
Incident type	The type of conflicts the subject vehicle has with other objects.	(i) "Rear-end, striking"
Precipitating event	The state of the environment or action at the beginning of the event.	(i) Other vehicle ahead—at a slower constant speed; or (ii) other vehicle ahead—decelerating.
Pre-incident maneuver	The last driving maneuver that the subject vehicle driver engaged in or was engaged in just prior to or at the time of the *Precipitating Event*.	(i) Going straight, constant speed; (ii) going straight, accelerating; (iii) decelerating in traffic lane; or (iv) maneuvering to avoid a vehicle.
Locality	The surroundings influencing traffic flow at the beginning of *Precipitating Event*.	(i) Interstate, bypass, or divided highway with no traffic signals; or (ii) bypass or divided highway
Event severity	The outcome of the event.	(i) Crash; (ii) near crash; or (iii) crash-relevant
Intersection influence	A judgment whether the subject vehicle's movement is under the influence of an intersection during the event.	(i) Yes, interchange; or (ii) No
Traffic flow	Roadway design presents at the start of the *Precipitating Event*.	(i) Divided or (ii) one-way traffic
Relation to junction	The spatial relation of the subject vehicle to a junction at the time of the start of the *Precipitating Event*.	(i) No junction

include driver-related and distraction-related variables. Exogenous predictors are environment-related predictors.

Variables selected and data coding criterion used are shown in Table 2.

TABLE 2 Information of Variables for Analysis

		Variable	Definition	Type and Value
Response Variable		Reaction Time	Time gap between the time point when leader vehicle's brake light first went on and the time when the follower driver first braked as response to leader driver's brake.	Continuous variable. The unit is seconds.
Predictors	Driver-related	Gender	Driver gender.	Binary variable: "M" for male drivers and "F" for female drivers.
		Age	Driver age. All drivers are divided into four groups based on their age: 16–19, 20–34, 35–54, 55 years old and up.	Categorical variable: "teen," "young," "middle," and "old" are assigned to four age groups, respectively.
	Environment-related	Visual Obstruction	Visual factor that may contribute to the event.	Binary variable: "0" for "no obstruction" and "1" otherwise.
		Weather	Weather condition when the event happened.	Binary variable: "0" for "no adverse weather" and "1" otherwise.
		Lighting	Lighting condition when the event happened.	Binary variable: "0" for "daylight" condition and "1" otherwise.
	Distraction-related	Distraction Presence	A driver is regarded as distracted if the driver "has chosen to engage in a secondary task that is not necessary to perform the primary driving task" [4].	Binary variable: "0" for normal driving (no distraction) and "1" for distracted driving.
		Distraction Scenario	The chronological order of leader's brake light onset and follower's distraction presence. Events are divided into 4 categories: (1) normal driving; (2) follower's distraction ended before leader braked; (3) follower's distraction began after leader braked; and (4) follower driver was distracted when leader braked.	Categorical variable: "S1," "S2," "S3," and "S4" are assigned to 4 defined categories, respectively.
		Distraction Duration	Time length that driver's distraction lasted.	Continuous variable. The unit is seconds.
		Secondary Task Type	Type of secondary task. Secondary tasks are divided into two categories: nonvisual and visual tasks.	Binary variable: "NV" for nonvisual secondary task and "V" for visual secondary task.

2.4 Analysis

2.4.1 Driving Feature Estimation

The NDS forward video and time-series data are used to estimate follower driver's (i) reaction time to leading vehicle braking lights and (ii) distraction duration.

The following are computed:

$r = t_2 - t_1$

$d = t_4 - t_3$

where

r = follower's reaction time,
d = follower's distraction duration,
t_1 = time point when the leader's brake first went on,
t_2 = time point when the follower's brake first went on,
t_3 = time point when the follower's distraction began, and
t_4 = time point when the follower's distraction ended.

Figure 1 is a snapshot of follower's speed profile in a brake-to-stop event from NDS database (*1*).

Figure 2 is the corresponding longitudinal acceleration profile of the event presented in Figure 1.

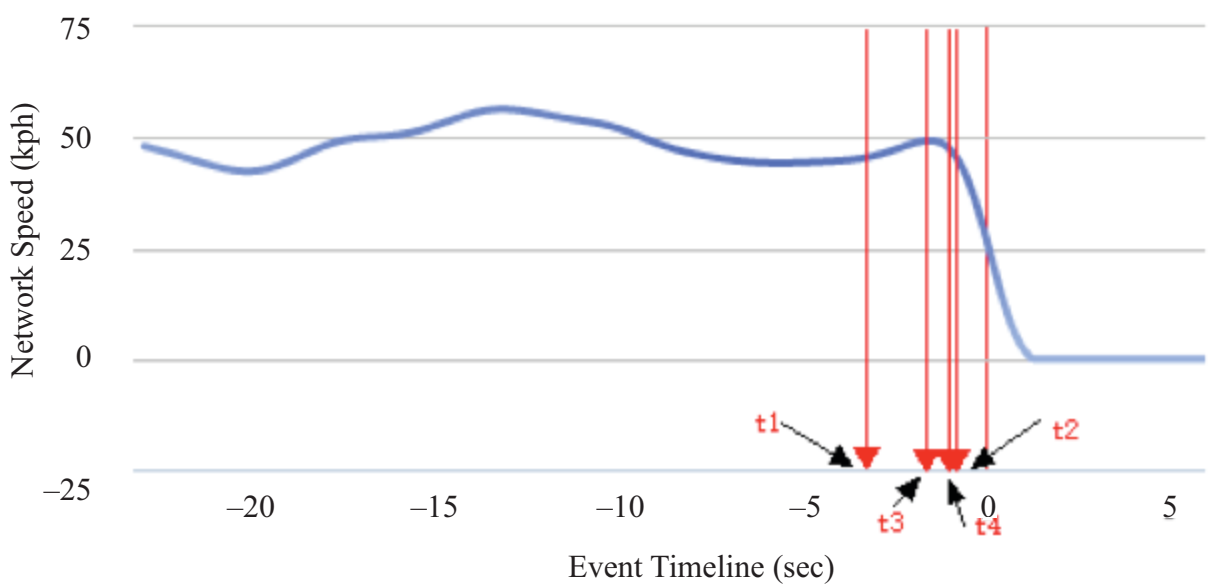

FIGURE 1 Follower driver's speed profile in the example brake-to-stop event. (Source: SHRP 2 InSight Website)

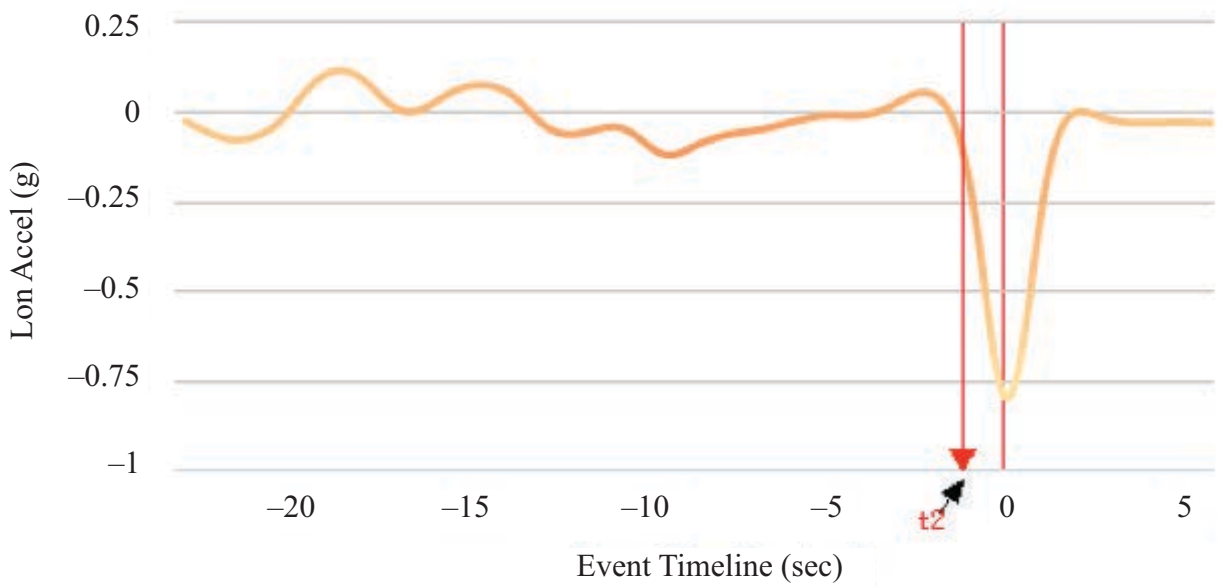

FIGURE 2 Follower driver's longitudinal acceleration profile in the example brake-to-stop event. (Source: SHRP 2 InSight Website)

Driving features mentioned above can help better picture the event. The leader braking point and duration are pretty straightforward from event detail data. Speed, acceleration, and pedal use time-series data are needed to determine follower braking time point.

2.4.2 Model Validation

In this study, linear model is used to test the association between reaction time and selected predictors. Two methods are used to deal with the issue of correlation among predictors that may disguise the significance of associations.

Causal Model. Causal modeling is an effective method to untangle correlated predictors. It can help select the more important predictors among correlated predictors which, if certain conditions are satisfied, can be interpreted as the causal structure of the correlated predictor set, thus containing the most "information" that could explain the variation in the response variable.

In this study, a causal model search will be conducted in Tetrad software (*8*) using both pattern search (*9*) and PAG search (*10*), and PC (*11*) and FCI algorithm (*12*) were chosen, respectively. Different from the PC algorithm, the FCI algorithm assumes the input dataset may have latent variables or sample selection bias, thus all relationships between variables are indeterminate. A structural equation model (SEM) is used to estimate the parameters in causal structures.

Linear Regression. Linear regression is used to test the proposed model structure in Figure 3. The proposed model structure assumes that (i) exogenous and endogenous predictors are independent from each other; (ii) driver-related predictors and distraction-related predictors are independent from each other; and (iii) distracted-related predictors are correlated and *Distraction Duration* is determined by *Distraction Scenario* and *Secondary Task Type*.

To verify the model structure above, the following steps are taken:

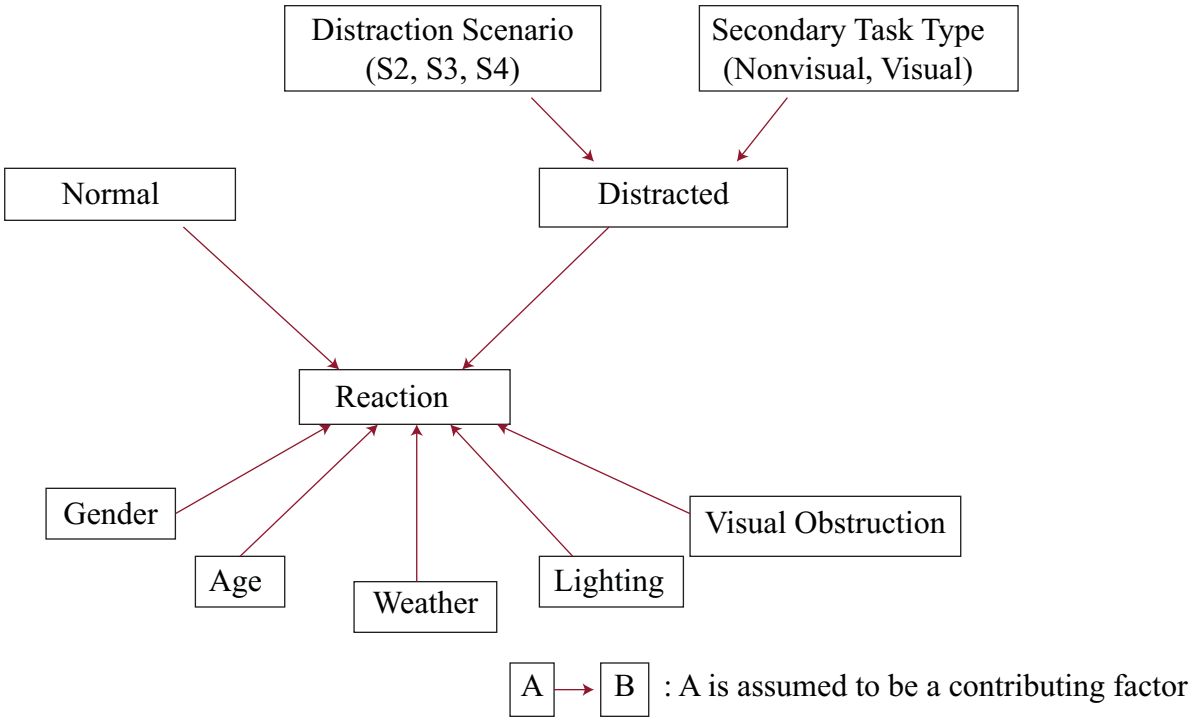

FIGURE 3 Proposed model structure.

i. Test the assumption that there is an association between *Reaction Time* and first-layer predictor$_i$ (*Distraction Duration*, *Gender*, *Age*, *Weather*, and *Lighting*).

This leads to the following linear model, *M*1:

$$\begin{aligned} &\textit{Reaction Time} \\ &\quad = \beta 0 + \beta 1 \times \textit{Distraction Duration} + \beta 2 \times \textit{Gender} + \beta 3 \times \textit{Age} + \beta 4 \\ &\quad\quad \times \textit{Weather} + \beta 5 \text{ x } \textit{Lighting} \end{aligned}$$

For each predictor in this model, the null hypothesis $\beta i = 0$ is tested. The assumption is valid when the null hypothesis is rejected.

ii. Test the assumption that Reaction Time and second-layer predictor *i* (Distraction Scenario and Secondary Task Type) are *d*-separated by Distraction Duration. This test is performed only within Distracted Driving group.

The following linear model, *M*2, is established:

$$\text{Residuals of } M1 = \beta 0 + \beta 1 \times \textit{Distraction Scenario} + \beta 2 \times \textit{Secondary Task Type}$$

For each predictor in this model, *t* null hypothesis $\beta i = 0$ is tested. The assumption is valid when the null hypothesis is not rejected.

iii. Test the assumption that there is an association between Distraction Duration and second-layer predictor$_i$ (*Distraction Scenario* and *Secondary Task Type*)

The following linear model, *M*3, is established:

$$\textit{Distraction Duration} = \beta 0 + \beta 1 \times \textit{Distraction Scenario} + \beta 2 \times \textit{Secondary Task Type}$$

For each predictor in this model, the null hypothesis $\beta i = 0$ is tested. The assumption is valid when the null hypothesis is rejected.

3. RESULTS

Originally, 130 events out of 36,832 records were extracted from the NDS database (*1*) by the filter in Table 1. After that, a data quality check process was conducted and 19 events were screened out because (i) event detail or time-series data are incomplete, (ii) same driver was involved in more than one event, or (iii) driver's first distraction occurred after his or her first reaction to leader vehicle's brake light onset. Events with issue (i) or (iii) were deleted directly. As for events with issue (ii), for each repeated driver, only one event was randomly selected for analysis in order to avoid correlation among events. However, as visual obstruction only presents in 3 of the 111 events left, those 3 events were excluded from the sample because such big difference in the sample size of difference levels of *Visual Obstruction* would result in biased analysis.

The final sample size in this study is 108 events, including 62 female and 46 male drivers. Among all 108 events, 45 events were coded as *Normal Driving* and 63 events were coded as *Distracted Driving*. In *Distracted Driving* group, 37 drivers were nonvisually distracted and 26 drivers were visually distracted.

Some basic statistics of reaction time in *Normal Driving* and *Distracted Driving* groups are shown in Table 3.

The frequency distribution of reaction time in different driving groups (Figure 4) visualized the characteristics of the statistics presented in Table 3.

Judging from the figure, the reaction time in *Distracted Driving* group tends to be longer and more spread out compared to *Normal Driving* group in the sample events studied.

A *t*-test was performed to test the null hypothesis that there is no statistically significant difference in reaction time between *Normal Driving* and *Distracted Driving* groups, and the results are shown in Table 4.

Since *p*-value in Table 4 is less than 0.05, the null hypothesis was rejected. That is, there is significant difference in the mean of reaction time between *Normal Driving* and *Distracted*

TABLE 3 Reaction Time in Different Driving Groups

	N	M	Mdn	SD
Normal Driving	45	1.669	1.290	1.254
Distracted Driving	63	2.192	1.956	1.317
N = number of events, M = mean, Mdn = median, SD = standard deviation				

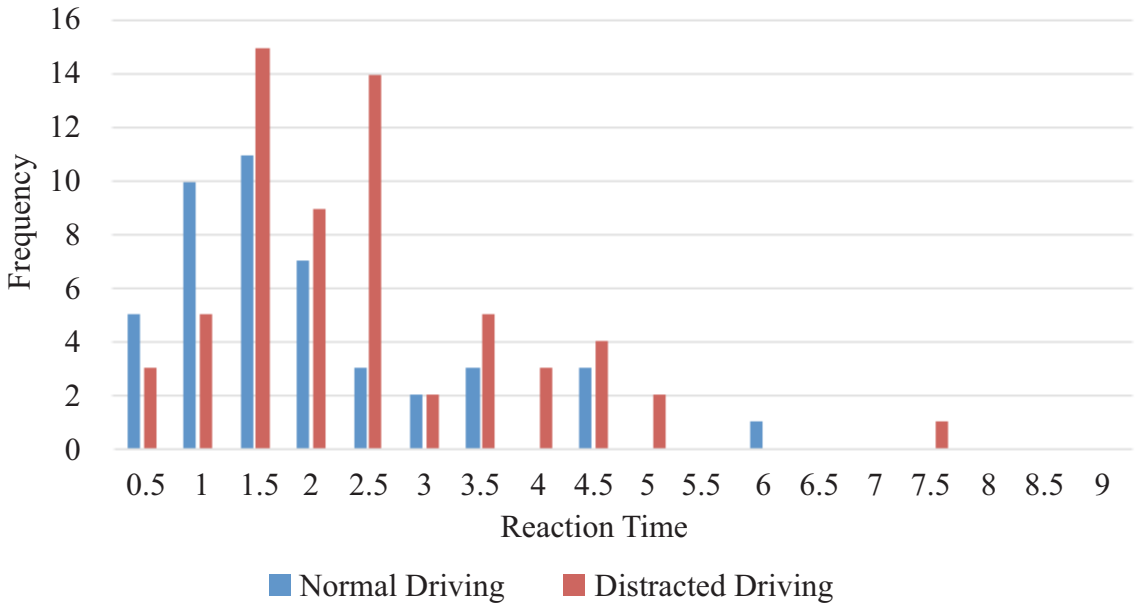

FIGURE 4 Frequency distribution of reaction time in normal and in distracted driving groups.

Driving groups at 0.05 level, indicating that driver distraction may have contributed to reaction time in the sample events studied.

3.1 Causal Modeling

Causal modeling was carried out using Tetrad software (*8*). Unfortunately, as Tetrad can conduct causal model search only for datasets consisting either entirely of continuous variables or entirely of discrete variable (*13*), data discretization was required for continuous variables, *Reaction Time* and *Distraction Duration*. Since the sample size is so limited, an appropriate way to discretize those two continuous variables above was not found; therefore, the attempt to conduct a causal model search in Tetrad was not pursued.

3.2 Linear Regression

All linear regression work was done using R software (*14*). The regression results of model *M1* are shown in Table 5.

TABLE 4 *t*-Test Result of Reaction Time in Normal and Distracted Driving Groups

Welch Two-Sample *t*-Test		
t = -2.0904	df = 97.627	p-value = 0.039
95 percent confidence interval: (–1.01860580, –0.02646722)		

TABLE 5 *M*1 Linear Regression Results

lm(formula = ReactionTime ~ factor(Gender) + factor(Age) + DistractionDuration + Weather + Lighting, data = data)

Residuals:

Min	1Q	Median	3Q	Max
–2.3480	–0.8583	–0.2692	0.5050	5.3122

Coefficients:

	Estimate	Std. Error	*t*-Value	Pr(>|*t*|)
(Intercept)	1.44920	0.39537	3.665	0.000397 ***
factor(Gender)M	0.14696	0.26190	0.561	0.575948
factor(Age)Old	0.40695	0.48725	0.835	0.405597
factor(Age)Teen	0.04204	0.43847	0.096	0.923806
factor(Age)Young	0.09372	0.38637	0.243	0.808840
DistractionDuration	0.12890	0.04315	2.987	0.003540 **
Weather	–0.21928	0.41441	–0.529	0.597879
Lighting	0.09328	0.31231	0.299	0.765815

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.294 on 100 degrees of freedom
Multiple *R*-squared: 0.09005; adjusted *R*-squared: 0.02636
F-statistic: 1.414 on 7 and 100 DF, *p*-value: 0.2081

The regression results in Table 5 indicate that, in the sample events studied, the usual reaction time of *Normal Driving* drivers is about 1.45 seconds, and *Distraction Duration* is the only factor that has association with *Reaction Time*. The longer the distraction lasted, the longer the reaction time would be.

Table 6 shows the regression results of model *M2*: As the regression results in Table 6 suggest, within *Distracted Driving* group, neither *Distraction Scenario* nor *Secondary Task Type* has significant association with the residuals of *M1*, indicating that the residuals of *M1* are independent of *Distraction Scenario* and *Secondary Task Type*. Thus, it can be concluded that *Reaction Time* and second-layer predictors, namely *Distraction Scenario* nor *Secondary Task Type*, are d-separated by *Distraction Duration*, thus neither *Distraction Scenario* nor *Secondary Task Type* has direct effect on *Reaction Time*.

The regression result of *M3* is shown in Table 7. According to the regression results in Table 7, only *Distraction Scenario* has a significant association with *Distraction Duration* in the sample events studied. Combined with the regression results of *M1* and *M2*, the conclusion can be drawn that, *Distraction Scenario* does not have a direct effect on *Reaction Time* but affects it through *Distraction Duration*. Drivers who were distracted when the leader braked tended to have longer distraction duration, thus resulting in longer reaction time. *Secondary Task Type* does not show a significant effect on *Reaction Time*.

The final model structure is presented in Figure 5. Solid arrows indicate those associations that turned out to be significant, and dashed arrows indicate those that were eventually removed from the model.

TABLE 6 *M*2 Linear Regression Results

lm(formula = Residuals ~ factor(DistractionScenario) + factor(SecondaryTaskType), data = datad)

Residuals:

Min	1Q	Median	3Q	Max
–1.3575	–0.6502	–0.1905	0.4490	3.8944

Coefficients:

	Estimate	Std. Error	*t*-Value	Pr(>|*t*|)
(Intercept)	0.38730	0.34557	1.121	0.267
factor(DistractionScenario)S3	0.09531	0.49935	0.191	0.849
factor(DistractionScenario)S4	–0.56486	0.36537	–1.546	0.127
factor(SecondaryTaskType)Visual	0.01316	0.26576	0.050	0.961

Residual standard error: 1.002 on 59 degrees of freedom
Multiple *R*-squared: 0.07338; adjusted *R*-squared: 0.02626
F-statistic: 1.557 on 3 and 59 DF, *p*-value: 0.2093

TABLE 7 *M*3 Linear Regression Results

lm(formula = DistractionDuration ~ factor(DistractionScenario) + factor(SecondaryTaskType), data = datad)

Residuals:

Min	1Q	Median	3Q	Max
–4.9840	–0.7934	0.2074	1.1762	3.6440

Coefficients:

	Estimate	Std. Error	*t*-Value	Pr(>|*t*|)
(Intercept)	2.0217	0.6589	3.068	0.00325 **
factor(DistractionScenario)S3	–0.4735	0.9522	–0.497	0.62084
factor(DistractionScenario)S4	4.0403	0.6967	5.799	2.78E–07 ***
factor(SecondaryTaskType)Visual	–0.4201	0.5068	0.829	0.41043

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.911 on 59 degrees of freedom
Multiple *R*-squared: 0.5239; adjusted *R*-squared: 0.4997
F-statistic: 21.64 on 3 and 59 DF, *p*-value: 1.415E–09

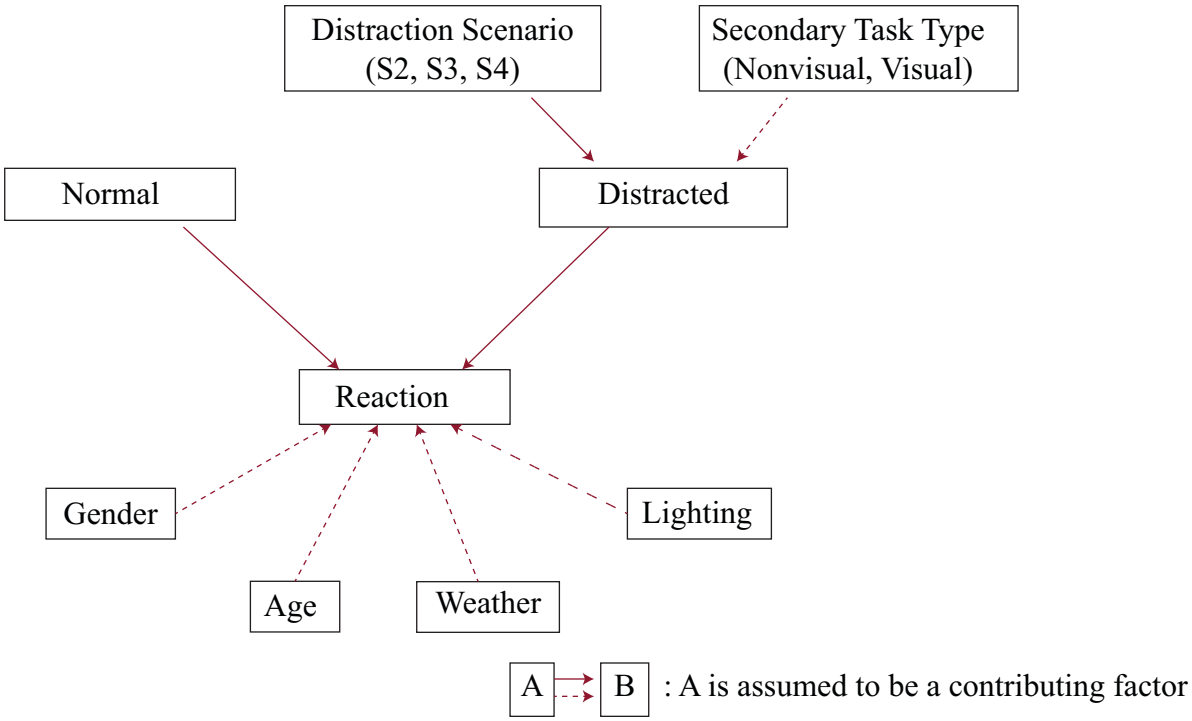

FIGURE 5 Final model structure validated.

4. DISCUSSION OF RESULTS

In this study, both qualitative and quantitative metrics of distraction depicting its presence, length, and type have been considered. According to the findings, among all factors selected only *Distraction Duration* and *Distraction Scenario* showed significant effects on reaction time. However, those two factors contain only a small part of the information that could explain the variation in reaction time. There must be other factors affecting reaction time. For example, some drivers are described by the video reductionists as "lost in thought" or "checking mirrors for lane-changing behavior" when the leader braked, but neither of them can be considered as a secondary task, that is, distracted.

The greatest two limitations to be overcome for further study are the access to radar data and the sample size. The radar data are important because they can provide information on situation kinematics. There are extensive studies showing that the relationship between reaction time and following headway determines whether the situation is critical, rather than the reaction time itself (*3*, *15*, *16*). To study how driver distraction can influence crash risk, one needs to explore the relationship among driver distraction, reaction time, and situation kinematics first.

In addition, if radar data are available, then the analysis of one fairly interesting case, where the follower driver was distracted after first reacting to the leader vehicle's brake light onset, can be conducted. This driving scenario is common because drivers may attempt to increase following distance to compensate for reduced attention to the road when performing a secondary task (*17*). In this case, the speed profile is always a multiphase procedure. Trajectory-based estimation (*18*) can be used to find the speed change points of both the leader and follower in every

braking phase, thus the follower's reaction time in each phase can be imputed. Study of these cases can help better understand the association among driver distraction, situation kinematics, and reaction time.

Besides reaction time, there are other driving features through which driver distraction affects crash risk that are worth studying. For example, another critical driving feature associated with rear-ending crashes occurs when a following driver brakes at rates substantially higher than the leading driver (*18*). Limited empirical studies in the past have shown some evidence suggesting drivers with longer reaction time tend to brake harder than the leader. Such a proposition can be verified if radar data are available.

5. CONCLUSION

This study aims to demonstrate the impact of driver distraction on reaction time and subsequently, its effect on the outcome of rear-ending events. Historically, obtaining information specific to driver behavior was challenging. However, with the availability of NDS data, one can quantify the association between driver distraction and reaction time. The subject driving feature, reaction time, can be estimated based on NDS time-series data, and association assumptions between driver distraction and reaction time can be tested by statistical techniques.

The test results showed that driver distraction duration is the primary direct cause of reaction time, with other factors having indirect effects mediated by distraction. Longer distraction duration and the distracted status when a leader braked tended to result in longer reaction times. However, although certain clues were found in this study, due to the limitation in sample size and lack of situation kinematics data, the exact mechanism of how distraction affects reaction time, still requires future efforts. This study has provided the methodology that can be adopted to study the association between driver behavior and driving features.

REFERENCES

1. Transportation Research Board of the National Academies. *The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset*. 2013. https://insight.shrp2nds.us.
2. National Center for Statistics and Analysis. Distracted driving 2014 (Traffic Safety Facts Research Note. Report No. DOT HS 812 260). Washington, D.C.: National Highway Traffic Safety Administration, April 2016.
3. Knipling, R. R. IVHS Technologies Applied to Collision Avoidance: Perspectives on Six Target Crash Types and Countermeasures. *Proceedings of the Third Annual Meeting of IVHS America*. IVHS America. pp. 249–259, 1993.
4. Klauer, S. G., T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. T. Ramsey. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. Report DOT-HS-810-594. NHTSA, U.S. Department of Transportation, 2006.
5. Klauer, S. G., F. Guo, B. G. Simons-Morton, M. C. Ouimet, S. E. Lee, and T. A. Dingus. Distracted Driving and Risk of Road Crashes Among Novice and Experienced Drivers. *New England Journal of Medicine*, 370(1), 2014, pp. 54–59.
6. Hickman, J. S., and Hanowski, R. J. An Assessment of Commercial Motor Vehicle Driver Distraction Using Naturalistic Driving Data. *Traffic Injury Prevention*, 13, 2012, pp. 566–574

7. SHRP2 Researcher Dictionary for Video Reduction Data, Version 3.4. Virginia Tech Transportation Institute Blacksburg, Virginia, 2015.
8. Tetrad (Version 5.2.1-3). Carnegie Mellon University, 2015.
9. Verma, T. S., and J. Pearl. Equivalence and Synthesis of Causal Models. *Proc., Sixth Conference on Uncertainty in Artificial Intelligence*, Cambridge, Mass., AUAI Press, Corvallis, 1991, pp. 220–227.
10. Richardson, T. A Discovery Algorithm for Directed Cyclic Graphs. *Proc., Twelfth International Conference on Uncertainty in Artificial Intelligence*, Portland, Ore. Morgan Kaufmann, San Francisco, Calif., 1996, pp. 454–461.
11. Spirtes, P., and C. Glymour, An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review*, Vol. 9, No. 1, 1991, pp. 62–72.
12. Spirtes P. An Anytime Algorithm for Causal Inference [C]//AISTATS. 2001.
13. *Tetrad Manual*. Tetrad, Pittsburgh, Pa., 2016. http://www.phil.cmu.edu/tetrad/new_manual.pdf. Accessed July 6, 2016.
14. R Core Team. R: A Language and Environment for Statistical Computing (Version 3.3.0). Vienna: R Foundation for Statistical Computing, 2016.
15. Brill, E. A. A Car-Following Model Relating Reaction Times and Temporal Headways to Accident Frequency. *Transportation Science,* Vol. 6, No 4, 1972, pp. 343–353.
16. Victor, T., J. Bärgman, M. Dozza, and H. Rootzén. Safer Glances, Driver Inattention, and Crash Risk: An Investigation Using the SHRP 2 Naturalistic Driving Study. *Proc., 3rd Conference of Driver Distraction and Inattention*, Gothenburg, Sweden, 2013.
17. *Understanding the Distracted Brain: Why Driving While Using Hands-Free Cell Phones Is Risky Behavior*. Report. National Safety Council, Itasca, Ill., 2010.
18. Chatterjee, I., and G. Davis. Use of Naturalistic Driving Data to Characterize Driver Behavior in Freeway Shockwaves. *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2434*, 2014, pp. 9–17.

Jingru Gao, Department of Civil, Environment, and Geo-Engineering, University of Minnesota, 500 Pillsbury Dr SE, Minneapolis, MN 55455; e-mail: gaoxx692@umn.edu.

Examination of Factors Determining At-Fault and Not-at-Fault Status in Multivehicle Conflicts Using the SHRP 2 Data

Raha Hamzeie
Peter T. Savolainen
Department of Civil, Construction, and Environmental Engineering
Iowa State University

Multivehicle crashes resulted in more than 14,000 fatalities in 2013, which accounted for more than 40 percent of all traffic fatalities across the United States. In a majority of these crashes one or more of the crash-involved drivers were identified as being an at-fault driver, meaning that they performed a leading error that contributed to the crash occurrence. Research has generally shown that the characteristics and behaviors of such drivers are different from those of not-at-fault drivers. The purpose of this study was to identify the factors associated with at-fault and not-at-fault status in a multivehicle conflict. The Naturalistic Driving Study data that have been developed as part of the second Strategic Highway Research Program were utilized to further investigate these factors. Multivehicle conflicts, including both crash and near-crash events, were investigated in association with driver characteristics and behaviors that were captured through different surveys. Mixed-effect logistic regression models were developed to examine and compare a multitude of data elements including driver demographic information, risk perception, driving behaviors, and sleep habits. Consequently, the factors associated with fault status were identified using the detailed data. The findings revealed that full-time workers, drivers who perceive tailgating as being high risk, and those who were not involved in any crashes during the three years prior to the study were less likely to be at fault. On the other hand, the likelihood of being at fault was found to be noticeably higher among drivers who were less risk averse, including those who believed that accelerating at the onset of yellow is low risk, as well as for those drivers who feel fatigued nearly every day.

Multivehicle crashes accounted for more than 14,000 fatalities in 2013, which represents over 40 percent of all traffic fatalities in the United States (*1*). In the majority of such collisions, one or both of the crash-involved drivers are generally found to have been at fault to some degree for causing the collision. For example, the National Motor Vehicle Crash Causation Survey suggests that human error is the critical reason for 93 percent of crashes (*2*). Prior research has shown a variety of differences between the circumstances, characteristics, and behaviors of the at-fault and not-at-fault drivers (*3*). Several research studies have tried to examine the effect of various factors on the odds of fault and concluded that male drivers and nonvehicle owners are more likely to be at fault in crashes (*4*). The license status of the driver was also found to play a significant role in crash-involvement as an at-fault driver. Research has shown that the probability of being at-fault in a fatal crash is 160 percent greater among those with suspended or revoked licenses. This increases to 173 percent for unlicensed drivers (*3*). Associations have been found not only between crash involvement, traffic conditions, and roadway geometry, but also between driver demographic and socioeconomic characteristics and crash involvement. At-fault drivers have been shown to be less prevalent among older populations, those working

from home, and those whose daily commute is less than 15 minutes (*5*). Median family income was also shown to be correlated with fault status (*5*). In addition, differences have also been observed with respect to the injury outcomes associated with the at-fault and not-at-fault parties (*6*). Consequently, further research into those driver, vehicle, and environmental factors that are associated with fault status could provide critical insights to better understand the contributory or causal factors associated with multivehicle collisions.

The Naturalistic Driving Study (NDS) data collected as a part of the second Strategic Highway Research Program (SHRP 2) provide a comprehensive database with unparalleled detail regarding event information (e.g., type of the incident, precipitating event, driving behavior, etc.), as well as detailed participant information and trip data.

The fidelity of these data has the potential to overcome some of the typical methodological limitations involved in assessing fault status. For example, prior research has demonstrated interrelationships between fault status and several potentially endogenous predictors, such as impaired driving or the nonuse of safety restraints (*7*). Such endogeneity has the potential to lead to biased estimates and erroneous conclusions as to the effects of specific factors. This endogeneity is largely reflective of the crash reporting process, wherein officers are tasked with determining which involved drivers are at fault.

The focus of the study presented herein is to assess those factors associated with at-fault and not-at-fault status among drivers involved in multivehicle conflicts, which include both crash and near-crash events over the duration of the NDS (*8*). Identification of such factors may help to inform subsequent policies and programs aimed at reducing traffic crashes, injuries, and fatalities. Ultimately, the findings of this study result in important insights to improve understanding of those factors that contribute to multivehicle conflicts. This includes the identification of a variety of high-risk behaviors that lead to a driver being at fault for a specific conflict. By identifying such factors, countermeasures, policies, and programs may be developed that target specific high-risk driver behaviors. For example, the study results could provide important insights that could be implemented as a part of driver training, public education/outreach, and similar programs. The identification of these factors would also have implications in broader research, highlighting some of the important driver-specific factors that should be considered in subsequent research utilizing the SHRP2 data. Failure to account for such factors would otherwise potentially lead to biased estimates as to the effects of other factors such as roadway geometry and traffic characteristics.

DATA

The data used for the purpose of this study were drawn from the publicly available portion of the data collected as part of SHRP 2. This subset of the NDS data includes aggregated summary data that excludes any personally identifying information. The information as to event details, as well as trip and vehicle attributes was available through various data sets developed as a part of the NDS. General driver-specific information, including demographic characteristics, was provided for each study participant. In addition, information about behaviors and perceptions was captured through several participant surveys. These included questionnaires regarding driver behavior, driver demographic characteristics, driving history, risk perception, and risk-taking, as well as surveys about driving knowledge and sleeping habits.

The purpose of this study was to examine the various parameters available from these data sources to identify those that are associated with a driver being found at fault in a multivehicle conflict. The integration of all available data sources using the unique identifiers resulted in a total of 1,360 conflicts. The event subset of the data included a fault variable, which indicated whether the subject driver committed a leading error that contributed to the conflict. In this case the leading error is the last event in the conflict causal chain. This variable was coded using the video data. Subsequently, a binary variable was introduced that took a value of one if the subject driver was at fault and zero otherwise. Table 1 provides a cross-tabulation detailing the severity of the conflicts (e.g., crash, near-crash, etc.) and fault statuses.

For each conflict, a series of data elements was available that explained different attributes of the involved drivers and other pertinent information. The 1,360 total conflicts identified through the data integration process corresponded to a total of 684 unique drivers. Table 2 provides the summary statistics for some salient variables for the entire sample of conflicts (meaning some drivers are included more than once), as well as for each unique driver. In other words, some drivers were involved in multiple conflicts in the entire sample. In these cases, all driver-specific characteristics are identical, while the event-specific attributes are variable. Running the summary statistics for the entire sample is based on the assumption that each observation is different from the other. On the other hand, unique drivers are indicative of the sample in which each driver is included once. In each case, these variables are binary indicators (i.e., yes/no variables). The mean value is the percentage of drivers for whom the variable took a value one. For example, this table indicates that when considering the entire sample nearly 48 percent of the involved drivers were female. However, when duplicate observations for drivers with multiple conflicts were removed, this percentage jumps to approximately 51 percent, which implies that there are male drivers who are involved in multiple conflicts.

As noted previously, the main purpose of this study was to identify those driver behaviors that are associated to a significant level with the likelihood of being involved in a conflict as an at-fault driver. Different questionnaires were available to assess drivers' behavior regarding risk perception and risk taking. However, due to the correlation between these responses, the authors opted to utilize the risk perception questionnaires in that it better reflects the driver behavior and action selection. In the risk perception questionnaire, participants were asked to report the associated risk of each of the behaviors on a scale of one to

TABLE 1 Sample Crosstab Between Conflict Type and Fault Status of the Involved Drivers

Conflict Type	Assigned At-Fault Status			
	Subject Driver	Other Driver(s)	Unknown	Total
Crash	67 (47%)	75 (52%)	1 (1%)	143
Crash relevant	3 (50%)	2 (33%)	1 (17%)	6
Near-crash	689 (58%)	484 (41%)	15 (1%)	1188
Nonsubject conflict	6 (26%)	15 (65%)	2 (9%)	23
Total	765 (56%)	576 (42%)	19 (1%)	1360

TABLE 2 Summary Statistics of Driver Demographic and Driver History Data

	Total Sample		Unique Drivers	
Variable	**Mean**	**Std. Deviation**	**Mean**	**Std. Deviation**
At-fault subject driver (Yes/No)	0.563	0.496	—	—
Full-time worker (Yes/No)	0.460	0.499	0.452	0.498
Unemployed (Yes/No)	0.083	0.276	0.083	0.277
No children at home (Yes/No)	0.739	0.439	0.734	0.442
Driver feels fatigued nearly every day (Yes/No)	0.163	0.369	0.155	0.362
Female (Yes/No)	0.485	0.500	0.507	0.500
Latino/Hispanic (Yes/No)	0.088	0.284	0.075	0.263
Education beyond high school (Yes/No)	0.913	0.283	0.898	0.303
Two-parent household (Yes/No)	0.590	0.492	0.624	0.485
Driver rental status (Yes/No)	0.743	0.437	0.749	0.434
Income 50,000+ (Yes/No)	0.629	0.483	0.635	0.482
No traffic violations in past 3 years (Yes/No)	0.576	0.494	0.55	0.498
No crashes in past 3 years (Yes/No)	0.641	0.48	0.675	0.468

seven. The responses are ordered from one to seven, reflecting the degree of associated risk with the listed behavior from the participants' standpoint. Figure 1 presents the percentage of the respondents within each category for each of the listed questions.

Consequently, driver behaviors and characteristics, as determined through various driver behavior questionnaires, were examined. These questionnaires provide insight as to how the general driving patterns, risk-taking behaviors, risk perception, and so forth. vary among at-fault and not-at-fault drivers. Given the dichotomous nature of the dependent variable, discrete outcome models were estimated to identify those factors unique to the at-fault versus not-at-fault drivers.

Prior research using the NDS data has shown that specific drivers tend to be at consistently higher (or lower) risk for involvement in conflicts such as crash and near-crash events (*9*). That is, there are a number of drivers who have several crash/near-crash events and others who have no such events. There are also likely a number of drivers who have a mixture of these event outcomes. Therefore, from an analytical standpoint, random effects logit models would provide a means by which to account for this potential correlation for drivers who were involved in multiple crash or near-crash events, whether they were at fault or not.

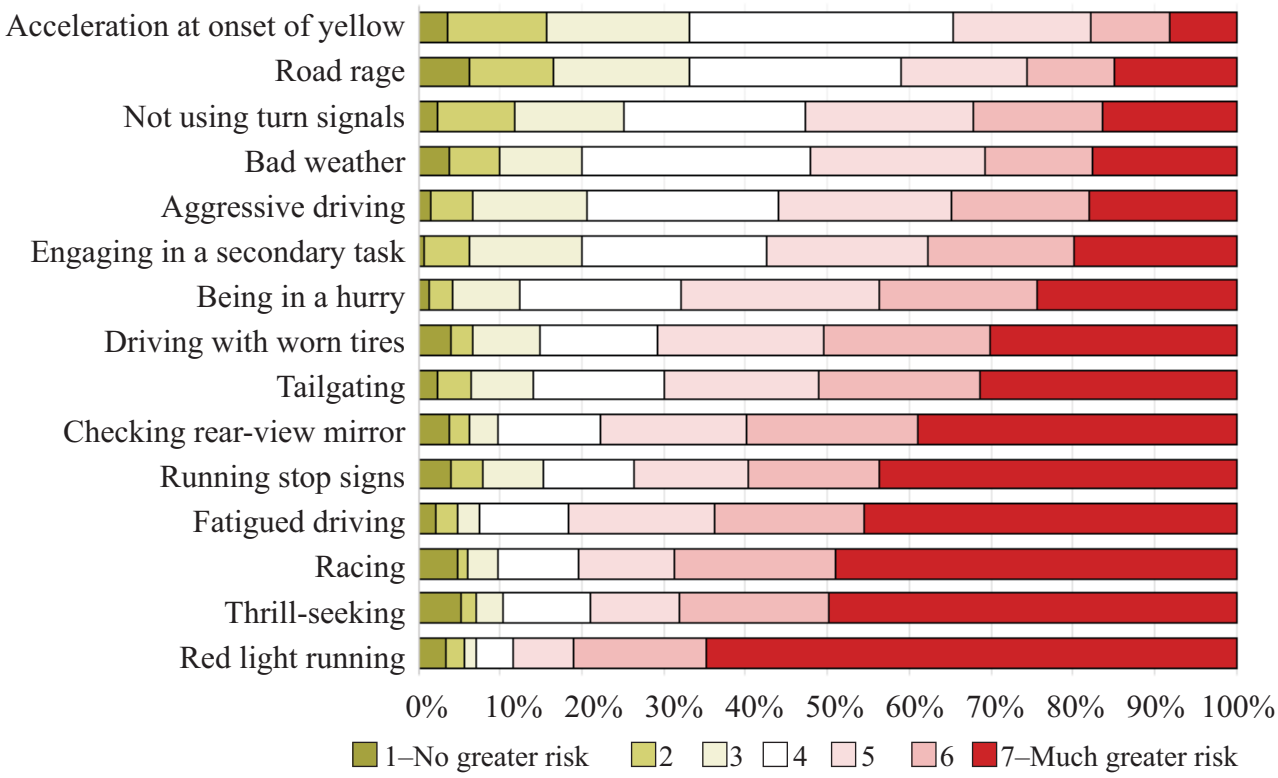

FIGURE 1 Percentage of respondents by the level of risk perception.

METHODOLOGY

The purpose of this study was to identify driver characteristics and other factors associated with each trip that distinguish between at-fault drivers and not-at-fault drivers. To this end, logistic regression models were estimated to examine different factors associated with each event. Logistic regression presents an appropriate modeling framework since the dependent variable is dichotomous in nature (fault status of the subject driver takes a value of one if the participant was at fault and zero otherwise). Under the logistic regression framework, the odds of a driver being at fault is related to a linear function of predictor variables as shown in Equation 1:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} \tag{1}$$

where p_i is the probability of participant i being at fault in a multivehicle conflict, the β_i terms represent a vector of estimable parameters, the x_i terms indicate a vector of measurable characteristics (e.g., age, sex, responses to the survey questions, etc.), and ε_i is an error term that is assumed to follow the logistic distribution.

One concern that arises within the context of this study is the anticipated correlation between different events or conflicts associated one single driver. For example, certain drivers may intrinsically be more or less likely to be at fault during conflicts as compared to other similar drivers. The logistic regression model assumes that the error terms (ε_i) are independently

and identically distributed, which is potentially problematic as there is expected to be potential correlation among the same drivers as noted above. This assumption can be relaxed by adding a driver specific parameter vector that varies randomly across different drivers and maintains same values for multiple events for a single driver. This vector allows the constant term to vary across different individuals, permitting the model to capture heterogeneity that is due to other unobserved factors. Under this setting, the probability of crash or near-crash involvement is then

$$p_i = \int_x \frac{\exp\left(\beta x_i + \varepsilon_i\right)}{1+\exp\left(\beta x_i + \varepsilon_i\right)} f\left(\beta|\varphi\right) d\beta \tag{2}$$

where $\beta|\varphi$ is the density function of β with φ referring to a vector of parameters of the density function, and all other terms as previously defined. This model structure is commonly referred to as a random effects (or random intercept) logistic regression model.

RESULTS AND DISCUSSION

A series of binary logistic regression models were developed to examine the effect of different attributes on the likelihood of being involved in a multivehicle conflict as an at-fault driver. Given the limited size of this exploratory sample of NDS data, variables were retained if they were statistically significant at a 10 percent significant level. Initially, several risk perception factors were found to be statistically significant. However, strong correlation was observed between the responses to various risk perception questions for individual drivers. This suggests that drivers fall into two general categories: (1) risk-averse drivers, who perceive various types of driving behaviors to be aggressive and high-risk; and (2) risk-prone drivers, who exhibited fairly consistent responses that suggested various illegal driving behaviors were of lower risk for crash involvement.

Table 3 presents the results of the pooled (i.e., naïve) and random effect logistic regression models that were found to provide the best fit given the available data elements. When examining

TABLE 3 Binary Logistic Regression Model for the Fault Status of the Involved Driver

	Fixed-Effect Model			Random Effect Model		
Parameter	**Coeff.**	**Std. Error**	***p*-Value**	**Coeff.**	**Std. Error**	***p*-Value**
Intercept	0.585	0.117	.000	0.632	0.419	0.074
Full-time worker	-0.257	0.111	.021	–0.325	0.12	0.017
Driver perceives tailgating as high risk	-0.490	0.119	.000	–0.517	0.128	0.000
Driver perceives acceleration at onset of yellow as low risk	0.773	0.342	.024	0.843	0.366	0.032
Driver feels fatigued nearly everyday	0.269	0.154	.080	0.307	0.168	0.900
No crashes in past 3 years	–0.198	0.117	.090	—	—	—
Log likelihood at convergence	–912.305			–909.708		
Restricted log likelihood	–932.027					
McFadden *R*-squared	0.021					

the model results, a positive coefficient is reflective of a variable that increases the likelihood of being the at-fault driver in a multivehicle conflict, whereas a negative coefficient is indicative of a variable that was more strongly associated with the not-at-fault drivers.

Drivers who had a full-time job were found to be less likely to be the at-fault driver in a multivehicle conflict as compared to the complementary group (i.e., those with any type of part-time job or those who are unemployed). This variable was found to be correlated with a number of other factors, such as income and level of education. Consequently, though these variables are not included in the final model, they do provide some explanation for the fact that full-time employees were less likely to be at fault. This is somewhat similar to the findings of previous studies, which had found the fault status to be related to the family median income. The coefficient for the full-time work status is –0.257 when using the pooled logistic regression model, while this jumps up to –0.325 in the random effects framework. This reveals that when controlling for the correlation among the same individuals, the effect is more pronounced.

Back to the survey results that were the primary variables of interest, drivers' perception regarding two of the behaviors were found to be significantly associated with the fault status. Those drivers who believe that tailgating is associated with much greater risk are significantly less likely to be at fault. Tailgating is associated with smaller headways between vehicles, resulting in limited time span for drivers to assess the situation, identify the most proper action, and react ultimately. This usually results in a rear-end strike, in which the driver of the following vehicle is at fault. In addition, driver decision while facing a yellow at intersections was found to have a significant association with fault status of drivers. Drivers who perceive accelerating at the onset of yellow as low risk are associated with a significant increase in the likelihood of involvement in multivehicle conflicts as the at-fault driver. Such a decision may lead to running the red light, which could potentially result in occurrence of angle crashes or near-crashes. In this case the fault is assigned to the driver who ran the red light. In addition to these types of collisions/conflicts, accelerating at the onset of yellow might result in rear-end crashes or need prompt reaction of drivers in case the drivers' decision varies between the following and leading vehicles. Although these variables and their associated coefficients can be interpreted individually, these findings demonstrate that these behaviors are reflective of the actual behavior of drivers and can help identify risk-averse drivers versus risk-prone individuals.

In addition to investigation of the risk perception questionnaire, a series of factors regarding sleeping habits were examined to identify any parameter with a meaningful relationship with fault status. An extensive investigation of the available data elements revealed that those drivers who feel fatigued nearly every day are more likely to be at fault in multivehicle conflicts. This is consistent with prior research, which had found drowsiness and sleepiness to be significantly associated with crash risk (*10*).

The information regarding driving history of the participants was also available in the analyzed dataset. Various parameters, such as violation history, crash history, and license age were investigated. Initially, several such factors were found to be associated with fault status. However, these factors were no longer significant when other similar variables were included in the model. A comparison of goodness-of-fit measures across various model formulations showed that the pre-study crash history of the involved drivers was the strongest factor among such variables. The results indicate those drivers who did not experience any crashes during last three years prior to the NDS were less likely to be identified as an at-fault driver. This measure may be interpreted as an indicator of driving proficiency or the degree of caution the subject driver

exhibits in driving. However, it is noteworthy that this parameter was not statistically significant in the random effects model. This suggests that certain unobserved driver characteristics may be correlated with prestudy crash history. It may also be reflective of the limited number of drivers who had no crash on their records and were not identified as being the at-fault driver.

CONCLUSION

Ultimately, this study identified several factors that were associated with the fault status of drivers involved in multivehicle conflicts. Pooled and random effects logistic regression models were estimated, with the latter allowing for consideration of correlation in the propensity of individual drivers being at fault across multiple conflict events. The results are generally consistent between two sets of models. However, prestudy crash history was not significant in the random effects model, which provides evidence that specific drivers may be more (or less) likely to be at fault in general due to certain unobserved factors that are common to these drivers.

The magnitude of the associations between various parameters and fault status can be discerned by examining the odds ratios. In this setting, the odds ratios indicate the probability of a driver being at fault divided by the probability of a driver not being at fault. Table 4 provides a summary of the odds ratios for each parameter included in the final model. Odds ratios greater than 1.0 are indicative of a variable associated with a higher risk of being at fault, whereas the opposite is true for odds ratios less than 1.0.

These results show that full-time workers are 28 percent less likely to be at fault in a conflict event. Drivers who perceive tailgating to be a high-risk behavior were even less likely to be at fault, as shown by 40 percent lower odds of being at fault. Additionally, driver behavior at the onset of yellow was found to have a very pronounced association with fault status. The odds ratio values reveal that drivers who believe that accelerating at the onset of yellow is low risk are twice as likely to be at fault as compared to those who perceive this action as high risk. Also, regarding sleep habits and condition, the findings show that the probability of being at fault is approximately 35 percent higher for drivers who reported to feel fatigued almost every day. Lastly, the pooled model showed an 18 percent reduction in the likelihood of being at-fault among drivers with no prior crashes on their record during the preceding three-year period. However, this effect was not significant in the random effects model.

TABLE 4 Odds Ratio for the Coefficients Estimated in the Fixed- or Mixed-Effect Binary Logistic Regression Models

Parameter	Fixed-Effect Model	Random Effect Model
Intercept	N/A	N/A
Full-time worker	0.77	0.72
Driver perceives tailgating as high risk	0.61	0.60
Driver perceives acceleration at onset of yellow as low risk	2.17	2.32
Driver feels fatigued nearly everyday	1.31	1.36
No crashes in past 3 years	0.82	—

The nature of this study design allowed for results that can be broadly generalized. In more than 97 percent of the cases the fault status was assigned to one of the involved drivers, which is close to that of prior research that has shown drivers to be the leading critical factor in multivehicle crashes (*2*). This study is unique due to the high resolution of the available data, which provided the opportunity to assess drivers' behavior in multivehicle conflicts considering the naturalistic driving data in association with drivers' behavior and characteristics. However, there are potential limitations that need to be addressed. First, several important data elements were not available. While most behavioral information were available for study participants, similar information was not available for other involved drivers who were not participants in the NDS. For instance, there is no information available as to how other involved drivers perceive the examined risk-taking behaviors or what demographic characteristics they had. In addition, exposure data have not been examined as part of this study that may have some associations with fault status. There were also a number of conflicts where either fault status or other data elements were missing. As shown in Table 1, approximately in one percent of the cases the fault was not assigned to any of the involved drivers. Also, there were some drivers for whom the risk perception questionnaire results or parts of demographic information, or both, were not available. Consequently, these events were discarded from the analysis dataset.

The findings of this study provide insights as to those factors that have associations with multivehicle conflicts and more importantly to those that are more prevalent among at-fault drivers. As mentioned previously, these results can generally help insurance companies to set and develop pricing strategies. In addition, these results can help to establish countermeasures such as traffic enforcement cameras, as well as to develop and improve programs like graduated drivers licensing and public education/outreach as a whole.

REFERENCES

1. National Highway Traffic Safety Administration, 2010. Fatality Analysis Reporting System: Query FARS Data, Available from http://www-Fars.nhtsa.gov/QueryTool/QuerySection/SelectYear.aspx. Accessed August 16, 2016.
2. *National Motor Vehicle Crash Causation Survey, Report to Congress*. National Highway Traffic Safety Administration, 2008, Department of Transportation HS 811 059.
3. Brar, S. S., 2014. Estimating the Over-Involvement of Suspended, Revoked, and Unlicensed Drivers as At-Fault Drivers in California Fatal Crashes. *Journal of Safety Research*, Vol. 50, pp. 53–58.
4. Harootunian, K., B. H. Y. Lee, and L. Aultman-Hall. 2014. Odds of Fault and Factors for Out-of-State Drivers in Crashes in Four States of the USA.
5. Lee, J., M. Abdel-Aty, and K. Choi. 2014. Analysis of Residence Characteristics of At-Fault Drivers in Traffic Crashes. *Safety Science*, Vol 68, pp. 6–13.
6. Russo, B. J., P. T. Savolainen, W. H. Schneider, and P. C. Anastasopoulos, 2014. Comparison of Factors Affecting Injury Severity in Angle Collisions by Fault Status Using a Random Parameters Bivariate Ordered Probit Model. *Analytic Methods in Accident Research*, Vol. 2, pp. 21–29.
7. Schneider, W. H., P. T. Savolainen, D. Van Boxel, and R. Beverley, 2012. Examination of Factors Determining Fault in Two-Vehicle Motorcycle Crashes. *Accident Analysis and Prevention*, Vol. 45, pp. 669–676.
8. Second Strategic Highway Research Program Naturalistic Driving Study Database. Transportation Research Board of the National Academies, 2013.

9. Savolainen, P. T., O. Smadi, R. Hamzeie, B. Russo, S. Knickerbocker, T. J. Gates, and J. J. Kay, 2015. *The Interrelationships Between Speed Limits, Geometry, and Driver Behavior.* Report to Michigan DOT from the Second Strategic Highway Research Program (SHRP 2) Implementation Assistance Program (IAP).
10. National Highway Traffic Safety Administration, 2006. *The Impact of Driver Inattention on Near-Crash/Crash Risk,* DOT HS 810 594.

Raha Hamzeie and Peter T. Savolainen, Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA 50011; e-mail: rhamzeie@iastate.edu; and pts@iastate.edu.

Effect and Influence of Different Factors on Driver Behavior When Vehicles Make Right Turns at Signalized Intersections

Jianqing Wu
Department of Civil and Environmental Engineering
University of Nevada, Reno

Roadway crashes have been a major concern for public safety. One of the main issues in traffic safety is the conflict between turning vehicles and other roadway users at intersections. Turning driver behavior has not been studied in depth due to limited data. In this paper, driver behavior was provided based on the second Strategic Highway Research Program Naturalistic Driving Study. A total of 211 right-turn crash and near-crash turning events at signalized intersections were used to analyze driver behavior. Each turning movement was divided into different stages by analyzing the sensor data and by reviewing the forward videos. Driver behavior was extracted from crash detailed information provided by the database. Prioritized influencing factors on crash severity were identified by using the random forest method. Driver behavior and speed were the top two significant influencing factors on crash severity. An ordered probit model was applied to the risk factors on turning crash severity at intersections. The results showed distraction behavior, high speed, and deceleration were associated with the most serious right-turn crashes at intersections. The relationship between driver behavior and traffic flow, roadway geometric design, and traffic signs were also analyzed in this paper. The result showed that in a "right turn on red" situation, 80% of drivers involved in the right-turn events did not fully stop before turning right. Driving operation varied with different geometric designs of intersections. Drivers were found to have the lowest frequency of improper operation when conflicting traffic flow was between 1,000 vph and 2,000 vph. As face videos and baseline data were not available for this research, more related data is still needed to evaluate the findings.

1. INTRODUCTION

Accidents at signalized intersections continue to represent a major highway safety problem in the United States, even though intersections comprise a small portion of the total highway system. In 2002, approximately 3.2 million intersection-related crashes occurred, representing 50 percent of all reported crashes (*1*). Conflicts between right-turn vehicles and other roadway users are a major issue for traffic safety at signalized intersections. About 20% of crashes that occurred at intersections involved right-turning vehicles (*2*). Right-turn vehicles continuously have to negotiate with the traffic environment, including vehicles from other directions, cyclists, and pedestrians, as there are no protected phases in traffic signals for right turning vehicles at most intersections in the United States. Right-turn drivers are expected to yield to pedestrians, cyclists, and oncoming vehicles when there are conflicts; however drivers may fail to see or react to other roadway users when they are influenced by various factors such as the surrounding traffic,

vehicle condition, or driver conditions. Right-turn-on-red (RTOR) has been widely used throughout the United States since 1980. RTOR allows motorists to turn right on a red signal after they have stopped and searched for pedestrians, cyclists and oncoming vehicles. The benefit of RTOR is that it could reduce driving delay at intersections and thus save fuel, but RTOR also increases the probability of conflicts between right-turn vehicles and other roadway users. If drivers turn right under right-turn-on-green (RTOG), they may only have to yield to pedestrians, cyclists and left-turn vehicles from the opposite direction. At RTOR, drivers may have to yield to pedestrians or cyclists, left-turn vehicles from the opposite direction, and oncoming vehicles from the left direction. Drivers have to handle more complex situations under RTOR compared to RTOG, so the safety benefit–cost tradeoff of RTOR at signalized intersections needs to be evaluated.

One of the key elements in the design and operation of signalized intersections is driver behavior. Driver behavior was identified as a sole or contributing factor in 94% of crashes in the United States (*3*). A previous study (*4*) showed about 96% of intersection-related crashes involved critical contributing factors attributed to drivers. Driver behavior has been an important contributing factor to intersection-related crashes. Appropriate driver operation and observation can help to reduce conflicts between right-turn vehicles and other road users. However, limited studies on the behavior of right-turn drivers have been performed because of the difficulty of collecting driver behavior data. Many countermeasures, such as speed limit signs and yield to pedestrians signs, have been applied to reduce conflicts between right-turn vehicles and other roadway users at signalized intersections; yet it is still not clear how the different countermeasures and strategies impact driver behavior, especially when crashes occur. Understanding the detailed influence of the different factors on driver behavior will help the development and application of traffic safety countermeasures at intersections.

The research purpose of this paper is to identify factors that play a role in the behavior of right-turn drivers at signalized intersections when crashes occur. The paper is organized as follows: first a brief literature review to summarize current driver behavior analysis, followed by explanation of the data source and driver behavior analysis, and finally the major findings and limitations of this study.

2. LITERATURE REVIEW

Previous studies (*5–8*) have investigated the influence of geometric design, speed limits and signs on driver operation behavior on freeways, but very limited research has focused on driver behavior at signalized intersections. Wortman and Matthias (*9*) studied driving operation at six signalized intersections in Arizona by using time-lapse photography. The results showed the mean deceleration rate was 11.6 feet per second per second when drivers approached intersections. The mean deceleration rate was not significantly different in daytime and nighttime. Sato and Akamatsu (*10*) employed four experimental vehicles with various sensors to measure driver behavior before a right turn at a specific intersection. The influence of traffic conditions on driver operation behavior were analyzed from data collected in eight weeks. The results suggest that drivers approached the target intersection in a car-following condition, and that the positions of the front and rear vehicles and the vehicle velocity influence the onset location and timing of releasing the accelerator pedal and covering the brake pedal. Harbluk et al. (*11*) studied the impact of cognitive distraction on drivers' behavior in an on-road experiment. The results of this study

indicate that even when in-vehicle devices are hands-free, significant changes in driver behavior may result due to the cognitive distraction associated with their use. Summala (*12*) first tested an explanation that drivers turning right simply focus their attention on the cars coming from the left and fail to see cyclists from the right early enough. This research used two cameras set up near an unsignalized T-intersection to collect right-turning-driver head movement during nonpeak hours. These current studies used two main methods to collect turning-driver behavior: driver behavior questionnaires and on-road experiments. It is not guaranteed that drivers answered the questionnaires accurately. On-road experiments could record drivers' real behavior accurately, but drivers may have different behaviors if they use experimental cars rather than their own vehicles. Naturalistic driving, also known as naturalistic observations, is a new approach among already-applied driver behavior research methods. Under this naturalistic observation approach, the behavior of road users is observed unobtrusively in a natural setting for a long period of time. The second Strategic Highway Research Program (SHRP 2) conducted the largest and most comprehensive Naturalistic Driving Study (NDS) ever undertaken. The study recruited more than 3,400 volunteer drivers, ages 16–80, at sites in six states: Florida, Indiana, North Carolina, New York, Pennsylvania, and Washington. All their trips are recorded for up to two years. Data include vehicle speed, acceleration, and braking; vehicle controls; lane position; forward radar; and video views to the front and rear of the vehicle and on the driver's face and hands. Several studies (*13–15*) used NDS to analyze the relationship between driver behavior and crash risk. The influence of different factors on driver behavior at signalized intersections has yet to be studied. This paper focused on studying the association of different factors with right-turn driver behavior contributing to crashes, which would be used to help select countermeasures with a better benefit–cost ratio or develop new, more effective countermeasures to change driver behavior in order to reduce crash severity.

3. DATA SOURCE AND DATA PROCESSING

3.1 Data Source

The SHRP 2 Safety Data were used in this research for analyzing drivers' behavior. SHRP 2 Safety Data, including the NDS data and the Roadway Information Database (RID), provided a continuous description of drivers, vehicles roadways and environments. The NDS data include time-series records from the sensors installed on the volunteer vehicles and multidirectional video clips. The RID contains comprehensive roadway and environmental data related to the NDS road network (*16*).

In total, 211 right-turn events at signalized intersections were received to support this research. The received data of each crash event included time-series recordings (sensor data records), forward videos, and crash detail tables. The time-series recordings included speed, acceleration, brake pedal position and so on. The speed was collected every 0.1 seconds, which can provide good samples for driver operation analysis. The corresponding forward videos showed weather conditions, traffic signs, and traffic signal information. The length of each forward video was no longer than 3 minutes before crash occurrence. The crash detail tables recorded driver behavior, roadway characteristics, collision type, and crash severity information. As pedestrians involved in right-turn events were very limited in the database, the sample size was not enough to

analyze only pedestrian-vehicle crashes. For this paper, each right-turn event included at least one right-turn vehicle. The crashes involved right-turn vehicles and pedestrians or right turn vehicles and other vehicles.

The SHRP 2 RID is a spatially enabled geodatabase to store, query, and manipulate geographic data of points, lines, and polygons. The roadway information was in the Geographic Information System (GIS) data format also located by the linear referencing system (LRS). The ESRI GIS software package ArcGIS was selected to process the SHRP 2 RID data. The LRS allowed geographic processing with ArcGIS linear reference functions. Only the RID data of Washington State were processed and used because of the comprehensive crash data (2006–2013) available from the state. The crash data were first selected by location, in the 300-feet-radius range of the intersection nodes, in order to identify the intersection crashes. The results were further queried by the number of injuries (>0) and subject vehicle actions (turning right).

3.2 Data Processing

3.2.1 Most Frequent Conflict Type

The most frequent conflict types were identified by analyzing historical crash data in the RID. Six different conflict types were identified to study right turn crashes: a. both right turn, b. right turn–left turn, c. right turn–pedestrian, d. right turn–straight, e. fixed object, and f. overturned. Right turn, right turn–pedestrian, right turn–straight, and right turn–left turn are the typical conflict types. Figure 1 demonstrates these conflict types. The historical right-turn crash data of Washington State in the RID database showed the two most frequent conflict types were between two right-turn vehicles and a right-turn vehicle with pedestrians. The percentage of conflicts between right-turn and left-turn vehicles was the lowest. In order to better understand the conflicts between two right-turn vehicles, two detailed types were divided: right turn–both moving and right turn–one moving, one stopped. The result of the statistics showed 83.4% of the conflicts between two right-turn vehicles were right turn–one moving, one stopped. The reason why right-turn vehicles stopped is most likely that right-turn drivers yielded to pedestrians at the crossing line. If right-turn drivers failed to observe the pedestrians before the stop bar, they might stop

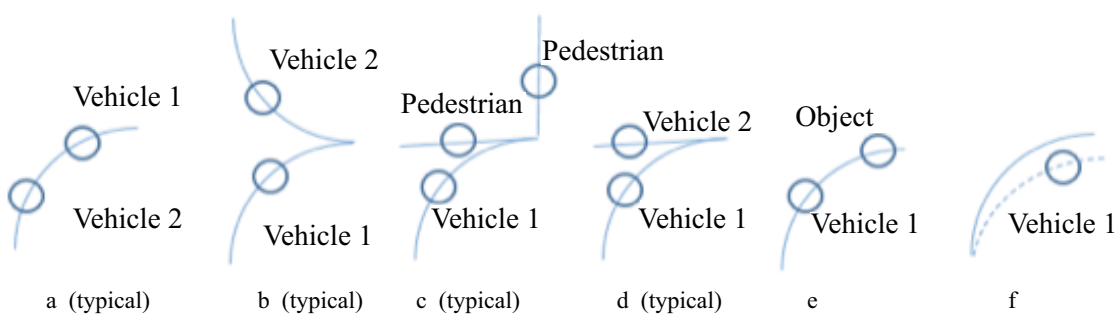

FIGURE 1 Conflict types.

to yield to pedestrians if they saw the pedestrians after they passed the stop bar. The following vehicles may not have enough reaction time to avoid rear-end crashes. This explains why there was a high percentage of conflicts between two right-turn vehicles and right-turn vehicle with pedestrians.

3.2.2 The Influence of Different Factors on Crash Severity

The association between different factors and crash severity was studied in order to identify the role of driver behavior on crashes.

The crash severity is categorical and ordered. In order to deal with an ordered categorical variable, the use of discrete response models, such as ordered logit ordered probit models are more appropriate. Many studies have implemented ordered probit models to analyze crash severity in the context of traffic safety. Quddus et al. (*17*) used ordered probit models to analyze motorcycle injury and vehicle damage severity. Abdel-Aty (*18*) performed ordered probit models to analyze the influence level of different factors on crash severity at multiple locations. Wang and Kockelman (*19*) used an ordered logit model to distinguish the effects of vehicle weight and type on occupant injury severity. Kockelman and Kweon (*20*) and Rifaat and Chin (*21*) used ordered probit models to examine the risk of different injury levels sustained under all crash types, two-vehicle crashes, and single-vehicle crashes. Duncan et al. (*22*) and Lemp (*23*) analyzed truck crash severity using ordered probit models. Ordered probit models, which assume standard normal distribution for F, are the most commonly used models to analyze crash severity. The ordered probit model has the following Equations 1, 2, and 3:

$$P_n(1) = \phi(\alpha_1 - \beta_j X_n) \tag{1}$$

$$P_n(j) = \phi(\alpha_1 - \beta_j X_n) - \phi(\alpha_{j-1} - \beta_j X_n) \quad j \neq 2, \ldots, j-1 \tag{2}$$

$$P_n(J) = 1 - \sum_{j=1}^{J-1} P_n(j) \tag{3}$$

where ϕ is the cumulative standard normal distribution function (*18*). For all the probabilities to be positive, $\alpha_1 < \alpha_2 < \alpha_{j-1}$ (*18*). The same probability equations can be written for ordered logit models. The model is estimated using maximum likelihood. In this study, the ordered probit model was performed by using "MASS" Package in R language.

In order to rank the importance of the variables, the random forest method based on decision trees was used. The random forest method is an ensemble learning method for regression (*24*). Random forests are a way of averaging multiple deep decision trees, trained on different parts of the same training set, with the goal of reducing the variance (*25*). Doing this comes at the expense of a small increase in the bias and some loss of interpretability, but generally greatly boosts the performance of the final model. Breiman (*26*) proposed to evaluate the importance of a variable X_m for predicting Y by adding up the weighted impurity decreases $p(t)\Delta i(s_t, t)$ for all nodes t where X_m is used, averaged over all N_T trees in the forest:

$$Imp\left(X_m\right)=\frac{1}{N_T}\sum_t\sum_{t\in T:\upsilon\left(S_t\right)=X_m}p\left(t\right)\Delta i\left(S_t,t\right) \tag{4}$$

where $p(t)$ is the proportion N_T/N of samples reaching t and $\upsilon(s_t)$ is the variable used in split s_t. When using the Gini index as impurity function, this measure is known as the Gini importance or mean decrease Gini. R language is programmed to identify the prioritized influencing factors of crash severity. In R language, for each tree, the prediction accuracy on the out-of-bag portion of the data is recorded. Then, the same is done after permuting each predictor variable. The difference between the two accuracies are then averaged over all trees and normalized by the standard error (*27*). Figure 2 shows the mean decrease Gini of different factors generated by R. In Figure 2, the speed and acceleration are the average speed and acceleration from 30s before crash occurrence to 5s before crash occurrence. The results showed driver behavior has the most significant influence on crash severity while construction zone has the least influence. The top four factors on crash severity were kept in the analysis model.

In the selected factors, there were some non-numerical variables. R treats all independent variables in the analysis as numerical. Numerical variables are interval or ratio scale variables whose values are comparable. However non-numerical variables will not be processed directly by

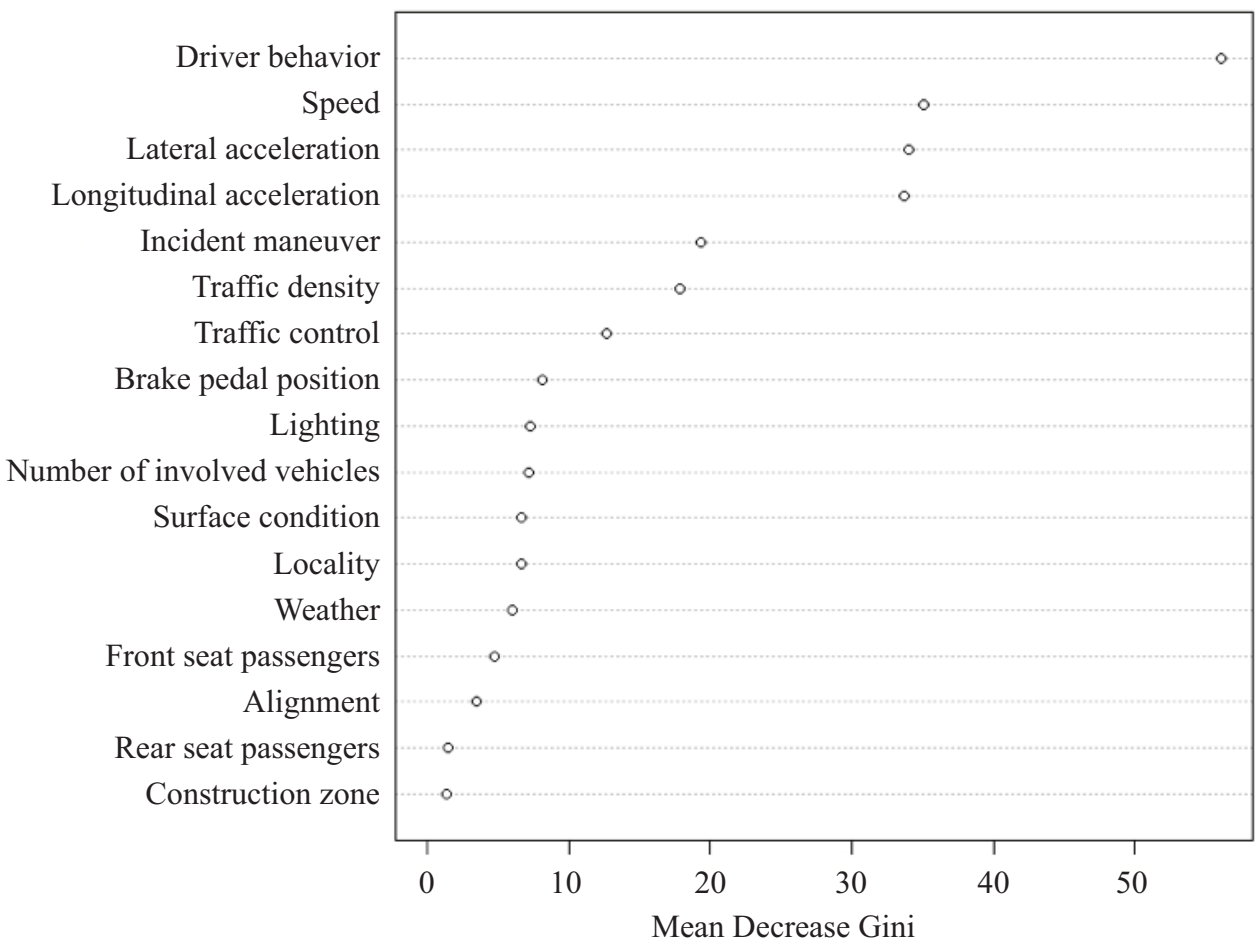

FIGURE 2 Mean decrease Gini of different factors in relation to crash severity.

the model. Dummy variables were created in this situation to "trick" the regression algorithm into correctly analyzing attribute variables in R. Table 1 shows explanatory variables used in the model.

3.2.3 Conflict Spots with Pedestrians

Drivers may focus more on finding a gap to merge into traffic and inadvertently ignore pedestrians. The key conflict spots with pedestrians were identified by reviewing forward videos and time-series data. Generally, there are two conflict spots (Spot 1 and Spot 2) between right turn vehicles and pedestrians, shown in Figure 3. In the whole right-turn movement, drivers pass two pedestrian crossing routes (AB route and CD route). If the traffic signal status is red, drivers may have conflicts with pedestrians using AB route. If the traffic signal status is green, drivers may have conflicts with pedestrians using CD route. In the received data, about 40% of vehicle conflicts involved pedestrians. Of these vehicles, 46% had conflicts with pedestrians using the CD route. The results also show only 50% of drivers yield to pedestrians when right-turning vehicles conflict with pedestrians.

The 211 trips were divided into different types to identify the key conflict spots. The position when drivers begin to slow down is defined as the beginning point of any right-turn movement. The location and speed of the beginning points are diverse; no two vehicles approached to intersection in the same exact manner. 400 feet before reaching conflict Spot 1 was selected as the beginning point, since most vehicles begin to decelerate at this position. The position when drivers

TABLE 1 Variables Used in the Model

Explanatory Variables	Description of Variables	Influencing Rank	Dummy Variables
Numerical Variables			
Speed	Average speed of 60 s to 30 s before crashes	2	
Lateral.Acceleration	Average lateral acceleration of 60 s to 30 s before crashes	3	
Longitudinal.Acceleration	Average longitudinal acceleration of 60 s to 30 s before crashes	4	
Categorical or dummy variables			
Driver.Behavior	None	1	0
	Signal violation		1
	Sign violation		2
	Distracted		3
Severity	Not a crash		
	Low-risk tire strike		
	Minor crash		
	Police-reportable crash		
	Most severe		

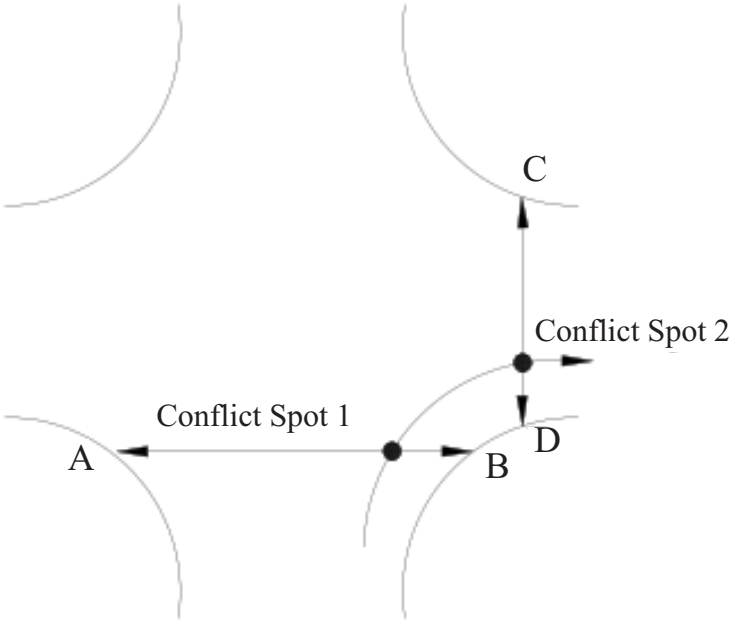

FIGURE 3 Conflict between right-turn vehicles and other road users.

pass conflict Spot 2 is defined as the ending point of right-turn movements. In the received data, the beginning of the video is often less than 400ft from conflict Spot 1. For these particular cases, the beginning of the video was selected as the beginning points. The right-turning movement was divided into different stages in order to better study driver observation behavior at varying locations. Each trip was divided into different stages by analyzing the speed and reviewing the front videos. The criteria for dividing a turn into different stages is shown in Table 2.

TABLE 2 Criteria for Dividing Stages of Right Turns

Stage ID	Description	Criterion	Note
1	Approaching the queue	Begin: Speed begins to drop	If there is no queue, the end point of Stage 1 is the moment when the vehicle arrives at the stop line.
		End: Vehicle reaches the end of queue	
2	Waiting in a queue	Vehicle stops in the queue	If there is no queue in front of the vehicle, Stage 2 and Stage 3 could be excluded.
3	Queue disappeared and vehicles moved to stop bar	Begin: Speed begins to increase	
		End: Vehicle stops at the stop bar	
4	Vehicle waiting for a gap to turn right	Vehicle stops at the stop bar	Vehicle may stop slightly before or after the stop line without other vehicles in front of it.
5	Vehicle turned right at intersection	Begin: Speed begins to increase	
		End: Vehicle finished turning at intersection	

When a driver is turning right they display different types of behaviors throughout the movement. Dividing drivers' behaviors into defensive or aggressive does not adequately represent the observations of the drivers. More detail is needed to effectively classify drivers' behaviors. In the received data, drivers were described as distracted, making improper turns and violating signals. No other detailed information was provided. So in this paper, improper operation was considered as the result of not enough drivers' observation (left side, right side and rear). High frequency of left-side observations and right-side observations are considered to be helpful for drivers to better observe pedestrians, and high frequency of distraction observations and improper operations are considered to be a higher risk for pedestrians.

4. DRIVER BEHAVIOR ANALYSIS

4.1 Crash Severity Analysis

The estimated results of the influence of the factors on crash severity are reported in Table 3. The output of R did not show all coefficients of dummy variables. R took dummy variables having a value of zero as reference variables and did not cover them in the output file. The influence of other variables was calculated based on reference variables. If the other dummy variable has a positive estimate value, it means the crash severity increases compared with the reference variable. If the other dummy factor has a negative estimate value, it shows the crash severity decreases compared with the reference variable. The output of R also shows the standard error and significance level (*P*-value) of each variable. For driver behavior, the results show that signal or sign violation and distraction result in increasing severity compared with no adverse driver behavior. Distracted behavior (with an estimate of 1.2887) has the most significant effect on crash severity. For driver operation, high speed and high deceleration increases the crash severity.

TABLE 3 Parameter Estimates of the Model

	Variables	Estimate	Std. Error	*z*-Value	Pr(>|*z*|)	Significance Code
(Intercept)		–2.1054	0.59683	–3.528	0.000419	***
(Driver.Behavior)1	Signal violation	0.33553	0.3793	0.885	0.376373	
(Driver.Behavior)2	Sign violation	0.90359	0.33738	2.678	0.0074	**
(Driver.Behavior)3	Distracted	1.2887	0.39818	3.236	0.00121	**
Speed		0.0134	0.00492	2.723	0.006475	**
Lateral.Acceleration		–1.42084	1.1574	–1.228	0.21959	
Longitudinal.Acceleration		–0.01752	0.57051	–0.031	0.975497	
Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

4.2 Right Turn on Red

RTOR has been common in the United States. It was reasoned that RTOR would reduce driver delay at intersections and thus save fuel; however, RTOR may cause some impact to traffic safety. Signal status (red, yellow, or green) is random when right-turn vehicles reach the stop bar. If the signal is green, right-turn vehicles do not have to yield to other vehicles; however, they still need to yield to pedestrians in the CD route (Figure 3). In some situations, drivers may need to observe vehicles coming from opposite left-turn direction. If the signal is red, right-turn vehicles need to yield to pedestrians in the AB route (Figure 3) and other vehicles coming from the left-through direction and the opposite left-turn direction. Drivers have to observe more road users under RTOR compared to RTOG.

Some vehicles had to stop before the stop bar if there was a queue to turn right at the intersection, but the location when vehicles stopped in a queue was different. This explained why the speed of right-turn vehicles fluctuated near the intersections, as seen in Figure 4. Whether RTOR or RTOG, the trends of speed first decreased then increased.

The definition of conflict between right-turn vehicles and pedestrians in front videos is that vehicles begin to turn right while pedestrians are crossing or preparing to cross. In conflicting trips, 56.2% of the 211 conflicts happened on RTOR. When the signal was red, drivers had to yield to conflicting vehicles coming from the left-side direction. Conflicting vehicles constituted a potential danger to right turning drivers. Appropriate right-turning gap acceptance is crucial for traffic safety at intersections (*12*). The result of a SHRP 2 Implementation Assistance Program project documented in another paper (*28*) submitted for the 2017 TRB Annual Meeting showed drivers often experienced anxiety when faced with this situation, and there were often impli-

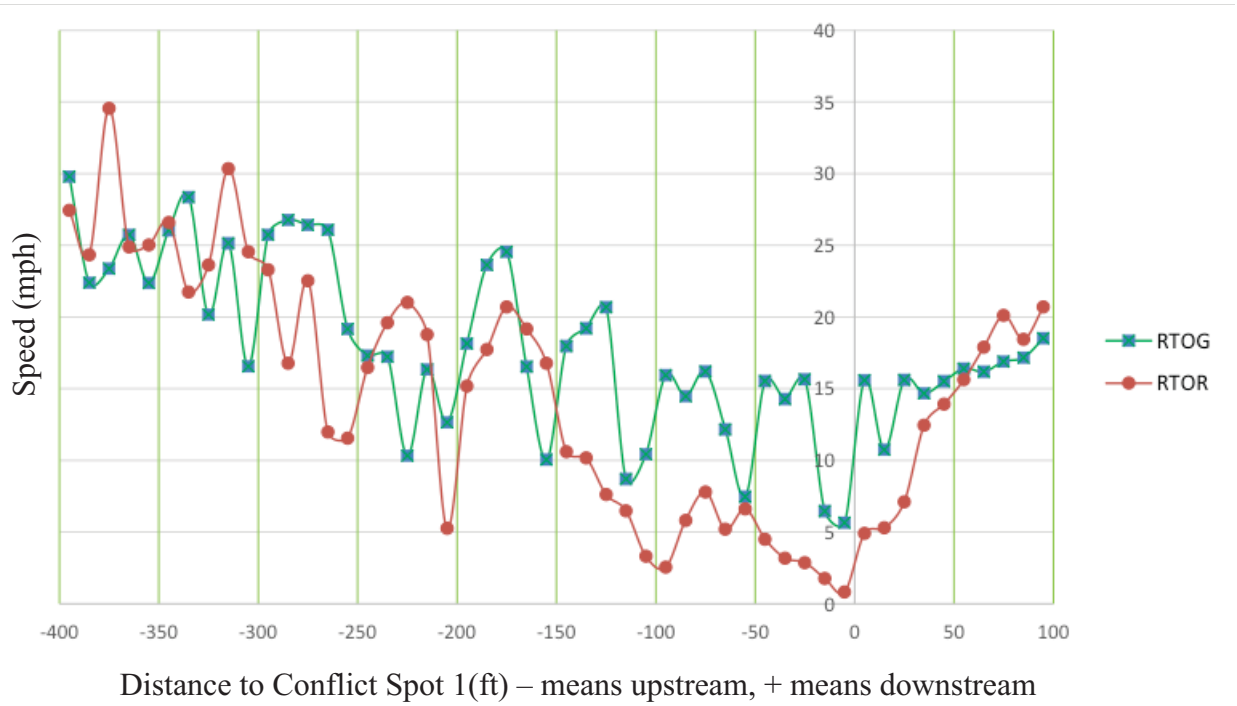

FIGURE 4 Driving speed of right-turn vehicles under different traffic signals.

cations when they forgot to check pedestrians on their right side. When the signal was green, drivers only needed to check if there were vehicles turning left from the opposite direction. In addition, the volume of conflicting vehicles was often low, therefore drivers would have more time to check for pedestrians on their right side.

RTOR laws specify that drivers must stop and yield to approaching vehicles and to all pedestrians within the intersections before turning right when the traffic signal is red. However, the speed in the time-series records showed in 80% of the crash/near-crash events the drivers did not fully stop when they reached the stop bar at RTOR; they may have slowed down but did not fully stop. Figure 4 shows that the average speed at conflict Spot 1 was lower under RTOR compared with the speed under RTOG. Also, drivers had high acceleration under RTOR when they left the intersections. Higher acceleration was helpful to merge into the volume but was also dangerous to pedestrians. Drivers might not have enough time to change their feet from the gas pedal to the brake pedal in response to the emergency. This could offer insight into why right-turn crashes at signalized intersections increased after the adoption of RTOR. The majority of these crashes occurred when the pedestrians or bicyclists were coming from the drivers' right side and remained undetected because drivers were looking to the left for a gap in traffic.

4.3 Channelization

Right-turn lanes provide space for the deceleration and storage of turning vehicles. There are many different types of right-turn lanes used in the United States. The most common types of right-turn lanes are the following:

1. Right-turn lane with lane line pavement marking,
2. Channelized right-turn lane with acceleration lane on cross road, and
3. Channelized right-turn lane without acceleration lane on cross road.

When there is channelization with dedicated downstream, vehicle speeds decrease slightly near the stop bar, then increase. As there is a dedicated downstream lane, drivers do not need to look for merging vehicles after yielding to pedestrians. They have the opportunity to accelerate in their own lane before merging with cross-street traffic. So channelization with dedicated downstream can be used to eliminate the influence of long queues if the volume of right turn vehicles is high at intersections.

Figure 5 shows the relationship between speed and the different channelization types. When there is a channelization without dedicated downstream, the speed decreased more than channelization with dedicated downstream. With this geometric design, drivers turned directly into a through-traffic lane on the cross street. Therefore, they needed to check for a gap in the conflicting traffic on the cross street before turning. When there is no channelization, the trend of average speed is similar to channelization without dedicated downstream.

This section only considered the role of different types of right-turn lanes on driver operation. The more detailed geometric information such as right-turn lane length, radius, width, and island types are not studied in this paper.

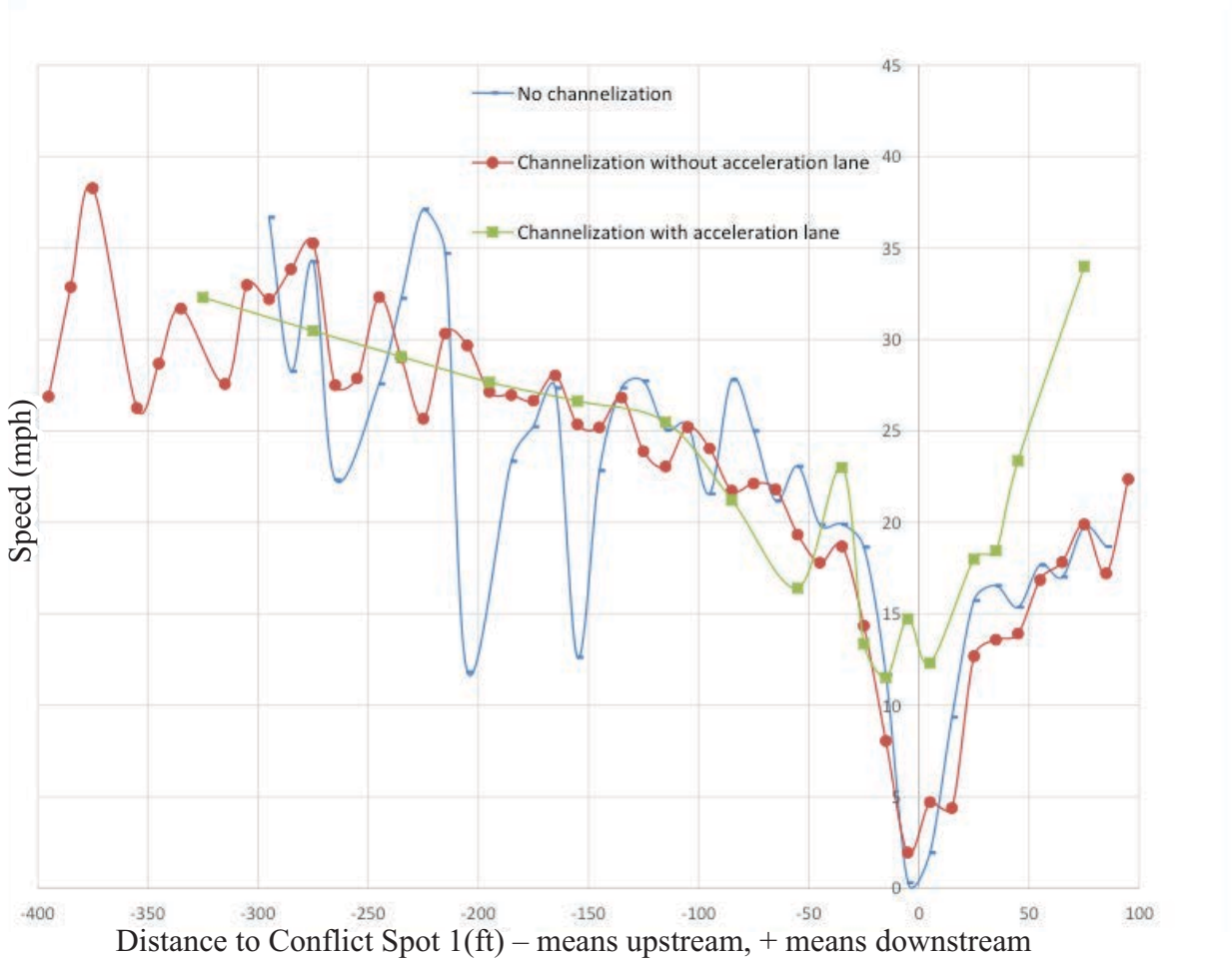

FIGURE 5 Driving speed of right-turn vehicles with different geometric designs.

4.4 Conflicting Traffic Flow

The association of conflicting traffic volume with driver operation was also studied in the 211 events. The speed distribution under different conflicting vehicle volume is shown in Figure 6. The result shows that there is no obvious difference when the conflicting volume is different. The trend of speed first decreases then increases. When the conflicting volume is between 1,000 vph to 2,000 vph, speed is lower before vehicles reach the stop bar compared with other situations. Speed is lowest near conflict Spot 1 during the whole right-turn movement. For low conflicting traffic volume (<1,000 vph), some drivers speed up when they approach conflict Spot 1.

When conflicting traffic flow was between 1,000 vph and 2,000 vph, drivers had the lowest frequency of improper operation. When conflicting traffic volume was high (>2,000 vph), the frequency of improper operation increased.

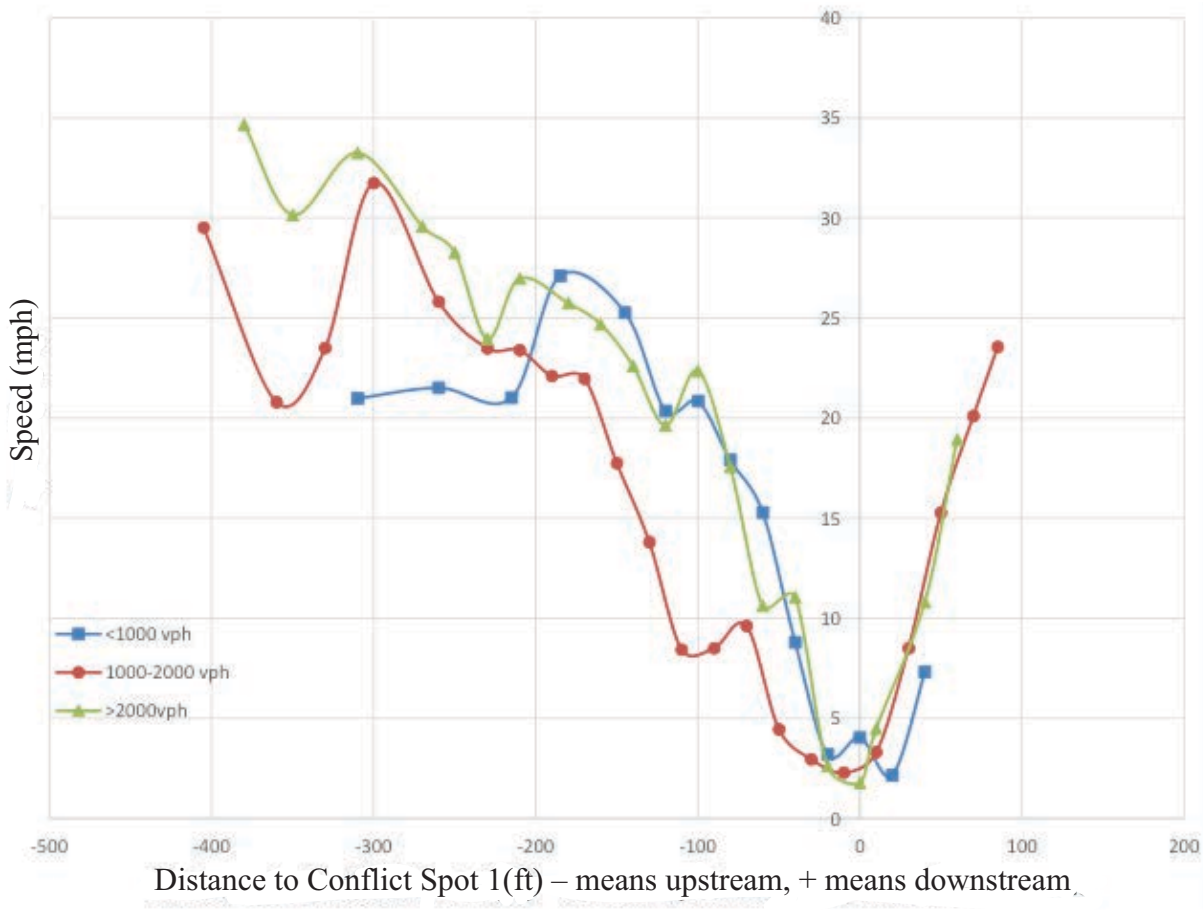

FIGURE 6 Driving speed of right-turn vehicles with different conflicting traffic volumes.

5. CONCLUSION AND DISCUSSION

This research developed a new method to divide right-turn movement into stages to study driver observation behavior. Similar methods could also be used to divide other turn movements, such as left turn movement and U-turn movement. This paper addressed the relationship between different factors and right-turn accident severity at intersections. The results show driver behavior and speed are the top two prioritized influencing factors. Crash severity will rise if the speed and deceleration increase before crashes. Approximately 80% of drivers did not fully stop when they reached the stop bar at RTOR. The speed when vehicles reached an intersection having channelization without dedicated downstream decreased more than at intersections having channelization with dedicated downstream. The results also showed that there was no obvious difference in speed when the conflicting traffic volume was different. Drivers had the lowest frequency of improper operation when conflicting traffic flow was between 1,000 vph and 2,000 vph.

The findings of this paper could eventually be used to guide engineers to choose or improve possible countermeasures to reduce right-turn crashes at signalized intersections. The NDS data were used to study right-turn crashes for the first time in this paper. While this research provides interesting findings there are limitations. There was no comparison between crash events and baseline events as there were no baseline events in the received data. Right-turn drivers' behaviors at unsignalized intersection are not included in this paper. The conflicts between right-turn vehicles and pedestrians/cyclists are not considered in this paper because of limited data. Since it was not possible to acquire crash location information for this study, historical crash data at these intersections were not available. As face videos were not available in this study, driver observation behavior was also not available. The results of the regression model may be affected by the high standard error of some variables. While this research begins to explore the factors that may influence driver behavior while turning right at signalized intersections, more data are still needed to evaluate the findings of this research. Finally, the influence of other factors such as roadway geometric design still needs to be studied in the future.

ACKNOWLEDGMENTS

This paper presents the main results of the research sponsored by SHRP 2 Safety Data Student Paper Competition Program and partial results of the research sponsored by SHRP 2 Implementation Assistance Program.

REFERENCES

1. Institute of Transportation Engineers and FHWA. The National Intersection Safety Problem. Issue Briefs, April, 2004. http://library.ite.org/pub/e26c787c-2354-d714-514a-8d415cd476eb. Accessed September 3, 2016.
2. Transportation Research Board of the National Academies. (2013). The 2nd Strategic Highway Research Program Naturalistic Driving Study Dataset. Available from the SHRP 2 NDS InSight Data Dissemination Website: https://insight.shrp2nds.us.
3. Evans, L. The Dominant Role of Driver Behavior in Traffic Safety. *American Journal of Public Health*, Vol. 86, Issue 6, 1996, pp. 784–786.
4. Tijerina, L., J. D. Chovan, J. Pierowicz, and D. L. Hendricks. Examination of Signalized Intersection, Straight Crossing Path Crashes and 13 Potential IVHS Countermeasures. Final Report. No. HS-808 143. 1994.
5. Rämä, P. Effects of Weather-Controlled Variable Speed Limits and Warning Signs on Driver Behavior. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1689, 1999, pp. 53–59.
6. Toledo, T., H. Koutsopoulos, and M. Ben-Akiva. Modeling Integrated Lane-Changing Behavior. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1857, 2003, pp. 30–38.
7. Salvucci, D. D., and A. Liu. The Time Course of a Lane Change: Driver Control and Eye-Movement Behavior. *Transportation Research Part F: Traffic Psychology and Behaviour* , 5.2, 2002, pp. 123–132.
8. Chatterjee, I., and G. Davis. \Use of Naturalistic Driving Data to Characterize Driver Behavior in Freeway Shockwaves. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2434. 2014, pp. 9–17.

9. Wortman, R. H., and J. S. Matthias. An Evaluation of Driver Behavior at Signalized Intersections. Arizona Department of Transportation, 1983.
10. Sato, T., and M. Akamatsu. Influence of Traffic Conditions on Driver Behavior Before Making a Right Turn at an Intersection: Analysis af Driver Behavior Based an Measured Data on an Actual Road. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10.5, 2007, pp. 397–413.
11. Harbluk, J. L., Y. I. Noy, and M. Eizenman. The Impact of Cognitive Distraction on Driver Visual Behaviour and Vehicle Control. No. TP# 13889 E. 2002.
12. Summala, H., E. Pasanen, M. Räsänen, and T. Sievänen. Bicycle Accidents and Drivers' Visual Search at Left and Right Turns. *Accident Analysis & Prevention*, 28.2, 1996, pp. 147–153.
13. Antin, J. F. Design of the In-Vehicle Driving Behavior and Crash Risk Study: In Support of the SHRP 2 Naturalistic Driving Study. Transportation Research Board, 2011.
14. Klauer, S. G., T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. 2006.
15. Campbell, K. L. The SHRP 2 Naturalistic Driving Study: Addressing Driver Performance and Behavior in Traffic Safety. *TR News,* No. 282, 2012.
16. Second Strategic Highway Research Program Roadway Information Dataset. Transportation Research Board of the National Academies, 2014.
17. Quddus, M. A., R. B. Noland, and H. C. Chin. An Analysis of Motorcycle Injury and Vehicle Damage Severity Using Ordered Probit Models. *Journal of Safety Research,* 33:4, 2002, pp. 445–462.
18. Abdel-Aty, M. Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models. *Journal of Safety Research,* 34:5, 2003, pp. 597–603.
19. Wang, X., and K. M. Kockelman. Occupant Injury Severity Using a Heteroscedastic Ordered Logit Model: Distinguishing the Effects of Vehicle Weight and Type. *Transportation Research Record*: *Journal of the Transportation Research Board*, No. 1908, 2005, pp. 195–204.
20. Kockelman, K. M., and Y.-J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis & Prevention*, 34:3, 2002, pp. 313–321.
21. Rifaat, S. M., and H. C. Chin. Accident Severity Analysis Using Ordered Probit Model. *Journal of Advanced Transportation*, 41:1, 2007, pp. 91–114.
22. Duncan, C., A. Khattak, and F. Council. Applying the Ordered Probit Model to Injury Severity in Truck–Passenger Car Rear-End Collisions. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1635, 1998, pp. 63–71.
23. Lemp, J. D., K. M. Kockelman, and A. Unnikrishnan. Analysis of Large Truck Crash Severity Using Heteroskedastic Ordered Probit Models. *Accident Analysis & Prevention,* 43:1, 2011, pp. 370–380.
24. Ho, T. K. Random Decision Forests. *Proceedings of the Third International Conference on Document Analysis and Recognition, Vol. 1*. IEEE, 1995.
25. Friedman, J., T. Hastie, and R. Tibshirani. *The Elements of Statistical Learning. Vol. 1*. Springer, Berlin, 2001.
26. Breiman, L. Manual on Setting Up, Using, and Understanding Random Forests, v3.1. Statistics Department, University of California, Berkeley, 2002.
27. Breiman, L. Random Forests. *Machine Learning,* 45:1, 2001, pp. 5–32.
28. Wu, J. Observation Behavior Analysis for Turning Drivers at Signalized Intersections. Submitted to Transportation Research Board, Paper No. 17-02852, 2016.

Jianqing Wu, Graduate Student, Department of Civil and Environmental Engineering, University of Nevada, Reno. 1664 North Virginia Street, MS258, Reno, NV 89557; e-mail: jianqingwu2015@gmail.com.

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the National Academies of Sciences, Engineering, and Medicine to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at www.national-academies.org.

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied committees, task forces, and panels annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. www.TRB.org

Learn more about the Transportation Research Board at www.TRB.org.

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001