

Utilizing crash and violation data to assess unsafe driving actions

Mohammad Mahdi Rezapour Mashhadi *, Shaun S. Wulff **,

Khaled Ksaibati ***

* University of Wyoming,

1000 E University Ave, Dept. 3295, Laramie, WY 82071, USA

Graduate Research Student, Department of Civil & Architectural Engineering

** University of Wyoming,

1000 E University Ave, Dept. 3295, Laramie, WY 82071, USA

Associate Professor, Department of Statistics

*** University of Wyoming,

1000 E University Ave, Dept. 3295, Laramie, WY 82071, USA

Ph.D., P.E., Director, Wyoming Technology Transfer Center



Article history:

Received: September 12, 2017

1st Revision: October 10, 2017

Accepted: October 28, 2017

DOI:

[10.14254/jsdtl.2017.2-2.3](https://doi.org/10.14254/jsdtl.2017.2-2.3)

Abstract: Wyoming has one of the highest crash rates in the United States and a higher fatality rate than the U.S. average. These high rates result from many factors such as the high traffic through I-80 and the mountainous areas of Wyoming. This study employed two approaches to study contributory factors to crashes in the most hazardous interstate, I-80, in Wyoming by employing crash and citation data sets. Different factors may contribute to different driver actions so it is important to consider these crash causes separately. Thus, multiple logistic regression models were used in this study to examine the differences in crash-contributing factors for three driver actions: driving too fast for conditions, improper lane change, and no improper driving. These driver actions account for about 70% of all the crash causes on this interstate. The same violations as the two driver actions, improper lane change and driving too fast for conditions, account for 42% of all the crashes. The literature has indicated that previous violations can be used to predict future violations, and consequently crashes. Therefore, these violations were identified to detect the groups that are at higher risk of involvement in crashes. The analyses indicated that there are substantial differences across different driver actions for crash and violation data. For instance, not-dry-surface conditions increased the estimated odds of driving too fast for conditions 33 times while it decreased the risk of no improper driving by an estimated 250%. Crash severity, number of vehicles, vehicle maneuver,

Corresponding author: Mohammad Mahdi Rezapour Mashhadi

E-mail: mrezapou@uwyo.edu

This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.



point of impact, driver condition, and speed compliance also impacted different driver actions differently. The results of violation analyses revealed that the interaction between types of vehicle and various variables were significant. For instance, nonresident truck drivers were more likely to violate all types of risky violations, which increased the estimated odds of crashes, compared with resident truck drivers. Recommendations based on the results are provided for policy makers to reduce high crash rate in the state.

Keywords: No Improper driving, improper lane change, drive too fast for conditions, violation, citation, enforcement, Traffic Safety, Road Safety, Multinomial logistic regression, vehicle maneuver, point of impact.

1. Introduction

Road traffic crashes take more than 1.2 million lives around the world each year and put a huge burden on the development of the world economy (World Health Organization, 2015). In 2015, there were 32,166 road fatal crashes in the U.S., which resulted in 35,092 deaths. Wyoming has the highest fatality rate per 100,000 people in the nation (National Highway Traffic Safety Administration, 2016). In Wyoming, the high fatality rate resulted from high truck traffic through I-80, mountainous areas and adverse-weather conditions during winter. Reasons for crashes can be assigned to three categories: drivers (94%), vehicle component failure (2%), and environment (2%) (Singh, 2015).

The reasons attributed to drivers can be categorized into recognition errors (such as inattention and inadequate surveillance), driver decision errors (such as driving aggressively and driving too fast), and performance errors (e.g. overcompensation and improper directional control) (National Highway Traffic Safety Administration, 2008). However, most of the traffic crashes are predictable and preventable (World Health Organization, 2015). Policy makers have practiced different countermeasures in the U.S., which can be divided mainly into 4 E's of safety: enforcement, engineering, education and emergency medical services (EMA). Wyoming has the highest fatality rate in the nation (National Highway Traffic Safety Administration, 2008), and it has the highest large truck crash per million vehicle mile travelled (MVMT). Also, it is in the bottom ten states of enforcement contribution (Weber & Murray, 2014). Therefore, the analyses of citation data, along with crash data, can help the policy makers to determine how to use their limited resources to reduce high crash and fatality rates in the state.

2. Background

The majority of the safety studies analyzed crash data as a whole, and so they did not detect the predictors unique to different crash types. However, few studies identified the unique contributory factors for each crash type.

Zou et al. (2017) studied the differences between single-vehicle and multiple-vehicle truck crashes in New York City. The results indicated that different factors impacting single-vehicle and multi-vehicle truck crash severity. It was also found that truck weight behaves differently for these two types of crashes. Bham et al. (2011) examined the factors that contribute to different collision types. The results indicated that contributory factors vary across different collision types and types of highways. Shinstine et al. (2016) investigated the factors associated with crash severity on Wyoming rural highways. Five different rural highway systems, such as global and interstate system, were used to develop different models. The results indicated that there are substantial differences across different rural highway systems thereby justifying separate analyses.

Many studies also were carried out to identify groups at higher risk of future crashes. Identification of the groups at higher risk can help policy makers to reduce crashes by changing their policies and targeting specific groups. Li and Baker (1994) found that conviction records can be used to identify groups at higher risks of being involved in fatal crashes. In another study, Elliott et al. (2001)

studied the ability of previous violations in predicting future offences and crashes. The results indicated that the drivers with previous tickets were at higher risk for future crashes.

Based on the literature review, there are substantial differences between crashes with different characteristics, so it is important to identify risk factors associated with different types of crashes to determine appropriate safety countermeasures. The literature also indicated that violation data can be used as a way to identify the groups at higher risk of violating laws and, consequently, being involved in crashes. The results of violation data analysis can help the highway patrol to target specific groups by identifying appropriate countermeasures. The countermeasures could reduce crashes in the most efficient ways.

This study was set forward to fulfill the following objectives:

1. Investigate contributory factors to crashes with different driver actions (crash data).
2. Investigate contributory factors to the violations that account for the highest percentage of crashes with the objective of identifying the groups at higher risk of getting involved in crashes (violation data).

Understanding contributory factors to driver actions and fulfilling the aforementioned objectives will help provide a better understanding of crash causation and consequently address the high crash rate in Wyoming.

3. Methods

This study was undertaken to investigate contributory factors to crashes on I-80 which has the highest number of crashes in Wyoming. For crash data, the outcome had four categories: driving too fast for conditions, no improper driving, improper lane change, and other types of driver actions. Violation data was also used to identify drivers who are at a higher risk of committing particular traffic law violations that are associated with the leading causes of crashes. Thus, the same levels were used for the violation data that were used for the crash data.

The response variable, denoted Y_{ij} for observation i and action type j , is used to denote driver action types. Thus, the response is assumed to have a multinomial distribution. Different predictors such as driver characteristics and environmental characteristics are used as the explanatory variables denoted by $x_{i1}, x_{i2}, \dots, x_{ip}$, where i indexes the observation (crashes) and p is the number of predictors. Multinomial logistic regression is used to model nominal outcomes with more than two levels (Hosmer et al., 2013). For the j multinomial categories, there are $j(j-1)/2$ pairs of categories and $j(j-2)/2$ sets of predictors (Kutner et al., 2004). By using j as a baseline category, $j-1$ comparisons are considered in relation to the reference category. The baseline category was other types of driver actions for the crash data and others types of citations for the citation data.

The logit for the j th comparison is:

$$\pi_{ijj} = \log_e \left[\frac{\pi_{ij}}{\pi_{ijj}} \right] = x_i' \beta_{jj}, j = 1, 2, \dots, J - 1, \quad (1)$$

where x_i is the vector of predictors for observation i and β_{jj} is the vector of regression coefficients for associated with these predictors for category j with reference category J . The $J - 1$ category probabilities can be obtain as:

$$\pi_{ijj} = \frac{\exp(x_i' \beta_{jj})}{1 + \sum_{k=1}^{J-1} \exp(x_i' \beta_{kj})}, j = 1, 2, \dots, J. \quad (2)$$

Odds ratios are commonly used to interpret the effects of the predictors on the response category. The odds ratio (OR) is the ratio of the odds obtained from the model probability for one combination of regressors relative to the odds for the model probability of another combination of regressors. This assumes the other predictor variables, not of interest, are constant across the comparison. As in Shinstine et al. (2016), consider a specific binary predictor $x_k = x_{ki}$ that is not involved in any interaction effect. The odds ratio for $x_k = 1$ compared with $x_k = 0$ for category j with reference category J is

$$OR = \exp(\beta_{(k)jj}) \quad (3)$$

Now, consider a comparison between two levels of a binary predictor x_k that is involved in a single interaction effect with the binary predictor x_t . When there is an interaction effect between x_k and x_t , the impact of the predictors x_k and x_t on the response cannot be interpreted separately.

For $x_k = 1, x_t = 0$ compared to $x_k = 1, x_t = 0$, the odds ratio is again given by equation (3). However, for $x_k = 1, x_t = 1$ compared with $x_k = 0, x_t = 1$, the odds ratio is given by

$$OR = \exp(\beta_{(k)jj} + \beta_{(kt)jj}) \quad (4)$$

For $x_k = 1, x_t = 1$ compared with $x_k = 0, x_t = 0$, the odds ratio is given by

$$OR = \exp(\beta_{(k)jj} + \beta_{(t)jj} + \beta_{(kt)jj}) \quad (5)$$

Furthermore, if x_k is involved in another pairwise interaction with a binary predictor, say h , then formulas (3) and (4) still hold, but with $x_h = 0$. Now, equations (3) and (4) are unknown since they depend upon the unknown regression coefficients (β_{jj}). Estimated odds ratios are calculated by plugging in the estimates of these regression coefficients ($\hat{\beta}_{jj}$) into these equations. The estimated odds ratio (OR) will be presented and discussed in the results section.

4. Data Preparation

Data was used from the interstate in Wyoming, I-80, which has the highest crash rate in the state. Crash data from 2011 to 2014 was obtained from the Wyoming Department of Transportation (WYDOT) using the critical analysis reporting environment (CARE). Variables used in this study were categorized into 6 characteristics: driver, crash, temporal, environmental, roadway, and vehicle. Driver characteristics included gender, age, residency, speed limit compliance, driver conditions, and citation record at the time of crash. Crash characteristics included point of impact, vehicle maneuver, traffic, and number of vehicles. Temporal characteristics included day of a crash, weekend or not, and time of crash, off peak or peak hours. Different variables were categorized under the environmental category such as weather conditions, road conditions, and lighting conditions. Roadway characteristics at the crash location included vertical and horizontal characteristics of the segment and the posted speed limit of the segment. Vehicle characteristics were divided into truck and non-truck vehicles. The driver actions category as a response was chosen to include the driver actions, including no improper driving, driving too fast for conditions and improper lane change, which account for 71% of all crashes. If any driver action did not belong to any of these three driver actions, it was categorized under the other category, which was used as the baseline for crash data analysis. A crash with no improper driving was titled under no improper driving.

Violation data from the same period, 2011-2014, was obtained from the Wyoming court. The targeted violations, improper lane change and driving too fast for conditions, were identified from among 1000 different types of violation. These violations accounted for the total of 79,738 citations. As for no improper driving, crash data, drivers/vehicles had no improper driving. Thus, no violation can be assigned to this type of driver action so this variable was not included in violation analysis. If any type of violation did not belong to these violations, it was categorized under "others". This category was chosen as a reference for the violation analysis.

5. Variable description

In order to obtain insights about general characteristics of crashes and violations, summary statistics of crashes and violations are presented in Table 1. Due to the high number of variables, only significant explanatory variables, responses, and the distributions of truck and non-truck crashes are presented in Table 1. As can be seen from this table, a high proportion of crashes (46%) are attributed to trucks. Also, the table indicates different types of responses: driving too fast for conditions (29%), no improper driving (29%), and failure to keep proper lane (13%) accounted for 71% of all types of

crashes. A large portion of the drivers (11%) were fatigued or sick at the time of crashes. The majority of all the crashes (58%) on this interstate occurred on not-dry-road conditions.

The summary statistics for violations are also presented in Table 1. As can be seen, although speed too fast for conditions and failure to keep proper lane account for 42% of all the crashes, the related violations: speeding too fast for conditions (1%) and failure to keep proper lane change (2%), account for only 3% of all the violations. The majority of the drivers (91%) were male and most of the citations (94%) were issued at peak hours.

Table 1: Summary of crash characteristics

Types of Variables		Variables/Categories	Frequency	Percent
Crash data				
Response		Driving too fast for conditions	1,761	29
		No improper driving	1,782	29
		Failure to keep proper lane	825	13
		Others	1,816	29
Predictors				
Crash characteristics	Crash severity	PDO	4,918	80
		Injury/fatality	1,266	20
	Number of vehicle	Single	4,581	74
		Multiple	1,603	26
	Point of impact	Stop/slow		
		Rear	1,424	24
		Front	2,967	49
		Rollover	1,318	22
	Vehicle maneuver	Sideswipe	325	5
		Going straight	4,777	78
Turn		535	9	
	Negotiating curves	455	7	
	Stop/slow	381	6	
Vehicle characteristics	Type of vehicle	Truck	2,590	46
		Non truck	3,011	54
Driver characteristics	Driver condition	Normal	5,344	89
		Fatigued/sick	665	11
	Speed compliance	Speed limit was complied	5,500	89
		Speed limit was not complied	684	11
Residency	Wyoming resident	1,602	26	
	Non Wyoming resident	4,537	74	
Environmental characteristics	Road conditions	Dry road conditions	2,624	42
		Not-dry-road conditions	3,560	58
Road characteristics	Posted speed limit	Greater than 65 mph	1,469	24
		Less than/equal 65 mph	4,603	76
Violation Data				
Response		Speed too fast for conditions	745	1
		Improper driving	378	0.5
		Fail to keep proper lane	1,874	2
		Others	76,740	96
Driver characteristic	Gender	Male	59,099	91
		Female	5,804	9
	Residency	Resident of Wyoming	16,209	20
		Non resident	63,528	80
Temporal characteristic	Day of week	Weekend	23,893	30
		Business days	55,844	70
	Hours of day	Peak hours	74,723	94
		Off-peak hours	5,014	6
Vehicle characteristics		Truck	10,525	13
		No truck	69,212	87

6. Results and Discussion

Tables 2-3 present the results from multinomial logistic regression (MLR) models along with odds ratios (OR), p-values, and upper and lower confidence limits (CL). Only significant variables are presented in these tables. The proportional odds assumption was evaluated to see if the regression coefficients could be assumed to be the same across all the categories for driver actions and for violations (Kutner et al., 2004).

The results of the test for both the crash data (Chi-square=2946, DF=2, p-value= <.0001) and the citation data (Chi-square=55, DF=12, p-value= <.0001) provide strong evidence against the assumption that the regression coefficients are the same across the categories. These results indicated that separate analyses of the categories would be justified and more straightforward for driver actions and for citations. Thus, separate logistic regressions were conducted for each response category (j) in relation to the reference category (j).

Contributory factors to driver actions, crash data

Table 2 presents the final model including contributory factors to driver actions. Various interaction terms which were believed to be meaningful were included in the model (e.g. vehicle type*weather). However, no significant interactions were identified. The response has four categories including driving too fast for conditions, no improper driving, failure to keep proper lane, and all others driver actions. The other driver action, others, was chosen as a reference and all the categories were compared with this category.

Driving too Fast for Conditions

The results in Table 2 indicate that compared with other types of driver actions, driving too fast for conditions was estimated to be 32% more likely to result in severe vehicle crashes (OR=1.32). Compared with other types of driver actions, driving too fast for conditions was estimated to be about 38% more likely to be single vehicle crashes ($1/0.73=1.38$) compared with multiple-vehicle crashes. This can be due to loss of control and going off road, which can result from involvement of only a single vehicle crash. Driving too fast for conditions occurred on less-than-optimal-weather and road conditions, so as can be expected, driving on not-dry road conditions increased the estimated odds of involvement in a crash with this driver action about 34 times. Vehicle maneuver is another variable which was divided into four categories including straight (reference), turn (left or right), negotiating curves, and stopping/slowning. Negotiating a curve, compared with going straight, increased the odds of getting involved in driving too fast for conditions by an estimated 60% compared with other types of driver actions. This also can result from loss of control on less-than-optimal-road conditions while negotiating curves. Point of impact was categorized into 4 categories including rear (reference), head on, sideswipe and rollover/jackknife. Compared with rear point of impact, head on collision was estimated to be 50% more likely to occur when a vehicle involved in driving too fast for conditions than when a driver was involved in other types of driver actions. This impact can be explained as drivers lose control, they go straight ahead and hit other subjects instead of being hit on the rear sides.

The driver conditions variable such as being fatigued/sick was significant with OR=0.15, indicating fatigued/sick drivers were more likely to be involved in other types of driver actions. This may be due to the fact that fatigued drivers were less likely to impose themselves on driving in adverse-weather and -road conditions. A higher posted speed limit than 65 mph (OR= 1.79) and not complying with posted speed limit (OR=1.64) were some of the variables that increased the estimated odds of being involved in a crash for drivers who drove too fast for conditions. These impacts may result from not having enough control over vehicles on less-than-optimal road conditions, which would be exacerbated when the drivers speed up or fail to comply with the posted speed limits.

No Improper Driving

Table 1 indicated that having no improper driving actions account for 29% of the causes of the crashes. However, the causes of this high percentage of crashes were not clear. Therefore, investigating the factors that contribute to this type of driver action could help policy makers address this type of driver action, and consequently reduce the high crash rate on interstate 80. Table 1 presents contributory factors to crashes in which the driver had no improper driving action, but are nevertheless involved in crashes. When a crash occurred due to no improper driving action, with no specific reasons, it was estimated to about 3 times more likely for a crash to involve in more than 1 vehicle (OR=2.98). Vehicles slowing or stopping increased the likelihood of involvement in no improper driving action

(OR=1.77). This can be an indication that while stopping/slowing is not a law violation, it can result in a crash in which vehicles were hit by the other vehicle.

Higher posted speed limits increased the likelihood of being involved in a crash resulting from no improper driving. This can be an indication that the speed of a vehicle is an important factor for this type of crash. This impact may result from the loss of control for a vehicle when the posted speed, and consequently vehicle speed goes up. Other variables such as crash severity, road conditions, head on point of impact, and driver condition were found to be significant. However, these variables were more likely to increase crashes with driver actions other than improper driving. When a crash occurred with a driver having no improper driving, the likelihood of a crash to be severe decreased, compared with other driver actions. This could be an indication of the impact of having any improper driving on crash severity. Decreased risk of no improper driving (OR=0.40) while drivers were driving on not-dry-road conditions may be due to the fact that these drivers were not involved in any type of improper driving and they took precautions while driving on not-dry-road conditions. For no improper driving crashes, drivers did not hit other vehicles from the rear side and other vehicle hit these vehicles which can result in decreased risk of involvement in head on collisions. Drivers with no improper driving were also less likely to drive while fatigued which can be a reason behind the negative estimate $\beta=-3.021$) of driver conditions.

Being fatigued/sick, compared with being under normal conditions, increased the estimated odds of crashes with other types of driver actions, an estimated 25 times (1/0.04), compared with this driver action. In contrast with driving too fast for conditions, improper driving decreased the likelihood of crash severity (OR= 0.68) and increased the likelihood of number of vehicles (OR=2.98).

Failure to Keep Proper Lane

Table 2 presents contributory factors to improper lane change. Similar to driving too fast for conditions, driving on not-dry road conditions increased the estimated odds of this driver action (OR=1.57). These types of crashes were more likely to occur when drivers negotiated curves compared to when they drive on straight segments. This may result from traction loss in a curve, which can result in failure to keep proper lane.

Driving under not normal conditions increased the estimated odds of being involved in this type of crash, which is estimated to be more than two times versus normal conditions, compared with other driver actions. This may result from the fact that these drivers lacked the judgment and depth perception needed to keep proper lane. Driving on segments with posted speed limit greater than 65 mph, increased the estimated odds of crashes with driver action of improper lane change compared with other types of driver actions. This can result from the difficulty associated with controlling vehicles and keeping vehicles in an appropriate lane while driving a vehicle at higher speeds.

Residency was a factor that impacts only failure to keep proper lane change driver action, compared with other types of driver actions. Nonresidents of Wyoming were an estimated 34% more likely to be involved in improper lane change compared with other types of driver actions. This may be due to lack of familiarity with the mountainous areas of Wyoming and difficulty associated with the keeping proper lane in these conditions. However, two factors, stop/slow (OR= 0.34) and non-speed limit compliance (OR=0.59) were more likely to impact crashes with driver actions than failure to keep proper lane.

Table 2: Multinomial Logistic Regression, parameter Estimates and Odds ratios for different Driver Actions, crash data

Variable	Categories	Reference	Estimates	P-value	OR	95% L CL	0.95 U CL
Drive too fast for conditions							
Intercept			-3.420	<0.0001			
Crash severity	Injury/fatality	PDO	0.2777	0.0156	1.32	1.054	1.652
Number of vehicle	More than 1	Single vehicle	-0.319	0.0047	0.72	0.583	0.907
Surface Conditions	Not-dry-road conditions	Dry weather conditions	3.523	<0.0001	33.88	23.217	49.434
Vehicle Maneuver	Negotiating curves	Straight	0.468	0.0037	1.60	1.164	2.193
	Stop/slow	Straight	-0.446	0.239	0.64	0.434	0.943
Point of impact	Head on	Rear	0.407	0.0001	1.50	1.222	1.846
Driver conditions	Fatigued	Normal	-1.892	<0.0001	0.15	0.059	0.303
Posted Speed Limit	Greater than 65 mph	Less than/equal 65 mph	0.580	<0.0001	1.79	1.488	2.142
Speed compliance	Speed limit was not complied	Speed limit was complied	0.495	0.0003	1.64	1.254	2.145
No Improper Driving							
Intercept			0.064	0.721			
Crash severity	Injury/fatality	PDO	-0.379	0.0016	0.68	0.543	0.866
Number of vehicle	More than 1	Single vehicle	1.0914	<0.0001	2.98	2.420	3.666
Road Conditions	Not-dry-road conditions	Dry weather conditions	-0.918	<0.0001	0.40	0.329	0.484
Vehicle Maneuver	Stop/slow	Straight	0.569	0.0009	1.77	1.262	2.475
Point of impact	Head on	Rear	-0.370	0.0008	0.69	0.556	0.857
Driver conditions	Fatigued	Normal	-3.207	<0.0001	0.04	0.022	0.076
Posted Speed Limit	Greater than 65 mph	Less than/equal 65 mph	0.509	<0.0001	1.66	1.352	2.048
Fail to keep proper lane							
Intercept			-1.973	<0.0001			
Road Conditions	Dry conditions	Dry weather conditions	0.452	0.0002	1.57	1.242	1.989
Vehicle Maneuver	Negotiating curves	Straight	0.345	0.050	1.41	0.996	2.002
	Stop/slow	Straight	-1.072	0.0023	0.34	0.172	0.681
Driver conditions	Fatigued	Normal	0.747	<0.0001	2.11	1.602	2.783
Posted Speed Limit	Greater than 65 mph	Less than/equal 65 mph	0.718	<0.0001	2.05	1.595	2.636
Speed compliance	Speed limit was not complied	Speed limit was complied	-0.530	0.0027	0.59	0.417	0.832
Residency	Non Wyoming residency	Wyoming resident	0.292	0.0114	1.34	1.068	1.678

Contributory factors to risky violations related to crash driver actions, citation data

The literature has shown that previous violations can be used to predict future offenses and, consequently, future crashes. So, the current section is set forward to use citation data, in addition to crash data, to identify the groups that are at higher risks of future crashes attributed to particular

violations. Different temporal and driver characteristics were used in this section. The interactions which seemed to be meaningful were identified and also included in the analyses. Two types of violations, "speed too fast conditions" and "fail to drive within single lane" were identified and included in the analyses. These violations were identified among about 1000 different types of citations. These violation are the same as driver actions: "Drive too fast for conditions" and "Fail to keep proper lane", which accounted for 42% of all the crashes.

Driving too Fast for Conditions

The summary statistics of crash data indicated that driving too fast for conditions accounted for 29% of the causes of crashes, so this violation related to this driver action was identified and included in the analysis. Table 3 presents contributory factors to this type of violation. The results indicated that drivers were estimated to be 45% more likely to violate driving too fast for conditions on weekends compared with other violation on business days.

Although the main effects of type of vehicle, residency, and hours of a day were significant, the effect of these predictors could not be separated due to the presence of the corresponding interaction terms. As explained in equation 1, consider truck driver (x_k) and residency as x_r . The estimated odds of getting involved in driving too fast citation for nonresident truck drivers is 2.2 times higher compared with resident truck drivers ($\exp(-0.132 + 0.923)$) during peak hours. This may be due to lack of familiarity of nonresident truck drivers with the mountainous areas of Wyoming. Also, compared with non-truck drivers driving during peak hours, truck drivers who were driving during off peak hours were estimated to be about 83% times less likely to be involved in a crash (OR= 1/0.55) assuming the driver is a non-resident, see equation 5. This can result from the hazard associated with truck driving at night during less than ideal environmental conditions (adverse weather/road conditions).

Failure to Keep Proper Lane

Failure to keep proper lane accounted for 13% of all the crashes on I-80 (see Table1). Similar to the driving too fast for condition violation, failure to keep proper lane was estimated to be 24% more likely to occur on weekend compared with business days (see Table 3). Compared with resident non-truck drivers, nonresident truck drivers were (OR=3.61) more likely to be involved in this type of violation during peak hours. Also, a truck driver driving during off peak hours was (1/0.60) less likely to be involved in this violation compared with no truck drivers during peak hours for a non-resident.

Table 3: Multinomial Logistic Regression: parameter estimates and odds ratios for different driver actions, citation data

Variable	Categories	Reference	Estimates	P-value	OR	95% L CL	0.95 U CL
Drive too fast for conditions (8)							
Intercept			-4.883	<0.0001			
Day of week	Weekend	Weekdays	0.374	<0.0001	1.45	1.291	1.616
Type of vehicles	Truck	Non truck	0.045	0.915	1.05	0.217	1.875
Residency	Resident	Non resident	-0.132	0.227	0.88	0.655	1.098
Hours of a day	Off peak hours	Peak hours	0.832	<0.0001	2.30	2.014	2.580
Truck*residency	Truck*non resident	No truck*resident	0.923	0.033	2.52	1.670	3.364
Truck*hours of a day	Truck*off peak hours	No truck*peak hours	-0.607	0.033	0.55	-0.013	1.103
Fail to keep proper lane (15)							
Intercept			-3.728	<0.0001			
Day of week	Weekend	Weekdays	0.215	<0.0001	1.24	1.134	1.344
Type of vehicles	Truck	Non truck	0.215	0.330	1.24	0.807	1.672
Residency	Resident	Non resident	-0.617	<0.0001	0.54	0.406	0.673
Hours of a day	Off peak hours	Peak hours	1.236	<0.0001	3.44	3.275	3.606
Truck*residency	Truck*non resident	No truck*resident	1.284	<0.0001	3.61	3.169	4.059
Truck*hours of a day	Truck*off peak hours	No truck*peak hours	-0.509	0.0004	0.60	0.320	0.882

7. Conclusions

This study examined factors that contribute to crashes by using crash data and violation data. The response for crash analysis included no improper driving, improper lane change, speed too fast for conditions, and failure to keep proper lane. For violation analysis, the response included two types of violations, speed too fast conditions and failure to keep proper lane which accounted for 42% of the causes of crashes. Before investigating contributory factors to different driver actions, crashes, and related violations, the proportional odds ratio test was carried out to see if the effects of the predictors were constant across the categories. The results of the test led to separate analyses of the categories. The results of crash data indicated that although there are similar predictors for some driver actions, there are also meaningful differences between different driver actions. As for violation data, although the same predictors were observed for the included categories, the estimated values are different. These important differences are discussed below.

The findings of this study provided new insights into crash and violation contributing factors that vary by types of violations and driver actions. While driving too fast for conditions was more likely to involve only single vehicle crashes, no improper driving was more likely to involve more than a single vehicle. Severe crashes were more likely for driving too fast for conditions. These results contrasted with the results obtained from no improper driving type of driver action, indicating that severe crashes were less likely to occur. While it was less likely for driving too fast for conditions to be involved in multiple-vehicles (OR=0.72), it was more likely for drivers with no improper driving to be involved in multiple-vehicle crashes (OR=2.98).

Surface road conditions were found to be significant for all the included driver actions. However, it should be noted that the signs and degrees varied across different driver action categories. As can be expected, the highest impact of road surface condition was observed for driving too fast for conditions (OR = 33.88) compared with failure to keep proper lane (OR = 1.57). In contrast, no improper driving crashes were less likely to occur when the road surface was not dry (OR = $1/0.4 = 2.50$).

The results of vehicle maneuver identified the driver actions that can increase the odds of being involved in crashes with different driver actions. Negotiating a curve was a factor that increased the estimated odds of crashes caused by driving too fast for conditions (OR = 1.60) and failure to keep proper lane (OR = 1.41). When stop/slowing type of maneuver occurred, as a vehicle maneuver for a crash, it was less likely that this maneuver result in driving too fast for conditions (OR = 1.56) and failure to keep proper lane change (OR = 3.92). However, it was more likely for this maneuver to result in no improper driving type of driver action.

On the other hand, while negotiating a curve, drivers were estimated to be 60% more likely to be involved in crashes due to driving too fast for conditions (OR = 1.60) and failure to keep proper lane (OR = 1.41). Although, there was an increase in the estimated odds of head on collisions due to driving too fast for conditions (OR = 1.50), there was a decrease in likelihood of head on collisions due to no improper driving when compared with other types of driver actions (OR = $1/0.69 = 1.44$). This may be due to the fact that driving too fast for conditions resulted from losing control, which impacted the front side of the vehicles while not having improper driving rear part of vehicles got impacted, rather than the front sides.

Driving under not-normal conditions increased the estimated odds of failure to keep proper lane. However, it was less likely for drivers to be under not normal conditions when they were involved in no improper driving and driving too fast for conditions, compared with normal driver conditions. Posted speed limit increased the likelihood of being involved in crashes with different driver actions. Non-speed compliance increased the estimated odds of involvement in driving too fast for conditions (OR = 1.64). However, it was more likely for non-speed compliance drivers to be involved in the driver actions other than failure to keep proper lane (OR = $1/0.59 = 1.70$). Residency was the only significant variable for failure to keep proper lane where it increased the estimated odds of the involvement in this type of driver action (OR = 1.34).

Based on the literature review, violation data can be used to identify the groups that are likely to violate the traffic laws in the future and, consequently, be involved in a crash. Therefore, after learning about contributory factors to crashes by using crash data, violation data was used to identify the groups that are at a higher risk of getting involved in crashes by including these particular violations. The results of the proportional odds ratio indicated that the violations were significantly different, and so were

analyzed separately. Those violations, driving too fast for conditions and failure to keep proper lane, were similar to driver action which resulted in more than 40% of all the crashes. Weekends driving increased the estimated odds of getting involved in crashes when driving too fast for conditions (OR = 1.45) and improperly changing lane (OR = 1.24).

The results also identified some important interactions. The results of interaction between vehicle and residency indicated that nonresident truck drivers were more likely to be involved in the two risky driving actions, 252%-361%, compared with resident non-truck drivers. The interaction results between vehicle type and hours of a day was an indication that truck drivers during off peak hours were less likely to violate the two risky-driving laws, compared with non-truck during peak hours.

8. Recommendations

This study aimed at identifying contributory factors to crashes by using crash and violation data. Crashes and violations were divided into different categories to obtain more accurate results. In order to reduce crash rate in the most efficient way, it is recommended for future studies to identify different types of crashes separately. It will help the policy makers to identify various causes of different crash types in detail. The results of this study can help different organizations in charge of traffic safety in Wyoming to target the correct groups based on driver actions. The following recommendations are made based on the results:

- Higher posted speed limit increased the odds of crashes for all three driver actions. More investigations/adjustments are recommended for posted speed limit.
- Negotiating a curve increased the odds of driving too fast for conditions and improper lane change. More warning signs and enforcement are recommended at the locations with high horizontal alignment. Enhance delineation and friction for horizontal curves can be done as well along with more enforcement at these locations, especially in winter for driving too fast for conditions.
- Not-dry-road conditions increased the odds of failure to keep proper lane, and especially driving too fast for conditions. More countermeasures, such as lowering speed limit, and setting up variable speed limit recommended during the winter season.
- Nonresident truck drivers were at higher risks of violating risky driving laws. More enforcement and new regulations/training are recommended for this group.
- More investigations are recommended for studying different driver actions for future studies. This study showed that there are significant differences between contributory factors across different driver actions and related violations.

Stopping/slowng type of vehicle maneuver was a significant and contributory factor for no improper driving type of crashes, which accounted for 29% of all crashes. It is recommended for WYDOT to set/adjust the minimum recommended speed limit along this corridor.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://jsdtl.sciview.net>

Funding

The Wyoming Technology Transfer (WYT2/LTAP) center at the University of Wyoming provided extensive resources to assist in the compilation of the data sets used. WYDOT provided the funding of this study. Special acknowledgement goes to WHP and WYDOT for providing the data used in this research.

Citation information

Mashhadi, M. M., Wulff, S. S., & Ksaibati, K. (2017). Utilizing crash and violation data to assess unsafe driving actions. *Journal of Sustainable Development of Transport and Logistics*, 2(2), 35-46. doi:10.14254/jsdtl.2017.2-2.3.

References

- Bham, G. H., Javvadi, B. S., & Manepalli, U. R. (2011). Multinomial logistic regression model for single-vehicle and multivehicle collisions on urban US highways in Arkansas. *Journal of Transportation Engineering*, 138(6), 786-797.
- Elliott, M. R., Waller, P. F., Raghunathan, T. E., Shope, J. T., & Little, R. J. (2001). Persistence of violation and crash behavior over time. *Journal of Safety Research*, 31(4), 229-242.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013a). *Applied logistic regression*. John Wiley & Sons.
- Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models* McGraw-Hill/Irwin.
- Li, G., & Baker, S. P. (1994). Prior crash and violation records of pilots in commuter and air taxi crashes: A case-control study. *Aviation, Space, and Environmental Medicine*, 65(11), 979-985.
- National Highway Traffic Safety Administration. (2008). National motor vehicle crash causation survey: Report to congress. *National Highway Traffic Safety Administration Technical Report DOT HS, 811, 059*.
- National Highway Traffic Safety Administration. (2016). 2015 motor vehicle crashes: Overview. *Traffic Safety Facts Research Note, 2016, 1-9*.
- SAS Institute. (2014). *SAS 9.4 output delivery system: User's guide* SAS institute.
- Shinstine, D. S., Wulff, S. S., & Ksaibati, K. (2016). Factors associated with crash severity on rural roadways in Wyoming. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(4), 308-323.
- Singh, S. (2015). *Critical reasons for crashes investigated in the national motor vehicle crash causation survey* (No. DOT HS 812 115).
- Weber, A., & Murray, D. C. (2014). Technical brief: Commercial motor vehicle Enforcement–Top 10 high performance states.
- World Health Organization. (2015). *Global status report on road safety 2015*. World Health Organization.
- Zou, W., Wang, X., & Zhang, D. (2017). Truck crash severity in New York City: An investigation of the spatial and the time of day effects. *Accident Analysis & Prevention*, 99, 249-261.



© 2016-2017, Journal of Sustainable Development of Transport and Logistics. All rights reserved.

This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share – copy and redistribute the material in any medium or format Adapt – remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution – You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Journal of Sustainable Development of Transport and Logistics (ISSN: 2520-2979) is published by Scientific Publishing House "CSR", Poland, EU and Scientific Publishing House "SciView", Poland, EU

Publishing with JSDTL ensures:

- Immediate, universal access to your article on publication
- High visibility and discoverability via the JSDTL website
- Rapid publication
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

Submit your manuscript to a JSDTL at <http://jsdtl.sciview.net/> or submit.jsdtl@sciview.net

