Using Hierarchical Tree-Based Regression Model to Predict Train-Vehicle Crashes at Passive Highway-Rail Grade Crossings

-- FINAL REPORT --

Contract No.

Submitted to
Southeastern Transportation Center (STC)

By

Xuedong Yan, Ph.D.
Stephen H. Richards, Ph.D., P.E.
The University of Tennessee

Center for Transportation Research
Department of Civil & Environmental Engineering
The University of Tennessee
Knoxville, TN 37996-4133

September 2009
# STC Research Project Description

**Project Title:** Using Hierarchical Tree-Based Regression Model to Predict Train-Vehicle Crashes at Passive Highway-Rail Grade Crossings  
**Principal Investigator:** Xuedong Yan, Ph.D. and Stephen H. Richards, Ph.D., P.E.  
**University:** The University of Tennessee  
**Telephone:** (865) 974-0298  
**Email Address:** xyan1@utk.edu  
**Project Start Date:** December 1st, 2008  
**End Date:** December 30th, 2009  
**Other Milestones, Dates:**  
March 1st, 2008 – A paper was submitted to the journal of Accident Analysis & Prevention

## Project Objective:

The objectives of this project are to apply HTBR models to predict train-vehicle crash frequencies for passive grade crossings controlled by crossbucks only or stop signs respectively and assess how the crash frequencies change after stop-sign treatment is applied at the crossbuck-only-controlled crossings.

## Project Abstract:

This study applies a nonparametric statistical method, Hierarchical Tree-Based Regression (HTBR), to explore train-vehicle crash prediction and analysis at passive highway-rail grade crossings. Using the Federal Railroad Administration database, the research focuses on 27 years of train-vehicle accident history in the United States from 1980 through 2006. A cross-sectional statistical analysis based on HTBR is conducted for public highway-rail grade crossings that were upgraded from crossbuck-only to stop signs without involvement of other traffic-control devices or automatic countermeasures. In this study, HTBR models are developed to predict train-vehicle crash frequencies for passive grade crossings controlled by crossbuck only and crossbucks combined with stop signs respectively, and assess how the crash frequencies change after the stop-sign treatment is applied at the crossbuck-only-controlled crossings.

## Task Description:

This project includes five major research tasks as described in the body of the proposal.  
- Task 1: Literature Review  
- Task 2: Data Preparation  
- Task 3: Statistical Modeling of Train-Vehicle Crashes  
- Task 4: Final Research Report

## Total Budget:

$ 25,000

## Student Involvement (Thesis, Assistantships, Paid Employment):

One Ph.D. students involved this project to assist in data preparation.

## Relationship to Other Projects:

No direct relationships

## Technology Transfer Activities:

A paper was published in the journal of Accident Analysis & Prevention

## Potential Benefits of Project:

The study results indicate that stop-sign treatment is an effective engineering countermeasure to improve safety at the passive grade crossings. Decision makers and traffic engineers can use the HTBR models to examine train-vehicle crash frequency at passive crossings and assess the potential effectiveness of stop-sign treatment based on specific attributes of the given crossings.

## TRB Keywords:

Grade Crossing; Hierarchical Tree-Based Regression; Annual Crash Frequency; Vehicle-Train Crashes; Crossbucks; Stop Signs.
ACKNOWLEDGEMENT

The work described in this final report was sponsored by the Southeaster Transportation Center from a U.S. Department of Transportation Grant through the University Transportation Centers Program. The recommendations of this study are those of the authors, and they do not represent views of U.S. DOT.
Executive Summary

Grade crossings are train-vehicle crash-prone areas due to potential points of conflict between roadway traffic and trains. Because of the substantial mass difference between train and vehicle, the train-vehicle crash injury and fatality rates are much higher than other types of traffic crashes. Compared to highway intersections, highway-rail grade crossings are more critical for crash modeling analysis and prediction. The objectives of this study are to apply HTBR models to predict train-vehicle crash frequencies for passive grade crossings controlled by crossbucks only or stop signs respectively and assess how the crash frequencies change after stop-sign treatment is applied at the crossbuck-only-controlled crossings. The advantages of HTBR models over other methods used here are that the trees allow one to identify homogeneous groups with high or low crash risk and construct rules for making predictions about individual cases. The tree classification rules are helpful for explaining the complex crash patterns based on crossing attributes, train/vehicle traffic characteristics, and safety countermeasures.

In this study, the target crossings are those public highway-rail grade crossings in which each crossing had been open and operating during the 27 years (1980-2006), and were controlled by crossbucks and subsequently upgraded to add stop signs without involvement in other TCDs or automatic countermeasures. 6,596 crossings are used for this study. At these crossings, during the research period there are 6,244 train-vehicle crashes that ever occurred: 4,154 crashes that occurred when the crossings were controlled by crossbucks only and 2,090 crashes that occurred after stop signs were applied.

Four HTBR models were constructed respectively for predicting crossbuck-only-controlled crashes (tree Model #1), predicting stop-sign-controlled crashes (tree Model #2), predicting annual crash frequency difference between periods before and after stop-sign application (tree Model #3), and assessing effectiveness of the stop-sign treatment (tree Model #4).

The results indicate that the significant independent factors of crash prediction for crossbuck-only-controlled crossings are different from those for stop-controlled crossings. The crossbuck-
only-controlled crossing crash frequency depends on five predictors: the number of vehicles per day, the number of trains per day, highway paved or not, the number of tracks, and railroad advance warning signs. The stop-sign-controlled crossing crash frequency depends on three predictors, including the number of trains per day, the number of vehicles per day, and crossing angle. In this research, the crash history at the crossings indicates that the annual crash frequency during the period when the crossings were controlled by crossbucks-only is higher than that during the period after the stop-sign installation. This finding supports that stop-sign treatment should be an effective and inexpensive method for passive grade-crossing safety improvement.

It was found that four important variables were identified for predicting annual crash frequency difference between periods before and after stop-sign application the stop-sign effect, including the maximum train speed, percentage of trucks in highway traffic, the number of vehicles per day, and distance between highway intersections and crossings. Furthermore, the stop-sign effectiveness assessment model was developed to estimate percentage changes in the number of crashes due to the stop-sign effect at a specific crossing during the whole observation period. The HTBR model identified six important predictors for assessing the stop-sign effect, including the number of vehicles per day, the number of tracks, highway paved or not, railroad advance warning signs, crossing angle, and maximum train speed. The model result is consistent with the results of crash prediction trees for crossings controlled by crossbucks only and stop signs.

This research illustrates the flexibility of HTBR application in exploring the safety effectiveness of stop-sign treatment based on specific crossing attributes. Decision makers and traffic engineers can use the HTBR models to examine train-vehicle crash frequency at passive crossings and assess the potential effectiveness of stop-sign treatment based on specific attributes of the given crossings.
# Table of Content

STC Research Project Description.......................................................................................................................... i
Acknowledgement...................................................................................................................................................... ii
Executive Summary...................................................................................................................................................... iii
Table of Content.......................................................................................................................................................... v
List of Tables ............................................................................................................................................................ vi
List of Figures ............................................................................................................................................................. vii
List of Figures............................................................................................................................................................. vii
Chapter 1 - Introduction .............................................................................................................................................. 1
Chapter 2 - Background .............................................................................................................................................. 4
Chapter 3 – Methodology .......................................................................................................................................... 6
  3.1. Train-vehicle crash data.................................................................................................................................. 6
  3.2. Hierarchical Tree-Based Regression............................................................................................................. 7
Chapter 4 – HTBR Modeling and Analyses .............................................................................................................. 9
  4.1 HTBR Model #1 - Predicting Crossbuck-Only-Controlled Crossing Crashes ................................................. 12
  4.2 HTBR Model #2 - Predicting Stop-Sign-Controlled Crossing Crashes ......................................................... 14
  4.3 HTBR Model #3 - Predicting Annul Crash Frequency Difference between Periods Before and After Stop-Sign Treatment ........................................................................................................ 16
  4.4 HTBR Model #4 - Assessing Safety Effectiveness of Stop-Sign Treatment ............................................. 18
Chapter 5 – Discussions and Conclusions ............................................................................................................ 21
  5.1 Passive Grade-Crossing Crash Frequency Prediction.................................................................................. 21
  5.2 Effectiveness of Stop-Sign Treatment .......................................................................................................... 22
  5.3 HTBR Vs. Count Regression Models ........................................................................................................... 24
  5.4 Limitation of the Study ................................................................................................................................. 26
References................................................................................................................................................................. 28
List of Tables

Table 1: The continuous input variables’ descriptions and related statistics ........................................... 9
Table 2: The nominal input variables’ descriptions and related statistics .................................................. 10
List of Figures

Figure 1: Histograms of target variables of the three HTBR trees .................................................. 11
Figure 2: HTBR Model #1 - Predicting crossbuck-only-controlled crossing crashes .................... 13
Figure 3: HTBR Model #2 - Predicting stop-sign-controlled crossing crashes .............................. 15
Figure 4: HTBR Model #3 - Predicting Annul Crash Frequency Difference between Periods Before and After Stop-Sign Treatment .......................................................... 17
Figure 5: HTBR Model #4 - Assessing Safety Effectiveness of Stop-Sign Treatment ............... 20
CHAPTER 1 - INTRODUCTION

Crash modeling analysis and prediction are crucial for developing mechanisms of traffic safety evaluation and management. Decision makers and traffic engineers can use the models to examine safety performance of transportation infrastructures and assess effectiveness of safety countermeasures. Using statistical models, the relationships between vehicle accidents and geometric design, traffic characteristics, and engineering factors have been extensively studied for highway intersections or segments and rail-highway grade crossings. Grade crossings are train-vehicle crash-prone areas due to potential points of conflict between roadway traffic and trains. Because of the substantial mass difference between train and vehicle, the train-vehicle crash injury and fatality rates are much higher than other types of traffic crashes. Therefore, compared to highway intersections, highway-rail grade crossings are more critical for crash modeling analysis and prediction.

A number of previous studies have been carried out to analyze and predict train-vehicle crashes at highway-rail grade crossings. Oh et al. (2006) reported that the earlier grade-crossing accident prediction methods, such as the Peabody Dimmick Formula, the New Hampshire Index, the National Cooperative Highway Research Program (NCHRP) Hazard Index, and US Department of Transportation (USDOT) Accident Prediction Equations, either lack descriptive capabilities because of their limited number of explanatory variables, or decline in accident prediction accuracy over time. McCollister and Pflaum (2006) used the logistic regression model and Federal Railroad Administration (FRA) databases to predict the probability of train-vehicle accidents, injuries, and fatalities. The model was used for cost analysis and estimation for various types of grade crossings based the significant risk factors related to crossing attributes. Washington and Oh (2006) explored the Bayesian methodology that incorporated expert judgment to rank countermeasure effectiveness on grade-crossing safety.

Because crash frequency data are nonnegative integer-valued random data, Poisson and negative binomial regression models have been commonly applied for grade-crossing safety analysis and prediction (Austin and Carson, 2002; Park and Saccomanno, 2005; Millegan et al., 2009). As
extensions of Poisson and negative binomial models, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models have also been used (Nam and Lee, 2006; Oh et al., 2006; Hu and Lee, 2008), accounting for crashes that are so low in frequency over some time period that they can be considered as a zero-accident state (i.e., many zero crash frequencies in the data). However, it is difficult to use the count regression models to directly test safety effectiveness of a specific safety countermeasure. Because the crash frequency difference between periods before and after the countermeasure application is not a nonnegative integer any more, the crash frequency difference cannot be treated as a dependent variable in these count models. Although other parametric regression tools such as linear models or generalized linear models can be used to model the before-after crash frequency difference as a function of multiple independent variables, there are some typical concerns associated with the parametric linear regression methods, as indicated by, e.g., Hadi et al. (1995), Mohamedshah et al. (1993), Karlaftis and Tarko (1998), and Karlaftis and Golias (2002). Firstly, these parametric methods rely heavily on various statistical assumptions, which could be hard to verify or unlikely to hold in applications. Secondly, the functional relationship between dependent and independent variables must be fully specified and typically assumed to be linear. Nevertheless, model misspecifications such as omitting relevant variables from the model or including irrelevant variables in the model are common (Washington, 2000). For example, one may simply assume linearity in modeling the relationship between accidents and exposure (such as traffic flow) that actually shows a nonlinear pattern (Lord, 2002; Hauer and Persaud, 1987). While linear regression does have certain flexibility to incorporate nonlinearity by allowing for transformations of variables, e.g., polynomial terms, linear approximation is often inadequate to cope with complicated nonlinear relationships, especially in high dimension. Thirdly, cross-product terms are used to formulate interactions among predictors. Nevertheless, interaction may occur in complex forms and of high orders, which renders interaction detection a daunting task in linear regression. Fourthly, the dummy variable approach of handling discrete predictors would result in a massive model when there are many discrete variables involved or when the number of categories becomes large.

In this study, a nonparametric statistical method, Hierarchical Tree-Based Regression (HTBR), is used to predict train-vehicle crash frequency at passive highway-rail grade crossings and
evaluate effectiveness of safety countermeasures. HTBR methods are data exploration models that are extremely valuable for understanding the data structure or relationships among variables (Washington, 2000). In contrast to parametric linear regression, HTBR facilitates a piecewise constant approximation to the underlying regression function. Such an approximation can be made to a high degree of accuracy with enhancement on trees, i.e., bagging, boosting, or random forests (see, e.g., Hastie, Tibshirani, and Friedman, 2001). As a nonparametric tool in nature, HTBR relies on few statistical assumptions and does not require a functional form to be specified. Usually no transformation on the predictors is necessary as HTBR results are invariant to monotone transformations. HTBR also excels in efficiently handling categorical predictors. Besides, the hierarchical tree structure automatically handles data with higher-order and complex interactions. The HTBR models have been successfully applied to highway safety studies for predicting rural road crash frequency (Karlaftis and Golias, 2002) and freeway crash frequency (Chang and Chen, 2005), comparing crash types at signalized intersections (Abdel-Aty et al., 2005), classifying intersection crashes (Qin and Han, 2008), analyzing rear-end crashes (Yan and Radwan, 2006), investigating roadway safety (Stewart, 1996) and traffic injury (Chang and Wang, 2006), examining the effects of on-ramp design factors on vehicle operation (Lederer et al., 2005), identifying roadway characteristics that influenced vehicle activity (Hallmark et al., 2002), and predicting trip generation (Washington and Wolf, 1997).

However, there are very few applications of HTBR in the study of grade crossing safety. The objectives of this study are to apply HTBR models to predict train-vehicle crash frequencies for passive grade crossings controlled by crossbucks only or stop signs respectively and assess how the crash frequencies change after stop-sign treatment is applied at the crossbuck-only-controlled crossings. The advantages of HTBR models over other methods used here are that the trees allow one to identify homogeneous groups with high or low crash risk and construct rules for making predictions about individual cases. The tree classification rules are helpful for explaining the complex crash patterns based on crossing attributes, train/vehicle traffic characteristics, and safety countermeasures.
CHAPTER 2 - BACKGROUND

The Manual on Uniform Traffic Control Devices (MUTCD) provides guidance on what traffic-control devices (TCDs) should be used at public passive highway-rail grade crossings. At minimum, one crossbuck sign should be used on each highway approach to every highway-rail grade crossing, alone or in combination with other traffic control devices to mark the location of the railroad tracks at the point where they cross the road (MUTCD, 2003). A stop sign is an optional TCD treatment at passive crossings. Stop-sign use is recommended at the discretion of the responsible state or local highway agency if highway-rail grade crossings have two or more trains per day and are without automatic traffic control devices (MUTCD, 2003). However, engineers and policy makers are not in complete agreement about whether stop signs are effective when used at highway-railroad grade crossings. The safety effectiveness of stop-sign treatment employed at passive crossings has been a controversial subject for many years (Lerner et al., 2002).

The National Transportation Safety Board (NTSB) (1998) suggests a broader use of stop signs at railroad-highway grade crossings and recommends stop signs as an interim device until intelligent transportation systems are developed to warn the driver. Eck and Shanmugam (1981) reported that upgrades from no signs or crossbucks to stop signs can significantly reduce crash rates at both low-volume and higher-volume highway-rail grade crossings. Sanders et al. (1978) found that stop signs are used more frequently in urban areas, and crossings that have stop signs tend to have higher train volumes; crash rates for stop-sign crossings are lower than those for crossbuck-only crossings for higher vehicle-train exposure values; and stop signs, when properly used, result in improved driver behaviors that help with detection and avoidance of trains. They suggested that stop signs should be applied selectively, only at hazardous passive grade crossings, and should not be used indiscriminately at all passive grade crossings. Additionally, in Canada, it was found that the stop-sign countermeasure can improve crossing safety performance by as much as 35% (Saccomanno et al., 2007). In sum, these study results support stop sign application at passive grade crossings.
On the other hand, some researchers do not suggest the use of stop signs. It has been argued by Bezkorovainy and Holsinge (1966) and Burnham (1994) that motorists frequently disregard stop signs at grade crossings. The high level of noncompliance might increase and carry over to other locations if stop signs are used indiscriminately (Lerner et al., 2002). Those observational studies show that the percentage of drivers not coming to a complete stop is higher than the percentage found at highway intersections. Nevertheless, one needs evidence to support that the high noncompliance rate correspondingly leads to a high crash rate at stop-controlled crossings. A recent study (Raub, 2006) examined 10 years of collision data in seven Midwestern states using the FRA crash database, and compared collision rates among four types of crossings: crossbucks, stop signs, flashing lights, and gates. It is reported that, compared to the other type of crossings, collision rates for crossings with stop signs are much higher, especially when using millions of crossing vehicles as the collision rate calculation base.

To date the safety benefit of the stop-sign treatment employed at passive crossings appears still unresolved and controversial. The best way to address this issue is to investigate how train-vehicle crash risk changed at crossings that were upgraded from Crossbuck-only to stop signs without using other TCDs or automatic countermeasures. Nevertheless, a before-and-after analysis of stop-sign treatment has not yet been conducted for passive grade crossings at a nationwide scale.
CHAPTER 3 – METHODOLOGY

3.1 Train-vehicle crash data

This is a cross sectional study using 27 years (1980-2006) of train-vehicle crash data retrieved from the Federal Railroad Administration (FRA). FRA keeps records of train-related accidents across the entire United States (FRA, 2008). The FRA database contains three sub-databases:

1. The Grade-Crossing Inventory database: it is a record of the current crossing inventory. Reference attributes in this database reflect the current state of each crossing.
2. The Grade-Crossing Inventory History database: it reflects the state changes of the crossing, including a reason for the update and an effective date for the change.
3. The Grade-Crossing Accident History database: it provides a record of accidents that have occurred at the crossings and the conditions at the time of the accident.

The three databases are linkable to each other by common crossing ID. In this study, the target crossings are those public highway-rail grade crossings in which each crossing had been open and operating during the 27 years (1980-2006), and were controlled by crossbucks and subsequently upgraded to add stop signs without involvement in other TCDs or automatic countermeasures. The Grade-Crossing Inventory was used to identify independent factors that reflect crossing-related attributes and train/vehicle traffic patterns. The Grade-Crossing Inventory History was used to identify when passive crossings were updated from crossbucks-only to stop sign control and ensure that during the research period the crossings were not involved in other control treatments. The Grade-Crossing Accident History database was applied to obtain the accident frequency for each target passive crossing during the period of specific control treatment (crossbucks-only or stop signs added).

It should be mentioned that although the FRA database is found to be the most complete and accurate datasets available for this research, several issues that potentially cause confusion arise when manipulating the data. For example, some redundant duplicates are found in the Grade-Crossing Inventory database. Also, some records in one database do not match well with these in
another separately maintained database. To ensure the appropriateness and quality of data, we consulted with the FRA staffs who were responsible for data maintenance. Those confusing data have been excluded from analysis in the study. Finally, 6,596 crossings are used for this study. At these crossings, during the research period there are 6,244 train-vehicle crashes that ever occurred: 4,154 crashes that occurred when the crossings were controlled by crossbucks only and 2,090 crashes that occurred after stop signs were applied.

3.2 Hierarchical Tree-Based Regression

HTBR classifies observations by recursively partitioning the predictor space. Due to its nonparametric nature and easy interpretation, HTBR has received wide popularity in various fields. In tree-structured representations, the entire dataset is represented by a root node. When a split is made, child nodes that correspond to partitioned subsets are formed. If a node is not to be split any further, it is called a terminal node that is associated with a group membership; otherwise, it is an internal node. The tree is constructed following a set of decision rules applied sequentially. Each decision rule is used to form branches (i.e. splitting) connecting the root node to the terminal node at a certain level of the tree. Based on the decision rules, the HTBR model’s explanation is straightforward for both analysis and predictive purposes.

The HTBR procedure creates a tree-based classification model using the CART (Classification and Regression Trees), CHAID (Chi-squared Automatic Interaction Detection), or QUEST (Quick, Unbiased, Efficient, Statistical Tree) algorithm. Among the three algorithms, QUEST is not appropriate for this study since it can be used only for the nominal dependent variable (crash frequency is obviously a continuous variable). Both CHAID and CART techniques can construct regression-type trees (continuous dependent variable), where each (non-terminal) node identifies a split condition, to yield optimum prediction. However, CART will always yield binary trees, which can sometimes not be summarized as efficiently for interpretation and/or presentation (Breiman et al., 1984). Therefore, CHAID algorithm is applied for this study because it allows multi-way splits of a node, based on a relatively simple algorithm that is particularly well suited for the analysis of larger datasets.
The CHAID algorithm, originally proposed by Kass (1980) and further developed by Magidson (1993), automatically determines the most significant split criteria based on Chi-square tests to measure the association between dependent variable and independent variables. The procedure consists of three steps: first preparing predictors, then merging categories, and finally selecting the split variable. The process continues until no further splits can be performed (given the alpha-to-merge and alpha-to-split values).

Preparing predictors is to create categorical predictors out of any continuous predictors by dividing the respective continuous distributions into a number of categories with an approximately equal number of observations. Merging categories is to cycle through the predictors to determine for each predictor the pair of (predictor) categories that is least significantly different with respect to the dependent variable. If the respective test for a given pair of predictor categories is not statistically significant as defined by an alpha-to-merge value, the categories are merged, and then this step is repeated. If the statistical significance for the respective pair of predictor categories is significant, then a Bonferroni adjusted p-value is computed for the set of categories for the respective predictor. Selecting the split variable is to choose the split the predictor variable with the smallest adjusted p-value, which indicates the strongest association between independent variable and target variable. If the smallest adjusted p-value for any predictor is greater than some alpha-to-split value, then no further splits will be performed, and the node is claimed as terminal. For further discussions of the CHAID algorithm, the reader is referred to SPSS 16.0 Algorithms (2008).

The HTBR analyses in this study were carried out using the SPSS software package (version 16.0; SPSS Inc., Chicago, IL, USA). The specifications for tree construction include: the maximum tree depth was set as 3 levels; the minimum number of cases for parent nodes was set as 100 and the minimum number of cases for child nodes was set as 50; the significance values for both splitting nodes and merging categories are set as 0.05. In order to select the best tree size, the CART convention is followed. More specifically, a large initial tree is first grown and pruned to obtain a sequence of nested subtrees. Then, the 10-fold cross-validation method was used to assess how well each subtree generalizes to a larger population and hence determine the best final tree model.
CHAPTER 4 – HTBR MODELING AND ANALYSES

Four HTBR models were constructed respectively for predicting crossbuck-only-controlled crashes (tree Model #1), predicting stop-sign-controlled crashes (tree Model #2), predicting annual crash frequency difference between periods before and after stop-sign application (tree Model #3), and assessing effectiveness of the stop-sign treatment (tree Model #4). For data inputs, the four tree models have different target (dependent) variables, but have the same predictor variables (independent variables) that represent crossing attributes and train/vehicle traffic characteristics. Totally, twelve predictor variables were used in each model, including six continuous variables and six nominal (categorical) variables. The input variables’ descriptions and related statistics for the four HTBR models are listed in Tables 1 and 2 for continuous variables and nominal variables respectively.

Table 1: The continuous input variables’ descriptions and related statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Descriptive</th>
<th>Input Type</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>XCrash_Annual_Freq</td>
<td>Annual crash frequency of crossbuck-only-controlled crossings</td>
<td>Target for Tree Model #1</td>
<td>0.058</td>
<td>0.200</td>
</tr>
<tr>
<td>SCrash_Annual_Freq</td>
<td>Annual crash frequency of stop-sign-controlled crossings</td>
<td>Target for Tree Model #2</td>
<td>0.032</td>
<td>0.122</td>
</tr>
<tr>
<td>Crash_Freq.Diff</td>
<td>Annual crash frequency difference between before and after stop-sign application</td>
<td>Target for Tree Model #3</td>
<td>-0.025</td>
<td>0.217</td>
</tr>
<tr>
<td>Sign_Effect</td>
<td>The portion of annual crash frequency change due to the stop-sign effect at a specific crossing during the whole observation period</td>
<td>Target for Tree Model #4</td>
<td>-0.138</td>
<td>0.590</td>
</tr>
<tr>
<td>TRAINS</td>
<td>The number of trains per day</td>
<td>Predictor</td>
<td>9.603</td>
<td>11.397</td>
</tr>
<tr>
<td>TRACKS</td>
<td>The number of tracks</td>
<td>Predictor</td>
<td>1.318</td>
<td>0.817</td>
</tr>
<tr>
<td>AADT</td>
<td>The number of vehicle per day</td>
<td>Predictor</td>
<td>344.679</td>
<td>875.615</td>
</tr>
<tr>
<td>MAXTTSPD</td>
<td>The maximum train speed (mph)</td>
<td>Predictor</td>
<td>39.603</td>
<td>18.170</td>
</tr>
<tr>
<td>PCTTRUK</td>
<td>Percentage of truck in highway traffic</td>
<td>Predictor</td>
<td>8.852</td>
<td>8.980</td>
</tr>
<tr>
<td>TRAFICLN</td>
<td>The number of traffic lanes</td>
<td>Predictor</td>
<td>1.814</td>
<td>0.428</td>
</tr>
</tbody>
</table>
Table 2: The nominal input variables’ descriptions and related statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Descriptive Information</th>
<th>Input Type</th>
<th>Levels</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEVELTYP</td>
<td>Type of development of the crossing surrounding</td>
<td>Predictor</td>
<td>1 = Open space</td>
<td>3657</td>
<td>55.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 = Residential</td>
<td>1661</td>
<td>25.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 = Commercial</td>
<td>656</td>
<td>9.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 = Industrial</td>
<td>525</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 = Constitutional</td>
<td>97</td>
<td>1.5</td>
</tr>
<tr>
<td>HWYPVED</td>
<td>Is highway paved?</td>
<td>Predictor</td>
<td>1 = Yes</td>
<td>3943</td>
<td>59.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 = No</td>
<td>2653</td>
<td>40.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 = Stoplines</td>
<td>91</td>
<td>1.4</td>
</tr>
<tr>
<td>PAVEMRK</td>
<td>Pavement markings</td>
<td>Predictor</td>
<td>2 = RR Xing symbols</td>
<td>152</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 = No markings</td>
<td>5357</td>
<td>81.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 = Stoplines and RR Xing symbols</td>
<td>996</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 = less than 75 ft</td>
<td>2678</td>
<td>40.6</td>
</tr>
<tr>
<td>HWYNEAR</td>
<td>Nearby highway intersections</td>
<td>Predictor</td>
<td>2 = 75-200 ft</td>
<td>45</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 = 200-500 ft</td>
<td>46</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4 = more than 500 ft</td>
<td>3827</td>
<td>58.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 = Yes</td>
<td>4498</td>
<td>68.2</td>
</tr>
<tr>
<td>ADVWARN</td>
<td>Railroad advance warning signs</td>
<td>Predictor</td>
<td>2 = No</td>
<td>2098</td>
<td>31.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 = 0-29</td>
<td>277</td>
<td>4.2</td>
</tr>
<tr>
<td>XANGLE</td>
<td>Smallest crossing angle</td>
<td>Predictor</td>
<td>2 = 30-59</td>
<td>1151</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3 = 60-90</td>
<td>5168</td>
<td>78.4</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>6596</td>
<td>100</td>
</tr>
</tbody>
</table>

In Model #4, the target variable (Sign_Effect) is calculated by the other three target variables, XCrash_Anual_Freq, SCrash_Anual_Freq, and Crash_Freq_Diff, as shown in Equation 1:

$$\text{Sign\_Effect} = \frac{\text{Crash\_Freq\_Diff}}{\text{SCrash\_Anual\_Freq} + \text{XCrash\_Anual\_Freq}}$$ (1)

Sign_Effect indicates the portion of annual crash frequency change due to the stop-sign effect at a specific crossing during the whole observation period. In the equation, the sum of SCrash_Anual_Freq and XCrash_Anual_Freq is used as denominator rather than XCrash_Anual_Freq. This strategy of calculation helps avoid the situation where XCrash_Anual_Freq is equal to zero. When both crash frequencies during crossbuck-only-controlled period and stop-sign-controlled period, zero is assign to the target variable of Sign_Effect.
a. Histogram of annual crash frequency of crossbuck-only-controlled crossings

b. Histogram of annual crash frequency of stop-sign-controlled crossings

c. Histogram of annual crash frequency difference between before and after stop-sign application

Figure 1: Histograms of target variables of the three HTBR trees
On average, the annual crash frequency when controlled by crossbucks-only (0.058 crashes/year/crossing) is higher than that when controlled by stop signs (M = 0.032 crashes/year/crossing). The average crash frequency difference between before and after stop sign application is -0.025 crashes/year/crossing. Histograms of target variables, XCrash_Annual_Freq, SCrash_Annual_Freq, and Crash_Freq_Diff in the three HTBR trees are respectively illustrated in Figures 1-a, b, and c. It can be observed that zero crash frequency accounts for a very high percent of crash data, and the stop-sign application leads to a lower crash frequency. Note that these crash observations cover a 27-year crash history of the studied crossings. Therefore, one can conclude that the train-vehicle crashes are rare events on the annual frequency scale. The following sections specify the HTBR results and analyze how the train-vehicle crash frequency is associated with the crossing attributes and traffic characteristics.

4.1 HTBR Model #1 - Predicting Crossbuck-Only-Controlled Crossing Crashes

The results of Model #1 are displayed in Figure 2, which are applied for crash prediction and classification at the crossbuck-only-controlled crossings. The final tree structure for XCrash_Annual_Freq involves five splitting variables, including AADT, TRAINS, HWYPVED, TRACKS, and ADVWARN.

The first optimal split in node 0 is according to AADT, which classifies crashes into three groups: if AADT is less than or equal to 40 vehicles/crossing, the tree predicts 0.032 crash/year/crossing; if AADT is greater than 40 vehicles/crossing but less than 400 vehicles/crossing, the crash frequency is 0.055 crashes/year/crossing; and if AADT is greater than or equal to 400 vehicles/crossing, the crash frequency is 0.093 crashes/year/crossing.

In the second level of the tree, TRAINS leads to the further splits. However, these splits are based on different thresholds among the three AADT groups (nodes 1, 2, and 3) in the first level of the tree. For the \( \leq 40 \) vehicles/crossing AADT group, TRAINS segments the data into two subgroups: less than or equal to three trains per day, or greater than three trains per day. For the 41-399 vehicles/crossing AADT group, TRAINS segments the data into three subgroups: one train per day, 2-15 trains per day, or greater than 15 trains per day. For the >399.
vehicles/crossing AADT group, TRAINS also segments the data into three subgroups: less than or equal to three trains per day, 3-15 trains per day, or greater than 15 trains per day. These classification rules display a clear trend that the crash frequency at the crossbuck-only-controlled crossings increases as AADT or TRAINS increases. The daily traffic volumes of both vehicles and trains play the most important role in predicting crash frequency at the crossings. This finding is consistent with research results in many previous crossing crash frequency modeling studies (Austin and Carson, 2002; Saccomanno et al., 2004; Oh et al., 2006).

**Figure 2: HTBR Model #1 - Predicting crossbuck-only-controlled crossing crashes**

In the third level of the tree, HWYPVED leads to the split in node 6 (41-399 vehicles/crossing and less than or equal to one train per day). It shows that the paved highways have a higher crash frequency than unpaved highways. TRACKS further splits the node 7 (41-399 vehicles/crossing
and 2-15 trains per day) into two groups: one track (node 14) and multiple tracks (node 15) at the crossings. These two crossing groups account for 42.4% of the total crossings. This finding indicates that the crossings with multiple tracks (0.083 crashes/year/crossing) almost double crash frequency than those with only one track (0.049 crashes/year/crossing). The Federal Highway Administration’s *Railroad-Highway Grade Crossing Handbook* uses TRACKS to calculate hazard index (FHWA, 1986), which means that a greater number of tracks leads to a higher crash index. Since multiple tracks increase the exposure that vehicles are blocked on the top of tracks, TRACKS should be considered as an important crossing feature presented in safety models.

Additionally, ADVWARN leads to the split in node 10 (>399 vehicles/crossing and >15 trains per day), but the split shows an abnormal crash pattern: the crossings with railroad advance warning signs have a higher crash frequency (0.123 crashes/year/crossing) while those without railroad advance warning signs have a lower crash frequency (0.078 crashes/year/crossing). A presumable explanation is that traffic engineers are more likely to choose potential dangerous crossings (for example, limited sight distance) for application of the railroad advance warning signs. Since ADVWARN is not a significant variable in any other groups, it implies the railroad advance warning signs would not improve safety at the crossbuck-only-controlled crossings.

### 4.2 HTBR Model #2 - Predicting Stop-Sign-Controlled Crossing Crashes

The results of Model #2 shown in Figure 3 are applied for crash prediction and classification at the crossings during the stop-sign-controlled period. The SCrash_Anual_Freq is dependent on three predictors, including TRAINS, AADT, and XANGLE.

The top of the tree shows that at the most important crash prediction factor for stop-sign-controlled crossings is the number of trains per day (TRAINs). It classifies the crashes into four groups: if TRAINS is less than or equal to 3 trains per day, the crash frequency is 0.011 crashes/year/crossing; if TRAINS is greater than 3 trains per day and less than 8 trains per day, the crash frequency is 0.030 crashes/year/crossing; if TRAINS is between 8-15 trains per day, the crash frequency is 0.042 crashes/year/crossing; and if TRAINS is greater than 15 trains per
day, the crash frequency is 0.061 crashes/year/crossing. This finding indicates that the higher train volume is associated the higher level of crash frequency at the stop-sign-controlled crossing.

Figure 3: HTBR Model #2 - Predicting stop-sign-controlled crossing crashes

In the second level of the tree, AADT leads to the further splits. The general trend is that within each group segmented by the number of trains, the crash frequency increases as the traffic volume of vehicles increases. However, the thresholds of AADT corresponding to different levels of train volume at crossings are different, as shown in nodes 5-16 in Figure 3. The crash frequency prediction is based on the various crash exposure combinations of vehicle and train traffic volumes. The difference in crossing crash prediction between crosbucks and stop signs is
that the daily traffic volume of vehicles is the most important predictor for crossbuck-only-controlled crossings while the daily train volume plays a more important role for stop-controlled crossings.

Additionally, XANGLE leads to a split in the third level of the tree. An interesting finding is that for the crossings with the lowest vehicle and train volumes, both sharp and right crossing angles between rail tracks and highways show slightly higher crash frequency than the 30-59 degree of crossing angles.

### 4.3 HTBR Model #3 - Predicting Annul Crash Frequency Difference between Periods Before and After Stop-Sign Treatment

In Model #3, as shown in Figure 4, the target variable (Crash_Freq.Diff) is the crash frequency difference between crossbuck-only-controlled period and stop-sign-controlled period, which is equal to SCrash_Anual_Freq minus XCrash_Anual_Freq. Thus, a negative value indicates that the annual crash frequency at a specific crossing decreases after the stop-sign application. Figure 1-c shows more negative values of Crash_Freq.Diff, which means that the stop-sign treatment leads to a lower crash frequency.

The results of Model #3 are applied for predicting and assessing the effect of stop-sign treatment based on the crossing characteristics. As shown in Figure 4, MAXTTSPD, PCTTRUK, ADVWARN, AADT, and HWYNEAR are significant factors determining the stop-sign effect on the grade crossing safety.

The top of the tree shows that the maximum train speed (MAXTTSPD) is the most important factor. The stop-sign treatment has an apparently better effect on crash reduction when stop signs are applied on crossings with lower train speed (lower than or equal to 40 mph) compared to crossings with higher train speed.

For the crossings with the lower train speed (node 1), the percentage of trucks in highway traffic seems to be nonlinearly associated the stop-sign safety effect. The crossings with 6-10 percent of
trucks achieved a larger amount of crash frequency reduction than those with lower or higher truck percentage. For the crossings with the higher train speed (node 2), if railroad advance warning signs were not applied, the stop-sign treatment barely had a safety effect; however, if the railroad advance warning signs were applied at the crossings, the crash frequency is reduced by 0.016 crashes/year/crossing. The railroad warning signs are used to give notice of crossings that are potentially hazardous to traffic. The combination application of stop signs and railroad warning signs might be helpful to increase drivers’ alertness level to potential hazard when approaching the higher train-speed crossings.

Figure 4: HTBR Model #3 - Predicting Annual Crash Frequency Difference between Periods Before and After Stop-Sign Treatment
The third level of the tree displays that for the lower train speed crossings with a lower truck percentage, those with higher AADT (greater than 179 vehicles/crossing) are associated with a larger amount of crash frequency reduction than those with lower AADT. Trucks can be 40 or more times heavier than the other vehicles in the traffic stream. Due to physical and operational characteristics of the heavy trucks, they can significantly affect traffic system performance and safety at grade crossings (Davey et al., 2008). There may be incompatibility issues in crossing design, which are perhaps ignored when the percentage of trucks is lower in highway traffic. Additionally, according to the Federal Safety Regulations, trucks hauling hazardous materials are required to completely stop at any passive grade crossings before proceeding across. The stop-sign use may efficiently reduce the non-stop truck drivers who forget or do not know the special regulation for truck safety.

For the lower train speed crossings with 6-10 percent of truck in highway traffic, HWYNEAR plays an important role on the stop sign’s safety effect. The effect was larger at crossings that were more than 500 ft from the adjacent intersections (-0.108) than at crossings that were nearer (-0.028). A previous study reported that the distance between intersection and grade crossing is an important factor for both intersection operation and crossing safety (Cho and Rilett, 2007). During peak hours, the queue at an adjacent intersection may be extended to the crossing. In this case, stop signs can help drivers stop first, and then judge if there is a safe clearance distance between the track and the queue end. It would contribute to avoiding the situation in which drivers are forced to stop on the top of tracks by the queue. On the other hand, the drivers, who drive from the adjacent intersection to the crossing, may be less willing to stop at the crossing if they have stopped at the intersection and/or they make a prejudgment about the crossing’s risk at the intersection. This case may provide a possible explanation why stop signs are less effective at crossings that are nearby intersections than at crossings that are 500 ft from intersections.

4.4 HTBR Model #4 - Assessing Safety Effectiveness of Stop-Sign Treatment

Model #4 is built for assessing the effect of stop-sign treatment based on the crossing characteristics. As shown in Figure 5, AADT, TRACKS, HWYPVED, ADVWARN, XANGLE, and MAXTTSPD are significant factors determining the stop-sign effect on the grade crossing
safety. The model results show that the values of Sign_Effect in most nodes are negative, except for nodes 13 and 17. It indicates that the stop-sign treatment generally is an effective safety countermeasure for reducing vehicle-train crashes.

The top of the tree shows that AADT is the most important factor to predict stop-sign treatment effect. It was classified into three groups: if AADT is less than or equal to 80 vehicles/crossing, the crash reduction rate is -8.9%; if AADT is greater than 80 vehicles/crossing but less than or equal to 400 vehicles/crossing, the crash reduction rate is -13.2%; and if AADT is greater than 399 vehicles/crossing, the crash reduction rate is -25.1%. Clearly the higher AADT, the larger effectiveness of stop sign.

For the crossings with the lower AADT (node 1), the number of tracks is significantly associated with the stop-sign safety effect. It is found that stop signs applied to single-track crossings can achieve more crash reduction (-13.3%) that those applied to multiple-track crossings (-3.9%). Further, the single-track crossings with 36-59 degrees of crossing angle (-16.8%) have a better effect of stop-sign treatment than those with 0-35 and 60-90 degrees of crossing angles (-8.8%). For the multiple-track crossings, stop signs are effective at crossings when train speed is lower than or equal to 55 mph (-12.8%), while they may increase train-vehicle crashes when train speed is higher than 55 mph (-12.8%).

For the crossings with the moderate AADT (node 2), stop signs applied to crossings with paved highways (-15.4%) are more effective than non paved highways (-7.5%). For the crossings with paved highways, stop signs are less effective if the train speed is higher than 40 mph (-8.5% vs. -18.4%).

For the crossings with the higher AADT (node 3), if railroad advance warning signs were applied, the crash reduction rate due to the stop-sign treatment may be up -30.1%. The railroad warning signs are used to give notice of crossings that are potentially hazardous to traffic. The combination application of stop signs and railroad warning signs might be helpful to increase drivers’ alertness level to potential hazard when approaching the higher train-speed crossings. However, this combination effect are more significant at paved crossings than unpaved crossings.
(-31.4% vs. -8.0%). On the other hand, if railroad advance warning signs were not applied, a stop sign is a significant countermeasure for crossings 60-90 degrees of crossing angles (-19.8%) but has no safety effect for those with 0-35 and 60-90 degrees of crossing angles (1.1%).

**Figure 5: HTBR Model #4 - Assessing Safety Effectiveness of Stop-Sign Treatment**
5.1 Passive Grade-Crossing Crash Frequency Prediction

In this study, a nonparametric-regression statistical method, HTBR, is applied to predict annual train-vehicle crash frequency at passive highway-rail grade crossings controlled by crossbucks-only or stop signs. The results indicate that the significant independent factors of crash prediction for crossbuck-only-controlled crossings are different from those for stop-controlled crossings. The crossbuck-only-controlled crossing crash frequency depends on five predictors: the number of vehicles per day, the number of trains per day, highway paved or not, the number of tracks, and railroad advance warning signs. The stop-sign-controlled crossing crash frequency depends on three predictors, including the number of trains per day, the number of vehicles per day, and crossing angle.

As the exposure of train-vehicle conflicts, the number of vehicles per day and the number of trains per day are the most important crash predictors for passive grade crossings. The daily vehicle volume is the most important crash predictor for crossbuck-only-controlled crossings while the daily train volume plays a more important role on crash prediction at stop-sign-controlled crossings. MUTCD only use train volume (if highway-rail grade crossings have two or more trains) as a guideline for stop-sign use (MUTCD, 2003). The finding of this study suggests that the vehicle volume should also be included into the guideline. For the crossbuck-only-controlled crossing crash prediction, some other noticeable findings include that the paved highways lead to a higher crash frequency than unpaved highways; multiple tracks at crossing result in almost two times more train-vehicle crashes than one track; and railroad advance warning signs may not significantly improve safety at the crossings.

Previous grade-crossing crash prediction studies based on generalized linear regression models reported that the crossing angle between rail track and highway is not a significant factor in affecting highway-rail crossing accident frequency (Austin and Carson, 2002; Oh et al., 2006).
However, the HTBR modeling result indicates that for the stop-sign-controlled crossings with the lowest vehicle and train volumes both 0-35 and 60-90 degrees of crossing angles have slightly higher crash frequencies than the 36-59 crossing angle. This finding demonstrates the capability of HTBR in identifying nonlinear relationship between the response and the predictors.

5.2 Effectiveness of Stop-Sign Treatment

As mentioned early, the safety effect of stop-sign treatment employed at passive crossings is a controversial issue. Several reasons may account for the inconsistent conclusions. The primary reason for nonuse or limited use of stop signs is that there are a high percent of drivers failing to come to a complete stop (Bezkorovainy and Holsinger, 1966; Burnham, 1994). However, the high level of noncompliance does not necessarily offset the effect of stop signs. Compared to crossbucks, adding stop signs can provide longer perception and reaction times to drivers approaching the crossings even for those who do not completely stop (Washington and Oh, 2006). Secondly, due to regional differences existing in drivers, environments, designs, and management factors, there may be different crash experiences related to the crossings in different regions. Oh et al. (2006) found that presence of stop signs is a significant factor affecting accident frequency in the USDOT accident formula, while it is found insignificant in a gamma model built on Korean crash data. Thirdly, simply comparing crash rates between crossbuck-only-controlled crossings and stop-sign-controlled crossings may cause bias in conclusion. STOP signs are more likely determined by engineering judgment to be used for dangerous crossbucks-controlled crossings such as limited sight distance or increment of train and vehicle volume. It is possible to observe a higher crash rate at the crossings selected for stop-sign treatment than those crossings without any treatment need. Therefore, a large-scale before-after study would be more robust to test the efficacy of stop-sign use.

In this research, the crash history at the crossings indicates that the annual crash frequency during the period when the crossings were controlled by crossbucks-only is higher than that during the period after the stop-sign installation. This finding supports that stop-sign treatment should be an effective and inexpensive method for passive grade-crossing safety improvement.
This analysis is consistent with several prior study conclusions in stop-sign usage at passive crossings (Sanders et al., 1978; Eck and Shanmugam, 1981). This research also illustrates the flexibility of HTBR application in exploring the safety effectiveness of stop-sign treatment based on specific crossing attributes. HTBR can be used to predict and estimate crash frequency change and change rate due to a safety countermeasure, but the count regression models cannot do that directly.

It was found that four important variables were identified for predicting annual crash frequency difference between periods before and after stop-sign application the stop-sign effect, including the maximum train speed, percentage of trucks in highway traffic, the number of vehicles per day, and distance between highway intersections and crossings. Furthermore, the stop-sign effectiveness assessment model was developed to estimate percentage changes in the number of crashes due to the stop-sign effect at a specific crossing during the whole observation period. The HTBR model identified six important predictors for assessing the stop-sign effect, including the number of vehicles per day, the number of tracks, highway paved or not, railroad advance warning signs, crossing angle, and maximum train speed. The model result is consistent with the results of crash prediction trees for crossings controlled by crossbucks only and stop signs.

The number of vehicles per day plays the most important role in the stop-sign effect. The effectiveness of stop sign is increasing as the vehicle traffic volume increases. For the lower traffic volume, stop-sign treatment is more effective at crossings with multi tracks than single tracks; and for the multiple-track crossings stop signs has no safety effect if maximum train speeds at the crossings are higher than 55 mph. This analysis is consistent with the finding that for the paved crossings with moderate vehicle traffic volume, stop signs are less effective if the train speed is higher than 40 mph. Lerner et al. (2002) reported that driver decision errors play an important role in safety at passive highway-rail crossings, and misjudging the speed of approaching trains would be more likely to occur as train speed increases. Stop signs could not essentially resolve this perpetual problem although they would help drivers slow down significantly, allowing them to come to a stop safely if necessary. Therefore, for the high train speed crossings, updating the level of grade-crossing control (i.e., flashing lights, automatic gates, and grade separation) may be more appropriate for reducing train-vehicle crashes.
A very interesting finding in this study is that for crossings with 400 or higher vehicles per day, the combination application of stop signs and railroad warning signs can achieve a great effect on crash reduction, especially for the crossings that have paved highways. Noyce and Fambro (1998) found that the railroad warning signs can increase driver awareness of passive crossings existence, cause drivers to approach passive crossings with additional caution, and reduce average vehicle speeds at passive grade crossings.

Another interesting finding is that stop signs have different effects for different crossing angles, depending on traffic volume, the number of tracks, with or without railroad warning signs. For the crossings with lower traffic volume and single track, 36-59 degrees of crossing angle have a better effect of stop-sign treatment than 0-35 and 60-90 degrees of crossing angles. However, for the crossings with higher traffic volume and without railroad warning signs, stop signs have a significant safety effect on very sharp crossing angles less than 35 degree but no effect on 36-90 degrees of crossing angles. Wigglesworth (2001) pointed out that at acute-angled crossings, it might be difficult for vehicle drivers to detect an oncoming train when it is approaching from one of the rear quadrants with the risk of “over-the-shoulder” collision. Especially, severe skewed-angle crossings can pose a challenge for older drivers to search as they near the track, who have limited head-and-neck flexibility (Lerner et al., 2002). Nevertheless, if crossing angles are close to 90 degree, the drivers’ hazard alertness level may be reduced so that the conflict risk between vehicle and train may also increase.

5.3 HTBR Vs. Count Regression Models

As illustrated above, HTBR helps identify homogeneous groups with high or low crash frequency and construct rules for predicting crash frequency, and it is also flexible in estimating effectiveness of safety countermeasures based on crossing attributes. It is straightforward for traffic engineers and decision makers to use the HTBR models. With given grade crossings’ control types (crossbucks or stop signs) and attributes related to those significant factors, predictions can be obtained based on the ‘if-then’ rules found in the HTBR tree structure. The prediction rules can be easily incorporated in a crash frequency prediction and stop-sign
treatment evaluation system. This process can assist traffic engineers and decision makers in identifying the best strategies to mitigate crash hazards in the highway-rail crossing design/planning stage.

However, by no means does it imply that HTBR is always a better tool for crash prediction and analysis than those parametric count regression models, such as Poisson and negative binomial (NB) models and their variants (ZIP and ZINB), which have been successfully applied for crash prediction and analysis. In a comparison of the results between HTBR and NB models for predicting crash frequency, Chang and Chen (2005) pointed out that it is difficult to determine which modeling approach is better than the other and HTBR is a good alternative. Using the FRA crossing database, Hal et al. (2009) have employed NB models to predict and analyze train-vehicle crashes at the passive grade crossings. In their approach, an NB model was developed for paved and unpaved highways separately. Comparing the NB regression results to the HTBR results in this study, the number of vehicles per day, the number of trains per day, the number of tracks, the maximum train speed, and percentage of trucks were identified as significant crash predictors in both models; the number of traffic lanes and develop type are significant in the NB models but not in the HTBR models; and railroad advance warning signs, crossing angle, and distance between highway intersections and crossings are important predictors in the HTBR models but not in the NB models. In fact, parametric count regression and HTBR are two quite different modeling processes. The general advantages of trees over count models (generalized linear models) include efficient handling of complex nonlinearity, interactions, as well as categorical predictors, and robustness to statistical assumptions and outliers. On the other hand, the count model is capable to catch linear patterns. Generally, HTBR models are especially valuable for explore structure, or relationships among variables but lack statistical inferences for evaluating the effect of predictors.

With regard to HTBR and count regression applications in crash frequency prediction, the following research topics that might be of great interest for future studies. Firstly, Washington (2000) suggested an approach, termed as iteratively specified tree-based regression (ISTBR), which is aimed to borrow the strength of both HTBR and linear regression by iteratively growing tree structures on the residuals obtained from a linear model. Secondly, Lee and Jin (2006)
proposed a tree procedure particularly designed for zero-inflated count data. Thirdly, other extensions such as multivariate adaptive regression splines (MARS; Friedman, 1990) can help improve the prediction accuracy of recursive partitioning considerably without loss of interpretability. In the future, these attractive approaches to count data are recommended for application to model crash frequency.

5.4 Limitation of the Study

Although FRA keeps the most complete data of highway-rail crossings and train-vehicle accidents across the entire United States, some important variables that may affect passive crossing safety are unavailable in its Grade-Crossing Inventory database, such as sight distance. The amount and type of sight distance provided at highway-rail grade crossings can have significant impacts on safety (Knapp, 1999; Oh et al., 2006). Lerner et al. (2002) reported that limited sight distance is the primary reason for using stop signs at passive crossings. This is based on the assumption that if a driver’s corner sight triangle is obstructed by any objects or constructions at the crossing, a stop sign should be used so that the driver can recognize that there is a need to stop. Since the available sight distances at crossings is not reported by FRA, this study cannot reflect the limited sight distance’s effect on crossing safety prediction and stop-sign treatment, which should be considered in the further studies when appropriate crash databases are available.

In this study, only passive crossings that were updated from crossbucks-only to stop sign control without involvement in other countermeasures are used for analysis in order to reduce the confounding effect of other factors. However, there could be some unreported crossing countermeasures when or after the stop-sign treatment was applied at crossings, such as clearing sight obstructions, resurfacing highway pavement, and repainting marking and signage. Omitting these countermeasures may lead to overestimating the efficacy of stop-sign use to some extent. In future studies, further efforts should be made to confirm whether a confounding effect exists due to other unknown factors.

Additionally, a relative large “slice” of time, i.e, 27 years was taken in this cross-sectional study.
However, considering the whole picture of at-grade crossing safety in U.S., crossings are getting safer and safer over years due to improvement in vehicle design, drivers’ safety awareness, and other factors. In a prior study by Millegan et al. (2009), the effect of year based on annual accident rates for the crossings that are controlled by crossbucks-only or stop sign was a heuristically explored. The annual accident rate was calculated as the total number of accidents occurring divided by the number (in thousands) of crossings in a particular year. It was found that between 1980 and 2005 the yearly accident rates for the stop-sign control are constantly lower than those for the crossbuck-only control, and there is no apparent overall increasing or decreasing pattern over time in either case. The reason that significant time series patterns is not identified in this dataset may be accounted for by the fact that dangerous crossbuck-only-controlled crossings are more likely to be chosen by experienced engineers for application of stop-sign treatment, and therefore, the hazard factors existing in these crossings may substantially overweight the general safety trend during the 27 years. Thus, in the current study, all observations have been treated as independent by heuristically ignoring the effect of time. The results can be certainly refined in future studies by applying more complex modeling tools such as random- or mixed-effect models that allow incorporation of the time factor.


