11-2016

Modeling Factors Contributing to Injury and Fatality of Run-off-Road Crashes in Ohio

Deogratias Eustace
University of Dayton, deustace1@udayton.edu

Omar Eid Almutairi
University of Dayton

Peter W. Hovey
University of Dayton, phovey1@udayton.edu

Advances in Transportation Studies an international Journal Section B 40 (2016)

ISBN 978-88-548-9700-7 – ISSN 1824-5463-16005 – DOI 10.4399/97888548970075 – pag. 53-68

Modeling factors contributing to injury and fatality of run-off-road crashes in Ohio

D. Eustace O. Almutairi P.W. Hovey

Department of Civil and Environmental Engineering and Engineering Mechanics, University of Dayton, 300 College Park, Dayton, OH 45469-0234, 937-229-2984
email: Deo.eustace@udayton.edu

subm. 16th October 2015 *approv. after rev. 21st March 2016*

Abstract

The main objective of this study was to determine the factors that contribute significantly to the levels of injury severity when run-off-road (ROR) crashes occur. This study used a 5-year crash data for years 2008 - 2012 from the state of Ohio. The decision tree model in conjunction with generalized ordered logit model was used to investigate characteristics of injury and fatality of run-off-road crashes in Ohio. The decision tree modeling was used for exploratory data analysis identified eight factors that explain a large amount of the variation in the response variable, injury severity. These predictor variables include road condition, ROR crash types, posted speed limit, vehicle type, gender, alcohol-related, road contour, and drug-related.. The results from the generalized ordered logit regression show that the following are significant factors in increasing the likelihood of ROR injury severity levels, i.e., incapacitating and fatal injuries: alcohol and drugs use, curves and grades, female victims, overturn/rollover crashes, ROR crashes on dry roadway surfaces. Additionally, buses, truck, and emergency vehicles, and ROR crashes on roadways with posted speed limits of 40 mph or higher increase the probability of injury severity.

Keywords – run-off-road, generalized ordered logit model, decision tree, classification tree, injury severity

1. Introduction

The Federal Highway Administration (FHWA) defines a run-off-road (ROR) crash also known as a roadway departure crash as a non-intersection crash which occurs after a vehicle crosses an edge line or a center line, or otherwise leaves the traveled way [1]. In such a crash, the vehicle may collide with a non-traversable obstacle or another vehicle travelling in the opposite direction or hit a pedestrian. An ROR crash may also end up with the vehicle overturning. The Federal Highway Administration (FHWA)'s Roadway Departure Safety Program reports that 56 % of all fatal crashes that occurred in the United States in 2013 involved the run-off-road crash types [1]. This is clear evidence that the likelihood of an ROR crash becoming fatal is very high, which makes this type of motor vehicle crashes to be one of the major public health problems.

The Ohio Department of Transportation (ODOT) vigorously campaign for the reduction of roadway departure crashes in the state of Ohio. ODOT [2] reports that in the period of 2007-2011 in the state of Ohio, there were a total of 292,446 ROR crashes that resulted into 2,920 fatalities and 124,491 injuries. In 2013, there were 990 traffic related fatalities in Ohio where 526 of those fatalities (53%) involved in run-off-road crashes [2]. Consequently, there is a need to identify important factors that contribute highly or increase the likelihood of the occurrences of ROR crashes in the state of Ohio, as they threaten the lives of drivers and passengers, and the economy of the state.

The main objective of this study was to determine the factors contributing to fatalities and serious injuries when ROR crashes occur in the state of Ohio. In other words, this study was designed to highlight major factors contributing to ROR crashes that likely cause fatalities and severe injuries in the state of Ohio. The study reported in this paper was set to develop a statistical model that quantifies the level of contribution by environmental, traffic, geometric, and driver behavior related factors that can be easily extracted from available crash data records. Traffic-related injury levels are usually categorized into fatal injury, incapacitating injury, non-incapacitating injury, possible/invisible injury, and no injury. In addition, a traffic crash severity is normally categorized based on the injury severity sustained by the most injured person in that crash, which may be named as a fatal crash, an injury crash, or property damage only (PDO) crash. A fatal crash is defined as a crash in which at least one death occurs, an injury crash is defined as a crash in which at least one injury occurs but no death, and a property damage crash is defined as a crash in which no injury occurs but only involving damage to a vehicle.

ROR crashes normally make up the majority of motor vehicle traffic fatalities because they are predominantly severe crashes [1]. For example, in 2011, nationwide, there were 15,307 ROR fatal motor vehicle crashes, which accounted for 51% of all fatal motor vehicle crashes in the United States and resulting into 16,948 fatalities [1]. It is reported that in the state of Ohio for the period 2007-2011 there was a total of 2,920 motor vehicle fatalities related to ROR crashes [2] while the total number of motor vehicle traffic fatalities in the same time period in Ohio was 5,564 [3], which means that ROR crashes accounted for 52.5% of all traffic-related fatalities in the state.

A number of researchers have studied factors contributing to the severity and occurrences of ROR crashes in recent years. A study by Hall and Zador [4] examined the locations of fatal rollover crashes in New Mexico and Georgia and found that sites associated with fatal crashes had sharper curves and steeper downhill sections. Zegeer et al [5] recommend that paying better attention during design by increasing the width of lanes, having less steep terrain, and reducing the curviness of two-lane undivided highways resulted in fewer ROR crashes.

Viner [6] used the 1991 Fatality Analysis Reporting System (FARS) and General Estimating System (GES) data to study the effects of rollovers on sideslopes and ditches and he found that rollover crashes on sideslopes and ditches were the principal cause of ROR driver fatalities as they contributed to about 25% of all ROR driver fatalities and noting that these types of ROR crashes most often occur on the outsides of horizontal curves. Newman et al. [7] attribute the occurrences of ROR crashes to factors that include (i) avoiding a vehicle, object or animal, (ii) inattentiveness due to distractions, drowsiness, or drugs, (iii) slippery pavement due to climactic conditions, (iv) too high a speed for the road contour (curve or hill) and various roadway features that can exacerbate driver inattentiveness such as narrow lanes, excessively curved road, and narrow or steep shoulders.

Spainhour and Mirsha [8] used traffic crash data from the state of Florida to investigate the effects of driver, roadway, vehicle, and environmental factors on fatal ROR crashes involving overcorrection. They found that among the contributory factors, alcohol use, speeding, inattention, and sleeping or fatigue were the principal factors. Liu and Subramanian [9] used FARS crash data for the period 1991-2007 to analyze factors related to fatal single-vehicle ROR crashes. Their study found that alcohol use, speeding, curved road segments, rural roads, higher posted speed limits, fewer number (one or two) of lanes, adverse weather condition, nighttime driving, two or more occupants in a vehicle, vehicle driven by a young driver and a male driver were significant factors that predicted higher probabilities of a vehicle being involved in ROR crashes.

Zhu et al. [10] used 1997 and 1998 FARS data for states of Alabama, Georgia and South Carolina to study single-vehicle fatal crashes that occurred on two-lane rural highways in these selected states. Their study found that lane width, shoulder width and type, horizontal curves and crest vertical curve and their interactions, traffic volume, driveway type, lighting conditions, and crash time as significant factors for fatal single-vehicle ROR crashes on rural two-lane highways. Roy and Dissanayake [11] used traffic crash data from the state of Kansas for the period of 1999-2008 to study and compare factors associated with ROR and non-ROR crashes in Kansas. They found that nighttime, weekends, adverse weather, rural area, gravel and curved roads, higher speed limits, wet and icy road surfaces, utility vehicles as common characteristics associated with ROR crashes in Kansas. In addition, falling asleep, medical condition, alcohol use, driving too fast for condition, strong winds, freezing rain, shoulders, ruts, holes, and bumps were found to be significant factors having a greater role in contributing to ROR crashes than Non-ROR crashes.

Peng and Boyle [12] investigated factors affecting commercial drivers in single-vehicle ROR crashes. They used large truck crash data from the state of Washington for years 2006-2009 in analyzing crash-related factors including driver, vehicle, roadway, and environmental factors. The analysis of the data reveal that the effect of truck driver distraction, inattention, speeding, seat belt non-usage, drowsiness and fatigue increase the likelihood of an ROR resulting into injury or fatality.

2. Methodology and data

2.1. Data

Traffic crash data were obtained from the Ohio Department of Public Safety (ODPS) crash database, which consist of all police-reported crashes that occur on Ohio's public roadways and streets. Traffic crash data for five years, from 2008 through 2012, were downloaded from the ODPS website. A file that contains ROR crash records only was created by querying and sorting records using a variable known as SEQUENCEEVENT1. This variable contains information on the events in sequence for the vehicle as it met its first harmful event. A particular crash event was classified as an ROR crash if in the SEQUENCEEVENT1 variable it was recorded as overturn/rollover, run-off-road right, run-off-road left, cross median/centerline, or crash with a fixed object. If a record did not contain either one of the five mentioned events, then it was categorized as a non-ROR crash. Thus, the dataset was split into two files, i.e., ROR crashes only file and non-ROR crashes only file. Some records with either missing variables or recorded as unknown were deleted from the ROR crashes only file to create a final file of ROR-related traffic crashes. A file containing a total of 384,505 records of run-off-road traffic crashes with complete crash-related information from the database of traffic crashes that occurred on Ohio's public roads and highways was extracted.

The characteristics of run-off-road-related traffic crashes that occurred on Ohio's public roads and highways between 2008 and 2012 after deleting records with missing, incomplete or unknown values are summarized in Table 1. Variables selected for the current study, which are shown in Table 1, were selected based on the experience and previous studies in the literature of traffic injury severity and on the basis of being included in the ODPS database.

Tab. 1 - Description of Ohio ROR crash data for 2008-2012

Variable	Code	Description	Frequency	Percent
Alcohol-related	0	No	338778	88.1
	1	Yes	45727	11.9
Drug-related	0	No	373816	97.2
	1	Yes	10689	2.8
Person type	1	Driver	290965	75.7
	2	Occupant	93532	24.3
	3	Pedestrian	8	.0
Gender of person	1	Female	152780	39.7
	2	Male	231725	60.3
Age of person	1	<20	101782	26.5
	2	20-25	76280	19.8
	3	26-64	186779	48.6
	4	65+	19664	5.1
Road contour	1	Straight Level	199058	51.8
	2	Straight Grade	71242	18.5
	3	Curve Level	51966	13.5
	4	Curve Grade	62239	16.2
Light condition	1	Daylight/Dawn/dusk	225863	58.7
	2	Dark - lighted roadway	55925	14.5
	3	Dark - Unlighted roadway/unknown	104134	26.7
Weather condition	1	Clear	158979	41.3
	2	Cloudy	93004	24.2
	3	Rain/fog/sleet/snow/wind/other	132522	34.5
Road condition	1	Dry	197100	51.3
	2	Wet/water	92377	24.0
	3	Snow/ice/mud/oil/slush/gravel/other	95028	24.7
Time of crash	1	Early morning/daytime (0400-1859)	245939	64.0
	2	Early night (1900-2259)	65857	17.1
	3	Late night (2300-0359)	72709	18.9
Day of week	1	Weekends	119352	31.0
	2	Weekdays	265153	69.0
Injury severity	1	No injury	263433	68.5
	2	Possible/non-incapacitating injury	100782	26.2
	3	Incapacitating/fatal injury	20290	5.3
Crash severity	1	Fatal	4138	1.1
	2	Injury	141034	36.7
	3	PDO	239333	62.2
ROR crash type	1	Overturn/rollover	5449	1.4
	2	Run-off-road right	186798	48.6
	3	Run-off-road left/cross median/centerline	144058	37.5
	4	Crash with fixed object	48200	12.5
Vehicle type	1	Passenger vehicles	353187	91.9
	2	Trucks/Buses	30640	8.0
	3	Motorcycles/motorized bicycles	283	0.1
	4	Emergency vehicles	395	0.1
Posted speed limit	1	< 40 mph	106286	27.6
	2	40-50 mph	58110	15.1
	3	55-70 mph	220109	57.2

Some other variables that may influence the ROR-related traffic injury severity levels any researcher would be interested to be tested in the severity model such as shoulder width, median width and type, presence of road markings, safety barriers, rumble strips, etc., are not included in the ODPS crash database.

2.2. Methodology

Classification tree (also known as decision tree) modeling was used in conjunction with generalized ordered logit (gologit) model; the basis of selecting these methods and their suitability to injury severity data are discussed. Classification tree model was used for exploratory data analysis (selection of significant independent variables) and generalized ordered logit model was used for model prediction/parameter estimation (using only the variables identified by the decision tree procedure).

Therefore, both classification tree modeling and generalized ordered logit modeling were used to investigate characteristics of injury and fatality of run-off-road crashes. A decision (classification) tree is a multivariate technique that traditionally has been used for data exploration (mining) and prediction and often used in business to model customer behaviors and in medicine to test best diagnosis of a disease [13]. According to Lavery [13]: "Tree algorithms, or simply trees, split a dataset (assign observations in the data set to groups) hierarchically (groups are then divided into subgroups) based on the ability of the X (explanatory) variables, associated with the observations, to predict the Y (response) variable. Tree analysis can be used in conjunction with, or as a replacement for, logistic or multiple regression, correspondence analysis, ANCOVA and neural nets."

In recent years, classification tree methods have been successfully applied in the field of traffic safety to analyze traffic injury/crash severity (e.g., [14-19]). Review of existing literature also shows that generalized ordinal logit (gologit) models used in traffic safety studies (e.g., [20-23]). Therefore, both methods used in the current study are gaining acceptance in analyzing traffic safety studies but in our literature review we did not find any study that utilized both methods together.

2.2.1. Decision tree modeling

Classification tree modeling works by dividing the dataset into small and more homogeneous subgroups by a set of "if-statements." A decision tree is a hierarchical model composed of discriminant functions, or decision rules, that are applied recursively to partition the entire dataset into pure, single class subsets. It divides the dataset based on the most predictive independent variable for the response variable. Trees use some statistical measurements in order to split the dataset into small and more homogeneous subgroups. The classification tree method selects an appropriate measurement based on the type of the response variable.

For a categorical predictor variable, the model splits the predictor variable categories (levels) into two groups of levels. The splits are computed by maximizing a LogWorth statistic and the G2 statistic, also known as log-likelihood ratio (Deviance). The LogWorth statistic, which is the log of the adjusted p-value for the chi-square test is a parameter usually used to grow as well as prune the decision tree model. It is used to indicate whether a particular predictor variable is significant or not. The larger the LogWorth value, the more significant the predictor variable is.

The model generally splits the node based on the larger LogWorth statistic and is computed as shown in Equation 1:

$$LogWorth = -log_{10}(p\text{-}value) \tag{1}$$

where the adjusted *p*-value is calculated by taking into account the number of different ways splits (partitions) can be made. The G^2 (deviance), which is the likelihood ratio for testing independence of the outcome and predictor variables, is essentially twice the change in the entropy. This entropy is computed as shown in Equation 2:

$$G^2 = 2\sum\left[f_o\, log\left(\frac{f_o}{f_e}\right)\right] \tag{2}$$

where:

f_o = observed frequency (counts of observations) in a node

f_e = expected frequency in a node

A candidate G^2 that has been chosen for splitting is computed as shown by Equation 3:

$$G^2_{test} = G^2_{parent} - \left(G^2_{right} + G^2_{left}\right) \tag{3}$$

where:

G^2_{parent} = deviance in parent node

G^2_{right} = deviance in the right-hand side child node

G^2_{left} = deviances error in the left-hand side child node

The fitted values are the estimated proportions within the groups. An example of a classification tree is depicted in Figure 1. A decision tree consists of two types of nodes, branch nodes (including the root node) and leaf (terminal) nodes. Node 1 (the top most node) in the decision tree also known as root node contains the entire sample dataset. Each of the remaining nodes (referring to Figure 1, nodes 2 through 5) contains a subset of the entire dataset. Each branch node is a “parent” to two “children” nodes. For example, node 1 is split to produce nodes 2 and 3 and thus node 1 becomes their parent and then node 2 is the parent to nodes 4 and 5.

Decision pathways are represented by lines linking a parent node to its children nodes. In essence, the decision tree starts by splitting the original dataset (at root node) into two subsets based on a particular attribute value test. This process is repeated on each derived subset (branch node) in a recursive manner known as recursive partitioning. This repetitive procedure stops when the subset at a node has all the same value of the response variable, or when splitting no longer improves the predictions and this node becomes a leaf or terminal node (e.g., nodes 3, 4, and 5 in Figure 1).

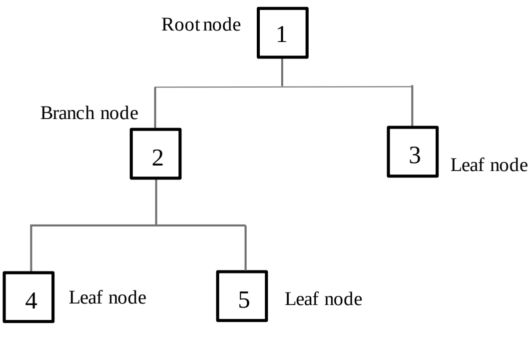


Fig. 1 - An example of a flow-chart structure of a decision tree

Validation is a technique used to examine how strong the independent (predictor) variables predict the response variable. Under this procedure, the entire dataset is divided into two sets. The first set (called training dataset) is used to build the tree model and the second set (called validation dataset) is used to evaluate the performance of the model built by using the first set of data. Detailed discussion on the use of decision tree method can be found in the SAS documentation [24].

The method used for assessing the decision tree model goodness of fit depends on the type of response variable in the dataset. The receiver operating characteristics (ROC) curve is used for assessing categorical responses. An ROC curve is a statistical tool in the form of a graphical plot that provides a complete and visually attractive way to assess the accuracy (power) of predictions [25]. It is a graphical plot of the sensitivity (true positive rate) vs. 1- specificity (false positive rate) for a binary classifier system, such as decision trees. According to Agresti [25], the area under the ROC curve is used to summarize the accuracy (predictive performance) of the analysis data. In essence, the area under the ROC curve estimates the probability that the predictions and the results are in agreement. It takes values from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test [26]. A value of 0.5 specifies that the predictions are essentially based on a random guess, that is, a model includes an intercept term only [13]. The area inside the ROC curve measures discrimination, that is, for case of the current study, the ability of the classification tree analysis to correctly identify the risk factors of ROR crashes.

2.2.2. Generalized ordered logit modeling

Crash injury severity represents a typical ordinal categorical data. The response variable, crash injury severity, as used in this study, it consists three levels after combining some of them together as shown in Table 1. The three levels in increasing severity were coded as 1 = no injury, 2 = possible injury and non-incapacitating injury, 3 = incapacitating injury and fatal injury. Therefore, let’s define k =1 as the lowest level value of injury severity variable, i.e., no injury.

Based on the existing literature, this paper attempts to model injury severity of ROR crashes using the generalized ordinal logit model, a procedure that has been recognized to be a more flexible modeling approach [20-23, 27]. This approach is capable of overcoming the limitations of both the conventional ordered logit/probit and the unordered methods [28]. The suitability of generalized ordered logit model in modeling injury severity data is due to its flexibility in its procedure as it is capable of relaxing the parallel-line assumption in the ordered logit model by allowing the variability of the regression parameter β across outcome levels, while maintaining the ordinal nature of the response variable (injury severity). This method is also known as the partial proportional odds model. The generalized ordered logit model can be expressed as shown in Equation 4:

$$P(Y \le y_j \mid x) = \frac{\mathrm{Exp}(\eta_j - x^T \beta_j)}{1 + \mathrm{Exp}(\eta_j - \mathrm{x}^\mathrm{T} \beta_\mathrm{j})} \tag{4}$$

where:

β_j = a vector of unknown regression coefficients
x = a vector of observed explanatory variables
x^{T} = a transpose vector of observed explanatory variables
η_j = unknown threshold or intercept parameters, satisfying the condition $\eta_1 \le \eta_2 \; \ldots \le \eta_k$
j = comparison groups; j = 1, 2,…, k-1 (in this study j = 1,2)

y_j = outcome in comparison group j
Y = ordinal response variable with k outcomes
k = the number of outcome levels (categories) of injury severity (in this study k = 1, 2, 3)
$P(Y \leq y_j)$ = cumulative probability of the event $(Y \leq y_j)$

Some of the commonly known models are in fact special cases of the generalized ordered logit model presented by Equation 4. When k = 2, the model becomes the logistic regression model; when k > 2, the model becomes equivalent to a series of binary logistics regressions where categories of the dependent variable are combined; and if the β's are the same for all values of j, it becomes the traditional ordered logit model. From Equation 4, the probabilities that Y (injury severity) will take on each of the values 1, 2, or 3 (i.e., the individual outcome groups) can be written using cumulative probability distribution as depicted in Equations 5 through 7:

$$P(Y = y_1 \mid x) = F(\eta_1 - x^T \beta_1) \tag{5}$$

$$P(Y = y_2 \mid x) = F(\eta_2 - x^T \beta_2) - F(\eta_1 - x^T \beta_1) \tag{6}$$

$$P(Y = y_3 \mid x) = 1 - F(\eta_2 - x^T \beta_2) \tag{7}$$

The weakness of the ordered logit model is that the regression coefficient, β, does not depend on the outcome comparison group, j, because the model restricts it to be the same value for all outcome levels (which means all equations of outcome levels have the same slope, i.e., parallel lines) regardless of the possibility of variations [27]. For the generalized ordered logit model, some of the regression coefficients, β_j, may be the same for some outcome levels but it allows others to vary if they violate the parallel line assumption. According to Williams [27] the parallel-lines assumption is more often violated as it does not hold up most of the time.

The results from this procedure are interpreted as explained here. For a variable of three outcome levels (as is in this paper), the three outcome levels are grouped into two comparison groups. As a result, two sets of outcome groups for each model are developed. Since k = 3, then for j = 1, outcome level 1 is compared with outcome levels 2 and 3; and for j = 2, outcome levels 1 and 2 are compared with outcome level 3. A positive regression coefficient estimate indicates that higher values on the predictor variable increase the likelihood of an injury being in the higher injury severity levels than the current one. Similarly, a negative coefficient estimate indicates that higher values on the predictor variable increase the likelihood of an injury being in the current or lower injury severity levels, meaning that it reduces the likelihood of being in higher injury severity levels. Alternatively, a negative coefficient can simply be interpreted as increasing the likelihood of being in the current or lower injury groups. The results from the generalized ordered logit model can be interpreted as explained in the following example. For instance, let's assume that a predictor variable "alcohol-related" is coded with two outcome levels (coded as 0 for No and 1 for Yes). If the results show that this predictor variable "alcohol-related" has a positive regression coefficient in the outcome group j = 2 (comparing no-injury & possible/non-incapacitating injuries against incapacitating/fatal injuries), it means that a person injured in a traffic crash where alcohol use is involved is likely to sustain higher injury levels (incapacitating and fatal injuries) than no-injury or possible/invisible injuries when compared to no alcohol-related crashes.

3. Results

A total of fourteen variables were selected for exploratory analysis to investigate characteristics of predictor variables of ROR traffic injury severity and screen out the most promising ones for use in the next step. A decision tree procedure in JMP software version 10 was used in developing the classification tree modeling. The generalized ordinal logit regression model which uses maximum likelihood estimation (MLE) method was applied to estimate statistically the effects of these variables in contributing to the occurrence of run-of-road traffic-related injury severity levels. The gologit2 procedure in the STATA software release 10 was used in this estimate. Predictor variables were tested at a 95% confidence level. It is noteworthy to mention that the Brant test for parallel lines assumption was performed and it was statistically confirmed that the assumption was violated and hence a justification of using the generalized ordered logit model as opposed to a simpler ordered logit model. In addition, a correlation analysis between the independent variables was performed and none of the variables showed significant correlations.

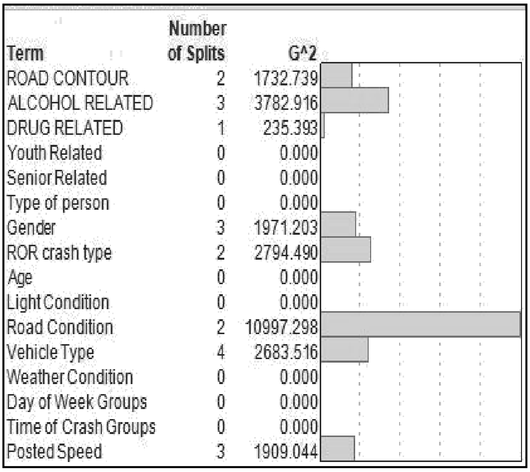
3.1. Results of decision tree modeling for predictor parameter screening

The main task for the decision tree modeling was to investigate the complex relationships between the injury severity and the fourteen predictor variables entered into the model by utilizing the decision tree's capability of identifying such relationships regardless of the form, i.e., either linear or nonlinear. The final product from this decision tree analysis was to identify the most significant predictor variables, which should be used in the regression modeling of predicting the injury severity of run-off-road crashes.

The dataset inputted in the JMP program consisted of 384,505 observations of run-off-road crashes, out of which 345,795 (90%) were randomly selected (automatically by the program) to form a training sample set and 38,710 observations (10%) were set aside as the validation sample set. The validation sample set is actually used to test the validity of the model developed using the training sample set. In the training sample, there were 236,822 no injury cases, 90,697 possible and non-incapacitating injury cases, and 18,276 incapacitating and fatal injury cases. The JMP program was first set to build a full tree where a total of 116 splits where observed. By assessing the R^2 plot it was observed that the value of R^2 remained about the same after 20 splits for both training and validation datasets. Therefore, the best tree size was found to be attained after about 20 splits and node splitting has to be stopped here (usually called tree pruning). By stopping the node splitting correctly, is one of the ways that help in protecting the model against both underfitting and overfitting.

The column contributions report from the JMP output as shown in Figure 2 suggests that there are eight independent variables which explain a large amount of the variation in the response variable in terms of the G^2 statistic values. These are the predictor variables that contributed in fitting the model. The eight predictor variables identified under this analysis include road condition, alcohol-related, ROR crash type, vehicle type, gender, posted speed, road contour, and drug-related. Based on this analysis, these are the predictor variables that were passed to the generalized ordered logit model to determine their effects on levels of injury severity of ROR crashes.

The training and evaluation models were evaluated by using the ROC curves and the results are summarized in Table 2. The ROC curve results indicate moderate ability to predict crash severity, better than random guess but not very accurate. However, the model does provide guidance on potential significant factors affecting injury severity. Both training and validation models predict better the injury group level 3 (i.e., incapacitating and fatal injuries), explaining the variations in the response variable by almost 71%.



Term	Number of Splits	G^2
ROAD CONTOUR	2	1732.739
ALCOHOL RELATED	3	3782.916
DRUG RELATED	1	235.393
Youth Related	0	0.000
Senior Related	0	0.000
Type of person	0	0.000
Gender	3	1971.203
ROR crash type	2	2794.490
Age	0	0.000
Light Condition	0	0.000
Road Condition	2	10997.298
Vehicle Type	4	2683.516
Weather Condition	0	0.000
Day of Week Groups	0	0.000
Time of Crash Groups	0	0.000
Posted Speed	3	1909.044

Fig. 2 - The contribution report for selecting significant predictor variables

Tab. 2 - ROC curves results for training and validation sample datasets

Sample	Injury Severity	Area Under ROC Curve
Training	Incapacitating/fatal (injury level 3)	0.7054
	Possible/non-incapacitating (injury level 2)	0.6083
	No injury (injury level = 1)	0.6410
Validation	Incapacitating/fatal (injury level 3)	0.7091
	Possible/non-incapacitating (injury level 2)	0.6095
	No injury (injury level = 1)	0.6418

3.2. Results of generalized ordered logit model for predictor parameter estimation

The eight predictor factors that were identified by the decision tree procedure were used in the generalized ordinal logit regression, which uses maximum likelihood estimation (MLE) technique. The gologit2 procedure in the STATA software release 10 was used to produce the predictor parameter estimates. All predictor variables tested in this model were found to be significant in predicting the injury levels. The results from STATA output are presented in Table 3. The goodness-of-fit statistics show that it a good model with a Pseudo R^2 value greater than zero justifying that the model has improved the likelihood when compared with base (intercept only) models. Road contour categories appear in both panels (with positive parameter estimates) in Table 3 indicating that these geometric features increase the likelihood of injuries of all levels. However, the effect of straight grade segments to ROR injuries is modest (due to small coefficient values). The curve level segments are the most hazardous segments, besides having the largest coefficient values, they increase from panel one (j = 1) to panel 2 (j = 2), meaning that these features tend to increase higher levels of injury severity, i.e. incapacitating and fatal injuries. Alcohol-related crashes (coded as 1) have a very strong effect on injury severity of crashes. This parameter has positive coefficients in both estimate panels with the larger coefficient in the second panel indicating that alcohol related crashes increase more the likelihood of sustaining incapacitating and fatal injuries.

Drug-related crashes (coded as 1) have also positive coefficients in both panels with increasing values indicating that drug use is significant in increasing the likelihood of injuries especially sustaining higher levels of injuries such as incapacitating and fatal injuries. Drug-related predictor parameter shows the same trends as those shown by the alcohol-related parameter to ROR traffic crashes.

Gender is significant with positive parameter estimates in both the first and second panels (but with a higher coefficient in the first panel) indicating that females are more likely to sustain injuries compared to males but they sustain more of possible/non-incapacitating than incapacitating/fatal injuries. Run-off-road crash type was modeled as a four-level category class variable. The “overturn/rollover” category (Type 1) was used as the reference level. The results in Table 3 show that run-off-road right (Type 2), run-off-road left/cross median/centerline (Type 3), and crash with fixed object (Type 4) all are found to increase the likelihood of no-injury crashes because all of them have negative coefficients in both panels. Likewise, vehicles involved in overturn/rollover crashes are the ones that increase the likelihood of injuries. In terms of road condition, dry roadway condition (code = 1) is the only one among the three road condition categories that were considered in this study, which increases the likelihood of injury crashes. For the other two conditions, i.e., wet/standing or running water (code = 2) and snow/ice/sand/mud/etc. (code = 3), they have negative coefficients in both panels indicating that they simply increase no injury crashes. Vehicle type (Type 2) variable is significant in both panels with positive coefficients and the larger value in the second panel. This indicates that vehicle type 2 (buses and trucks) increases the likelihood of sustaining all kinds of injuries especially incapacitating and fatal injuries. But motorcycles and motorized bicycles tend to increase the possibility of no injury ROR crashes. However, a very interesting result is for emergency vehicles (Type 4), with large positive coefficients in both panels indicating that ROR crashes involving emergency vehicles increase the likelihood of sustaining all kinds of injuries almost equally, ranging from possible/non-incapacitating injuries to incapacitating/fatal injuries. Due to emergency nature in which these vehicles are operated, typically driven at high speeds, this may be one of the reasons why their crashes tend to involve severe injuries.

The parameter estimate for posted speed limit is found to be significant for posted speed limit of 40 mi/h and higher (code = 2 for 40-50 mi/h and code = 3 for 55+ mi/h) in both panels with positive coefficients (also increasing from code 2 to code 3) indicating that higher posted speed limits are more likely to increase the likelihood of sustaining injuries related to ROR crashes. Especially for the 55-70 mph posted speed limits group having a higher parameter coefficient in panel 2 than in panel 1, indicates that ROR crashes in these high speed limits tend to increase incapacitating and fatal injuries.

Comparing results from the current study with other previous studies of run-off-road traffic crashes, we observe fairly notable agreements of significant contributing factors and some few differences as well. However, sometimes it is important to note the differences in the objectives of the studies being compared because these also affect the findings and methods used in the analyses. For example, some researchers may be studying significant factors affecting the injury severity of drivers only or all occupants; others may be studying only fatal crashes versus all crashes, while others may be investigating factors affecting the occurrences of crash frequencies. In addition, the statistical experimental study unit may be different as well due to the setup and objective of the study, for example, the study unit can be a vehicle occupant, a driver, a vehicle or the crash incident.

Tab. 3 - Parameter estimate results

Variable	Coeff.	Std. Err	z	P>|z|	[95% Conf. Interval]
PANEL 1: J = 1 (no-injury vs. possible, non-incapacitating, incapacitating and fatal injuries)					
Road Contour: Straight Grade	0.073	0.010	7.49	0.000	0.054 - 0.093
Road Contour: Curve Level	0.241	0.011	22.52	0.000	0.220 - 0.262
Road Contour: Curve Grade	0.198	0.010	19.53	0.000	0.178 - 0.218
Alcohol Related: Yes	0.544	0.011	49.33	0.000	0.522 - 0.565
Drug Related: Yes	0.573	0.021	27.37	0.000	0.532 - 0.614
Gender: Female	0.328	0.007	44.27	0.000	0.314 - 0.343
ROR Crash Type: Right	-1.702	0.033	-52.27	0.000	-1.766 - -1.638
ROR Crash Type: Left/Cross Median/Centerline	-1.676	0.033	-51.18	0.000	-1.74 - -1.612
ROR Crash Type:Fixed Objects	-1.948	0.034	-57.68	0.000	-2.014 - -1.881
Road Condition: Wet/Water	-0.401	0.009	-45.14	0.000	-0.418 - -0.384
Road Condition: Snow/Ice	-0.844	0.010	-87.16	0.000	-0.863 - -0.825
Vehicle Type: Trucks/Buses	0.131	0.014	9.49	0.000	0.104 - 0.158
Vehicle Type: Motorcycles	-0.634	0.153	-4.15	0.000	-0.934 - -0.335
Vehicle Type: Emergency Vehicles	1.036	0.109	9.47	0.000	0.821 - 1.250
Posted Speed: 40-50 mph	0.230	0.012	19.56	0.000	0.207 - 0.253
Posted Speed: 55-70 mph	0.370	0.009	42.03	0.000	0.353 - 0.387
Constant	0.633	0.033	19.06	0.000	0.568 - 0.698
PANEL 2: J = 2 (no-injury, possible, non-incapacitating vs. incapacitating and fatal injuries					
Road Contour: Straight Grade	0.072	0.021	3.48	0.000	0.031 - 0.113
Road Contour: Curve Level	0.283	0.020	13.81	0.000	0.242 - 0.323
Road Contour: Curve Grade	0.190	0.020	9.41	0.000	0.151 - 0.23
Alcohol Related: Yes	0.966	0.018	53.52	0.000	0.931 - 1.001
Drug Related: Yes	0.715	0.029	24.45	0.000	0.658 - 0.773
Gender: Female	0.145	0.016	9.23	0.000	0.114 - 0.175
ROR Crash Type: Right	-0.720	0.041	-17.54	0.000	-0.801 - -0.64
ROR Crash Type: Left/Cross Median/Centerline	-0.580	0.042	-13.94	0.000	-0.662 - -0.499
ROR Crash Type: Fixed Objects	-1.009	0.047	-21.32	0.000	-1.102 - -0.916
Road Condition: Wet/Water	-0.582	0.020	-29.68	0.000	-0.620 - -0.544
Road Condition: Snow/Ice	-1.169	0.026	-45.28	0.000	-1.219 - -1.118
Vehicle Type: Trucks/Buses	0.863	0.022	38.70	0.000	0.82 - 0.907
Vehicle Type: Motorcycles	-0.153	0.310	-0.49	0.622	-0.76 - 0.455
Vehicle Type: Emergency Vehicles	1.048	0.141	7.45	0.000	0.772 - 1.324
Posted Speed: 40-50 mph	0.190	0.025	7.69	0.000	0.141 - 0.238
Posted Speed: 55-70 mph	0.427	0.018	23.28	0.000	0.391 - 0.463
Constant	-2.644	0.044	-60.29	0.000	-2.73 - -2.558
Model goodness-of-fit statistics					
Number of Observations 384,505 Likelihood Ratio (LR) Chi-Square 20850.65 Prob > Chi-Square 0.0000			Log likelihood -280130.4 Pseudo R^2 0.0480		

Significant factors that increase the likelihood of run-off-road injury severity, which have been identified in the current study and are in agreement with some previous studies include curves [4-6, 7, 9-11], alcohol involvement [8-9, 11], higher posted speed limits [9, 11]. Weather condition was not significant in our study, but it was significant in some previous studies [9, 11], adverse roadway condition such as slippery roads, snow and ice on the roadway surfaces, etc., were significant factors [9, 11], but they were not significant in the current study, while it is dry roadway condition, which was found to be significant factor in increasing the likelihood of injuries especially severe injuries (incapacitating and fatal injuries). In the same sense, a study by Lee et al. [29] found that wet surface conditions decreased the severity of crashes for all types of crashes in the state of Florida. The run-off-road crash type was considered as a variable in the current study and was found to be significant but it was not included in most of the previous studies, which were reviewed. An interesting finding from the current study is that only overturning/rollover as an ROR type of crash is the only one that increases the chances of injuries while all others mostly end up into property damage only crashes (i.e., no injury to people involved in the crashes).

Another interesting finding that was not studied previously is the inclusion of emergency vehicles as a separate vehicle type in the analysis. The current study found that when emergency vehicles (i.e., police cruiser cars, fire trucks, and ambulances) get involved into run-of-the-road crashes, they mainly cause higher levels of injury severity. This may make sense due to the high speeds that these kinds of vehicles are always involved in when they are called for emergency incidents.

The time of crash, which did not even pass the decision tree analysis, and therefore it was not a significant factor in the current study, however it happened to be a significant factor in some previous studies of ROR crashes [e.g., 10-11].

Although weather condition was not significant in the current study, but it may be correlated with roadway condition, which was determined to be significant factor in the current study but the effect of roadway condition in this study was found to be the opposite of what some previous studies reported.

4. Conclusion

Run-off-road crashes have become a major concern in the state of Ohio as they continue to cause fatalities and major injuries to motorists.

All the predictor variables that the decision tree modeling found significant were also confirmed to be significant by the generalized ordered logit model.

The crash injury severity results analyzed in this paper reveal several issues that lead into severe injuries. Significant predictors of injury severity include road condition, run-off-road crash type, posted speed limit, vehicle type, gender, alcohol-related, road contour and drug-related.

The results of this study determined that the most severe ROR-related injuries occurred on roads when their surfaces were dry. The presumable reason is that drivers tend to over speed or drink and drive when the weather is good and road surfaces are in good/dry condition and hence when they get involved in-run-off-road crashes the odds of sustaining severe injuries increase. Likewise, drivers tend to be more cautious when driving on wet or snow/ice covered road surfaces and when they run-off-road, they are already driving at lower speeds and the collision impacts end up causing no-injury or minor injuries.

Roadway curves and grades are features found to increase the likelihood of injuries of all levels while grades tend to have moderate influence. The horizontal curves on level road segments were determined to be the most hazardous locations. Alcohol-and drugs-related crashes have very strong effects on injury severity of crashes as they tend increase more the likelihood of sustaining incapacitating and fatal injuries.

Like many studies, females were found to be more likely to sustain injuries compared to males but they sustain more of possible/non-incapacitating than incapacitating/fatal injuries. Also, in terms of run-off-road crash types, overturn/rollover crashes were found to increase the chances of injuries while run-off-road left/cross median/centerline and crash with fixed objects all increased the likelihood of no-injury crashes.

For roadway condition, dry roadway condition was the only one that increased the likelihood of injury crashes while wet/standing or running water and snow/ice/sand/mud/etc. simply increased no injury crashes.

In terms of vehicle type, buses, trucks and emergency vehicles increased the likelihood of sustaining all kinds of injuries especially incapacitating and fatal injuries. A roadway with posted speed limits of 40 mi/h or higher tends to increase the likelihood of sustaining injuries related to ROR crashes.

ODOT should improve curve delineation, check and improve friction treatments in curves to reduce run-off-road crashes on horizontal curves. In addition, ODOT can provide edge line and shoulder rumble strips where they are not installed and provide barriers to shield fixed objects, trees, shrubs, and slopes especially on locations with a history of high frequency of such crashes across the state. In addition, where possible, median barriers can be provided on locations determined to be susceptible with cross median/centerline crashes. ODOT needs to up their current traffic campaign of “every move you make keep it safe” by making it more visible and reaching more Ohioans. Major educational campaigns should warn drivers to be more careful when approaching horizontal curves, when driving on high speed roads, stressing on driving at a speed appropriate for the prevailing conditions and without forgetting drunken driving campaigns. Targeted enforcement campaign to deter alcohol/drug abuse and driving should also be part of the campaign.

References

1. Federal Highway Administration (FHWA). (2016). *Roadway Departure Safety. United States*. Department of Transportation, Washington, D.C.
   Website: http://safety.fhwa.dot.gov/roadway_dept/. Accessed January 10, 2016.
2. Ohio Department of Transportation (ODOT). (2015). *Roadway Departure Crashes: Stay Alive, Stay on the Road*.
   Website:http://www.dot.state.oh.us/Divisions/Planning/ProgramManagement/HighwaySafety/SHSP/Pages/EmphasisAreas.aspx. Accessed December 27, 2015
3. National Highway Traffic Safety Administration (NHTSA). (2013). *Traffic Safety Facts Ohio 2007-2011*. NHTSA, United States Department of Transportation, Washington, D.C., 2013.
4. Hall, J.W., Zador, P. (1985). Survey of Single Vehicle Fatal Rollover Crash Sites in New Mexico. *Transportation Research Record: Journal of the Transportation Research Board,* No. 819, pp 1-8.
5. Zegeer, C., Reinfurt, D.W., Hummer, J., Herf, L., Hunter, W. (1988). Accident Effects of Sideslope and Other Roadside Features on Two-Lane Roads. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1195, pp. 20-23
6. Viner, J.G. (1995). Rollovers on Sideslopes and Ditches. *Accident Analysis and Prevention*, 27, pp. 483-491.
7. Neuman T.R., Pfefer R., Slack K.L., Hardy K.K., Council F., McGee H., Prother L., Eccles K. (2003). *Guidance for Implementation of the AASHTO Strategic Highway Safety Plan*, Vol. 6: A Guide for

Addressing Run-Off-Road Collisions, NCHRP Report 500, Transportation Research Board, Washington, D.C.
8. Spainhour, L.K., Mishra, A. (2008). Analysis of Fatal Run-off-the-road Crashes Involving Overcorrection. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2069, pp. 1-8.
9. Liu, C., Subramanian, R. (2009). *Factors Related to Fatal Single-Vehicle Run-off-road Crashes*. Pub. DOT HS 811 232, National Highway Traffic Safety Administration, United States Department of Transportation, Washington, D.C.
10. Zhu, H., Dixon, K.K., Washington, S., Jared, D.M. (2010). Single-vehicle Fatal Crash Prediction for Two-lane Rural Highways in the Southeastern United States. *Proceedings of the 89th TRB Annual Meeting*, Washington, D.C., January 10-14.
11. Roy, U., Dissanayake, S. (2011). Comparison of Factors Associated with Run-off-road and Non-run-off Road Crashes in Kansas. *Journal of the Transportation Research Forum*, 50, pp. 69-86.
12. Peng, Y., Boyle, L.N. (2012). Commercial Driver Factors in Run-off-road Crashes. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2281, pp. 128-132.
13. Lavery, R. (2012). An Animated Guide: Regression Trees in JMP and SAS Enterprise Minor. *Proceedings of the 25th Annual Conference of Northeast SAS Users Group (NESUG)*, Baltimore, MD, November 11-14.
14. Yan, X. and Radwan, E. (2006). Analyses of Rear-End Crashes Based on Classification Tree Models. *Traffic Injury Prevention*. l. 7, pp. 276-282.
15. Montella, A., Aria, M., D´Ambrosio, A., Mauriello, F. (2011). Data Mining Techniques for Exploratory Analysis of Pedestrian Crashes. *Transportation Research Record: Journal of Transportation Research Board*, No. 2237, pp. 107-116.
16. Montella, A., Aria, M., D´Ambrosio, A., Mauriello, F. (2012). Analysis of Powered Two-wheeler Crashes in Italy by Classification Tree and Rules Discovery. *Accident Analysis and Prevention,* 49, pp. 58-72.
17. Kashani, A., Mohaymany, A., Ranjbari, A. (2012). Analysis of Factors Associated with Traffic Injury Severity on Rural Roads in Iran. *Journal of Injury and Violence Research*, 4, pp. 36-41.
18. Kashani, A., Rabieyan, R., Besharati. (2014). A Data Mining Approach to Investigate the Factors Influencing the Crash Severity of Motorcycle Pillion Passengers. *Journal of Safety Research*, 51, pp. 93-98.
19. Griselda, L., Juan, O., Joaquín, A. (2012), Using Decision Trees to Extract Decision Rules from Police Reports on Road Accidents." *Procedia – Social and Behavioral Sciences,* 53, pp. 106-114.
20. Wang, X., Abdel-Aty, M. (2008). Analysis of Left-Turn Severity by Conflicting Pattern Using Partial Proportional Odds Models. *Accident Analysis and Prevention,* 40, pp. 1674-1682.
21. Wang, Z., Chen, H., Lu, J.J. (2009). Exploring Impacts of Factors Contributing to Injury Severity at Freeway Diverge Areas. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2102, pp. 43-52.
22. Mergia, W.Y., Eustace, D., Chimba, D., Qumsiyeh, M. (2013). Exploring Factors Contributing to Injury Severity at Freeway Merging and Diverging Locations in Ohio. *Accident Analysis and Prevention*, 55, pp. 202-210.
23. Michalaki, P., Quddus, M. A., Pitfield, D., Huetson, A. (2015). Exploring the factors affecting motorway accident severity in England using the generalized ordered logistic regression model. *Journal of safety research*, 55, 89-97.
24. SAS Institute Inc. (2015). *JMP Documentation: Specialized Models*. Cary, NC.
25. Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* 2nd Edition, John Wiley and Sons, Inc., New York, US.
26. SAS Institute Inc. *SAS Help and Documentation*. Version 9.1, Cary, NC, 2003.
27. Williams, R. (2006). Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Response Variables. *The STATA Journal,* 6, pp. 58-82.

28. Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A. (2011). The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, 43, pp. 1666-1676.
29. Lee, J., Nam, B., Abdel-Aty, M. (2015). Effects of pavement surface conditions on traffic crash severity. *Journal of Transportation Engineering*, 141(10), 04015020.