An Application of Decision Tree Models to Examine Motor Vehicle Crash Severity Outcomes

by Jill M. Bernard Bracy

Classification and Regression Tree (CART) and chi-square automatic interaction detection (CHAID) decision tree models are estimated and compared to examine the effect of driver characteristics and behaviors, temporal factors, weather conditions, and road characteristics on motor vehicle crash severity levels using Missouri crash data from 2002 to 2012. The CHAID model is found to significantly better discriminate among severity outcomes, and results suggest that the presence of alcohol, speeding, and failing to yield lead to many fatalities each year and likely have interactive effects. Decision rules are used to identify changes in driving policies expected to reduce severity outcomes.

INTRODUCTION

Motor vehicle crashes are a leading cause of death in the United States (Centers for Disease Control and Prevention 2015), cost Americans approximately $277 billion annually in lost wages, rehabilitation, and medical care (Blincoe et al. 2010), and render serious psychological burdens such as grief, stress, depression, guilt, and travel anxiety for victims and their families (Mayou et al. 1993). Because of these devastating effects, many researchers have attempted to better understand the factors that affect motor vehicle crash severity, yet relatively little research has employed decision trees as the methodological approach used to gain greater insight into relationships among variables in the crash data.

Accordingly, decision tree models are estimated to gain greater insight into interactive effects in motor vehicle crash data and to examine how crash severity differs with numerous possible explanatory variables. Classification and Regression Tree (CART) and chi-square automatic interaction detection (CHAID) decision tree models are estimated using Missouri crash data from 2002 to 2012, and factors examined include driver characteristics and behaviors, temporal factors, weather conditions, road characteristics, and crash severity levels. The explanatory power of the CART and CHAID models are compared using the area under the receiver operating characteristic (ROC) curve on a hold-out sample. To provide a context for understanding the relative reduction in overall risks, outcomes are examined to determine annual upper and lower bounds on the changes in the number of drivers involved in fatal, injury, or property damage only crashes if selected contributory circumstances are eliminated. Decision rules are then used to identify changes in driving policies likely to reduce severity outcomes.

LITERATURE REVIEW

Relatively little research has employed such an approach, yet Savolainen et al. (2011) remarked that decision tree models are an effective data mining technique. Abdel-Aty and Keller (2005) claimed that tree-based regression improves the understanding of the importance of specific factors on individual levels of severity. Oh (2006) concluded that variables associated with injury severity levels may not be the cause of severity, and additional research in this area is necessary. Abay (2013) called for a more encompassing and alternative model specification for injury severity data analysis.

Sohn and Shin (2001) developed decision tree, neural network and logistic regression models to assess the factors that affect traffic crash severity in Korea. The classification tree identified the six factors used in the neural network and logistic regression models (accident mode, road width, car shape, speed before accident, violent driving, and protective device). Model results revealed protective device (i.e., safety belt use or helmet improperly worn) as the most influential variable for classification of crash severity.

Bayam et al. (2005) examined existing literature that used CART models to predict the occurrence of a crash or non-crash, given driver, roadway, vehicle, and other variables. The authors suggested the small sample size to be the cause of the poor predictive power in the test data; and, as a result, findings were not robust enough to be generalizable. However, the authors claimed that a larger data set "could be quite useful for this type of application" (p. 623), and concluded that data mining should be used to discover unknown relationships for crashes for senior and teenage drivers.

Abdel-Aty and Keller (2005) hypothesized that crash injury levels were impacted by crash and intersection specific characteristics. Expanding upon Abel-Aty (2003), the authors developed ordered probit models to assess 33,592 crashes that occurred in 832 intersections in Florida in 2000 and 2001. The study presented a hierarchical tree-based regression model to estimate the expected crash frequency for each crash severity level. Results indicated that the most significant factors for no-injury crashes, possible injury, non-incapacitating injury and incapacitating injury are traffic volume of the major road, the number of lanes on the minor road, the number of exclusive right-turn lanes, and the average daily traffic on the minor road, respectively. The authors concluded that the models should be developed for each level of severity as opposed to predicting the overall severity level, and the tree-based regression improves the understanding of the importance of specific factors on individual levels of severity.

Yan and Radwan (2006) used the classification tree approach to investigate factors of rear-end crashes that occur at signalized intersections. The 2001 Florida crash data used were restricted to two-vehicle, rear-end collisions, and the striking driver was the at-fault party. Model results suggested that drivers under the age of 21 and over 75 have the greatest risk of rear-end collisions. As a result, the authors recommended speed limit reduction to 40 mph at signalized intersections, enforcement for reducing alcohol intoxicated drivers, and additional education for drivers under the age of 21 for reducing rear-end crashes at signalized intersections, and concluded that the classification trees are an appropriate approach in investigating crash propensity.

Chang and Wang (2006) developed a CART model to examine the impact of gender, age, sobriety condition, crash location, vehicle type, contributing circumstance, and collision type on crash severity using 26,831 observations from crashes occurring in 2001 in Taipei, Taiwan. Model results illustrated an initial split based on vehicle type, and suggested that bicyclists, motorcyclists, and pedestrians have the highest risk, and contributing circumstance, collision type, and driver action are important in determining crash severity. The authors concluded by calling for future work in comparing CART model results with traditional models such as ordered probit and logistic regression models.

Abellán et al. (2013) developed decision trees to analyze traffic crash severity for motorcyclists in Granada, Spain. The authors extracted single-vehicle crash observations that occurred on two-lane rural highways from 2003 to 2009 for a total of 1,801 observations, and identified the following rules as having a high risk of a severe injury outcome for motorcyclists: when only one occupant was involved in a single vehicle crash, when at-fault motorcyclists were involved in a run-off-road crash in favorable weather conditions, when male motorcyclists were involved in a run-off-road crash as the result of driver characteristics, and when male motorcyclists were involved in a run-off-road crash in favorable weather. Findings inferred these additional rules to be a high risk of killed/seriously injured crashes on two-lane rural highways when no safety barriers are in place: motorcyclists with no-restrained site distance, crashes in the evening in good weather conditions with no lighting, and crashes with pedestrians during favorable weather when the driver is male.

The authors concluded that the method allowed for a high number of rules to be identified, and the method could be extrapolated for studies on other datasets.

Eustace et al. (2014) employed decision tree models in conjunction with generalized ordered logit models to examine factors that contribute to injury severity for run-off-road crashes in Ohio. Important interactions identified by the decision tree model included: females on higher posted speed limits have higher risk of injury; males with drug involvement and a higher posted speed limit have a higher risk of injury; alcohol use on a road with speed limits over 40 mph have higher risk of injury; and male drivers in crashes on wet road surfaces have higher risk of injury. The authors concluded that not only does the decision tree model analysis identify significant factors of injury severity, it also allows for the detection of multi-level interactions.

Recently, Kahn et al. (2015) used decision trees in conjunction with ordinal discrete choice models to analyze crash severity of cross-median crashes occurring in Wisconsin from 2001 to 2007. The authors claimed that the tree models revealed further information about the crash severity outcome and offered advantages regarding variable redundancy issues.

Drawing upon prior crash severity literature (presented in Table 1), variables suggested to affect crash severity include age, gender, number of occupants, speed limit, light conditions, weather conditions, road conditions and characteristics, and contributing circumstances.

Table 1: Variables Suggested by Reviewed Literature to Affect Crash Severity

Variables	Reviewed Literature
Age	Kuhnert et al. (2000); Abdelwahab and Abdel-Aty (2001); Abdelwahab and Abdel-Aty (2002); Bédard et al. (2002); Khattak et al. (2002); Abdel-Aty (2003); Khattak and Rocha (2003); Delen et al. (2006); Lu et al. (2006); Schneider et al. (2009); Haleem and Abdel-Aty (2010); Rifatt et al. (2011); Yasmin and Eluru (2013)
Gender	Kuhnert et al. (2000); Abdelwahab and Abdel-Aty (2001); Abdel-Aty and Abdelwahab (2004a); Abdel-Aty and Abdelwahab (2004b); Ulfarsson and Mannering (2004); Delen et al. (2006); Islam and Mannering (2006); Savolainen and Ghosh (2008); Schneider et al. (2009); Malyshkina and Mannering (2010); Schneider and Salovainen (2011); Eustace et al. (2014)
Number of Occupants	Renski et al. (1999); Oh (2006)
Speed Limit	Renski et al. (1999); Khattak et al. (2002); Oh (2006); Gårder (2006); Malyshkina and Mannering (2010); Savolainen and Ghosh (2008); Haleem and Abdel-Aty (2010); Zhu and Srinivasan (2011); Yasmin and Eluru (2013)
Light Conditions	Klop and Khattak (1999); Rifatt and Tay (2009); Wang et al. (2009); Haleem and Abdel-Aty (2010); Khattak et al. (2002)
Weather Conditions	Khattak et al. (1998); Abdel-Aty (2003); Wang et al. (2009)
Road Conditions & Characteristics	Khattak et al. (1998); Krull et al. (2000); Lu et al. (2006); Rifatt and Tay (2009); Quddus et al. (2010); Zhu and Srinivasan (2011)
Contributing Circumstances	Chang and Wang (2006); Bernard and Sweeney (2015)

Many researchers have reported age as a significant factor in influencing crash severity (Delen et al. 2006; Kuhnert et al. 2000). Khattak and Rocha (2003) found that young drivers increase the risk of higher injury severity in single-vehicle crashes, while others have suggested that older drivers have higher risks of severe injury, given a crash occurence (Abdelwahab and Abdel-Aty 2002; Bédard et al. 2002; Abdel-Aty 2003; Schneider et al. 2009; Rifaat et al. 2011; Yasmin and Eluru 2013).

Chang and Wang (2006) found that contributing circumstances and driver actions are critical in determining crash severity. Studies reported that passenger presence increases the risk of injury (Savolainen and Ghosh 2008; Schneider et al. 2009), severity increases as the number of vehicle passengers increase (Renski et al. 1999; Oh 2006), and distracted drivers have a higher risk of greater severity (Zhu and Srinivasan 2011).

Many studies also reported that alcohol presence significantly increases the risk of severe injury (Khattak et al. 1998; Renski et al. 1999; Krull et al., 2000; Bédard et al. 2002; Khattak et al. 2002; Kockelman and Kweon 2002; Abdel-Aty 2003; Zajac and Ivan 2003; Donnell and Mason 2004; Delen et. al 2006; Rifaat and Tay 2009; Schneider et al. 2009; Wang et al. 2009; Moudon et al. 2011; Yasmin and Eluru 2013) and fatality (Islam and Mannering 2006; Rifaat et al. 2011). And driving at speeds too fast for conditions (Bédard et al. 2002; Rifaat and Tay 2009), speeding (Khattak et al. 1998; Khattak and Rocha 2003; Schneider et al. 2009), and higher speed limits (Renski et al. 1999; Khattak et al. 2002; Oh 2006; Gårder 2006; Malyshkina and Mannering 2010; Savolainen and Ghosh, 2008; Haleem and Abdel-Aty 2010; Zhu and Srinivasan 2011; Yasmin and Eluru, 2013) significantly increase the risk of severe injury. Importantly, results revealed that the interaction between higher speed limits and alcohol increase the risk of crash severity (Yan and Radwan 2006; Eustace et al. 2014).

Wang et al. (2009) found that favorable weather decreases crash severity and Abdel-Aty (2003) reported that adverse weather increases severity. Yet, Khattak et al. (1998) found adverse weather to significantly decrease the risk of severe injury for crashes; and Delen et al. (2006) indicated that weather conditions are not influential in crash severity. Lu et al. (2006) claimed that road condition has the greatest influence on crash severity; however, Jiang et al. (2013) concluded that improved road quality does not essentially reduce injury severity. Khattak et al. (1998), Rifaat and Tay (2009), and Quddus et al. (2010) reported that wet/slippery road surface decreases the risk of severe injury; yet, Krull et al. (2000) and Zhu and Srinivasan (2011) found that dry surfaces increase the risk of severity. Finally, studies reported that dark, unlit conditions increase crash severity (Klop and Khattak 1999; Rifaat and Tay 2009; Haleem and Abdel-Aty 2010), favorable lighting conditions decrease severity at freeway diverge areas (Wang et al. 2009), dusk (over dark) conditions reduce the risk of severe injury at unsignalized intersections (Haleem and Abdel-Aty 2010), and darkness increases the risk of greater injury severity for older drivers (Khattak et al. 2002).

To build upon prior research, two decision tree models are estimated and compared to examine the effect of and intricate relationships among suggested explanatory variables, to provide a context for understanding the relative reduction in overall risks associated with reducing the frequency of driver behaviors that contribute to the likelihood of different crash severity outcomes, and to provide better information for transportation policy that will enhance transportation safety efforts.

DATA

The Missouri State Highway Patrol Traffic Division collects and preserves crash report data, and codes and classifies the reports for entry into the Statewide Traffic Accident Records System (STARS) database (Missouri State Highway Patrol 2012). This study uses three relevant datasets from the STARS database: accident level data, vehicle level data and personal level data, and the years 2002 to 2012 are combined into a single dataset containing 3,902,742 observations. When considering motor vehicle drivers with a Missouri issued driver's license who contributed to a reported crash, cross tabulation results identify 1,264,905 observations in the dataset with the crash severity distributed as 0.6% fatal, 28.1% injury, and 71.3% property damage only. The explanatory variables included in this analysis are presented in Table 2.

Table 2: Explanatory Variables Included in the Analysis

Driver Characteristics	**Categorical Variable**
Age	Young (<21 years-old); Middle (≥21 and <55 years-old); Mature (≥ 55 years-old); Unknown
Gender	Male; Female; Unknown
Vehicle Occupants	**Numerical Variable**
Total Number of Occupants	1 to 149
Contributing Circumstances	**Binary Variable**
Alcohol	Present = 1; Not Present = 0
Animal(s) in Roadway	Present = 1; Not Present = 0
Distracted/Inattention	Present = 1; Not Present = 0
Drugs	Present = 1; Not Present = 0
Failed to Yield	Present = 1; Not Present = 0
Following Too Close	Present = 1; Not Present = 0
Improper Backing	Present = 1; Not Present = 0
Improper Lane Usage/Change	Present = 1; Not Present = 0
Improper Passing	Present = 1; Not Present = 0
Improper Turn	Present = 1; Not Present = 0
Improperly Stopped	Present = 1; Not Present = 0
Other[1]	Present = 1; Not Present = 0
Overcorrected	Present = 1; Not Present = 0
Physical Impairment	Present = 1; Not Present = 0
Speed - Exceeds Limit	Present = 1; Not Present = 0
Too Fast for Conditions	Present = 1; Not Present = 0
Vehicle Defects	Present = 1; Not Present = 0
Violation Stop Sign/Signal	Present = 1; Not Present = 0
Vision Obstructed	Present = 1; Not Present = 0
Wrong Side - Not Passing	Present = 1; Not Present = 0
Wrong Way (One Way)	Present = 1; Not Present = 0
Location	**Categorical Variable**
Crash Location	On Roadway; Off Roadway
Road Characteristics	**Categorical Variable**
Road Conditions	Other/Unknown; Wet; Snow; Ice: Dry
Road Alignment	Unknown; Curve; Straight
Road Profile	Unknown; Hill/Grade; Crest; Level
Road Surface	Unknown; Asphalt; Gravel; Brick/Dirt/Sand/Multi-Surface, Concrete
Speed Limit	15 or 20mph; 25 or 30mph; 35 or 40mph; 45 or 50mph; 55 or 60mph; 65 or 70mph; Unknown

(Table 2 continued)

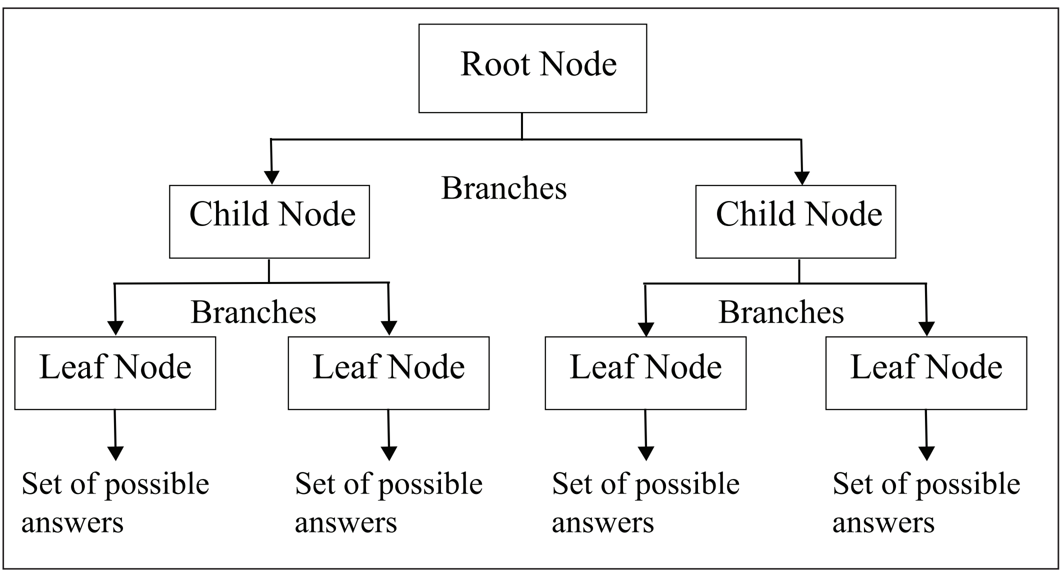
Environmental Factors	**Categorical Variable**
Weather Conditions	Cloudy; Rain; Snow; Sleet; Freezing Rain; Fog/Mist; Indeterminate; Clear
Light Conditions	Indeterminate; Dark-Streetlights On; Dark-Streetlights Off; Dark-No Streetlights; Daylight
Dependent Variable	**Categorical Variable**
Injury Severity	Fatal; Injury; Property Damage Only

Initial model runs suggested quasi-separation, which was resolved by combining the following variables with similar magnitudes into the *Other* variable: *Improper Signal*, *Improper Start from Park*, *Improperly Parked*, *Driver Fatigue/Asleep*, *Failed to Dim Lights*, *Failed to Use Lights*, *Improper Towing/Pushing*, *Improper Riding/ Clinging to the Vehicle Exterior*, *Failed to Secure Load/Improper Loading*, *Object/Obstruction in the Roadway*.

METHODOLOGY

Decision tree models may be used for classification of occurrences into pre-specified groups, prediction of values of a dependent variable based on values of independent variables, and data exploration in model building. The tree is built by applying decision rules sequentially that split a larger heterogeneous population into smaller more homogeneous subsets (termed nodes) based on the single, most predictive input factor (Eustace et al. 2014) as illustrated in Figure 1 (Bayam et al. 2005). Subset purity is measured and evaluated to determine the best split for the subset (Mingers 1989b), and factors deemed statistically homogenous, with respect to the target outcome, are combined (Trnka 2010). Splitting continues for each node until no more splits are possible or until pre-defined stopping parameters (e.g., maximum tree depth or minimum number of records in branch) are reached.

Figure 1: Structure of a Decision Tree (Bayam et al. 2005)



Decision trees have several advantages over other models including the following: nonlinear relationships between variables do not affect performance; the data partitioning yields insights into input/output relationships; each path of the decision tree contains an estimated risk factor; missing values are accommodated automatically; and the output is simple to understand and interpret (Bernard Bracy 2017). However, overfitting of the model can occur if the learning algorithm fits data that are irrelevant, resulting in a model that may not be generalizable (Bayam et al. 2005). To avoid overfitting and improve generalization, pruning may be used to remove lower-level splits that do not significantly contribute to the generalized accuracy of the model (Mingers 1989a).

Decision tree algorithms, including CART and CHAID, build and prune decision trees in different ways. CART creates binary trees by splitting records at each node, and builds larger trees that are pruned back to mitigate overfitting of the model. CHAID creates wider, non-binary trees (often with many terminal nodes connected to a single branch) and automatically prunes the decision tree to avoid overfitting (Bayam et al. 2005).

Both CART and CHAID trees are estimated, the discriminatory performance of each algorithm is evaluated, and the model with the greatest discriminatory power is identified. The models' performances are compared by calculating and evaluating the classification accuracy and the area under the ROC curve (AUC) values for each model. The dataset is randomly partitioned into a training set (75%) to estimate the model and a testing set (25%) to assess model classification accuracy, and partitioning was completed prior to estimating both models so that identical observations are used. Classification accuracy for each category of the outcome variable is determined by dividing the number of predicted outcomes by the number of observed outcomes (as illustrated in Tables 3 and 4). IBM SPSS 22.0 and IBM SPSS Modeler 15.0 are used to estimate the decision tree models.

The CART methodology employs the algorithm proposed by Breiman et al. (1984), where nodal splitting criteria are set to a minimum value of 100 records in a parent branch and a minimum of 50 records in a child branch as the stopping criteria. The Gini coefficient is used as the impurity measure for the categorical targets, the maximum tree depth is set to 15 branches, and the tree is pruned by merging leaves on the same branch using a value of one as the maximum difference in risk in standard errors. The resulting CART decision tree model finds 23 variables significant, includes 948,679 observations in the training set and 316,784 in the testing set, and results in a classification accuracy of 72.32% and 72.30% for the training set and the testing set, respectively.

The CART classification accuracy for the training set is determined by dividing the number of severity outcomes predicted by the model by the total observed severity outcomes in the testing set. As illustrated in Table 3, the model correctly classified 12.6% of injury outcomes, determined by dividing 33,743 by 268,465 (0+33,743+234,722); and correctly classified 96.8% of property damage outcomes, calculated by dividing 652,654 by 674,191 (0+21,837+652,354). The overall percent correct for the training set, 72.32%, is found by summing the total occurrences of the predicted fatal and observed fatal outcomes (0), the predicted injury and observed injury outcomes (33,743), and the predicted property damage and observed property damage outcomes (652,354), and then dividing by the number of observations in the testing set (948,679).

Table 3: CART Classification Accuracy for the Training Set

	Predicted			
Observed	Fatal	Injury	Property Damage	Percent Correct
Fatal	0	1,760	4,263	0.0%
Injury	0	33,743	234,722	12.6%
Property Damage	0	21.837	652,354	96.8%
Overall Percentage	0.0%	6.0%	94.0%	72.32%

The CHAID methodology employs the algorithm proposed by Kass (1980), where nodal splitting criteria are set to a minimum value of 100 records in a parent branch and a minimum of 50 records in a child branch, and the maximum tree depth is set to 15 branches. The Pearson measure is used as the chi-square measure to test for independence for categorical targets, and the significance level for both splitting and merging is set to 0.05. The final estimated CHAID decision tree model suggests 30 variables are significant, includes 948,679 observations in the training set and 316,784 in the testing set, and results in a classification accuracy of 73.06% and 73.00% for the training set and the testing set, respectively.

The CHAID classification accuracy for the training set, illustrated in Table 4, is also determined by dividing the number of severity outcomes predicted by the model by the total observed severity outcomes in the testing set. The model correctly classified 23.6% of injury outcomes, determined by dividing 63,279 by 268,465 (0+63,279+205,186); and correctly classified 93.4% of property damage outcomes, calculated by dividing 629,793 by 674,191 (0+44,398+629,793). The overall percent correct for the training set, 73.06%, is found by summing the total occurrences of the predicted fatal and observed fatal outcomes (0), the predicted injury and observed injury outcomes (63,279), and the predicted property damage and observed property damage outcomes (629,793), and then dividing by the number of observations in the testing set (948,679).

Table 4: CHAID Classification Accuracy for the Testing Set

	Predicted			
Observed	Fatal	Injury	Property Damage	Percent Correct
Fatal	0	3,175	2,848	0.0%
Injury	0	63,279	205,186	23.6%
Property Damage	0	44,398	629,793	93.4%
Overall Percentage	0.0%	11.7%	88.3%	73.06%

The AUC is a widely recognized measure of discriminatory power (Worster et al. 2006) and quality of probabilistic classifiers (Vuk and Curk 2006), which measures the classifiers' performance across the entire range of potential outcome distributions (Vuk and Curk 2006), and is equal to the probability that a classifier will rate a randomly chosen positive outcome higher than a randomly chosen negative outcome (Fawcett 2006). The AUC values are determined by calculating the area under the ROC curves, which are constructed by plotting the true positive rate against a false positive rate for subsets of the observations (Fawcett 2006). The AUC results for the CART and CHAID's capabilities to predict a fatal outcome relative to non-fatal outcomes and to predict a property damage only outcome relative to injury outcomes are presented in Table 5.

Table 5: Accuracy Comparison of CHAID and CART Models

Decision Tree Approach	Classification Accuracy Training Set	Classification Accuracy Testing Set	AUC Value Fatal vs. Nonfatal Training Set	AUC Value Fatal vs. Nonfatal Testing Set	AUC Value Non-injury vs. Injury Training Set	AUC Value Non-injury vs. Injury Testing Set
CHAID	73.06%	73.00%	0.899	0.898	0.717	0.717
CART	72.32%	72.30%	0.759	0.761	0.667	0.667

The AUC values are compared to determine if there is a significant difference between the models' abilities to predict (1) a fatal outcome relative to property damage and injury only outcomes and (2) a property damage only outcome relative to fatal and injury outcomes by calculating a critical ratio *z*, defined by Hanley and McNeil (1983). A statistically significant difference between the models' ability to predict a fatal outcome relative to a non-fatal outcome is not found ($z = 0.38$; $p = 0.35197$); yet, a statistically significant difference between the models' ability to predict a property damage only outcome relative to an injury or fatal outcome does exist ($z = 77.15$; $p = <0.00001$).

Because of its greater classification accuracy and higher AUC values, as illustrated in Table 5, the CHAID model better discriminates than the CART model between crash severity outcomes and is carried forward to compare findings here to prior literature findings and to examine current Missouri driving rule and policy implications.

ANALYSIS AND RESULTS

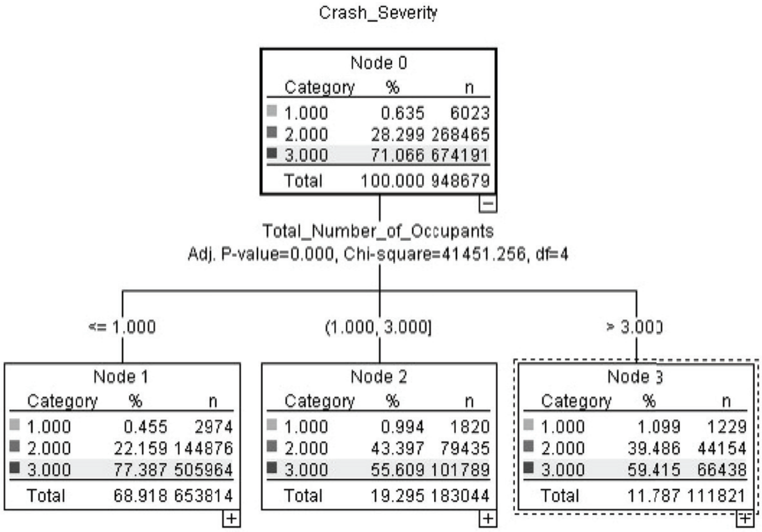
The factors with the greatest predictor importance for crash severity (i.e., the relative importance of each predictor in estimating the model) are calculated using the CHAID model results. The model determines predictor importance by computing the reduction in variance of the target attributable to each predictor via a sensitivity analysis (Saltelli 2002; Saltelli et al. 2004). The estimated CHAID tree identifies the factors with the greatest predictor importance for crash severity as total number of occupants, speed limit, speeds that exceed the limit, alcohol, failed to yield, violation of a stop sign or signal, physical impairment, driving on the wrong side of the road when not passing, crash location, and improper backing. To illustrate the insights afforded by the estimated CHAID decision tree and to provide a context within which to evaluate reductions in motor vehicle crash risk, decision rules focusing on the variables with the greatest predictor importance in the CHAID model are examined.

Number of Occupants

The CHAID model identifies *total number of occupants* as the best predictor to form the first branch of the decision tree, since it has the greatest importance in estimating the model and partitions the training set into three branches characterized as single occupant, two or three occupants, or more than three occupants. Findings suggest that as the total number of occupants involved in a crash increase, so does the probability that a fatal outcome will occur. This result is consistent with prior research findings that crash severity probabilities increase as the number of vehicle passengers increase (Renski et al. 1999; Oh 2006).

As illustrated in Figure 2, the probability that a fatal outcome (Category 1) will occur increases as the number of occupants involved in the crash increases: 0.455% for single occupant crashes, 0.994% for crashes involving two or three occupants, and 1.099% for crashes involving more than three occupants. Interestingly, the probability that an injury outcome (Category 2) will occur does not necessarily increase as the number of occupants increases. When increasing the total number of occupants from a single occupant to two or three occupants, the likelihood of an injury outcome increases from 22.159% to 43.397%; yet, when increasing the number of occupants to more than three occupants, the likelihood of an injury outcome decreases to 39.486%. Both findings indicate nonlinearity, and illustrate the importance of using the CHAID decision tree for analysis of non-linear effects.

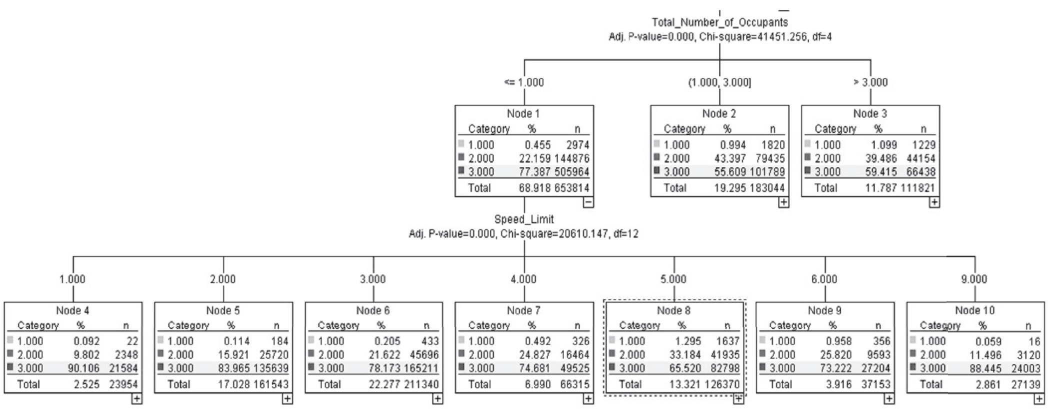
Figure 2: First Branch of CHAID Tree – Total Number of Occupants



Speed Limit

The CHAID model identifies speed limit as the second most important predictor variable, and serves as the second branch for single occupant crashes. As illustrated in Figure 3, the probability of a fatal or injury outcome increases for speed limit zones of up to 55 mph and 60 mph for single occupant crashes. Yet, as model results suggest, a change from 55mph and 60 mph to 65 mph and 70 mph decreases the likelihood that the outcome will be fatal or injurious, which could be attributed to the type of roads in which this speed limit is typically present in Missouri (e.g., interstates). This finding further solidifies the importance of using CHAID decision trees to analyze non-linear effects.

Figure 3: Single Occupant Crash Branch Two – Speed Limit



Zone 1 = 05 mph and 20 mph; zone 2 = 25 mph and 30 mph; zone 3 = 35 mph and 40 mph; zone 4 = 45 mph and 50 mph; zone 5 = 55 mph and 60 mph; zone 6 = 65 mph and 70 mph; and zone 9 = Unknown

This study's results also are consistent with previous research findings that higher speed limits significantly increase the risk of severe injury outcomes (Renski et al. 1999; Khattak et al. 2002; Oh 2006; Gårder 2006; Malyshkina and Mannering 2010; Savolainen and Ghosh 2008; Haleem and Abdel-Aty 2010; Zhu and Srinivasan 2011; Yasmin and Eluru 2013).

Speeds - Exceed Limit

Crashes involving driving at speeds that exceed the posted limit are more likely to cause fatal and injury outcomes for each partition of the number of occupants. For single occupant crashes, model results indicate that driving at speeds that exceed the limit in zones of 35 mph or 40 mph and 65 mph or 70 mph increases the chance of a fatal outcome from 0.133% to 3.689% and from 0.760% to 4.746%, respectively. For crashes with two or three occupants, additional results suggest that driving at speeds that exceed the limit in zones of 35 mph or 40 mph and 45 mph or 50 mph increases the chance of a fatal outcome from 0.233% to 4.671% and 0.568% to 6.534% respectively. Finally, for crashes with more than three occupants, driving at speeds that exceed the limit increases the chance of a fatal outcome occurring as speed limit zones increase: 25 mph or 30 mph equals 3.409%; 35 mph or 40 mph equals 6.902%; 45 mph or 50 mph equals 8.543%; 65 mph or 70 mph equals 13.223%. These findings, which suggest that driving at speeds that exceed the limit have a greater risk of injury, are consistent with prior research (Khattak et al. 1998; Renski et al. 1999; Khattak et al. 2002; Khattak and Rocha 2003; Gårder 2006; Oh, 2006; Savolainen and Ghosh 2008; Schneider et al. 2009; Haleem and Abdel-Aty 2010; Malyshkina and Mannering 2010; Zhu and Srinivasan 2011; Yasmin and Eluru 2013).

Importantly, this study identifies interactions between speeding and other circumstances. For example, for single occupant crashes, results indicate that a young driver (under the age of 21) driving at speeds that exceed the limit in a speed limit zone of 25 mph to 30 mph and 65 mph to 75 mph has a lesser chance of a fatal outcome (0.676% and 3.049%) than older drivers (2.063% and 1.105%, respectively for middle aged drivers and mature drivers in 25 mph/30 mph zones and 5.714% for both older groups in 65 mph/75 mph zones). For crashes with two or three occupants, driving at speeds that exceed the limit in a speed zone of 35 mph or 45 mph during dark, but lit conditions increase the likelihood of a fatal outcome from 2.607% to 8.108%, yet decreases the likelihood of an injury outcome of 70.142% to 66.366% when compared with driving at speeds that exceed the limit during other lighting conditions. Additionally, for two or three occupant crashes, males driving at speeds that exceed the posted limit of 45 mph or 50 mph have a greater chance of a fatal outcome (7.950%) relative to their female counterparts (1.010%). Finally, for crashes involving two or three occupants, a 20.870% chance of a fatal outcome and a 68.216% of an injury outcome results when driving at speeds that exceed the limit while under the influence of alcohol.

Alcohol

Crashes that occur while driving under the influence of alcohol have greater crash severity regardless of the number of occupants involved in the crash, as supported by prior literature; yet, this study shows that its importance is more prevalent for crashes involving multiple occupants. The presence of alcohol represents the second split in the decision tree for two and three occupant crashes, and model results show that alcohol presence increases the probability of a fatal outcome and an injury outcome from 0.778% to 5.175% and 42.411% to 62.486%, respectively, and for more than three occupant crashes, where alcohol presence increases the probability of a fatal outcome and an injury outcome from 0.856% to 7.023% and 38.676% to 59.273%, respectively.

Additionally, results reveal dangerous interaction effects between alcohol and other variables. For example, for single occupant crashes, driving under the influence of alcohol in a speed limit zone of 65,mph or 70,mph increases the probability of a fatal outcome from 0.820% to 3.053%

and of an injury outcome from 24.742% to 42.215%, compared with similar circumstances when alcohol is not present. When adding speeding to alcohol use at such high speeds, the risk of a fatal outcome and an injury outcome increases to 6.024% and 56.024%, respectively.

For crashes involving two or three occupants, results suggest that a crash occurring when alcohol is present increases the probability of a fatal outcome from 0.763% to 5.181% and the probability of an injury outcome from 42.380% to 62.327% compared with crashes when alcohol is not present. Moreover, adding speeding when on a hill or a crest to this scenario increases the probability of a fatal outcome and injury outcome to 20.882% and 65.429% respectively.

When a crash involves three or more occupants, the probability of a fatal outcome increases from 0.866% to 7.103% and an injury outcome increases from 38.752% to 60.276% when alcohol is present; when speeding is included, the chance of a fatal and an injury outcome increases to 17.221% and 65.558%, respectively. Finally, when adding a dark light condition (with no streetlights or streetlights off) to this scenario, the chance of a fatal outcome increases to 26.627%, and an injury outcome increases to 62.130%.

Failing to Yield

Crashes involving failing to yield are also more likely to cause severe outcomes, as confirmed by prior research (Peek-Asa et al. 2010), and failure to yield has important interaction effects with other characteristics. For instance, when failing to yield is present and a single occupant on-roadway crash in a speed limit zone of 65,mph or 70,mph occurs, model results indicate that the chance of a fatal or injury outcome increases from 0.471% and 19.260% to 0.972% and 25.791%, respectively, than if failing to yield is not present. For crashes with two or three occupants, drivers who fail to yield in a speed limit zone of 65,mph or 70,mph have a greater chance of a fatal outcome (4.708%) and injury outcome (53.861%) than if the driver yielded properly.

DISCUSSION

To provide a context for understanding the relative reduction in overall risks associated with reducing the frequency of driver behaviors that contribute to the likelihood of different crash severity outcomes, historic outcomes are examined to determine annual upper and lower bounds on the changes in the number of drivers involved in fatal, injury, or property damage only crashes if selected contributory circumstances might be individually eliminated. The analysis provides a context in which to employ counterfactual arguments to provide reasonable bounds on how the total number of crash severity outcomes might change with measures designed to reduce the frequency of occurrence of the significant contributing factors. The annual bounds are calculated for each severity outcome by dividing the number of outcomes by the number of effective years multiplied by the proportion of the training set (11 * 0.75).

Considering the contributing circumstances that have the greatest predictor importance for severe crash outcomes, lower and upper bounds for changes in the annual number of drivers involved in each of the three severity outcomes are determined by 1) removing the contributing circumstance for each driver and assuming the crash still occurs with severity outcome probabilities now determined by the outcome probabilities of the complementary node (a ceteris paribus lower bound) and 2) removing the contributing circumstance and alternatively assuming that the driver is not involved in a crash at all (an upper bound). This bounding technique presumes that no casual relationships exist among contributing circumstances in estimating the lower bounds and, alternatively, that the removed contributing circumstance was solely responsible for causing the accidents in estimating the upper bounds.

Table 6 presents the lower and upper bounds of the reductions in the annual numbers of drivers involved in fatalities, injury, and property damage outcomes associated with the six most important contributing circumstances. As illustrated in Table 6, the elimination of the specific contributing circumstance clearly changes the distribution of the number of drivers involved in the three outcomes. For example, alcohol involvement has significant detrimental effects on the number of Missouri drivers involved in fatal outcomes. When eliminating alcohol as a contributing circumstance and assuming the crash then does not occur, 191 fewer annual driver contributions towards fatal crashes might be prevented. When eliminating alcohol as a contributing circumstance and assuming the crash still does occur, the estimated severity outcomes are redistributed and at least 135 fatal accident outcomes per year might be avoided.

Table 6: Estimated Annual Reductions in the Number of Drivers Involved in Each Severity Outcome if a Contributing Circumstance is Eliminated

	Fatal		**Injury**		**Property Damage Only**		**N[1]**
Contributing Circumstance	Est Lower Bound	Est Upper Bound	Est Lower Bound	Est Upper Bound	Est Lower Bound[2]	Est Upper Bound	
Speed - Exceed Limit	107	133	477	1,344	-801	1,325	2,802
Alcohol	135	191	841	2,741	-1,418	3,187	6,119
Failed to Yield	43	88	1412	6,779	-1,455	15,268	22,135
Violation - Stop Sign/Signal	16	39	692	2,133	-708	2,956	5,128
Wrong-Side	67	110	157	1,065	-224	1,212	2,388
Physical Impairment	11	36	427	1,215	-437	1,190	2,442

[1]N = Number of estimated cases per year and equal to the sum of the estimated upper bounds.
[2]A negative value for property damage only outcome represents an increase for the least severe outcome, given the assumption that the crash still occurs.

Still, the interaction effects of variables identified by the CHAID model are important when analyzing crash severity data. For example, it is readily discovered that driving while under the influence of alcohol, driving at speeds that exceed the limit, failing to yield, driving on the wrong side of the road, violating a stop sign or signal, and driving while physically impaired lead to a significant number of fatalities each year in Missouri. The effect of these factors on the probability of a severe outcome is dependent upon other variables, including the number of vehicle occupants involved in the crash, the speed limit, actual driving speed, lighting conditions and driver's age. Thus, this study concludes that policy makers should consider the interaction of driver related contributory circumstances and other conditions when formulating future legislation intended to reduce the number of fatal outcomes.

CONCLUSIONS

Key findings presented in the analysis have important implications for possible changes in the current Missouri Driver Guide - Rules of the Road (Missouri Department of Revenue 2014). Drawing upon these findings, policy recommendations are identified and discussed for the contributing circumstances that greatly increase the likelihood of more severe outcomes of motor vehicle crashes: speed limit, driving at speeds that exceed the limit, and alcohol use.

Speed Limit/Speed - Exceed Limit

The Missouri Driver Guide states, "Speed limit signs indicate the maximum speed allowed by law, and do not mean that all parts of the road can be safely driven at those speeds under all conditions. The speed limit is the maximum allowable speed in ideal conditions" (Missouri Department of Revenue 2014, p. 37); and it is recommended that driving speed be adjusted as appropriate for changes in road conditions and characteristics, visibilities, other road users, and weather conditions. As previously suggested, the interaction of speed limit and driving at speeds that exceed the limit increase the likely severity of crash outcomes, which is confirmed by the statements made. For example, driving at speeds that exceed the posted limit of 35 mph to 45 mph during dark, but lit conditions have an increased likelihood of a fatal outcome than when speeding during other lighting conditions. As a result, it is recommended that patrol units be aware that dark conditions increase the probability of severe outcomes and adjust accordingly.

Additionally, the likelihood of a fatal crash is higher when driving on the wrong side of the road in speed limit zones of 45 mph to 60 mph and when failing to yield in a speed limit zone of 65 mph or 70 mph than if these contributing circumstances are not present. Following successful application in North Carolina and California, many states have adopted innovative strategies to reduce wrong-way driving, such as lowering the height of "Do Not Enter" and "Wrong Way" signs, increasing the size of signage, locating signage on both sides of the exit travel lane, changing lighting and minor ramp geometrics, and illuminating "Wrong Way" signs that flash when a wrong-way vehicle is detected (Zhou and Rouholamin 2014). As a result, the study may infer that in higher speed limit zones, preventive measures to reduce driving on the wrong side of the road and failing to yield, such as prominent signage, are of great importance.

Alcohol

Driver alcohol use is one of the most significant predictors of crash severity. Currently under Missouri law, drivers who are found guilty of driving while intoxicated (DWI) may be subject to paying a fine, having his/her license revoked, or being imprisoned (Missouri Department of Revenue 2014). Moreover, if someone is injured or killed because of driving under the influence of alcohol, the driver may "spend two to seven years in jail, pay a $5,000 fine, and/or lose your driver license for five years" (Missouri Department of Revenue 2014 p.77). Because of the large increase in the probabilities of injury and fatal outcomes when driving under the influence of alcohol, these laws may not be stringent enough in the prevention of drinking and driving given the clear large increase in the likelihood of severe outcomes. Additionally, Missouri law currently requires any person guilty of a second alcohol intoxication-related traffic offense to install an ignition interlock device on all vehicles operated by the offender before reinstating driving privileges (Missouri Department of Transportation 2013). Since drivers with a blood alcohol concentration above the legal limit that are involved in fatal crashes are six times more likely to have a prior DWI conviction (U.S. Department of Transportation (2014), to deter multiple offenses from occurring, all DWI first-time offenders could be required to use ignition interlocks.

Research suggests that injuries and fatalities from impaired driving can be prevented through community-based approaches (DeJong and Hingson 1998; Holder et al. 2000; Shults et al. 2009). The Missouri Department of Revenue encourages such approaches through reporting drunk drivers by calling 911 and providing law enforcement with the license plate number of the vehicle, a physical description of the car and driver, and the vehicle's location (Missouri Department of Revenue 2014). However, to reduce the number of DWI drivers on Missouri roadways, this study recommends that this process be simplified and that a hotline and/or web-notification mechanism be considered (with possible rewards) for reporting DWIs.

To further reduce DWIs, Missouri law enforcement agencies implement sobriety checkpoints at temporary, random locations (Reynolds 1989). Research indicates that high-profile enforcement efforts, specifically frequent sobriety checkpoints, are effective in reducing alcohol-related fatal crashes (Elder et al. 2002), and recent studies found such checkpoints reduce the number of fatal outcomes by 20% (Shults et al. 2009). As described earlier, a strong interaction is found between high speed limits, alcohol intoxication, and crash severity. As a result, this study recommends that future DWI checkpoints might be located at on-ramps to high speed highways and interstates to reduce the number of intoxicated drivers driving at high speeds.

LIMITATIONS AND FUTURE RESEARCH

Limitations to this research exist and may be resolved through future research endeavors. First, this study considers data compiled from the entire state of Missouri, and the general findings may not be appropriate in specific differentiated locations throughout the state. Future research may address this limitation by partitioning data into smaller regions of Missouri (urban, rural, suburban, county, zip code, and other meaningful partitions) and by examining regional factors and their effect on injury severity to contribute to more localized legislation. Second, this study considers only Missouri data. Future research may apply the same methodological approach to additional state crash datasets to assess policy implications for various locations. Third, additional or alternate variables may be considered in future research to examine other factors that may contribute differentially to crash severity. These include variables such as seasonality, peak driving times, highway class, rural versus urban location, crash type, and vehicle type. Additionally, future studies may compare the decision tree models to other methodological approaches, such as multinomial logit and ordinal logit and probit models. Finally, future research may apply the methodological techniques presented here to other modes of transportation and assess safety measures, risk, and disruptions beyond roadways.

References

Abay, Kibrom A. “Examining Pedestrian-Injury Severity Using Alternative Disaggregate Models.” *Research in Transportation Economics* 43 (1), (2013): 123-136.

Abdel-Aty, Mohamed. “Analysis of Driver Injury Severity Levels at Multiple Locations Using Ordered Probit Models.” *Journal of Safety Research* 34 (5), (2003): 597-603.

Abdel-Aty, M. and H. Abdelwahab. “Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison.” *Journal of Transportation Engineering* 130 (2), (2004a): 204-210.

Abdel-Aty, M. and H. Abdelwahab. “Modeling Rear-End Collisions Including the Role of Driver’s Visibility and Light Truck Vehicles Using a Nested Logit Structure.” *Accident Analysis & Prevention* 36 (3), (2004b): 447-456.

Abdel-Aty, M. and J. Keller. “Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections.” *Accident Analysis & Prevention* 37 (3), (2005): 417-425.

Abdelwahab, H. and M. Abdel-Aty. “Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections.” *Transportation Research Record: Journal of the Transportation Research Board* 1746 (1), (2001): 6-13.

Abdelwahab, H. and M. Abdel-Aty. “Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas.” *Transportation Research Record: Journal of the Transportation Research Board* 1784 (1), (2002): 115-125.

Abellán, J., G. López, and J. De OñA. "Analysis of Traffic Accident Severity Using Decision Rules via Decision Trees." *Expert Systems with Applications* 40 (15), (2013): 6047-6054.

Bayam, E., J. Liebowitz, and W. Agresti. "Older Drivers and Accidents: A Meta-Analysis and Data Mining Application on Traffic Accident Data." *Expert Systems with Applications* 29 (3), (2005): 598-629.

Bédard, M., G. Guyatt, M. Stones, and J. Hirdes. "The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities." *Accident Analysis & Prevention* 34 (6), (2002): 717-727.

Bernard, J.M. and D.C. Sweeney II. "Contributing Circumstances Impact on Missouri Teenage Driver Crash Fatalities." *Journal of the Transportation Research Forum* 54 (1), (2015): 5-21.

Bernard Bracy, J.M. "The Estimation, Ensembling and Comparison of Methodological Techniques Used in Motor Vehicle Crash Severity Research." *International Journal of Business, Marketing and Decision Sciences* 10 (1), (2017): In press.

Blincoe, L., T. Miller, E. Zaloshnja, and B. Lawrence. "The Economic and Societal Impact of Motor Vehicle Crashes." Report no. DOT HS-812-013. Washington, DC: National Highway Traffic Safety Administration, 2010.

Breiman, L., J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and Regression Trees*. CRC press. Boca Raton, FL: 1984.

Centers for Disease Control and Prevention. "Injury Prevention & Control: Motor Vehicle Safety." Last modified April 22, 2015. Accessed May 18, 2015. http://www.cdc.gov/motorvehiclesafety/.

Chang, L. and H. Wang. "Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques." *Accident Analysis & Prevention* 38 (5), (2006): 1019-1027.

DeJong, W. and R. Hingson. "Strategies to Reduce Driving Under the Influence of Alcohol." *Annual Review of Public Health* 19 (1), (1998): 359-378.

Delen, D., R. Sharda, and M. Bessonov. "Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks." *Accident Analysis & Prevention* 38 (3), (2006): 434-444.

Department of Transportation (US). "Traffic Safety Facts 2013 Data: Alcohol-Impaired Driving." *National Highway Traffic Safety Administration (NHTSA).* 2014. Accessed May 27, 2016. http://www-nrd.nhtsa.dot.gov/Pubs/812102.pdf.

Donnell, Eric T., and John M. Mason. "Predicting the Severity of Median-Related Crashes in Pennsylvania by Using Logistic Regression." *Transportation Research Record: Journal of the Transportation Research Board* 1897 (1), (2004): 55-63.

Elder, R.W., R.A. Shults, D.A. Sleet, J.L. Nichols, S. Zaza, and R.S. Thompson. "Effectiveness of Sobriety Checkpoints for Reducing Alcohol-Involved Crashes." *Traffic Injury Prevention* 3 (4), (2002): 266-274.

Eustace, D., O. Almutairi, P. Hovey, and G. Shoup. "Using Decision Tree Modeling to Analyze Factors Contributing to Injury and Fatality of Run-Off-Road Crashes in Ohio." *Transportation Research Board 93rd Annual Meeting*, 14-5668. (2014).

Fawcett, Tom. “An Introduction to ROC Analysis.” *Pattern Recognition Letters* 27 (8), (2006): 861-874.

Gårder, Per. “Segment Characteristics and Severity of Head-On Crashes on Two-Lane Rural Highways in Maine.” *Accident Analysis & Prevention* 38 (4), (2006): 652-661.

Haleem, K. and M. Abdel-Aty. “Examining Traffic Crash Severity at Unsignalized Intersections.” *Journal of Safety Research* 41 (4), (2010): 347-357.

Hanley, James A. and Barbara J. McNeil. “A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived from the Same Cases.” *Radiology* 148 (3), (1983): 839-843.

Holder, H.D., P.J. Gruenewald, W.R. Ponicki, A.J. Treno, J.W. Grube, R.F. Saltz, R.B. Voas et al. “Effect of Community-Based Interventions on High-Risk Drinking and Alcohol-Related Injuries.” *Jama* 284 (18), (2000): 2341-2347.

Islam, S. and F. Mannering. “Driver Aging and its Effect on Male and Female Single-Vehicle Accident Injuries: Some Additional Evidence.” *Journal of Safety Research* 37 (3), (2006): 267-276.

Jiang, Ximiao, Baoshan Huang, Xuedong Yan, Russell L. Zaretzki, and Stephen Richards. “Two-Vehicle Injury Severity Models Based on Integration of Pavement Management and Traffic Engineering Factors.” *Traffic Injury Prevention* 14 (5), (2013): 544-553.

Khan, G., Bill, A.R., and Noyce, D.A. “Exploring the Feasibility of Classification Trees Versus Ordinal Discrete Choice Models for Analyzing Crash Severity. *Transportation Research Part C: Emerging Technologies* 50, (2015): 86-96.

Kass, G.V. “An Exploratory Technique for Investigating Large Quantities of Categorical Data.” *Applied Statistics* (1980): 119-127.

Khattak, A.J., M.D. Pawlovich, R.R. Souleyrette, and S.L. Hallmark. “Factors Related to More Severe Older Driver Traffic Crash Injuries.” *Journal of Transportation Engineering* 128 (3), (2002): 243-249.

Khattak, A.J. and M. Rocha. “Are SUVs “Supremely Unsafe Vehicles?”: Analysis of Rollovers and Injuries with Sport Utility Vehicles.” *Transportation Research Record: Journal of the Transportation Research Board* 1840 (1), (2003): 167-177.

Khattak, A.J., P. Kantor, and F.M. Council. “Role of Adverse Weather in Key Crash Types on Limited-Access Roadways: Implications for Advanced Weather Systems.” *Transportation Research Record: Journal of the Transportation Research Board* 1621 (1), (1998): 10-19.

Klop, J.R. and A.J. Khattak. “Factors Influencing Bicycle Crash Severity on Two-Lane, Undivided Roadways in North Carolina.” *Transportation Research Record: Journal of the Transportation Research Board* 1674 (1), (1999): 78-85.

Kockelman, Kara Maria and Young-Jun Kweon. “Driver Injury Severity: An Application of Ordered Probit Models.” *Accident Analysis & Prevention* 34 (3), (2002): 313-321.

Krull, K. A., A. J. Khattak, and F. M. Council. “Injury Effects of Rollovers and Events Sequence in Single-Vehicle Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 1717 (1), (2000): 46-54.

Kuhnert, P.M., K. Do, and R. McClure. “Combining Non-Parametric Models with Logistic Regression: An Application to Motor Vehicle Injury Data.” *Computational Statistics & Data Analysis* 34 (3), (2000): 371-386.

Lu, G.X., D.A. Noyce, and R.J. McKendry. “Analysis of the Magnitude and Predictability of Median Crossover Crashes Utilizing Logistic Regression.” In *Transportation Research Board 85th Annual Meeting*, 06-2681. 2006.

Malyshkina, N.V. and F.L. Mannering. “Empirical Assessment of the impact of Highway Design Exceptions on the Frequency and Severity of Vehicle Accidents.” *Accident Analysis & Prevention* 42 (1), (2010): 131-139.

Mayou, R., B. Bryant, and R. Duthie. “Psychiatric Consequences of Road Traffic Accidents.” *BMJ: British Medical Journal* 307 (6905), (1993): 647.

Mingers, John. “An Empirical Comparison of Pruning Methods for Decision Tree Induction.” *Machine Learning* 4 (2), (1989a): 227-243.

Mingers, John. “An Empirical Comparison of Selection Measures for Decision-Tree Induction.” *Machine Learning* 3 (4), (1989b): 319-342.

Missouri Department of Revenue. “Missouri Driver Guide.” 2014. http://dor.mo.gov/pdf/DriverGuide.pdf.

Missouri Department of Transportation. “Ignition Interlock.” 2013. (Accessed May 27, 2016). http://www.modot.org/safety/ignitioninterlock.htm.

Missouri State Highway Patrol. *Missouri Traffic Crashes.* Jefferson City: Public Information and Education Division, 2012.

Moudon, Anne Vernez, Lin Lin, Junfeng Jiao, Philip Hurvitz, and Paula Reeves. “The Risk of Pedestrian Injury and Fatality in Collisions with Motor Vehicles, a Social Ecological Study of State Routes and City Streets in King County, Washington.” *Accident Analysis & Prevention* 43 (1), (2011): 11-24.

Oh, J.T. “Development of Severity Models for Vehicle Accident Injuries for Signalized Intersections in Rural Areas.” *KSCE Journal of Civil Engineering* 10 (3), (2006): 219-225.

Peek-Asa, Corinne, Carla Britton, Tracy Young, Michael Pawlovich, and Scott Falb. “Teenage Driver Crash Incidence and Factors Influencing Crash Injury by Rurality.” *Journal of Safety Research* 41 (6), (2010): 487-492.

Quddus, M.A., C. Wang, and S.G. Ison. “Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models.” *Journal of Transportation Engineering* 136 (5), (2010): 424-435.

Renski, H., A.J. Khattak, and F.M. Council. “Effect of Speed Limit Increases on Crash Severity: Analysis of Single-Vehicle Crashes on North Carolina Interstate Highways.” *Transportation Research Record: Journal of the Transportation Research Board* 1665 (1), (1999): 100-108.

Reynolds, Scott. “Use of Sobriety Checkpoints to Combat Drunk Drivers: Knowing When to Say When.” *Missouri Law Review* 54 (2), (1989): 1-16.

Rifaat, S.M. and R. Tay. “Effects of Street Patterns on Injury Risks in Two-Vehicle Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 2102, (1), (2009): 61-67.

Rifaat, S.M., R. Tay, and A. de Barros. “Effect of Street Pattern on the Severity of Crashes Involving Vulnerable Road Users.” *Accident Analysis & Prevention* 43 (1), (2011): 276-283.

Saltelli, Andrea. “Making Best Use of Model Evaluations to Compute Sensitivity Indices.” *Computer Physics Communications* 145 (2), (2002): 280-297.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, 2004.

Savolainen, P.T., F.L. Mannering, D. Lord, and M.A. Quddus. “The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives.” *Accident Analysis & Prevention* 43 (5), (2011): 1666-1676.

Savolainen, P. and I. Ghosh. “Examination of Factors Affecting Driver Injury Severity in Michigan’s Single-Vehicle—Deer Crashes.” *Transportation Research Record: Journal of the Transportation Research Board* 2078 (1), (2008): 17-25.

Schneider, W. and P. Savolainen. “Comparison of Motorcyclist Injury Severity Among Various Crash Types.” *Transportation Research Record* 2265, (2011): 70-80.

Schneider, W.H., P.T. Savolainen, and K. Zimmerman. “Driver Injury Severity Resulting from Single-Vehicle Crashes Along Horizontal Curves on Rural Two-Lane Highways.” *Transportation Research Record: Journal of the Transportation Research Board* 2102 (1), (2009): 85-92.

Shults, R.A., R.W. Elder, J. L. Nichols, D. A. Sleet, R. Compton, S. K. Chattopadhyay, and Task Force on Community Preventive Services. “Effectiveness of Multicomponent Programs with Community Mobilization for Reducing Alcohol-Impaired Driving.” *American Journal of Preventive Medicine* 37 (4), (2009): 360-371.

Sohn, So Young, and Hyungwon Shin. “Pattern Recognition for Road Traffic Accident Severity in Korea.” *Ergonomics* 44 (1), (2001): 107-117.

Trnka, A. “Classification and Regression Trees as a Part of Data Mining in Six Sigma Methodology.” *Proceedings of the World Congress on Engineering and Computer Science*, vol. 1. 2010.

Ulfarsson, G.F. and F.L. Mannering. “Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents.” *Accident Analysis & Prevention* 36 (2), (2004): 135-147.

Vuk, Miha, and Tomaz Curk. “ROC Curve, Lift Chart and Calibration Plot.” *Metodoloski zvezki* 3 (1), (2006): 89-108.

Wang, Z., H. Chen, and J. Lu. “Exploring Impacts of Factors Contributing to Injury Severity at Freeway Diverge Areas.” *Transportation Research Record: Journal of the Transportation Research Board* 2102 (1), (2009): 43-52.

Worster, Andrew, Jerome Fan, and Suneel Upadhye. “Understanding Receiver Operating Characteristic (ROC) Curves.” *Cjem* 8 (1), (2006): 19-20.

Yan, X. and E. Radwan. "Analyses of Rear-End Crashes Based on Classification Tree Models." *Traffic Injury Prevention* 7 (3), (2006): 276-282.

Yasmin, S. and N. Eluru. "Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash severity." *Accident Analysis & Prevention* 59, (2013): 506-521.

Zajac, Sylvia S., and John N. Ivan. "Factors Influencing Injury Severity of Motor Vehicle–Crossing Pedestrian Crashes in Rural Connecticut." *Accident Analysis & Prevention* 35 (3), (2003): 369-379.

Zhou, H. and M.P. Rouholamin. "Proceedings of the 2013 National Wrong-Way Driving Summit." (2014).

Zhu, X. and S. Srinivasan. "A Comprehensive Analysis of Factors Influencing the Injury Severity of Large-Truck Crashes." *Accident Analysis & Prevention* 43 (1), (2011): 49-57.

Jill Bernard Bracy *earned a Ph.D. in business administration with an emphasis in logistics and supply chain management with a marketing concentration, and serves as an assistant teaching professor in marketing and the assistant director of program development for the Center for Transportation Studies at the University of Missouri-St. Louis. She serves as vice president of operations for the Council of Supply Chain Management Professionals-St. Louis Roundtable, is a former Boeing Fellow, and is currently pursuing research interests in transportation safety.*