# Safety Evaluation of Horizontal Curves on Rural Undivided Roads

**By**

**Ghazan Khan, Ph.D.**(1)
Research Associate, Traffic Operations and Safety (TOPS) Laboratory
Department of Civil and Environmental Engineering University of Wisconsin-Madison
2205 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706
Phone: (608) 698-8519, Fax: (608) 262-5199
Email: gkhan@wisc.edu

**Andrea R. Bill**
Associate Researcher, Traffic Operations and Safety (TOPS) Laboratory
Department of Civil and Environmental Engineering University of Wisconsin-Madison
2205 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706
Phone: (608) 698-8519, Fax: (608) 262-5199
Email: gkhan@wisc.edu

**Madhav Chitturi, Ph.D.**
Assistant Researcher, Traffic Operations and Safety (TOPS) Laboratory
Department of Civil and Environmental Engineering University of Wisconsin-Madison
2205 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706
Phone: (608) 262-3649, Fax: (608) 262-5199
Email: mchitturi@engr.wisc.edu

**David A. Noyce, Ph.D., P.E.**
Professor, Director, Traffic Operations and Safety (TOPS) Laboratory
University of Wisconsin-Madison, Department of Civil and Environmental Engineering
1204 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706
Phone: (608) 265-1882
E-mail: noyce@engr.wisc.edu

**TRB Manuscript No: 13-4435**

Prepared for the 92nd Annual meeting of the
Transportation Research Board, Washington, D.C. January, 2013
Length of Paper: 4,998 words, 2 Figures + 7 Tables @ 250 words each
7,248 equivalent words

(1) Corresponding author

## ABSTRACT

The objective of this research was to develop total crash and fatal/injury crash prediction models for rural horizontal curves on undivided roads, with focus on three distinct aspects. The first was an emphasis on assembling a high quality large dataset. Crash prediction models were developed using a dataset of 11,427 rural horizontal curves on Wisconsin State Trunk Network roads with over 13 different parameters and four distinct types of crash dataset.

The second focus area was to use regression tree analysis in creating a simple model of horizontal curve safety aimed at practitioners of systemic road safety management and creating subsets of data which warranted further analysis. Regression tree results identified curve radius of approximately 2,500 feet as a significant point below which there is a marked increase in crashes on horizontal curves.

The third focus area of this research was to compare horizontal curve crash prediction models using different crash datasets. Models based on crash dataset with and without crashes in the proximity of intersections were compared. The results show that when crashes on horizontal curves are selected where crash report forms indicate the presence of a horizontal curve, crashes in proximity of intersections do not impact model results significantly; therefore, the inclusion of such crashes would increase the size of dataset benefiting model development.

# INTRODUCTION

In the United States, approximately one-quarter of highway fatalities occur on horizontal curves (1). The average crash rate for horizontal curves is about three times the average crash rate for highway tangents (2). Research indicates that there is greater propensity for severe crashes at horizontal curves as stated in the Texas Transportation Institute's horizontal curve signing handbook (3). Persaud et al. stated that motor vehicle crashes happen more frequently and are more severe on horizontal curves (4). Horizontal curves are necessary element of highways however, they are also likely to cause safety hazards to road users because of the changes in driver expectancy and vehicle handling maneuvers. Schneider et al. provided two explanations from driver awareness perspective; that the driver may be unaware of the approaching horizontal curve, or the driver underestimates the radius or sharpness of the curve (5). In another study, Schneider et al. states that horizontal curves may reduce the driver's available sight distance and reduce vehicle-handling capabilities (6). Therefore, improving safety at horizontal curves is an essential part of an overall safety management plan, which presents the need for developing crash prediction models especially with respect to horizontal curves. The objectives of this research were to develop crash prediction models for different conditions and crash data in order to understand the impacts of various geometric features on horizontal curve safety and gain more insight into this critical safety problem.

# LITERATURE REVIEW

The Federal Highway Administration (FHWA) published a document on providing low-cost safety treatment for horizontal curves signifying the importance of safety at horizontal curves (7). Although there has been some research in the past on safety at horizontal curves, the availability of high quality and large dataset has been the Achilles' heel in past research studies. Literature shows that safety at horizontal curves has been studied from a number of difference perspectives. Different crash types have been used in developing crash prediction models and modification factors e.g. truck-related, motorcycle, run-off-the-road, non-intersection related crashes etc. (5, 6, 8, 9, 10, 11). However, what is not clear is the difference in horizontal curve safety with respect to different types of crash dataset at the same location.

A review of literature shows that run-off-the-road and head-on crashes accounted for 87 percent of all fatal crashes at horizontal curves (2). Another report states that 76 percent of the curve-related fatal crashes involve single vehicles leaving the roadway and striking roadside objects such as trees, utility poles, or rocks (7). The effect of geometric features such as shoulder width may contribute significantly to safety at horizontal curves; an area which has not seen much research in the literature (12). Furthermore, the primary focus of horizontal curve-related safety research has been on two-lane rural roads given that about 75 percent of all curve-related fatal crashes occur in rural areas, and more than 70 percent are on two-lane secondary highways which are mostly local roads (7, 13). Therefore, the focus of this research was also on rural roads; however, all rural roads were considered as part of the dataset rather than just two-lane roads.

## Horizontal Curve Safety Influencing Factors

Many research studies have been conducted to investigate the relationship between crash frequency, severity, and geometric attributes of horizontal curves. Some key factors and research findings are summarized in Table 1.

**TABLE 1 Literature Review Summary of Horizontal Curve Safety Influencing Factors**

| Horizontal Curve Safety Influencing Factors | | |
|---|---|---|
| **Author** | **Factor** | **Summary** |
| Zegeer et. al. (14) | Curve Radius and Degree of Curvature | A 500-ft radius curve is 200% more likely to produce a crash than an equivalent tangent section, and a 1,000-ft radius curve is 50% more likely to produce a crash than an equivalent tangent section. |
| Schneider et al. (5, 6) | | When curves become sharper the model predicts an increase in truck crashes on horizontal curves. The radius and degree of curvature significantly influence motorcycle crashes on horizontal curves |
| Voigt and Krammes (15), Council (16) | | The degree of curvature and radius are significant variables influencing crash rate on horizontal curves. |
| Khan et al. (12) | | Crash rates decreases as radius increases. |
| Miaou and Lum (17) | | Truck crash involvement increases as horizontal curvature (Degree of Curvature) increases. |
| Schneider et al.(6), Zegeer et al. (9) | Curve Length | Curve length as a significant factor for Truck crash involvement. A horizontal curve with a length of 31 m (100 ft.) and a radius of 31 m (100 ft.) on a roadway segment would be expected to have an accident rate over 28 times as high as a tangent section on the same roadway |
| Schneider et al. (5, 6), Khan et al. (12) | Traffic Volume | The increase in passenger vehicle Average Daily Traffic (ADT) is associated with an increase in truck and overall crashes on curves. Also the total ADT also affects motorcycle crashes on curve. |
| Schneider et al.(6), Zegeer et al. (9), Khan et al. (12) | Shoulder Width | Shoulder width is a significant variable that affects crashes on curve. |
| Hallmark (18) | Tangent length before curve | Crash rates on curves with long preceding tangent lengths will be more dangerous when the curve is located on a downgrade of 5% or more, and tangent lengths more than 200 meters. |
| Fitzpatrick et al. (19) | Driveway Density (Curves and Tangent) | There is no significant difference in crash rates on horizontal curves and tangents with same driveway density. |

### Horizontal Curve Crash Prediction Models

Research studies in the past have focussed on developing crash prediction models for horizontal curves predominently using generalized linear models. Caliendo developed a crash prediction model based on a four lane, median divided roads in Italy using ADT, curve length, intersection presence, and radius as factors (20). Schneider et al. developed a model for truck crashes on horizontal curves using length, truck ADT, passenger vehicle ADT, and degree of curvature (5). Persaud et al. developed a model including AADT, length of curve, and curve radius as parameters (6). Other studies have developed crash prediction models for horizontal curves using limited variables. Bonneson et al. developed horizontal curve crash prediction models for multilane highways using radius and speed limit data (21, 22). Fizpatrick developed a crash prediction model for freeways using only the degree of curvature as an independent variable and assuming zero degree as the base condition (23). The Highway Safety Manual (HSM) also provides several Crash Modification Factors (CMFs) for horizontal curves however the standard error values are unknown making the results unreliable (24).

## RESEARCH OBJECTIVE

In light of the literature review, the main objective of this research was to develop crash prediction models to evaluate the effects of various geometric features on safety at horizontal curves. There were three main focus areas in this research aimed at adding to the current knowledge and building upon past research. The first was an emphasis on assembling a high quality large dataset with various roadway and geometric variables (posted speed, advisory speed, pavement type, etc.) to gain further understanding and insight into safety issues at horizontal curves. The use of a high quality comprehensive dataset would provide a better chance to develop accurate models. The second focus area pertained to the use of regression tree analysis to improve the development of crash prediction models and explore applications in systemic safety management. The third focus area was to research the differences in safety on horizontal curves with respect to different types of crash dataset.

## DATA COLLECTION AND PROCESSING

One of the main features of this research was an emphasis on assembling a comprehensive, high-quality, and large dataset. Horizontal curve, crash, and various roadway data elements from the Wisconsin Department of Transportation (WisDOT) roadway safety management database consisting of roadway, mobility, pavement data, were assembled details of which are described in the next sections.

### Horizontal Curve Data

WisDOT maintains horizontal curve information including attributes such as radius, degree of curvature, length, route, county, and mile markers for the start and end points of each curve. The data were collected on Wisconsin State Trunk Network (STN) roads from WisDOT Photolog dataset which has a scale of 0.01 miles (52.8 ft.) using an automated algorithm in a Geographic Information System (GIS) environment. The automated algorithm analyzed the angle between subsequent Photolog points (every 0.01 miles) to calculate curve attributes (25). The data were mapped using the Photolog Lane Mile (PLM) routes which were created to enable the integration of Photolog-based data with other WisDOT GIS database (26).

One of the drawbacks of using an automated algorithm to detect horizontal curves was the inclusion of potential tangent sections with very large radii in the dataset. Therefore, as a starting point, the dataset was trimmed by selecting curves with radius less than 10,000 ft. and greater than 200 ft. The lower end choice was based on manual review of locations almost all of which were intersections turns. The resulting dataset included 30,185 potential horizontal curve locations on the STN roads in Wisconsin. The dataset included separate records for curves in each direction of a highway on both divided and undivided roadways which was a significant departure from general practice in the past because it provided the opportunity to analyze detailed differences in horizontal curves safety.

Figure 1 shows the breakdown of the curve dataset in terms of location, type of highway, and the presence of sign data. The sample size of curve datasets as shown in Figure 1 signifies the strength of this research in assembling a large dataset. The focus of this research was on rural curves on undivided roads in view of the literature and objectives defined which totaled 20,842 curve locations. This included 27 curves on rural multilane roads which were included in the analysis with the belief that the use of travel-way width variable would account for the difference between curves on multilane and two-lane roads in rural areas. A total of 99 horizontal curves had one or more data elements missing therefore the final sample size was 20,743. The analysis of other curve types would be conducted later as part of a larger curve safety evaluation project.

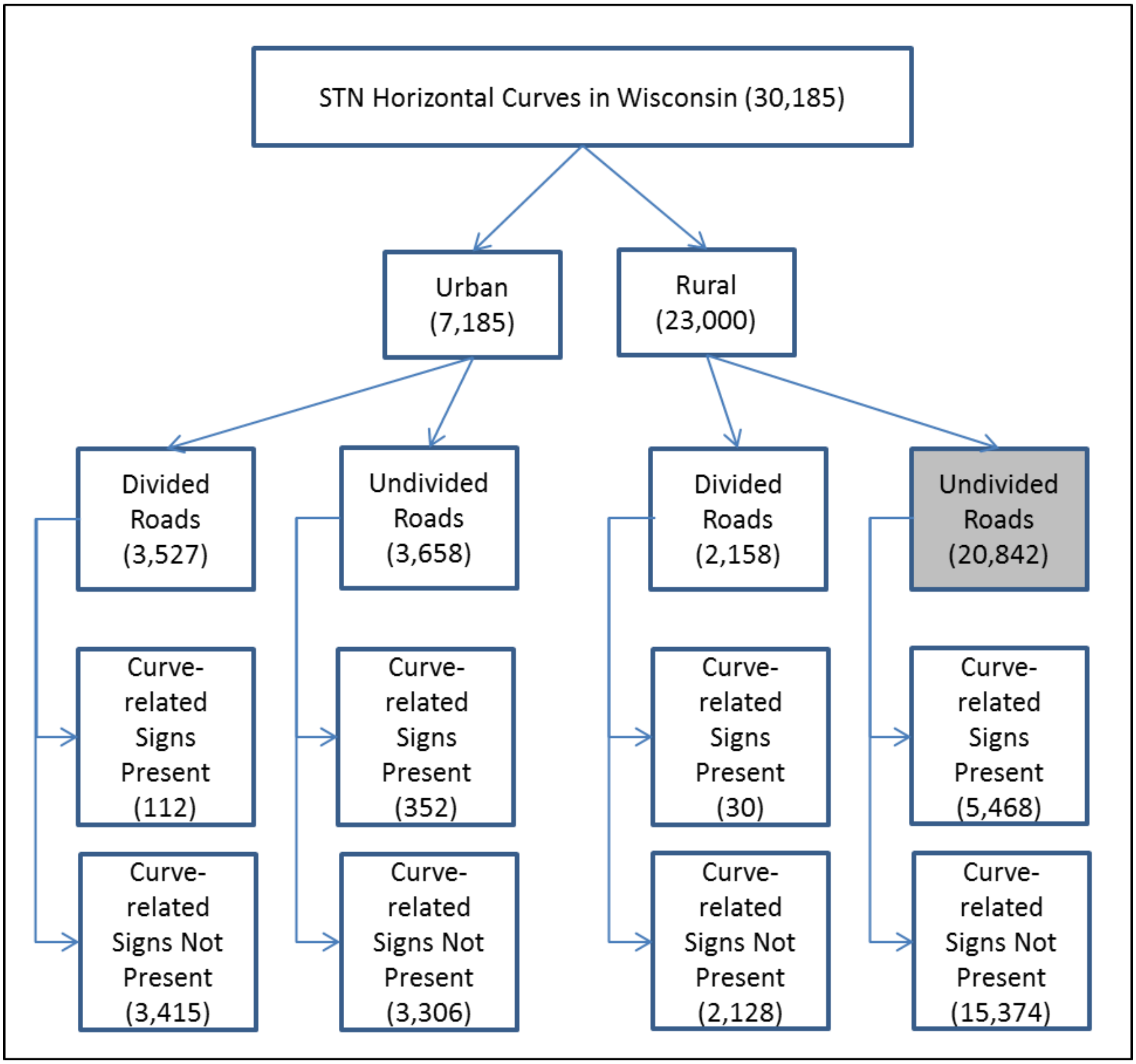


**Figure 1 Details of Horizontal Curve Dataset on Wisconsin STN Roads**

## Crash Data

Crashes on horizontal curves in Wisconsin for the five year period between 2006 and 2010 were obtained. A 200 foot buffer downstream of the horizontal curves was specified to capture crashes that may have ended outside the proximity of the curves. Deer and other animal-related crashes were removed from the analysis because it is difficult to identify an engineering countermeasure to deal with such crashes. Wisconsin experiences a large number of deer-related crashes each year and it is a common practice to remove these crashes from analysis. One question facing the authors was how to identify crashes most relevant to horizontal curve safety; the answer to which was not clear from the literature. Therefore, a decision was made to assemble several different crash dataset to compare the results as indicated in the objectives. The differences in the dataset were based upon two fields in the Wisconsin crash report forms (MV4000). The first field identified crashes within 150 ft. of an intersection or driveway; and the second field noted the presence of a horizontal curve at the point of impact of a crash as identified by the reporting officer. Furthermore, separate dataset were created for total and fatal/injury crashes (injury crashes included incapacitating and non-incapacitating injuries).

**Roadway Data**

Horizontal curve geometric attributes were available as part of the dataset maintained by WisDOT. Horizontal curve roadway data e.g. traffic and truck volumes, shoulder and travel-way widths, posted speeds, pavement information, were obtained from the WisDOT road safety management database. Furthermore, advisory speed data were obtained from sign database maintained by WisDOT which contains GIS points for each sign location on Wisconsin STN roads.

The individual datasets, namely horizontal curve geometric attributes, crash, roadway data elements, and signs are maintained at WisDOT using linear referencing system with an intended accuracy of 0.01 miles (52.8 ft.). The datasets were overlaid and merged together in a GIS environment using data integration techniques developed by Khan et al. (26). Crash and roadway data elements were aggregated for each individual horizontal curve location. The most relevant variables were selected for analyzing horizontal curve safety and data were checked for errors and missing elements. Table 2 shows the descriptive statistics of different variables and crash dataset on undivided rural curves in Wisconsin. Additionally, a number of categorical variables were created using existing data which are described below:

$RST_{Base}$ = right shoulder type paved *(base condition)*,
$RST_R$ = right shoulder type rumble,
$RST_U$ = right shoulder type unpaved,
*DiffPSAS* = Difference between posted and advisory speeds,
$PVT_{Base}$ = asphalt pavement *(base condition)*,
$PVT_C$ = concrete pavement,
$PVT_{RM}$ = road-mix pavement,
$UT_{Base}$ = upstream tangent > 2600 feet *(base condition)*,
$UT_1$ = upstream tangent 0 – 600 feet,
$UT_2$ = upstream tangent 601 – 1200 feet, and
$UT_3$ = upstream tangent 1201 – 2600 feet.

**TABLE 2 Descriptive Statistics of Continuous Variables**

| Variable Name | Mean | Median | Std. Dev. |
|---|---|---|---|
| **Curve Radius (ft.) *(R)*** | 2920.4 | 2280.0 | 2012.6 |
| **Curve Length (ft.) *(L)*** | 914.8 | 739.0 | 619.2 |
| **Historical AADT (*AADT)*** | 1337.8 | 1000.0 | 1121.9 |
| **Truck Percentage (%) *(TRK)*** | 10.9 | 11.0 | 4.1 |
| **Travel Way Width (ft.) *(TWD)*** | 11.7 | 12.0 | 0.9 |
| **Left Shoulder Width (ft.) *(LSW)*** | 6.1 | 6.0 | 2.8 |
| **Right Shoulder Width (ft.) *(RSW)*** | 6.1 | 6.0 | 2.8 |
| **Average IRI (mm/meter) *(IRI)*[a]** | 1.8 | 1.6 | 0.8 |
| **Pavement Surface Age (Yrs.) *(PSage)*** | 13.1 | 12.0 | 8.8 |
| **Upstream Tangent Length (ft.)** | 2854.9 | 1000.0 | 6545.5 |
| **Posted Speed (mph) *(PS)*** | 31.7 | 25.0 | 10.1 |
| **Difference between Posted and Advisory Speeds (mph) *(DiffPSAS)*** | 0.5 | 0.0 | 3.1 |
| ***HORC*[b]** | 0.3 | 0.0 | 0.9 |
| ***KABHORC*[c]** | 0.1 | 0.0 | 0.4 |
| ***HORC_N*[d]** | 0.3 | 0.0 | 0.7 |
| ***KABHORC_N*[e]** | 0.1 | 0.0 | 0.3 |
| ***ALL*[f]** | 0.7 | 0.0 | 3.3 |
| ***KABALL*[g]** | 0.2 | 0.0 | 0.7 |
| ***ALL_N*[h]** | 0.5 | 0.0 | 1.3 |
| ***KABALL_N*[i]** | 0.1 | 0.0 | 0.4 |

[a]International Roughness Index (27)
[b]Sum of crashes where horizontal road terrain at point of impact was a curve as identified by crash report form (*HORC* dataset = 7,024 crashes).
[c]Sum of fatal and injury (K, A, and B) crashes where horizontal road terrain at point of impact was a curve as identified by crash report form (*KABHORC* dataset = 1,973 crashes).
[d]Sum of crashes where horizontal road terrain at point of impact was a curve as identified by crash report form and distance from closest intersection or driveway was greater than 150 ft. (*HORC_N* dataset = 5,631 crashes).
[e]Sum of fatal and injury (K, A, and B) crashes where horizontal road terrain at point of impact was a curve as identified by crash report form and distance from closest intersection or driveway was greater than 150 ft. (*KABHORC_N* dataset = 1,628 crashes).
[f]Sum of all crashes located on horizontal curves using mile marker information regardless of crash report form information (*ALL* dataset = 15,097).
[g]Sum of all fatal and injury (K, A, and B) crashes located on horizontal curves using mile marker information regardless of crash report form information (*KABALL* dataset = 3,592).
[h]Sum of all crashes located on horizontal curves using mile marker information regardless of crash report form information and where distance from closest intersection or driveway was greater than 150 ft. (*ALL*_N dataset = 10,072).
[i]Sum of all fatal and injury (K, A, and B) crashes located on horizontal curves using mile marker information regardless of crash report form information and where distance from closest intersection or driveway was greater than 150 ft. (*KABALL_N* dataset = 2,545)

## STATISTICAL METHODOLOGY

Poisson regression has been traditionally used in crash data modeling but the constraint on equality of mean and variance has driven researchers to consider the Negative Binomial (NB) regression methodology. One way of dealing with the traditional Poisson model restrictions is to use the same estimating functions for the mean, but to base inference on the more robust Quasi-Poisson regression.

### Quasi-Poisson Model

Quasi-Poisson uses the mean regression function and the variance function from the Poisson Generalized Linear Model (GLM) but leaves the dispersion parameter unrestricted. Thus, the dispersion parameter is not assumed to be fixed at 1 but is estimated from the data which leads to the same coefficient estimates as the standard Poisson model but inference is adjusted for over-dispersion. Consequently, Quasi-Poisson does not correspond to models with fully specified likelihoods and its Akaike Information Criterion (AIC) does not have traditional meaning.

### Negative Binomial Model

Another way to model over-dispersed count data is to assume NB distribution for which there can be a gamma mixture of Poisson distributions. One parameterization of its probability density function is:

$$f(y;\mu,\theta)=\frac{\Gamma(y+\theta)}{\Gamma(\theta)\,y!}\cdot\frac{\mu^{y}\theta^{\theta}}{(\mu+\theta)^{y+\theta}} \qquad (1)$$

with mean $\mu$ and shape parameter $\theta$; $\Gamma()$ is the gamma function. It has variance $V(\mu)=\mu+\frac{\mu^2}{\theta}$. When $\theta$ goes to infinity, Negative Binomial approaches a Poisson distribution.

### Akaike Information Criterion

The AIC is a measure of the relative goodness of fit of a statistical model which loosely describes the tradeoff between the accuracy and complexity of the model. In the general case, the AIC is

$$AIC = 2k - 2\ln(L) \qquad (2)$$

where

k = the number of parameters in the statistical model, and

L is the maximized value of the likelihood function for the estimated model.

### Variance Inflation Factor

Variance Inflation Factor (VIF) quantifies the severity of multicollinearity in regression analysis by calculating a factor by which variance in regression coefficient is inflated due to multicollinearity (28). Generally, a VIF value of greater than four requires further review of the coefficients and a value greater than 10 is considered as an indication of serious multicollinearity (28).

**Regression Tree using GUIDE**

Regression trees are machine-learning methods for constructing prediction models through recursive partitioning of data which can be graphically represented as a decision tree. Regression trees are specific to continuous or ordered discrete dependent variables as compared to classification trees which are designed for finite number of unordered values. There are several algorithms in literature which implement regression tree with different strengths and weaknesses (29). The regression tree algorithm used in this research was GUIDE (Generalized, Unbiased, Interaction Detection and Estimation). GUIDE offers advantages in terms of unbiased splits (removing bias in splits due to large differences in sample sizes) and options of fitting complex node models, as compared to other regression tree algorithms e.g. Classification and Regression Tree (CART) (29, 30).

## MODEL DEVELOPMENT, RESULTS, AND DISCUSSIONS

**Regression Tree Model**

There were two main reasons to conduct regression tree analysis. The first reason was to provide a simple model and basic understanding of horizontal curve safety for use in systemic road safety management process. The second reason was to help trim the horizontal curve dataset to remove possible tangent sections with very high radius identified by automated algorithms. The aim was to identify a cut-off radius value to determine a more realistic subset of horizontal curves which would warrant more attention with respect to horizontal curve safety. In this respect, Figures 2(a) and 2(b) show the results of GUIDE piecewise constant regression tree models for rural horizontal curves on undivided roads in Wisconsin using *HORC* and *HORC_N* crash dataset.

At each intermediate node, an observation (individual horizontal curve record) goes to the left branch only if the condition is satisfied. The values in italics at each terminal node show the mean number of crashes for the five year period from 2006 to 2010 for the set of horizontal curves at that node. The results show that for *HORC* and *HORC_N* dataset, the mean number of crashes reduce significantly for horizontal curves with radius greater than 2499 foot and 2515 foot, respectively. Therefore, greater emphasis should be put on curves with radius less than 2,500 foot. The radius of 2,500 foot can also be used as a cut-off value to identify the most critical horizontal curves to develop crash prediction models. Additionally, for radii less than approximately 2500 feet, traffic volume becomes an additional significant factor in identifying curves which experience more crashes.

The results in Figure 2(a) and 2(b) are fairly consistent between the two crash dataset and illustrate a simple model which is easy to interpret and provides vital clues regarding safety on horizontal curves in terms of radius and traffic volumes. Furthermore, such results can be readily used in initial steps of systemic road safety management procedures by practitioners.

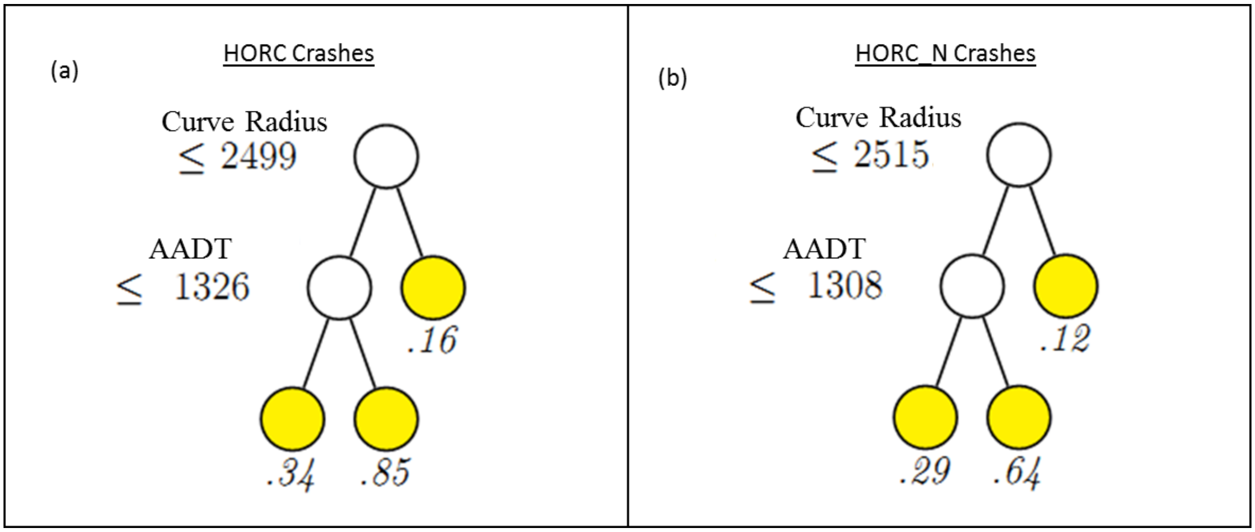


**Figure 2 GUIDE regression tree models for (a) *HORC* crash dataset (b) *HORC_N* crash dataset**

**Negative Binomial Crash Prediction Models**

For more detailed evaluation of rural undivided horizontal curves, both Quasi-Poisson and NB models were fitted using R GLM framework (31). A correlation matrix was developed to identify and remove correlated variables. The process of model development started with the specification of a base model and the final crash prediction models were generated based on the results of stepwise regression using AIC as the model selection criteria. The final Poisson model was refitted with Quasi-Poisson method to get the adjusted standard errors and significance levels. The Quasi-Poisson and NB models were compared with each other using ten-fold cross validation. Based on the cross validation score and ease of interpretation, the NB models were selected as the best models to be used in the final results. Finally, VIF test was performed for each model to check for multicollinearity in regression coefficients.

The complete rural undivided horizontal curve dataset contained 20,743 curves out of which 14,348 curves had radius greater than 1,660 feet (Curve Class A, degree of curvature 0.0 - 3.45) and 6,395 had radius less than 1,660 feet (Curve Class B-F, degree of curvature > 3.45) (12). Crash prediction models were compared for curves belonging to Curve Class A and Curve Class B-F curves. The results showed that models for Curve Class A were inconsistent with normal expectations, e.g., curve radius coefficient was positive, etc. The comparisons confirmed that as radius increased beyond a reasonable limit, the results could not be trusted as curves with large radii probably tend to behave as tangent sections. Therefore, a decision was made to select a cut-off distance for radius based on the results of GUIDE regression tree models. Horizontal curves with radius less than or equal to 2,500 feet were selected for developing NB crash prediction models (11,427 curves).

The results of crash prediction models using different crash dataset are presented in the next section using variables defined in data collection and processing section and Table 2. Right and left shoulder widths; right and left shoulder types were correlated variables; therefore, one was removed from the analysis depending upon statistical significance.

*Horizontal Curve Crash Prediction Models using HORC and KABHORC Crash Data*

The NB crash prediction models for total and fatal/injury crashes on horizontal curves using *HORC* and *KABHORC* crash dataset are presented in equation 3 and equation 4, respectively and Table 3. The results show curve radius, curve length, and natural log of AADT as highly significant variables ($p<0.0001$) with coefficients signs and magnitude in line with findings in literature. Left shoulder width is significant for *HORC* dataset and shows reduction in crashes as width increases; but is not significant for the severity model (*KABHORC* dataset). For right shoulder type, unpaved shoulder shows increase in crashes whereas rumble strips show decrease in crashes as compared to base condition of paved shoulder; however the results for rumble strips are not significant at $p = 0.05$.

The coefficient for average IRI shows that as the value decreases (pavement smoothness increases), there is an increase in crashes on horizontal curves. A possible explanation could be reduction in pavement friction as pavement becomes too smooth leading to increase in crashes. The *DiffPSAS* variable shows that as the difference between posted and advisory speed limit on the curve increases, more crashes are expected. This is a very important result because the difference determines the type of sign to be placed at a curve, hence an important finding of this research. The tangent length upstream of a curve was used as a categorical variable where the base condition was a tangent length greater than 2,600 feet (approx. 0.5 miles). The results show that compared with base conditions, less crashes are expected as tangent length decreases which points to possible driver expectancy issues as they approach the first horizontal curve after a long tangent section.

**Table 3 Crash Prediction Models for Horizontal Curves on Rural Undivided Roads – Using *HORC* and *KABHORC* Crash Dataset**

| *HORC Crash Dataset (Total crashes)* | | | | |
|---|---|---|---|---|
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -4.6703 | 0.2007 | -23.270 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0008 | 0.0000 | -21.550 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0007 | 0.0000 | 20.680 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.7072 | 0.0280 | 25.260 | 0.000 |
| Left Shoulder Width (ft.) ***(LSW)*** | -0.0237 | 0.0093 | -2.550 | 0.011 |
| Right Shoulder Type – Rumble ***($RST_R$)*** | -0.3529 | 1.1189 | -0.320 | 0.752 |
| Right Shoulder Type – Unpaved ***($RST_U$)*** | 0.1621 | 0.0538 | 3.010 | 0.003 |
| Average IRI ***(IRI)*** | -0.0821 | 0.0240 | -3.410 | 0.001 |
| Difference between Posted and Advisory Speed (mph) ***(DiffPSAS)*** | 0.0119 | 0.0045 | 2.640 | 0.008 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.4121 | 0.0459 | -8.970 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.3449 | 0.0556 | -6.200 | 0.000 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.1536 | 0.0521 | -2.950 | 0.003 |
| AIC = 19458 | | | | |
| *KABHORC Crash Dataset (Fatal/Injury Crashes)* | | | | |
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -5.2170 | 0.3212 | -16.240 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0007 | 0.0001 | -11.630 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0006 | 0.0000 | 12.280 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.5875 | 0.0410 | 14.330 | 0.000 |
| Right Shoulder Type - Rumble ***($RST_R$)*** | -15.5216 | 2199.6843 | -0.010 | 0.994 |
| Right Shoulder Type - Unpaved ***($RST_U$)*** | 0.2386 | 0.0826 | 2.890 | 0.004 |
| Average IRI ***(IRI)*** | -0.1149 | 0.0392 | -2.930 | 0.003 |
| Difference between Posted and Advisory Speed (mph) ***(DiffPSAS)*** | 0.0148 | 0.0070 | 2.130 | 0.033 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.5601 | 0.0730 | -7.670 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.4282 | 0.0882 | -4.850 | 0.000 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.2056 | 0.0804 | -2.560 | 0.011 |
| AIC: 8896 | | | | |

$$\mu_i = \exp\left[-4.67 - 0.0008 * R + 0.0007 * L + 0.707 * \ln(AADT) - 0.023 * LSW - 0.35 * RST_R + 0.16 * RST_U - 0.082 * IRI + 0.011 * DiffPSAS - 0.41 * UT_1 - 0.34 * UT_2 - 0.15 * UT_3\right. \quad (3)$$

$$\mu_i = \exp\left[-5.21 - 0.0007 * R + 0.0006 * L + 0.587 * \ln(AADT) - 15.52 * RST_R + 0.23 * RST_U - 0.115 * IRI + 0.014 * DiffPSAS - 0.560 * UT_1 - 0.428 * UT_2 - 0.205 * UT_3\right. \quad (4)$$

*Horizontal Curve Crash Prediction Models using HORC_N and KABHORC_N Crash Data*

The results of the NB crash prediction models for total and fatal/injury crashes on horizontal curves using *HORC_N* and *KABHORC_N* crash dataset are presented in equation 5, equation 6, respectively, and Table 5. The difference between these models and the models in Table 3 (equation 3 and equation 4) is the exclusion of crashes occurring within 150 feet of an intersection or driveway.

The model in Table 4 for *HORC_N* crash dataset is similar to the model in Table 3 for *HORC* crash dataset in terms of variables with slight differences in the magnitude of coefficients. The model in Table 4 for *KABHORC_N* crash dataset compared with model in Table 3 for *KABHORC* crash dataset shows that the *DiffPSAS* variable is replaced by left shoulder width and the $UT_3$ variable is insignificant. Overall, the comparisons are interesting because they show that when the crash report form indicates the presence of a horizontal curve at the point of impact, the inclusion of crashes in proximity of intersections is justified to increase the size of dataset. Therefore, the models in Table 3 (equation 3 and equation 4) are recommended for use.

**Table 4 Crash Prediction Models for Horizontal Curves on Rural Undivided Roads – Using *HORC_N* and *KABHORC_N* Crash Dataset**

| *HORC_N Crash Dataset (Total crashes)* | | | | |
|---|---|---|---|---|
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -4.7147 | 0.2120 | -22.240 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0006 | 0.0000 | -16.760 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0007 | 0.0000 | 19.510 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.6461 | 0.0295 | 21.920 | 0.000 |
| Left Shoulder Width (ft.) ***(LSW)*** | -0.0308 | 0.0099 | -3.120 | 0.002 |
| Right Shoulder Type - Rumble ***($RST_R$)*** | -0.3017 | 1.1198 | -0.270 | 0.788 |
| Right Shoulder Type - Unpaved ***($RST_U$)*** | 0.1508 | 0.0568 | 2.660 | 0.008 |
| Average IRI ***(IRI)*** | -0.0688 | 0.0254 | -2.710 | 0.007 |
| Difference between Posted and Advisory Speed (mph) ***(DiffPSAS)*** | 0.0140 | 0.0047 | 2.940 | 0.003 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.2510 | 0.0492 | -5.110 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.1948 | 0.0589 | -3.300 | 0.001 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.0525 | 0.0557 | -0.940 | 0.346 |
| AIC: 17559 | | | | |
| *KABHORC_N Crash Dataset (Fatal/Injury Crashes)* | | | | |
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -5.1178 | 0.3370 | -15.180 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0005 | 0.0001 | -9.260 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0006 | 0.0001 | 10.980 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.5542 | 0.0472 | 11.740 | 0.000 |
| Left Shoulder Width (ft.) ***(LSW)*** | -0.0323 | 0.0160 | -2.010 | 0.044 |
| Right Shoulder Type - Rumble ***($RST_R$)*** | -15.5043 | 2201.6 | -0.010 | 0.994 |
| Right Shoulder Type - Unpaved ***($RST_U$)*** | 0.1943 | 0.0930 | 2.090 | 0.037 |
| Average IRI ***(IRI)*** | -0.1245 | 0.0423 | -2.950 | 0.003 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.4481 | 0.0788 | -5.690 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.3222 | 0.0947 | -3.400 | 0.001 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.1147 | 0.0869 | -1.320 | 0.187 |
| AIC: 7998 | | | | |

$$\mu_i = \exp[-4.71 - 0.0006 * R + 0.0007 * L + 0.646 * \ln(AADT) - 0.0308 * LSW - 0.301 * RST_R + 0.15 * RST_U - 0.068 * IRI + 0.014 * DiffPSAS - 0.251 * UT_1 - 0.194 * UT_2 - 0.052 * UT_3 \quad (5)$$

$$\mu_i = \exp[-5.117 - 0.0005 * R + 0.0006 * L + 0.554 * \ln(AADT) - 0.032 * LSW - 15.50 * RST_R + 0.194 * RST_U - 0.124 * IRI - 0.448 * UT_1 - 0.322 * UT_2 - 0.114 * UT_3 \quad (6)$$

*Horizontal Curve Crash Prediction Models using ALL and KABALL Crash Data*

The NB crash prediction models for total and fatal/injury crashes on horizontal curves using *ALL* and *KABALL* crash dataset are presented in equation 7, equation 8, respectively and Table 5. The difference between these models and the models in Table 3 and Table 4 is that crashes were identified on horizontal curves using their mile markers regardless of whether the crash report forms indicated the presence of a horizontal curve at the point of impact.

The results show curve radius, curve length, and natural log of AADT as highly significant variables ($p<0.0001$) with coefficients signs and magnitude in line with findings in literature. Posted speed shows up as a significant variable which was missing from previous models showing that speed is an important factor in overall crash occurrence. The sign of *DiffPSAS* coefficient in Table 5 for model based on *ALL* crash dataset suggests that crashes increase as the difference between posted and advisory speed reduces which is counter to the results in previous models and warrants further investigation. The model based on *KABALL* crash dataset in Table 5 shows less crashes on concrete and road-mix pavement as compared to base condition of asphalt. A possible explanation could be issues related to pavement friction, however further investigation is required.

**Table 5 Crash Prediction Models for Horizontal Curves on Rural Undivided Roads – Using *ALL* and *KABALL* Crash Dataset**

| *ALL Crash Dataset (Total crashes)* | | | | |
|---|---|---|---|---|
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -4.8738 | 0.1759 | -27.710 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0007 | 0.0000 | -22.580 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0004 | 0.0000 | 13.080 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.7502 | 0.0208 | 36.050 | 0.000 |
| Posted Speed (mph) ***(PS)*** | 0.0169 | 0.0017 | 10.120 | 0.000 |
| Average IRI ***(IRI)*** | -0.0417 | 0.0205 | -2.030 | 0.043 |
| Difference between Posted and Advisory Speed (mph) ***(DiffPSAS)*** | -0.0144 | 0.0047 | -3.040 | 0.002 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.4830 | 0.0422 | -11.440 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.3912 | 0.0509 | -7.680 | 0.000 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.1202 | 0.0480 | -2.500 | 0.012 |
| AIC: 26064 | | | | |

| *KABALL Crash Dataset (Fatal/Injury Crashes)* | | | | |
|---|---|---|---|---|
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -5.7248 | 0.3032 | -18.880 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0006 | 0.0000 | -13.080 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0004 | 0.0000 | 9.580 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.6789 | 0.0368 | 18.460 | 0.000 |
| Posted Speed (mph) ***(PS)*** | 0.0132 | 0.0024 | 5.450 | 0.000 |
| Right Shoulder Type - Rumble ***($RST_R$)*** | -14.6861 | 1335.5095 | -0.010 | 0.991 |
| Right Shoulder Type - Unpaved ***($RST_U$)*** | 0.1645 | 0.0754 | 2.180 | 0.029 |
| Average IRI ***(IRI)*** | -0.0900 | 0.0350 | -2.570 | 0.010 |
| Pavement Type – Concrete ***($PVT_C$)*** | -0.3146 | 0.1527 | -2.060 | 0.039 |
| Pavement Type - Road Mix ***($PVT_{RM}$)*** | -0.2228 | 0.1149 | -1.940 | 0.052 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.5617 | 0.0643 | -8.740 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.4508 | 0.0786 | -5.740 | 0.000 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.1802 | 0.0715 | -2.520 | 0.012 |
| AIC: 11494 | | | | |

$$\mu_i = \exp\left[-4.873 - 0.0007 * R + 0.0004 * L + 0.750 * \ln(AADT) + 0.0169 * PS - 0.041 * IRI - 0.014 * DiffPSAS - 0.483 * UT_1 - 0.391 * UT_2 - 0.0120 * UT_3 \right. \tag{7}$$

$$\mu_i = \exp\left[-5.724 - 0.0006 * R + 0.0004 * L + 0.678 * \ln(AADT) + 0.013 * PS - 14.68 * RST_R + 0.164 * RST_U - 0.090 * IRI - 0.314 * PVT_C - 0.222 * PVT_{RM} - 0.561 * UT_1 - 0.450 * UT_2 - 0.180 * UT_3 \right. \tag{8}$$

*Horizontal Curve Crash Prediction Models using ALL_N and KABALL_N Crash Data*

The NB crash prediction models for total and fatal/injury crashes using *ALL_N* and *KABALL_N* crash dataset are presented in equation 9, equation 10, respectively and Table 6. The difference between these models and the models in Table 5 (equation 7 and equation 8) is the exclusion of crashes occurring within 150 ft. of an intersection or driveway.

The results in Table 6 for *ALL_N* crash dataset compared with results in Table 5 for *ALL* crash dataset shows the *DiffPSAS* variable is replaced by left shoulder width, pavement age becomes a significant variable (older pavement leading to more crashes), and the $UT_3$ variable becomes insignificant. The model in Table 6 for *KABALL* crash dataset is different from model in Table 5 for *KABALL_N* crash dataset with some variables interchanging between the models. Overall, the results suggest that when selecting crashes on horizontal curves using mile markers only, crashes outside the proximity of an intersection show significant variables which are more relevant to horizontal curve safety; unlike results from *HORC*-based crash dataset where the inclusion of crashes in proximity of intersection did not results in significant differences. Therefore, the models in Table 6 are recommended for use.

**Table 6 Crash Prediction Models for Horizontal Curves on Rural Undivided Roads – Using *ALL*_N and *KABALL_N* Crash Dataset**

| *ALL_N Crash Dataset (Total crashes)* | | | | |
|---|---|---|---|---|
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -4.7354 | 0.1819 | -26.030 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0005 | 0.0000 | -15.390 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0005 | 0.0000 | 17.180 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.6502 | 0.0244 | 26.690 | 0.000 |
| Posted Speed (mph) ***(PS)*** | 0.0075 | 0.0016 | 4.730 | 0.000 |
| Left Shoulder Width (ft.) ***(LSW)*** | -0.0221 | 0.0079 | -2.790 | 0.005 |
| Average IRI ***(IRI)*** | -0.0704 | 0.0241 | -2.920 | 0.003 |
| Pavement Surface Age (Yrs.) ***(PSage)*** | 0.0058 | 0.0022 | 2.650 | 0.008 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.2818 | 0.0421 | -6.700 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.1675 | 0.0501 | -3.340 | 0.001 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | 0.0034 | 0.0473 | 0.070 | 0.942 |
| AIC: 21844 | | | | |
| *KABALL_N Crash Dataset (Fatal/Injury Crashes)* | | | | |
| Variable Name | Estimate | Std. Error | z value | Pr(>\|z\|) |
| (Intercept) | -5.3204 | 0.2995 | -17.760 | 0.000 |
| Curve Radius (ft.) ***(R)*** | -0.0004 | 0.0001 | -8.320 | 0.000 |
| Curve Length (ft.) ***(L)*** | 0.0005 | 0.0000 | 11.320 | 0.000 |
| Log of Historical AADT ***(AADT)*** | 0.6019 | 0.0417 | 14.430 | 0.000 |
| Left Shoulder Width (ft.) ***(LSW)*** | -0.0324 | 0.0140 | -2.310 | 0.021 |
| Right Shoulder Type - Rumble ***(RSTR)*** | -15.7386 | 2203.0 | -0.010 | 0.994 |
| Right Shoulder Type - Unpaved ***(RSTU)*** | 0.1966 | 0.0826 | 2.380 | 0.017 |
| Average IRI ***(IRI)*** | -0.1299 | 0.0374 | -3.470 | 0.001 |
| Upstream Tangent (0-600 ft.) ***($UT_1$)*** | -0.4368 | 0.0695 | -6.280 | 0.000 |
| Upstream Tangent (601-1200 ft.) ***($UT_2$)*** | -0.2962 | 0.0829 | -3.570 | 0.000 |
| Upstream Tangent (1201-2600 ft.) ***($UT_3$)*** | -0.0640 | 0.0754 | -0.850 | 0.396 |
| AIC: 9502 | | | | |

$$\mu_i = \exp[-4.73 - 0.0005 * R + 0.0005 * L + 0.650 * \ln(AADT) + 0.007 * PS - 0.022 * LSW - 0.070 * IRI + 0.005 * PSage - 0.281 * UT_1 - 0.167 * UT_2 + 0.003 * UT_3 \quad (9)$$

$$\mu_i = \exp[-5.32 - 0.0004 * R + 0.0005 * L + 0.601 * \ln(AADT) - 0.032 * LSW - 15.73 * RST_R + 0.196 * RST_U - 0.129 * IRI - 0.436 * UT_1 - 0.296 * UT_2 - 0.064 * UT_3 \quad (10)$$

## SUMMARY AND CONCLUSIONS

The objectives of this research were to develop total and fatal/injury crash prediction models for rural horizontal curves on undivided roads, with focus on three distinct aspects. The first was an emphasis on assembling high quality large dataset for accurate model development. As a result, many new variables were included in model development, e.g., pavement roughness, pavement type, difference between posted and advisory speeds, shoulder widths and types, etc. which provide an important contribution to current knowledge. Also curves in both directions were analyzed separately. Interestingly, certain variables were not statistically significant in any models such as travel-way width, truck volume. The coefficients of most variables in presented models have correct signs and are reasonable in magnitude with strong statistical significant signifying the robustness of the models. The resulting crash prediction models can be used in a variety of horizontal curve related safety analyses.

The second focus area was to use regression tree analysis in exploring horizontal curve safety from a different perspective. The results of regression tree analysis were useful in two ways; creating subsets of data which warranted further analysis and a simple model aimed at practitioners of systemic road safety management. The results show that there is a marked increase in the number of crashes on horizontal curves with radius less than 2,500 feet and traffic volume greater than approximately 1300 vehicles per day.

The third focus area of this research was to compare horizontal curve crash prediction models using different crash dataset to analyze the differences in selecting crashes using different criteria. Table 7 presents the total and fatal/injury crash prediction models based on various crash datasets described previously.

**Table 7 Comparison of Total and Fatal/Injury Crash Prediction Models using Different Crash Datasets**

| **Crash Dataset** | **Crash Prediction Models (Total Crashes)** |
|---|---|
| *HORC* | $\mu_i = \exp[-4.67 - 0.0008 * R + 0.0007 * L + 0.707 * \ln(AADT) - 0.023 * LSW - 0.35 * RST_R + 0.16 * RST_U - 0.082 * IRI + 0.011 * DiffPSAS - 0.41 * UT_1 - 0.34 * UT_2 - 0.15 * UT_3$ **(3)** |
| *HORC_N* | $\mu_i = \exp[-4.71 - 0.0006 * R + 0.0007 * L + 0.646 * \ln(AADT) - 0.0308 * LSW - 0.301 * RST_R + 0.15 * RST_U - 0.068 * IRI + 0.014 * DiffPSAS - 0.251 * UT_1 - 0.194 * UT_2 - 0.052 * UT_3$ **(5)** |
| *ALL* | $\mu_i = \exp[-4.873 - 0.0007 * R + 0.0004 * L + 0.750 * \ln(AADT) + 0.0169 * PS - 0.041 * IRI - 0.014 * DiffPSAS - 0.483 * UT_1 - 0.391 * UT_2 - 0.0120 * UT_3$ **(7)** |
| *ALL_N* | $\mu_i = \exp[-4.73 - 0.0005 * R + 0.0005 * L + 0.650 * \ln(AADT) + 0.007 * PS - 0.022 * LSW - 0.070 * IRI + 0.005 * PSage - 0.281 * UT_1 - 0.167 * UT_2 + 0.003 * UT_3$ **(9)** |
| **Crash Dataset** | **Crash Prediction Models (Fatal/Injury Crashes)** |
| *KABHORC* | $\mu_i = \exp[-5.21 - 0.0007 * R + 0.0006 * L + 0.587 * \ln(AADT) - 15.52 * RST_R + 0.23 * RST_U - 0.115 * IRI + 0.014 * DiffPSAS - 0.560 * UT_1 - 0.428 * UT_2 - 0.205 * UT_3$ **(4)** |
| *KABHORC_N* | $\mu_i = \exp[-5.117 - 0.0005 * R + 0.0006 * L + 0.554 * \ln(AADT) - 0.032 * LSW - 15.50 * RST_R + 0.194 * RST_U - 0.124 * IRI - 0.448 * UT_1 - 0.322 * UT_2 - 0.114 * UT_3$ **(6)** |
| *KABALL* | $\mu_i = \exp[-5.724 - 0.0006 * R + 0.0004 * L + 0.678 * \ln(AADT) + 0.013 * PS - 14.68 * RST_R + 0.164 * RST_U - 0.090 * IRI - 0.314 * PVT_C - 0.222 * PVT_{RM} - 0.561 * UT_1 - 0.450 * UT_2 - 0.180 * UT_3$ **(8)** |
| *KABALL_N* | $\mu_i = \exp[-5.32 - 0.0004 * R + 0.0005 * L + 0.601 * \ln(AADT) - 0.032 * LSW - 15.73 * RST_R + 0.196 * RST_U - 0.129 * IRI - 0.436 * UT_1 - 0.296 * UT_2 - 0.064 * UT_3$ **(10)** |

## Comparison of Curve Crash Prediction Models using *HORC*, *KABHORC*, *HORC_N*, and *KABHORC_N* Crash Data

The four crash datasets, namely *HORC*, *KABHORC*, *HORC_N*, and *KABHORC_N* include crashes where the crash report form indicates the presence of a horizontal curve. However, *HORC* and *KABHORC* crash datasets include crashes occurring within 150 feet of an intersection which are excluded from *HORC_N*, *KABHORC_N* crash datasets.

A comparison of curve crash prediction models for total crashes (*HORC* vs. *HORC_N*) as presented in Table 7 (equations 3 and 5) shows that the models are almost the same in terms of variables with slight differences in the magnitude of coefficients. A comparison of curve crash prediction models for fatal/injury crashes (*KABHORC* vs. *KABHORC_N*) as presented in Table 7 (equations 4 and 6) shows slight differences where *DiffPSAS* variable is replaced by left shoulder width.

Overall, the comparison results show that when crash report form indicates the presence of a horizontal curve, the inclusion of crashes in the proximity of intersections do not impact model results much and could be included in the analysis to increase the size of the dataset. Although intuition dictates that horizontal curve crashes in the proximity of intersections should be excluded because they could be intersection-related, it may not be the case all the time given that the identification of such crashes is based on reporting officer's judgment and may result in exclusion of crashes relevant to horizontal curve safety.

**Comparison of Curve Crash Prediction Models using *ALL*, *KABALL*, *ALL_N*, and *KABALL_N* Crash Data**

The four crash datasets, namely *ALL*, *KABALL*, *ALL_N*, and *KABALL_N* include crashes which were identified on horizontal curves using mile markers regardless of whether the crash report forms indicated the presence of a horizontal curve at the point of impact. However, *ALL* and *KABALL* crash datasets include crashes occurring within 150 feet of an intersection which are excluded from *ALL_N*, *KABALL_N* crash datasets.

A comparison of curve crash prediction models for total crashes (*ALL* vs. *ALL_N*) as presented in Table 7 (equations 7 and 9) shows some differences where *DiffPSAS* is replaced by left shoulder width and pavement age becomes a significant variable (older pavement leading to more crashes). The sign of *DiffPSAS* coefficient in equation 7 shows that crashes increase as the difference between posted and advisory speed reduces which is counter to the results in previous models. A comparison of curve crash prediction models for fatal/injury crashes (*KABALL* vs. *KABALL_N*) as presented in Table 7 (equations 8 and 10) also shows some differences where posted speed is replaced by left shoulder width and pavement type variable becomes insignificant in equation 10.

Overall, the comparison results suggest that when crashes on horizontal curves are selected based on mile markers only regardless of crash report information, the dataset without crashes in the proximity of an intersection show significant variables which are more relevant to horizontal curve safety as compared to *HORC*-based crash datasets where the inclusion of crashes in proximity of intersections did not result in significant differences. Therefore, caution should be observed in including crashes in proximity of intersections in such conditions.

## FUTURE WORK

The development of high quality large dataset as described in this research will lead the way to the development of additional crash prediction models for other types of horizontal curves. Furthermore, the crash prediction models are the first step to developing a comprehensive set of horizontal curve CMFs in the future to be used in safety evaluations.

## ACKNOWLEDGEMENTS

The authors would like to thank Michael He at the Traffic Operations and Safety Laboratory for his contributions to the statistical analysis. The authors also recognize the Wisconsin Department of Transportation especially Rebecca Szymkowski and Andrew Schilling for their continued support of this and other significant research efforts.

## REFERENCES

1. United States Department of Transportation, Federal Highway Administration, Horizontal Curve Safety. http://safety.fhwa.dot.gov/roadway_dept/horicurves/cmhoricurves/#fn1. Accessed May 12, 2011.
2. NCHRP Report 500 Volume 7: A guide for reducing collisions on horizontal curves
3. Horizontal Curve Signing Handbook. Texas Transportation Institute. 2007.

4. Persaud, B., Retting, R. and Lyon, C. Guidelines for Identification of Hazardous Highway Curves. In Transportation Research Record: Journal of the Transportation Research Board, No. 1717, TRB, National Research Council, Washington, D.C., 2000, p. 14-18.
5. Schneider, W., Zimmerman, K., Van Boxel, D., and Vavilikolanu, S. Bayesian Analysis of the Effect of Horizontal Curvature on Truck Crashes Using Training and Validation Data Sets. In Transportation Research Record: Journal of the Transportation Research Board, No. 2096, TRB, National Research Council, Washington, D.C., 2009, p. 41-46.
6. Schneider, W., Zimmerman, K., Savolainen, P., and Moore, D. Effects of Horizontal Curvature on Single-Vehicle Motorcycle Crashes along Rural Two-Lane Highways. In Transportation Research Record: Journal of the Transportation Research Board, No. 2194, TRB, National Research Council, Washington, D.C., 2010, p. 91-98.
7. Low costs treatments of horizontal curve safety. Federal Highway Administration. 2006.
8. Miaou, Shaw-Pin and Harry Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. Transportation Research Record 1407. pp. 11-24. 1993
9. Zegeer, C., Stewart, R., Safety Effects of geometric improvements on horizontal curves. Transportation Research Board. No. 1356. Washington D.C, 1992. pp.11-19.
10. Tsyganov, A.R., N.M. Warrenchuk, and R. B. Machemehl. Driver Performance and Safety Effects of Edge Lines on Rural Two-Lane Highways. Transportation Research Board 88th Annual Meeting Compendium of Papers DVD-ROM. Washington, D.C., (2009).
11. Srinivasan, R., Baek, J., Carter, D., Persaud, B., Lyon, C., Eccles, K., Gross, F., Lefler, N. Safety Evaluation of Improved Curve Delineation. Report No. FHWA-HRT-09-045, Federal Highway Administration, Washington, D.C. (2009).
12. Khan, G., A. R. Bill, M. Chiturri, and D. A. Noyce. Horizontal Curves, Signs, and Safety. In publication, Transportation Research Record: Journal of the Transportation Research Board.
13. Harwood, D. W., F. M. Council, E. Hauer, W. E. Hughes, and A. Voigt. Prediction of the Expected Safety Performance of Rural Two-Lane Highways. Report FHWA-RD-99-207. FHWA, Washington, D.C., 2000.
14. Zegeer, C. V., R. J. Stewart, F. M. Council, and D. W. Reinfurt. Cost-Effective Geometric Improvements for Safety Upgrading of Horizontal Curves. Report FHWA-RD-90-021. FHWA, Washington, D.C., 1991.
15. Voigt, A., and Krammes, R,. An operational and safety evaluation of alternative horizontal curve design approaches on rural two-lane highways. International Symposium on Highway Geometric Design Practices, 1995.
16. Council, F., Safety Benefits of Spiral Transitions on Horizontal Curves on Two-Lane Rural Roads. In Transportation Research Record: Journal of the Transportation Research Board, Volume 1635, pp. 10-17. 1998
17. Miaou, Shaw-Pin and Harry Lum. Statistical Evaluation of the Effects of Highway Geometric Design on Truck Accident Involvements. Transportation Research Record 1407. pp. 11-24. 1993
18. Hallmark. Low Cost Strategies to Reduce Speed and Crashes on Curves. Iowa Highway Research Board Proposal. 2007
19. Fitzpatrick, K., D. Lord, and B. Park. Horizontal Curve Accident Modification Factor with Consideration of Driveway Density on Rural, Four-Lane Highways in Texas. Transportation

Research Board 88th Annual Meeting Compendium of Papers DVD-ROM. Washington, D.C., (2009).

20. Caliendo, C. M. Guida, A. Parisi. A Crash-Prediction Model for Multilane Roads. Accident Analysis & Prevention, Vol. 39, 2007, pp. 657-670.
21. Bonneson, J., K. Zimmerman, K. Fitzpatrick. Roadway Safety Design Synthesis. FHWA/TX-05/0-4703-P1. Texas Transportation Institute. 2005.
22. Bonneson, J. and M. Pratt. Calibration Factors Handbook: safety prediction models calibrated with Texas highway system data. Texas Transportation Institute, 2008.
23. Fitzpatrick, K., Lord, D., and Park, B. Evaluating Safety Effects of Ramp Density and Horizontal Curve for Freeways Using Texas Data. 2009.
24. Highway Safety Manual. American Association of State Highway and Transportation Officials, Washington D.C. 2010.
25. Zhixia Li, Madhav V. Chitturi, Andrea R. Bill, David A. Noyce (2012) "Automated Identification and Extraction of Horizontal Curve Information from GIS Roadway Maps", Transportation Research Board 91th Annual Meeting, January 2012, Washington D.C., Paper 12-3971.
26. Khan, G., K. R. Santiago-Chaparro, M. Chiturri, and D. A. Noyce. Development of Data Collection and Integration Framework for Road Inventory Data. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2160, No. -1, 2009, pp. 29-39.
27. Sayers, M.W., T. D. Gillespie, and W.D. Paterson. Guidelines for the Conduct and Calibration of Road Roughness Measurements. World Bank Technical Paper No. 46, The World Bank, Washington D.C., 1986.
28. Fox, J. and Monette, G. Generalized Collinearity Diagnostics. *Journal of American Statistical Associated*. Vol. 87, 1992. pp. 178-183
29. Loh, W.-Y. (2011), Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.1, 14-23.
30. Loh, W.-Y. (2002), Regression trees with unbiased variable selection and interaction detection, Statistica Sinica, vol. 12, 361-386.
31. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org. Accessed March 1, 2011.