Assessing the Likelihood of a Lane Departure Event using Naturalistic Driving Study Data

Shauna Hallmark
Associate Professor of Civil, Environmental, and Construction Engineering, Institute for Transportation, Iowa State University, 2711 South Loop Dr. Suite 4700, Ames, Iowa 50010, corresponding author

Linda Ng Boyle
Associate Professor, Depts of Industrial and Systems Engineering, Civil & Environmental Engineering, Box 352650, Seattle, WA 98195

Yu Qiu
Graduate Research Assistant. Department of Statistics, Iowa State University, Snedecor Hall, Ames, Iowa 50011

Submitted to the 3rd International conference on Road Safety and Simulation. September 14-15. Indianapolis, IN. USA.

ABSTRACT

While lane departures constitute a substantial number of motor vehicle crashes and a disproportionate number of fatalities, the factors and events associated with lane departure crashes are complex and not well understood. The use of data from naturalistic driving studies can provide insights into these factors in the timeframe before and after an event occurs as well as on the impact of existing ambient conditions and current roadway conditions on these crash types.

An evaluation of lane departure incidents was conducted using data from a field operational test on a road departure curve warning system conducted by UMTRI. Although a specific system was being evaluated, drivers were able to drive in their normal setting in all road and weather conditions during a baseline condition (when the system was not activated). Hence, naturalistic data were available and reduced for 44 drivers on rural two-lane roads. There were 22 incidents involving a vehicle leaving the lane on the right and 51 incidents departing the lane on the left on rural two-lane roads available for analysis. Because no near-crashes or crashes occurred, these incidents were used as a safety surrogate for lane departure crashes. Further, data for which no incidents occurred were also extracted and used for normal driving conditions.

Corresponding roadway, environmental, and driver variables were used in a logistic regression analysis. Separate models were developed for right-side and left-side lane departures. The analysis evaluated the relationship between curve characteristics, shoulder type, driver age and gender, percent of time the driver has exceeded the posted speed limit or speed limit on curves, lane width, and time of day.

Keywords: naturalistic driving studies, lane departure, run-off-road

INTRODUCTION

Background

The US DOT - Federal Highway Administration (FHWA) (2011) estimates that 58% of roadway fatalities are lane departures, while 40% of fatalities are single-vehicle run-off-road (ROR) crashes. Neuman et al. (2003) also estimated that 39% of national fatal crashes are single-vehicle run-off-road (SVROR) crashes. Addressing lane-departure crashes is therefore a priority for National, State, and local agencies.

However, there are several issues with crash-based safety analyses (Songchitruksa and Tarko, 2006). Events with similar traffic, weather, and roadway conditions are quite rare and as a result, safety analyses must depend on small sample sizes. Crash reporting across sites is also inconsistent making comparisons quite difficult. The data are also limited in the type of roadway, environmental, and driver characteristics recorded. Timeliness of the crash data is also an issue, particularly for evaluating crash countermeasures. An appropriate study requires data before and after the treatment is implemented, requiring several years of crash data.

Some researchers have addressed limitations in crash data by utilizing crash surrogates as a measure of risk. Reduction in violations has been used to evaluate the effectiveness of red-light running countermeasures, such as red light running camera enforcement (Retting et al., 2007; Bonneson et al., 2004; Garber et al., 2005; Fitzsimmons et al, 2009). Change in speed is frequently used to assess the effectiveness of treatments assuming that if speeds are reduced, crashes will also be reduced. Time to collision (TTC) or separation time has also been used as crash surrogates (Archer, 2001; Songchitruksa and Tarko, 2006; Hayward, 1972; Burgett and Miller, 2001).

Several types of crash surrogates have been used to evaluate lane departures. Lane deviation is one measure used as a crash surrogate to assess the likelihood of ROR crashes (LeBlanc et al., 2006) and the likelihood of crashes due to distraction (Donmez et al., 2006). Several studies have used lateral placement to assess countermeasures so more immediate measures of safety than reduction in crashes can be obtained. Porter et al (2004) used lateral placement and speed to evaluate centerline rumble strips. Pratt et al (2006) used vehicle lateral position and change in vehicle separation to evaluate the impact of centerline and edge-line rumble strips. Miaou (2001) developed a method to estimate roadside encroachment frequency and the probability distribution for the lateral extent of encroachments using an accident based prediction model. Miles et al (2006) recorded the number of erratic and avoidance maneuvers that occur with placement of advance stop line rumble strips to determine how drivers respond to the devices. Taylor et al (2005) observed vehicle placement relative to the edge line using single versus double paint lines to delineate presence of shoulder rumble strips. Hallmark et al (2010) used lateral position to evaluate the effectiveness of edge line rumble stripes.

Most of the studies that have used lane departures or deviation as a crash surrogate have used roadway based data collected using a video data collection setup or road tubes. The advantage of using a roadway based measuring system is the availability of exposure information over a long period of time. The main disadvantage it is usually only applied in a few locations due to resource constraints and the number of variables that can be collected is limited to a few roadway and prevailing environmental conditions. Driver variables usually cannot be included. Other alternatives for collection of lane position data are driving simulators and naturalistic driving studies (NDS). These data collection techniques provide additional variables and can be applied over a number of roadways. However, cost for conducting these studies can increase quite rapidly.

Project Scope

The purpose of this study was to explore use of existing NDS data to assess the feasibility of answering lane departure research questions. The knowledge gained can be applied to the full-scale study Strategic Highway Research Program (SHRP) 2 and other large scale naturalistic driving studies. SHPR 2 will encompass a rich naturalistic dataset on 1,950 instrumented passenger vehicles that can be used to answer a number of research questions including lane departures.

The value of naturalistic data is the ability to observe safety critical as well as non-safety critical events over time; something that is not possible in crash data only. Moreover, researchers are given the opportunity to observe drivers in their natural driving setting and interactions among different traffic, road, and environmental conditions can be assessed and incorporated in various analysis tools. The disadvantage of naturalistic data is the lack of control for establishing causality of the circumstances leading to an event. However, the SHRP 2 data collection addresses this downfall by providing data over a prolonged period (at least one year). Hence, there is a high likelihood of observing the same roadway type and conditions for commuters and trip repeaters.

Another strength of naturalistic driving studies is the substantial number of normal driving data that will be available. This can provide information on how often drivers are exposed to a particular set of circumstances. Currently, there is no realistic method to obtain exposure data for an individual driver, and it is even difficult to obtain detailed exposure for a cohort of drivers.

This study attempts to demonstrate the benefits of naturalistic data using lane departure and normal driving events that were identified and extracted from a NDS dataset from the University of Michigan Transportation Research Institute (UMTRI). Logistic regression analyses were then used to predict the likelihood of a lane departure as influenced by driver, roadway, and environmental factors that can only be gathered from naturalistic data.

DATA

Studies show that there are many relevant roadway factors that impact lane departure crashes including the presence and characteristics of horizontal curves, vertical curvature, lane width, shoulder type, shoulder width, median type, driveway density, shoulder rumble strips, centerline rumble strips, and roadway delineation and signing (Miaou et al, 1993; Hauer et al, 2004; Luediger et al, 1988; Council, 1998; Vogt and Bared; 1998; Zegeer et al, 1992; Deng et al., 2006; Garder and Davies, 2006; Corkle et al., 2001; Miles, 2005; Sun et al., 2007; Donnell et al., 2006). Environmental factors identified were pavement surface condition, roadway lighting, and precipitation (Deng et al., 2006; Shankar et al., 1998; McLaughlin et al., 2009).

Those roadway, driver and vehicle variables that were likely to contribute to the occurrence and severity of lane departures and by extension lane departure crashes were extracted from the NDS as available. Driver factors which have been correlated to lane departure crashes include influence of alcohol or drugs, speed, age, gender, and distraction/inattention (Dissanayake, 2003; McGinnis et al., 2001; Khattak and Hummer, 1998; McLaughlin et al., 2009; Ulmer et al., 1997; Williams et al., 1997). Another benefit of naturalistic data is the ability to quantify these driver factors.

Data source

Data were extracted from a field operational test for a road departure curve warning (RDCW) system conducted by the University of Michigan Transportation Research Institute (UMTRI). The 11 vehicles (same make and model) in the study included an instrumentation packages that encompassed a variety of sensing systems, including a forward video and driver face video, forward and side radar, and global positioning system (GPS). The RDCW system also utilized a lane tracking system that calculated lane position based on vehicle position relative to lane lines or roadway edge (LeBlanc et al., 2006).

RDCW data included 78 drivers that were evenly split by gender and age. Although the purpose of the RDCW data was to test the collision warning system, data were collected for a one-week period prior to activation of the system for each driver to use as a baseline. During the first week of driving, the system was recording data but alerts were not provided to the driver. As a result, the first week of data collection reflected naïve driving with no in-vehicle warning system alerts. Data used in this project were from that first week of naïve driving.

Data were requested from UMTRI for rural 2-lane roadways for the time periods when a vehicle was likely to have left its lane as well as periods of normal driving. UMTRI provided a database and forward imagery for 44 different drivers. The database contained a number of fields with data from the instrumentation system, such as lateral acceleration, forward speed, etc. Each row of data represented 0.1 second of data. GPS data provided vehicle position that can be overlaid with aerial imagery or roadway data. Lane width and vehicle position in relationship to the lane was provided through the lane tracking system. Forward imagery was provided and was usually available at 2 Hz (2 per s or 1 image per 5 rows of vehicle trace data).

Data Reduction

Lane departures were selected as the crash surrogate of interest. Data were partioned into events by lane departures and normal driving. The occurrence of a lane departure event was determined by calculating vehicle wheel path using vehicle offset, lane width, and track width from the NDS lane tracking system. A lane departure was defined as a vehicle wheel path crossing over the right (right-side lane departure) or left (left-side lane departure) lane line and encroaching upon either the shoulder or the adjacent lane by 0.1 m or more. The threshold 0.1 m was used as a buffer because there is some uncertainty in estimation of wheel path. In all cases, the vehicle departed the lane and then returned to the initial lane of travel without losing control. It should be noted that some of the left-side lane departures for either curve direction, may have been drivers intentionally crossing the centerline (i.e. "cutting the curve"). In future studies, it may be possible to ascertain this from the driver face video and from driver hand position on the steering wheel. However, for this study the team did not have access to this information.

The data reduction resulted in a total of 22 right-side lane departure and 51 left-side lane departure events for two-lane rural roads. It also resulted in over 113,000 observations (0.1 s data frames) of normal driving.

Data for each lane departure event were aggregated to a level for a logistic regression analysis. The start point for each lane departure was defined as the point that the vehicle began deviating from its path towards the edge of the lane. The end point of the event was the point after the vehicle returned to the roadway and corrected its path. The start and end times were noted at those points and data were summarized into a single observation. A lane departure event included time spent drifting from the roadway or lane, time off the roadway or lane, and time returning to the original lane of travel. The length of time did vary for each lane departure event and a variable was included in the model to account for differences that time interval may have on the probability of a right or left lane departure.

Data for which no lane departure had occurred were used to represent normal driving data. Driving traces received from UMTRI included segments of a vehicle trip along a rural 2-lane road. Sections of the driving trace when no lane departure event occurred were divided into epochs where roadway and environmental conditions were consistent. When a change in roadway characteristics occurred, a new epoch was created. For instance, data along a tangent section would be marked as one epoch if the roadway cross-section did not change. If cross section changed such as a major change in shoulder width, a new epoch was created. Vertical curvature was not available so this could not be included as a changing roadway characteristic. When the vehicle encountered a curve, a new epoch would be created that contained all of the vehicle activity on the curve. At the end of the curve, a new epoch would be created for the next tangent section. Data could not be partitioned by driver characteristics because dynamic driver characteristics were not available and static driver variables such as age and gender did not change. In most cases, environmental conditions were consistent across a roadway section, so it was not necessary to consider changes in environmental conditions for a particular trace. Data were summarized for each epoch. The length of each epoch was different because drivers spent different amounts of time driving on a particular type of roadway. The number of 0.1 s intervals for each epoch was included as a factor in the analysis to account for any differences that intervals may have on the outcome.

Several roadway variables were included with the NDS dataset. Lane width was calculated from the lane tracking system. Radius was also reported with the NDS data but did not appear to be accurate. A number of other variables that were not provided in the UMTRI data were extracted or created from either the UMTRI data or from other available data sources. Other data sources included aerial imagery, a roadway database, and a crash database for Michigan.

All of the data elements that the team determined were important from the literature and could be obtained from one of the available databases (vehicle data, aerial imagery, roadway data, forward imagery, and crash database) were extracted. For instance, curve radius and direction were determined by overlaying the vehicle database with aerial imagery and determining the start and end point in the vehicle data that corresponded to each curve, while curve radius was measured using the aerial imagery.

A qualitative assessment of the quality of pavement markings, and shoulder type were obtained from a review of the forward image. Shoulder width was estimated from the forward view by using the known distance of lane width as a reference. Annual average daily traffic (AADT) was obtained from the Michigan Department of Transportation. The forward view was used to tabulate the number of on-coming vehicles that passed the subject vehicle during the segment. On-coming vehicle density (vehicles/meter) was calculated using this information. Driveway density was calculated by identifying the number of driveways for each segment and then dividing the number of driveways by total segment length resulting in the variable "*DwyDensity.*"

Driver age and gender were available in the NDS dataset. The fraction of time a driver spent traveling over the posted or advisory speed limit was calculated for each driver using all of the observations of data that were available for that driver. Posted speed limit and advisory speed for the segment were determined either through the Michigan road database or from a review of the forward imagery. Driver speed for each time interval was compared against the posted or advisory speed. The intervals where a driver exceeded the posted or advisory speed by 5 or 10 mph were divided by total intervals for that vehicle trace. The resulting variable was fraction of time a driver spent traveling 5 or 10 mph over the speed limit. Time spent traveling over the posted or advisory speed was used as a measure of driver aggressiveness.

Time of day was identified as nighttime or daytime based on time and the forward view. Environmental conditions were assessed but only dry roads were present for the data obtained. No overhead street lighting was present on any of the roadways where NDS data were available.

Lane departure crash density was calculated by overlaying segments with the Michigan crash database (2000 to 2006). The number of lane departure crashes was summed and divided by the total segment length resulting in the variable, lane departure crash density (crashes per meter). This resulted in the variable "*CrashDensity.*"

A number of driver variables, such as driver distraction or driver glance location, were not be available since driver face video was not included due to IRB limitations. This type of data will be available in the full-scale SHPR 2 study and are typically available in other NDS.

METHODS

Separate logistic regression models were developed for right-side and left-side lane departures. For each model, data recorded for lane departures were used as the cases, while records that included no lane departures were considered controls (or normal driving). Each lane departure event or normal driving epoch was modeled as one observation. As noted earlier, the number of 0.1 s intervals for each epoch was included as a factor in the analysis to account for any differences that interval length may have on the outcome. A list of the explanatory variables considered for the analysis is shown in Table 1. Both models were created using the LOGISTIC procedure in the SAS/STAT 9.2 software package.

The response variable for lane departure (Z) was coded as 0 if there is no lane departure (normal driving) and 1 if a lane departure occurred (either right or left side departure). Given that Z is a Bernoulli outcome with p=P(Z=1) as the probability of occurrence of a lane departure, the odds that a lane departure happens is p/(1-p). In order to link the odds of a lane departure to the matrix of explanatory variables, X, the logit link function was used. Hence, a connection between the probability of a lane departure and the linear combination of explanatory variables (X's) using equation 1:

$$\text{logit}(p) = \log(p/(1-p)) = \beta_0 + \beta_1 X_1 + \ldots\ \beta_k X_k \qquad \text{(eq.1)}$$

Stepwise selection was used to determine which variables were relevant and should be included in the model. For each step, a covariate was added to the model if the significance level for entry was met (0.1 was used). Then the chi-square statistic was computed. If the covariate satisfied the significance level (0.1), it was included in the model. The Akaike Information Criteria (AIC) and Schwarz criterion (SC) were used to compare models and determine which variables to include in the final model.

Table 1: Explanatory Variables Used in Analysis

Variable	description	Type	Values
driver	driver ID, Included to account for repeated measurements		NA
Age	driver age category	categorical	0: 20 to 30 years old 1: 31 to 59 years old 2: 60 to 70 years old
Gender	driver gender	categorical	1: male 2: female
Curve	type and direction of curve	categorical	0: tangent 1: right curve 2: left curve
Radius	curve radius (meters)	continuous	98 to 1,717 tangent indicated as 9999
LaneWidth	lane width (meters)	continuous	3.0 to 4.7
AADT	volume (vpd)	continuous	11 to 57410
ShldWidth	shoulder width (meters)	continuous	0.8 to 5.0
Density	on-coming vehicles per meter	continuous	0.0 to 0.5
PvmMarking	pavement marking condition	categorical	0: highly visible 1: visible 2: obscure
TimeOfDay	time period	categorical	0: day 1: dusk/night
CrashDensity	lane departure crashes per meter	continuous	0.0 to 0.029
DwyDensity	driveways per meter	continuous	0.0 to 0.027
OvrSpd5	Fraction of time driver traveled 5 mph over the speed limit	continuous	0.0 to 0.90
OvrSpd10	Fraction of time driver traveled 10 mph over the speed limit	continuous	0.0 to 1.0
ShldType	shoulder type	categorical	1: paved 3: gravel 4: earth 6: no shoulder 7: partially paved

Only a small sample of left- and right-side lane departures was available (51 and 22, respectively). As a result, it was not possible to evaluate the significance of all variables and test correlations between variables. In order to build a model that best represented the data, the decision to remove variables from the model was based on whether it was expected that there would be correlation among input variables. Maximum likelihood (ML) method was used to calculate the coefficient estimates, and the Wald statistic was used to test the significance of each explanatory variable.

A variable labeled "Observation" was also included in the model to indicate the frequency of occurrence of each observation. Odds ratios were used to assess whether a specific condition was more or less likely to result in a lane departure. An odds ratio greater than 1 indicated that the odds of a lane departure occurring are higher, and an odds ratio less than 1 revealed lower odds. Hosmer and Lemeshow Goodness-of-Fit Test is used and Large Chi-Square values (and small p -values) indicate a lack of fit of the model.

RESULTS

Left-Side Lane Departures

There were several variables considered initially, but not included in the final model for several reasons. The covariates "AADT", "ShoulderType", and "RoadSurf" were not significant after the stepwise model selection was employed. "Density", "CrashDensity", "DwyDensity", and "OvrSpd5" were signficant but their coefficents were quite small indicating no practical signficance (parameter estimate << 0.001). As a result, these variables were removed from the model. As one would anticipate, "Curve" and "Radius" are highly correlated with each other. Thefore, only "Radius" was included as it provided a better model fit. The final model for the left-side lane departure is:

$$\log\left(\frac{P(LD)}{1-P(LD)}\right) = -0.3097 + 0.5746 * I_{\text{Age}}(0) + 0.4118 * I_{\text{Age}}(1) + 0.5197 * I_{Gender}(1) - 0.00025 * Radius - 0.7282 * LaneWidth + 0.3193 * ShoulderWidth - 0.9096 * I_{PvmMarking}(0) + 0.2320 * I_{PvmMarking}(1) - 0.6147 * I_{TimeOfDay} - 1.4494 * OvrSpd10 \qquad \text{(eq. 2)}$$

where P(LD) indicates the probability that a left-side lane departure occurs. The odds ratio (OR) estimates are shown in Table 2.

From the model above, the relationship between any change in explanatory variable and the change in the probability of getting left-side lane departure can be determined. For a non-categorical variable, let $\hat{\beta}_i$ be the estimation of coefficient for *ith* explanatory variable, as the *ith* explanatory variable increases one unit, the odds of a left-side lane departure will change $e^{\hat{\beta}_i}$. Since the exponential function is an increasing function, the positive sign of $\hat{\beta}_i$ means the increase in the odds of left-side lane departure occuring and a negative sign of $\hat{\beta}_i$ means the a decrease in the odds of a left-side lane departure. For example, for explanatory variable, *LaneWidth*, a one meter increase in lane with, the estimated odds of a left-side lane departure multiply by 0.483 times; that is, they decrease by 0.517 times.

Table 2: Results for the Left-Side Lane Departure Model

Variable	Condition	Estimate	Std Error	p-value	OR 95 percent lower	OR estimate	OR 95 percent upper
Intercept		-0.3097	0.3013	0.3040			
Age	0 vs 2	0.5746	0.0529	<.0001	1.602	1.776	1.970
Age	1 vs 2	0.4118	0.0528	0.0360	1.361	1.510	1.674
Gender	1 vs 2	0.5197	0.0423	<.0001	1.548	1.682	1.827
Radius		-0.00025	3.662E-6	<.0001	1.00	1.00	1.00
LaneWidth		-0.7282	0.0726	<.0001	0.419	0.483	0.557
ShldWidth		0.3193	0.0229	<.0001	1.316	1.376	1.439
PvmMarking	0 vs 2	-0.9096	0.1180	<.0001	0.32	0.403	0.507
PvmMarking	1 vs 2	0.2320	0.0876	.0081	1.062	1.261	1.497
TimeOfDay	0 vs 1	-0.6147	0.0373	<.0001	0.503	0.541	0.582
OvrSpd10		-1.4494	0.1052	<.0001	0.191	0.235	0.288

The coefficient estimates for agegroups 0 (20 to 30 year olds) and 1 (31-59 year olds) are reported in comparison to agegroup 2 (60 to 70 year olds). Hence, the odds of a left-side lane departure for drivers aged 20 to 30 years old compared to drivers aged 60 to 70 is given by:

$$\exp(\text{age} = 0 \text{ vs } 2) = \exp(0.5746) = 1.78 \qquad \text{(eq. 3)}$$

Consequently drivers aged 20 to 30 are 1.78 times more likely to be involved in a left lane departure than older driver. Similarly, the odds of a left-side lane departure for middle aged drivers (age = 1) compared to older drivers (age = 2) is 1.51. This indicates that middle-aged drivers 31 to 59 years old are 1.51 times more likely to be involved in a lane departure than older drivers. Middle aged drivers have 0.85 times the odds of being involved in a lane departure compared to their counterparts aged 20 to 30 years old. Conversely, the odds for a younger driver compared to a middle aged driver are 1/0.85 = 1.18.

Based on similar calculations, males are 1.68 times more likely to be involved in a lane left lane departure than females. The negative coefficient for “Radius” indicates that the odds of a left side lane departure decrease as radius increases. A very large radius value of “9999” was used for tangent sections and the variable was modeled as a continuous variable. For each 30.48 meter increase (approximately 100 feet) in radius the odds of having a left side lane departure decrease by 0.99. So a 100 foot increase in radius results in an approximate 1% decrease in the odds of a lane departure.

The positive coefficient for shoulder width indicates that as shoulder width increases, the odds of a left lane departure also increase. This result was unexpected since increased shoulder width has generally been correlated to a decrease in lane departure crashes.

Highly visible lane markings (*PvmMarking* = 0) had much lower odds (0.403) of a left lane departure than lane markings indicated as obscure (*PvmMarking* = 2) while moderate visible lane markings (*PvmMarking* = 1) had higher odds (1.062) of having a lane departure than obscure markings (*PvmMarking* = 2).

As noted in Table 2, so daytime crashes have 0.54 times the likelihood of a left lane departure than nighttime crashes or nighttime crashes have 1/0.54 = 1.85 times the odds than during daytime hours.

The negative coefficient for the explanatory variable "*OvrSpd10*" indicates that drivers who spend a greater fraction of their time traveling at 10 or more mph over the posted or advisory speed have a lower odds of a left lane departure. This result is somewhat counter-intuitive, however. The opposite effect was found for right side lane departures so drivers who regularly speed may be more likely to stay towards the right-side of their lane.

Right-Side Lane Departures

Equation 4 describes the final model for right-side lane departure events. The covariates "*OvrSpd5*", "*Gender*", "*ShldWidth*", "*TimeOfDay*", and "*DwyDensity*", were not statistically significant. The covariate, "*CrashDensity*", was statistically significant but the parameter estimate did not provide a practical significance and was removed from the model. As with the left-side lane departure model, "Curve" and "Radius" are correlated. For this model, a better fit model resulted using the "Radius". The final model for a right-side lane departure is given by:

$$\log\left(\frac{P(RD)}{1-P(RD)}\right) = 8.1914 - 1.7341 * I_{\text{Age}}(0) - 1.1016 * I_{\text{Age}}(1) - 0.0003 * Radius - 0.0001 * AADT - 2.1367 * LaneWidth - 5.3996 * Density + 0.2273 * I_{PvmMarking}(0) - 1.9341 * I_{PvmMarking}(1) + 2.4446 * OvrSpd10 + 2.4232 * I_{ShldType}(1) - 1.2771 * I_{ShldType}(3) - 3.2313 * I_{ShldType}(4) + 2.4946 * I_{ShldType}(6) \qquad \text{(eq. 4)}$$

where *P(RD)* indicates the probability that a right-side lane departure occurs. The odds ratio estimates are shown in Table 3.

Table 3: Results for the Right-Side Lane Departure Model

Variable	Condition	Estimate	Std Error	p-value	OR 95 percent lower	OR estimate	OR 95 percent upper
Intercept		8.1914	0.4943	<.0001			
Age	0 vs 2	-1.7341	0.1337	<.0001	0.136	0.18	0.229
Age	1 vs 2	-1.1016	0.0891	<.0001	0.279	0.33	0.396
Radius		-0.0003	6.768E-6	<.0001	1.00	1.00	1.00
AADT		-0.0001	7.115E-6	<.0001	1.00	1.00	1.00
LaneWidth		-2.1367	0.1321	<.0001	0.091	0.12	0.153
Density		-5.3996	1.4526	0.0002	<0.001	0.01	0.078
PvmMarking	0 vs 2	0.2273	0.1086	0.0364	1.015	1.26	1.553
PvmMarking	1 vs 2	-1.9341	0.0959	<.0001	0.120	0.15	0.174
OvrSpd10		2.4446	0.1151	<.0001	9.198	11.53	14.441
ShldType	1 vs 7	2.4232	0.0914	<.0001	9.431	11.28	13.496
ShldType	3 vs 7	-1. 2771	0.0786	<.0001	0.239	0.28	0.325
ShldType	4 vs 7	-3.2313	0.1219	<.0001	0.031	0.04	0.050
ShldType	6 vs 7	2.4946	0.1646	<.0001	8.775	12.12	16.731

Results are interpreted similar to that of the left-side lane departure model. The negative coefficient for "*LaneWidth*" indicates that for each meter increase in lane width, the odds of a right lane departure decreases by 0.882 times. Similarly, as on-coming traffic density and volume (*AADT*) increase, the odds of a right side lane departure decrease which may be due to improved lane keeping.

The negative coefficient for "*Radius*" indicates that the odds of a right side lane departure decrease as radius increases. A very large radius value of "9999" was used for tangent sections and the variable was modeled as a continuous variable. For each 30.48 meter increase (approximately 100 feet) in radius the odds of having a right-side lane departure decrease by 0.99. So a 100 foot increase in radius results in an approximate 1% decrease in the odds of a right lane departure.

The odds of a right-side lane departure for drivers aged 20 to 30 years old compared to drivers aged 60 to 70 is 0.18 times the odds of being involved in a right lane departure than older drivers. Similarly, the odds of a right-side lane departure for middle aged drivers (age = 1) compared to older drivers (age = 2) is 0.33 indicating that middle aged drivers 31 to 59 years old are are less likely to be involved in a lane departure than older drivers. And the odds of a left-side lane departure for middle aged drivers compared to younger drivers is 1.88.

The impact of highly visible pavement markings (*PvmMarking* = 0) versus obscure pavement markings (*PvmMarking* = 2) is given by 1.25 indicating that right lane departure were more likely to occur when highly visible pavement markings were present although this result is not consistent with the concept that better lane delineation will result in fewer lane departures. Alternatively the impact of visible pavement markings, (*PvmMarking* = 1), compared to obscure pavement markings, (*PvmMarking* = 2), is 0.15 so right side lane departures are much less likely with visible pavement markings.

The model also indicates a strong positive relationship exists between the amount of time a driver spent driving 10 or more miles per hour over the posted speed limit and the likelihood of a right lane departure.

Shoulder type was also relevant in the model. The coefficients indicate that paved shoulders (*Shldtype* = 1) are more likely to have a right lane departure than partially paved (*ShldType* = 7) while gravel and earth shoulders are more likely to have a lane departure than partially paved. A positive coefficient for no shoulders (*ShldType* = 6) versus partially paved indicates that a right side lane departure was much more likely when no shoulder was present than when shoulders were partially paved. Other relationships between shoulder types are provided in Table 4. As indicated, all shoulder types had less likelihood of a right lane departure than no shoulder. Paved shoulders were more likely to result in a right lane departure than gravel, earth, or partially paved shoulders. Although this is counterintuitive, it may be due to the fact that drivers are less likely to lane keep with a paved shoulder since there is less risk of a severe outcome if the tire leaves the travel way. Paved shoulders have been shown to reduce number of crashes (Hallmark et al, 2010) so the impact of a paved shoulder may be a less severe outcome to a lane departure.

Table 4: Comparison of Lane Departure Likelihood by Shoulder Type

	Gravel	**Earth**	**No shoulder**	**Partially paved**
Paved	40.44	282.05	0.931	11.28
Gravel	—	6.98	0.023	0.28
Earth	—	—	0.003	0.04
No shoulder	—	—	—	12.12

CONCLUSIONS AND DISCUSSION

Naturalistic data collection can provide insights on lane departures that may not have been gained otherwise. In this study, binary logit models were used to examine the likelihood of a right or left side lane departures. This could not have been possible using crash data that does not have a control (or normal driving) outcome. We do recognize that a larger number of departures may have been observed in a controlled setting, but the breadth of explanatory variables incorporated into the model would not have been as comprehensive.

This study demonstrated that in addition to age and gender, the radius of curvature impacted the likelihood of a lane departure. Although studies on age and gender have clearly been documented, this study also brings to light the impact of the road. Although it may seem obvious that greater radii would results in increase lane departure, studies have not actually captured the degree to which radius, lane and shoulder width, and even pavement marking may influence lane departures. This study also brings to light the differences between a right and left lane departure.

Left side lane departures were less likely as lane width and curve radius increase. They were also less likely in daytime compared to nighttime (OR = 0.54) and were more likely for males compared to females (OR = 1.68). Younger drivers (age 20 to 30 years old) were more likely to have a left side lane departure than older drivers, aged 60 to 70 (OR = 1.776) and were slightly more likely to be involved in a left lane departure (OR = 1.18) than their middle aged counterparts (ages 31 to 59). Middle aged drivers were more likely to be involved than their older counterparts, however (OR = 1.51). Results indicate that an increase in shoulder width increases the odds of a left-side lane departure although shoulder width has generally been correlated to a decrease in crash rate. Pavement marking condition was also relevant. The amount of time a driver spends at 10 or more mph over the speed limit decreased the odds of a left lane departure. Since the opposite result was found for right-side lane departures, it is speculated that drivers who speed may tend to stay towards the right side of their lane.

The right-side lane departure model indicated that an increase in lane width, radius, on-coming vehicle density, and an increase in volume, the odds of a right side lane departure decrease which may be due to improved lane keeping. The amount of time a driver spend traveling at 10 or more mph over the posted or advisory speed increased the odds of a right-side lane departure. Pavement marking condition and shoulder type were also relevant variables.

Results of the study indicated several relationships which are not intuitive. This may be due sample size. Correlation between variables was examined but correlation with variables that were not considered may have been present. Additionally, the impact of some variables may be different than what was expected. For instance, an increase in shoulder width resulted in an increase in left side lane departures. While a wider shoulder may decrease crash risk or severity if a driver leaves the roadway, a driver may be less likely to lane keep when a wide shoulder is present than with a narrow shoulder. The left lane departure model also indicated that drivers who spend more time traveling over the speed limit are less likely to have a lane departure. Aggressive drivers may be more likely to lane keep even though the consequences of leaving their lane are more likely to be severe. The opposite effect was found in right side lane departures where an increased amount of time traveling over the speed limit resulted in an increased odds of having a right side lane departure. Results that were counterintuitive in the right side lane departure model include an increase in the odds of a lane departure as lane width increases and presence paved shoulder had higher odds of a right lane departure than any other type of shoulder. This may be due to drivers paying more attention and lane keeping better when lanes are narrow or no shoulders are present.

STUDY LIMITATIONS

The study provided useful information which can be used to better understand why lane departures occur and the outcomes do demonstrate the value of using naturalistic data that could not have been observed otherwise. However, there are several limitations which should be acknowledged. First, the sample size was limited due to the available data which may have some consequences for the statistical models. For example, the coefficients for several covariates were not as intuitive as expected and the small sample size may not been sufficient to develop a robust model. Results may also have been affected by correlations which were not noted in the model. A larger dataset, such as the one being collected as part of the SHRP 2 program can solidify the results more concretely.

ACKNOWLEDGEMENTS

The team would like to thank the Second Strategic Highway Research Program (SHRP 2) for funding this project, the Midwest Transportation Consortium (MTC), a Tier 1 university transportation center, for partial funding of three students participating in this project, the University of Michigan Transportation Research Institute (UMTRI) for providing naturalistic driving study data and for their assistance in working with the data, and for the Michigan DOT for providing roadway and crash data. Any results and conclusions resulting from the use of this data represent the work and opinions of the authors and do not reflect the opinion or endorsement of UMTRI, MTC, SHRP 2, or the Michigan DOT.

REFERENCES

Archer, J. *Traffic Conflict Technique: Historical to Current State-of-the-Art*. Institutionen for Infrastruktur. 2001.

Bonneson, J., and K. Zimmerman. *Red-Light-Running Handbook: An Engineer's Guide to Reducing Red-Light-Related Crashes.* Texas Transportation Institute, Texas A&M University, College Station. 2004.

Burgett, A., and R.J. Miller, Jr. *A New Paradigm for Rear-end Crash Prevention Driving Performance.* National Highway Traffic Safety Administration. 2001.

Corkle, J., M. Marti, and D. Montebello. Synthesis on the effectiveness of rumble strips. Report no. MN/RC-2002-07. Minnesota Department of Transportation. 2001.

Council, F. M. "Safety Benefits of Spiral Transitions on Horizontal Curves on Two-Lane Rural Roads." *Transportation Research Record 1635*. pp. 10-17. 1993.

Deng, Z., J.N. Ivan, and P. Garder. "Analysis of Factors Affecting the Severity of Head-On Crashes: Two-Lane Rural Highways in Connecticut." *Journal of the Transportation Research Board*, No. 1953. pp. 137-146. 2006.

Dissanayake, S. "Young Drivers and Run-Off-The-Road Crashes." *Proceedings of the Mid-Continent Transportation Research Symposium*. 2003.

Donmez, B., L. Boyle, and J. Lee. "The Impact of Driver Distraction Mitigation Strategies on Driving Performance." *Human Factors.* Vol. 48, No. 4. pp. 785-804. 2006.

Donnell, E. T., and J.M. Mason. *Predicting the Frequency of Median Barrier Crashes on Pennsylvania Interstate Highway.* Accident Analysis and Prevention. Vol. 38. pp. 590-599. 2006.
FHWA. Roadway Departure Safety. Federal Highway Administration. http://safety.fhwa.dot.gov/roadway_dept. Accessed April 2011.

FHWA. Stop Red Light Running Facts and Statistics, U.S. Department of Transportation. Federal Highway Administration. August 3, 2006. http://safety.fhwa.dot.gov/intersections/redl_facts.htm. Accessed April 2011.

Fitzsimmons, E.J., S. L. Hallmark, M. Orellana, T. McDonald, and D. Matulac. "Investigation of Violation Reduction at Intersection Approaches with Automated Red Light Running Enforcement Cameras in Clive, Iowa Using a Cross-Sectional Analysis." ASCE Journal of Transportation Engineering. Volume 135, Issue 12. pp. 984-989. December 2009.

Garber N.J, J.S. Miller, S. Eslambolchi, R. Khandelwal, K.M. Mattingly, K.M. Sprinkle, and P.L. Wachendorf. *An Evaluation of Red Light Camera (Photo-Red) Enforcement Programs in Virginia: A Report in Response to a Request by Virginia's Secretary of Transportation*. Virginia Transportation Research Council, Charlottesville, Virginia. January 2005.

Garder, P. and M. Davies. "Safety Effect of Continuous Shoulder Rumble Strips on Rural Interstates in Maine". *Proceedings,* 85th Annual Meeting of the Transportation Research Board, Washington, D.C. 2006.

Hallmark, S.L., T. McDonald, Y. Tian, and D.J. Andersen. *Safety Benefits of Paved Shoulders In Iowa*. August 2010. http://www.intrans.iastate.edu/research/detail.cfm?projectID=832537926

Hauer, E., F.M. Council, and Y. Mohammedshah. "Safety Models for Urban Four-Lane Undivided Road Segments." *Transportation Research Record*, No. 1897. pp. 96-105. 2004.

Hayward, J. "Near-miss Determination Through Use of a Scale of Danger." *Highway Research Record*, No. 384. pp. 24-34. 1972.

Khattak, A. J., and J.E. Hummer. "Analysis of Run-off-the-Road Collisions on Two-Lane Rural Highways in North Carolina." *ITE Journal*, Vol. 68, No. 10. p. 18. 1998.

LeBlanc, D., J. Sayer, C. Winkler, R. Ervin, S. Bogard, J. Devonshire, M. Mefford, M. Hagan, Z. Bareket, R. Goodsell, and T. Gordon. *Road Departure Crash Warning System Field Operational Test: Methodology and Results.* University of Michigan Transportation Research Institute, Ann Arbor, Michigan. 2006.

Lamm, R., E.M. Choueiri, J.C. Hayward, and A. Paluri. "Possible Design Procedure to Promote Design Consistency in Highway Geometric Design on Two-Lane Rural Roads." *Transportation Research Record*, Vol. 1195. pp. 111-122. 1988.

McGinnis, R. G., M.J. Davis, and E.A. Hathaway. "Longitudinal Analysis of Fatal Run-off-Road Crashes, 1975 to 1997." *Journal of the Transportation Research Record*, Vol. 1746. pp. 47-58. 2001.

Neuman, T.R., R. Pfefer, K.L. Slack, K. K. Hardy, F. Council, H. McGee, L. Prothe, and K. Eccles. NCHRP 500. Volume 6. A Guide for Addressing Run-Off-Road Collisions. 2003.

McLaughlin, S. B., J.M. Hankey, S.G. Klauer, and T.A. Dingus, T. A., *Contributing Factors to Run-Off-Road Crashes and Near-Crashes*, National Highway Traffic Safety Administration, DOT HS 811 079. January 2009.

Miaou, S. *Estimating Roadside Encroachment Rates with the Combined Strengths of Accident and Encroachment-Based Approaches*. FHWA-RD-01-124. 2001.

Miaou, S. P. and H. Lum. "Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements." *Transportation Research Record* No. 1407. pp. 11-24. 1993.

Miles, J., M. Pratt, and P. Carlson. "Evaluation of Erratic Maneuvers Associated with Installation of Rumble Strips." *Journal of the Transportation Research Record,* 1973. pp. 73-79. 2006.

Miles, 2005 J.D. Miles, P.J. Carlson, M.P. Pratt, and T.D. Thompson. Traffic Operational Impacts of Transverse, Centerline, and Edgeline Rumble Strips (0-4472-2). TxDOT. http://tti.tamu.edu/documents/0-4472-2.pdf . 2005.

Porter, R., E. Donnel, and K. Mahoney. "Evaluation of Effects of Centerline Rumble Strips on Lateral Vehicle Placement and Speed." *Journal of the Transportation Research Record,* 1862. pp. 10-16. 2004.

Pratt, M. J. Miles, and P. Carlson. "Evaluation of Operational Impacts of Installation of Centerline and Edge Line Rumble Strips." *Journal of the Transportation Research Record*, 1973. pp. 80-88. 2006.

Retting, R.A., S.A. Ferguson, and C.M. Farmer. *Reducing Red Light Running Through Longer Yellow Signal Timing and Red Light Camera Enforcement: Results of a Field Investigation.* Insurance Institute for Highway Safety. 2007.

Shankar, V. N., R. B. Albin, , J.C. Milton, and F.L. Mannering. "Evaluating Median Crossover Likelihoods with Clustered Accident Counts: An Empirical Inquiry Using the Random Effects Negative Binomial Model." *Transportation Research Record*, No. 1635 pp. 44-57. 1998.

Songchitruksa, P. and A. Tarko. "The Extreme Value Theory Approach to Safety Estimation. *Accident Analysis and Prevention*, 38. 811-822. 2006.

Sun, X., J. Park, V.O. Tekell Jr., and N.S. Ludington. "Evaluation of Pavement Edge Lines on Narrow Highways with Low Traffic Volume in Louisiana." *Journal of the Transportation Research Record*, No. 1989. pp. 115-122. 2007.

Taylor, W., G. Abu-Lebdeh, and S. Rai. "Effect of Continuous Shoulder Rumble Strips and Pavement marking on Lateral Placement of Vehicles." *Journal of the Transportation Research Record,* 1911. pp. 105-112. 2005.

Vogt, A., and J. Bared. "Accident Models for Two-Lane Rural Segments and Intersections." *Transportation Research Record*, No. 1635. pp. 18-29. 1998.

Ulmer, R. G., A.F. Williams, and D.F. Preusser. "Crash Involvements of 16-year-old Drivers." Journal of Safety Research Vol. 28 No. 2. pp. 97-103. 1997.

Williams A.F., A.K. Lund, D.F. Preusser and R.D. Bloomberg. " Results of a seat belt use law enforcement and publicity campaign in Elmira, New York". *Accident Analysis and Prevention* 19, pp. 243–249. 1997.

Zegeer, C. V., R. Stewart, F.M. Council, D.W. Reinfurt, and E. Hamilton. "Safety Effects of Geometric Improvements on Horizontal Curves." *Transportation Research Record*, No. 1356. pp. 11-19. 1992.