Incorporating Truck Crash Modeling into a Methodology for Evaluating the Relative Need for Truck Route Improvements
ABSTRACT

In July 1997 a study entitled “Freight Movement and Intermodal Access in Kentucky” was initiated for a two-year period by the Kentucky Transportation Center (KYTC) with the University of Kentucky’s (UK) Department of Civil Engineering. A methodology for evaluation of highway access for trucks along specific truck routes was developed (Aultman-Hall et al. 1999). In all, 81 routes used for access between 46 facility sites and the National Highway System were evaluated. These routes represented approximately 800 miles of highway.

The methodology involves tabulation of problem truck miles and problem truck points for the following point and continuous features of the route: lane width, shoulders, railway crossings, grade, safe speed on horizontal curves, offtracking on horizontal curves, intersection turning radii, and stopping sight distance. Using criteria developed in brainstorming sessions and from reference sources, each of the above features were graded as “preferred”, “adequate” or “less than adequate” for truck access. For example, a 12 foot lane width was considered preferred for truck access while 10 foot lanes were considered less than adequate. In order to obtain the weighted sum of problem truck points along a route with respect to a particular feature, the number of less than “preferred” points were weighted by the number of trucks per day that passed that point. The “less than adequate” points were weighted twice that of “adequate” points. For continuous features, such as lane width, the problem truck miles were obtained by weighting the summation by the length of the individual sections with less than “preferred” features. Sections or points that are graded “preferred” with respect to a particular feature do not contribute to the sum of problem truck points or miles for that particular feature. In this way, the problem route features are weighted by the number of trucks as well as the section length as a measure of relative urgency.

Through use of the problem trucks per day and problem truck miles per day, specific sections or routes can be compared on a feature by feature basis to determine the urgency of needed improvements. However, there is also a need for a measure of the overall route quality in which all features are combined into one weighted route measure. Some consideration was given during the original project to weighting each of the features such that a composite quantitative measure could be determined. However, in the end no composite measure was developed. The objective of this research project is to use the truck crash histories along the routes where features have already been documented to develop a truck route safety model based on the problem truck points and miles for each feature according to those that correlate to actual crash problems. The composite measure would be used as a relative urgency measure when different truck routes are being considered for improvements.

Using linear, negative binomial and logistic regression this research analysis did not find a useful relationship between preferred, adequate and less than adequate roadway features and the truck crash history. The results further indicated that the adequate features had no impact on truck crash rates compared to the preferred features. Furthermore some less than adequate geometric features were found to have a positive effect on truck crash rates. The only significant and consistent predictor of crash rates was ADT. This would suggest that the relative urgency measure for deciding which truck routes to improve should be traffic volume.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures and Tables</td>
<td>ii</td>
</tr>
<tr>
<td>1.0 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2.0 Literature Review</td>
<td>2</td>
</tr>
<tr>
<td>2.1 The Importance of Truck Safety Research</td>
<td>3</td>
</tr>
<tr>
<td>2.2 How Geometric Roadway Features Affect Truck Safety</td>
<td>5</td>
</tr>
<tr>
<td>2.3 Quantifying Truck Safety Using Statistical Models</td>
<td>8</td>
</tr>
<tr>
<td>2.3.1 Linear Regression</td>
<td>8</td>
</tr>
<tr>
<td>2.3.2 Poisson Regression</td>
<td>10</td>
</tr>
<tr>
<td>2.3.3 Negative Binomial Regression</td>
<td>13</td>
</tr>
<tr>
<td>2.3.4 Logistic Regression</td>
<td>14</td>
</tr>
<tr>
<td>2.3.5 Loglinear Modeling</td>
<td>16</td>
</tr>
<tr>
<td>2.3.6 Artificial Intelligence Techniques</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Accounting for Exposure</td>
<td>18</td>
</tr>
<tr>
<td>2.5 Summary of Background Information</td>
<td>20</td>
</tr>
<tr>
<td>3.0 Database Description</td>
<td>21</td>
</tr>
<tr>
<td>3.1 Facility Sites and Truck Routes</td>
<td>22</td>
</tr>
<tr>
<td>3.2 Problem Truck Point and Problem Truck Mile Methodology</td>
<td>24</td>
</tr>
<tr>
<td>3.3 Database Construction</td>
<td>26</td>
</tr>
<tr>
<td>3.3.1 Crash Data</td>
<td>27</td>
</tr>
<tr>
<td>3.3.2 Data for Problem Truck Points</td>
<td>28</td>
</tr>
<tr>
<td>3.3.3 Data for Problem Truck Miles</td>
<td>30</td>
</tr>
<tr>
<td>4.0 Analysis and Results</td>
<td>30</td>
</tr>
<tr>
<td>4.1 Descriptive Data Information</td>
<td>31</td>
</tr>
<tr>
<td>4.2 Data Correlation</td>
<td>33</td>
</tr>
<tr>
<td>4.3 Regression Results</td>
<td>34</td>
</tr>
<tr>
<td>4.3.1 Ordinal Logistic Regression</td>
<td>35</td>
</tr>
<tr>
<td>4.3.2 Linear Regression</td>
<td>37</td>
</tr>
<tr>
<td>4.3.3 Negative Binomial Regression</td>
<td>39</td>
</tr>
<tr>
<td>4.4 A Closer Look at the Result Regarding Railroad Crossings</td>
<td>41</td>
</tr>
<tr>
<td>4.5 Summary of Results</td>
<td>43</td>
</tr>
<tr>
<td>5.0 Conclusions and Recommendations</td>
<td>44</td>
</tr>
<tr>
<td>6.0 References</td>
<td>46</td>
</tr>
<tr>
<td>Appendix: A Description of Geometric Features Included in This Study</td>
<td>49</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES AND TABLES

Figure 1. Kentucky Truck-Generating Sites and Clusters of Sites Selected for Study.  23

Figure 2. Histogram of Truck Crashes per Route Mile. ................................. 32

Table 1. Summary of geometric roadway characteristics found to affect truck safety.  7
Table 2. Descriptive statistics. ................................................................. 31-32
Table 3. Correlation table. ................................................................. 34
Table 4. Frequency distribution for truck crash per truck route length. .............. 35
Table 5. Results of ordinal logistic regression. ........................................... 36
Table 6. Summary of significant geometric predictors – ordinal logistic regression. 36
Table 7. Results of linear regression. .......................................................... 38
Table 8. Summary of significant geometric predictors – linear regression. ......... 38
Table 9. Results of negative binomial regressions. ........................................ 40
Table 10. Summary of significant geometric predictors
              – negative binomial regression. ..... 40
Table 11. Truck routes exhibiting “adequate” railroad crossings. ....................... 42
Table 12. Statistically significant predictors of truck crashes per mile of route length. 43
Incorporating Truck Crash Modeling into a Methodology for Evaluating the Relative Need for Truck Route Improvements

By: Stewart Robertson and Lisa Aultman-Hall

University of Kentucky
Department of Civil Engineering
Lexington, Kentucky

June 2001
1.0 INTRODUCTION

Between 1997 and 1999, the Kentucky Transportation Center (KTC) conducted a study entitled “Freight Movement and Intermodal Access in Kentucky” (SPR 98-189), in which researchers developed a method to evaluate the access for large trucks between truck-traffic-generating sites and the National Highway System (NHS). The routes evaluated were non-NHS routes that were likely to contain deficiencies in geometric design assumed detrimental to truck traffic. Researchers developed the problem truck point and problem truck mile methodology to convert the geometric design features to a relative urgency rating for design deficiencies and needed improvements for truck traffic. Each geometric roadway feature included in the study was graded as “preferred”, “adequate” or “less than adequate” for truck access and was then summed and weighted based on truck volume. The continuous geometric roadway features were also weighted based on deficient section length. Aultman-Hall et.al. (1999) described the methodology that was used to evaluate the non-NHS truck routes for the KTC freight study as well as provided a description of all the geometric roadway features evaluated in the study. A summary of this methodology is provided in Section 3 of this paper. A description of the geometric roadway features evaluated in the study as well as grading criteria for “preferred”, “adequate” and “less than adequate” locations is provided in the appendix of this paper. The geometric feature data and truck crash histories collected for the KTC freight study form the basis for this research.

The objective of this study is twofold. The first objective is to further develop the problem truck point and problem truck mile methodology by quantitatively evaluating the truck crash prediction ability of the two levels of severity (“adequate” versus “less than adequate”) within a problem truck point or problem truck mile roadway feature. In the original KTC research, “less than adequate” locations or sections were weighted twice that of “adequate” locations or sections in order to allow for an aggregate summation of geometric feature deficiency along a whole truck route. The objective here is to determine the relative influence of “adequate” locations or sections versus “less than adequate” locations or sections based on their contributing influence on truck crash rates. The second objective is to evaluate the relative influence of different geometric roadway features on truck crashes to get an aggregate measure of truck route safety for non-NHS
routes in Kentucky based on the truck crash histories, problem truck points, and problem truck miles documented in the KTC freight study. The objectives were accomplished by first evaluating different statistical analysis techniques in order to choose one or more techniques suitable for modeling truck crash occurrence. Then, various types of regression models were created using individual problem truck point and problem truck mile values as explanatory (independent) variables. The dependent variable was truck crash frequency or truck crash rate.

The next section of this report describes the important insights gained from a review of past studies that are relevant to this research. Following this, a description of the methodology and research approach is provided in Section 3. Section 4 describes the regression models that were developed and discusses the modeling results. A concluding discussion is provided in Section 5 in which the major findings and implications of this research are summarized and ways in which truck safety could be further analyzed in the future are presented.

2.0 LITERATURE REVIEW

A literature review was conducted with the purpose of providing an overall perspective on the major findings of past research work related to truck crashes and the impact of geometric roadway features on the occurrence and severity of truck crashes. The literature review also provided information concerning the methodology of past truck safety research and the viable types of statistical analysis. The literature review was conducted through facilities of the Kentucky Transportation Center library at the University of Kentucky, the Transportation Research Information Services (TRIS) online search, the Insurance Institute for Highway Safety, and CD-ROMS containing papers presented at Transportation Research Board meetings. A review of background literature reveals that significant research has been performed on various facets of crashes involving trucks. Common research objectives include the comparison of the safety of trucks to other vehicles, comparison of truck crashes on different road types, comparison of crashes of different truck configurations, and comparison of truck crashes in different weather conditions. However, a comparatively small amount of research has been devoted to creating safety models for truck crashes using geometric features of roadways.
These studies have often produced conflicting results regarding which geometric roadway features are significant in predicting crashes and which types of statistical analysis are feasible for modeling truck crashes.

This section begins by illustrating the importance of truck safety research, especially on two-way, two-lane roads, the main focus of this research. Next, a discussion of previous truck safety research efforts and the geometric roadway factors that were found to affect truck safety in these studies is presented. The remaining subsections present information on how past research studies have quantified truck safety, including incorporation of exposure, using various regression techniques and statistical models.

2.1 THE IMPORTANCE OF TRUCK SAFETY RESEARCH

Large trucks accounted for 4% of registered vehicles and 8% of miles driven in the United States in 1999 (Federal Highway Administration, 2000). However, large trucks were involved in 12% of all passenger vehicle occupant deaths and 23% of passenger vehicle occupant deaths in multiple-vehicle crashes in 1999 (Insurance Institute for Highway Safety, 2001). Almost one out of every four passenger vehicle occupant deaths in multiple-vehicle crashes occur in crashes with large trucks, yet trucks account for only 8% of miles driven. Data from Kentucky suggest a similarly disturbing trend in large truck safety. Agent and Pigman (1997) indicate that in Kentucky, trucks are involved in 7% of all accidents, but involved in approximately 13% of fatal accidents. It must be noted that in the National Highway Traffic Safety Administration (NHTSA) report entitled “Traffic Safety Facts 1996: Large Trucks”, police reported one or more errors or other factors related to the driver’s behavior for the other vehicle driver and none for the truck driver in 71% of two-vehicle fatal crashes involving a large truck and another vehicle. Nevertheless, based on the number of trucks on the road and the amount they travel, large trucks account for more than their share of highway deaths. Part of this undoubtedly stems from the size of trucks relative to passenger cars. However, the size of trucks is what allows them to play the role that they play in the nation’s economy – freight transportation. While it is infeasible to ban trucks from our nation’s highways because of their size, it is important to research truck crashes in order to better understand
the factors that affect truck safety. The results of these research efforts can lead to appropriate countermeasures being taken to reduce the trend of trucks being unsafe for the roads on which they travel.

Many factors are involved in truck crash rates. McGee (1986) produced a list of factors hypothesized to be important in truck safety. This list was produced by examining past literature, interviewing a panel of researchers experienced in truck accident studies, interviewing representatives of various operating offices of the Federal Highway Administration (FHWA), and examining accident data. All geometric roadway elements together was just one of the 28 factors listed as being important to truck safety. Included in geometric elements were curves, vertical gradient, passing zones, interchanges, intersections, work zones, lane width, and shoulder width. All of the 28 factors listed were believed to affect truck accident rates, although to varying degrees of significance. It was theorized that an enormous sample of accidents and exposure would be needed to obtain a relationship including all of the factors. McGee narrowed down the list of 28 by producing a list of key variables that influence truck safety. This modified list did not include geometric elements. The author proposed that the key variables that should dictate the experimental design and sampling requirements for truck safety are truck type, truck length, truck trailer type, truck weight, driver type, driver age, and highway type. While it has been documented that non-roadway elements do contribute to truck crashes, the focus of the research described in this paper is determining what geometric roadway deficiencies contribute to truck crashes on non-NHS truck routes in Kentucky in order to improve a model intended to prioritize roads for upgrading.

Most of the non-NHS routes studied are older two-way, two-lane roads (TWTL). Research by Donaldson (1986) points out that older TWTL roads are often riddled with substandard geometric design for the accommodation of large trucks. The interaction of narrow lanes, deficient superelevation, unspiraled and severe horizontal curves, and severe grades on TWTL roads acts to compromise the safety of the roadway users. Whatever shortcomings these designs have for passenger cars, their adverse effects on large trucks are speculated to be greater.

Truck safety is a serious issue. All the factors that contribute to truck crashes are worthy of research to develop models that can accurately describe the occurrence of truck
crashes. This paper describes research that focuses on determining which geometric roadway features contribute to truck crashes, and to what relative degree different levels of severity within a certain geometric roadway feature contributes to truck crashes. With better understanding of factors that lead to truck crashes, a more effective method to prioritize roadway improvements can be developed to reduce the occurrence of truck crashes.

2.2 HOW GEOMETRIC ROADWAY FEATURES AFFECT TRUCK SAFETY

Many geometric roadway design features can potentially be the cause of a truck crash. However, some geometric features are repeatedly shown to predict truck crashes to a higher degree than others. This subsection examines what is known about how geometric features affect truck safety.

Agent and Pigman (1991) conducted an evaluation of highway geometrics related to large trucks. The objective of the study was to determine the extent of highway safety and geometric problems associated with trucks using Kentucky’s highways. A second objective was to identify criteria that could be used in identifying roadway sections that cannot safely accommodate large trucks. As part of the study, an extensive literature review was conducted to summarize information that could be used as a guide for determining truck safety criteria. Using information gained in the literature review, a list of factors that impair truck safety was produced. The list included all geometric features used in the research described in this paper except grade. Agent and Pigman incorporated roadway profile information by including length of vertical curves in their study. Using these roadway criteria, Agent and Pigman identified numerous locations in Kentucky where truck traffic was hindered by geometric design deficiencies.

In a later study, Pigman and Agent (1999) conducted a detailed analysis of truck accidents in Kentucky and recommended countermeasures to reduce the number and severity of truck accidents. The data was obtained from police reports for fatal accidents in which a truck was involved for the period 1994 through 1997. Each accident was reviewed and classified by type of accident and causative factor. Based on this, the authors produced a list of potential countermeasures to address problems associated with heavy truck accidents. Among the roadway countermeasures recommended were
widening pavement in curves, installing advance warning at traffic signals on high-speed roadways, installing related signing at steep grades, constructing truck climbing lanes at locations with steep grades and high truck volumes, and placing active warning devices in advance of curves where accidents have occurred involving overturning trucks. These recommendations indicate the truck safety implications of lane width, truck stopping sight distance, grades, and safe truck speed on curves – all factors that will be used in this study to attempt to create truck safety models using regression. These recommendations also show the importance of roadway geometrics in truck safety.

Garber and Joshua (1991) developed truck safety models with geometric roadway features for three different highway types. For undivided highways, the exact geometric characteristics that were statistically significant depended on the type of regression technique used to create the model. Lane width, slope change rate, shoulder width, and horizontal curvature were all found to have varying degrees of significance, depending on the regression model. Vogt and Bared (1998) obtained a similar result in their study of modeling accidents at intersections. Lane width, shoulder width, horizontal curvature, and vertical gradient all could be used effectively to model accidents at intersections on two-lane rural roadways. Wright and Burnham (1985) predicted the accident rates of trucks (accidents/truck-mile/year) by using four factors – percent of total roadway mileage with two lanes, percent of two-lane mileage with substandard vertical curves, percent of two-lane mileage with substandard horizontal curves, and percent of two-lane mileage with substandard pavement width. Mohamedshah et.al. (1993) found that the truck accident involvement rate (accidents/kilometer/year) of two-lane rural roadways depended on shoulder width and horizontal curves with a degree of curvature greater than 6°, as well as traffic characteristics such as nontruck average annual daily traffic (AADT) and average daily truck traffic. Vertical gradient was not found to be statistically significant in contributing to the truck accident model, although the authors theorized that inadequate data might have led to this result. Two horizontal curvature variables were found to not be statistically significant – percentage of roadway section with horizontal curves with a degree of curvature between 1° and 3° and percentage of roadway section with horizontal curves with a degree of curvature between 3° and 6°. Therefore, this
study illustrates the importance of having separate explanatory variables representing
different levels of severity within a particular geometric roadway feature.

Some researchers have obtained entirely different results in modeling truck accidents with geometric roadway features. Miaou et.al. (1993) found no geometric characteristics to be statistically significant in predicting truck accidents. For rural two-lane undivided arterials, the best model included the variables AADT per lane, percent trucks, horizontal curvature, length of curve, and shoulder width. However, all models selected failed to pass the chi-square test at a 5% significance level. Saccomanno and Buyco (1988) also found no significant relationships between roadway variables and truck accident involvement rates.

From the results of previous studies, it is clear that the manner in which geometric roadway characteristics affect truck safety is still not completely understood. Gaining a better understanding of these relationships would allow safety researchers to understand which countermeasures will lead to notable reductions in truck crashes. Table 1 summarizes the geometric characteristics that were noted in the literature review as significantly affecting truck safety. Horizontal curvature and lane width appear to have a less disputed impact on truck safety than other geometric variables.

Table 1. Summary of geometric roadway characteristics found to affect truck safety.

<table>
<thead>
<tr>
<th>Geometric Roadway Characteristics</th>
<th>Agent and Pigman</th>
<th>Pigman and Agent</th>
<th>Garber and Joshua</th>
<th>Vogt and Bared</th>
<th>Wright and Burnham</th>
<th>Mohamadshah et.al.</th>
<th>Miaou et.al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal curvature</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Sight distance</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intersection turning radii</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Railway crossings</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bridges</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane width</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulders</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope change rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical curvature</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Characteristics listed in *italics* are available for the research described in this paper.
2.3 QUANTIFYING TRUCK SAFETY USING STATISTICAL MODELS

Studies of relationships between truck crashes and geometric roadway variables have employed different statistical models. Most of the statistical models were developed using conventional linear regression, Poisson regression, negative binomial regression, logistic regression, or loglinear models. Some investigation has also been made into the applicability of artificial intelligence techniques for predicting truck crashes. This subsection addresses the complications involved in quantifying the relative safety of various geometric roadway features using statistical models.

2.3.1 LINEAR REGRESSION

Multiple linear regression has been used more than other regression techniques in the past to predict truck crashes. Garber and Joshua (1991) used a linear regression model to investigate the major factors associated with large truck accidents and the role of traffic characteristics, geometric variables, and roadway type. In linear regression, the expected number of truck accident involvements, \( E(y_i) \), is related to \( n \) traffic and geometric variables, \( x_{ij} \), in the following form:

\[
E(y_i) = f(x, \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n
\]

where: \( x_j \) are geometric variables
\( \beta_j \) are model parameters to be estimated.

The geometric factors studied included intersections, grades, curves, stopping sight distance, roadside hazards, speed differentials, shoulders, and slope change rate. Accident data was obtained from Virginia police accident report forms between 1980 and 1985. The authors chose the best model by examining the Akaike’s information criterion (AIC). Model selection using AIC is based on the concept of entropy maximization. The AIC is given by:

\[
AIC(k) = -2 \log L(\Theta_k) + 2k
\]

where: \( \log L(\Theta_k) = \log_e \) [maximum likelihood]
\( k = \) number of free parameters from the model.

Smaller values of AIC indicate better models. According to Garber and Joshua, multiple linear regression models did not seem to describe adequately the relationship between large-truck accident involvement rates (accidents/truck-miles) and associated traffic and
geometric variables, based on the AIC value and the \( R^2 \) goodness of fit parameter. Linear regression models were developed for the annual truck accident involvement rate as a linear function of highway- and traffic-related variables. One model was developed for each of three roadway types. For undivided 4-lane and 2-lane highways, the geometric characteristics that were included in the best model include lane width, shoulder width, slope change rate, and horizontal curvature. Average daily traffic was also an explanatory variable in the best model as was percentage of trucks, although the coefficient for percentage of trucks had a negative sign. This indicates that as the percentage of trucks increases, the truck accident rate decreases. The AIC value for the chosen model was 288.46. The \( R^2 \) value was 0.6832. Models were also developed for both divided 4-lane highways and interstate highways. The AIC values for these two models were 245.21 and 866.39 respectively while the \( R^2 \) values for these two models were 0.2187 and 0.2317 respectively. The Garber and Joshua study indicates that linear regression modeling may be more appropriate for undivided highways than divided highways and interstates. This is of interest for the research described in this paper because most of the roadways in the KTC freight study are undivided, non-NHS roadways.

Many of the truck routes included in this study are two-lane rural roads. Mohamedshah et.al. (1993) developed truck accident models for two-lane rural roads as well as interstate highways. The variables considered for model development were nontruck average annual daily traffic (AADT) per lane, truck average daily traffic per lane, shoulder width, horizontal curvature, and grade as the independent variables. Truck involvement rate per kilometer per year was used as the dependent variable. Using a linear regression model for two-lane rural roads, the truck involvement rate per kilometer per year was best predicted using the variables AADT, truck ADT, shoulder width, and horizontal curves with a degree of curvature greater than 6°. The \( R^2 \) value equaled 0.415; therefore, much of the variation in the predicted truck accident rates for this model remains unexplained. All the variables were statistically significant at a significance level of \( \alpha = 0.05 \) (p=0.000). The linear model had a higher \( R^2 \) value than other nonlinear models.
Saccomanno and Buyco (1988) considered several methods for calibrating statistical models of truck accident rates by examining previous research. The authors concluded that the use of multiple linear regression for predicting the causes of truck accidents may be flawed due to the relationship not reflecting linear behavior. Also, multiple linear regression models cannot account for the non-negative nature of accident occurrence. The authors pointed out that many linear regression models for predicting accidents have negative constants (y-intercept values), indicating that for a roadway of favorable traffic and geometric characteristics, a negative accident rate could be obtained from the model. However, it is impossible to observe a negative accident rate.

Wright and Burnham (1985) used multiple linear regression to study the effect of geometric roadway features on truck accident and injury rates for Georgia data from 1981. The research showed that the accident rates for trucks could be predicted by four factors – percent of total mileage with two lanes, percent of two-lane mileage with substandard horizontal curves, percent of two-lane mileage with substandard vertical curves, and percent of two-lane mileage with substandard pavement width. With an $R^2$ value of 0.36, much of the variation in the predicted truck accident rates for this model remains unexplained.

Mixed results have been obtained using multiple linear regression to model truck crash data using geometric roadway features. The $R^2$ value for the models cited here have ranged from 0.36 to 0.68. Negative constants (y-intercept values) can be seen as a drawback to linear models since they are an impossibility for actual crash rates. However, some evidence exists that linear regression is more appropriate than other regression techniques for modeling truck crashes on undivided roadways, the predominant road type in this study.

2.3.2 POISSON REGRESSION

The occurrence of vehicle crashes can be described by the Poisson process. The basic assumption of Poisson processes is that the number of times an event occurs within an observed time interval is independent and the expected values depend on independent variables and estimated parameters. Therefore, Poisson regression can be used to model
vehicle crashes with independent variables including geometric roadway features and traffic characteristics.

Garber and Joshua (1991) found that Poisson regression models did describe adequately the various relationships between truck accidents and geometric roadway variables. The Poisson model assumed that the total number of truck accidents on a road section i during a 1-year period, \( y_i \), follows a Poisson distribution. Furthermore, the expected number of truck accident involvements, \( E(y_i) \), is related to \( n \) traffic and geometric variables, \( x_{ij} \), in the following form:

\[
E(y_i) = f(x, \beta) = \beta_0 \cdot x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_n^{\beta_n}
\]

where: \( x_j \) are geometric variables

\( \beta_j \) are model parameters to be estimated.

Unlike the linear regression models, this model suggests that the variance of \( y_i \) involves explanatory variables \( x \) raised to powers of unknown parameters \( \beta \). The authors chose the best model by examining the AIC value. For undivided highways, the Poisson model that best describes the expected number of truck accident involvements included slope change rate, average daily traffic, and truck percentage as explanatory variables (AIC = 62.06). Geometric characteristics such as lane width, shoulder width, and horizontal curvature were not included in the best Poisson regression model. The undivided highway model was a better fit for the data than either the divided highway model or the interstate highway model (AIC values of 83.56 and 407.48 respectively). That the Garber and Joshua undivided highway Poisson model performed better than either of the divided highway Poisson models is a similar result as that obtained in the same study using linear regression, indicating that crash patterns on undivided roadways may be easier to describe using regression techniques and geometric roadway characteristics than crash patterns on divided roadways.

Miaou et.al. (1993) performed a study to establish empirical relationships between truck accidents and highway geometric design. Data from the Highway Safety Information System (HSIS\(^1\)) were used to build and evaluate models for different highway types. A Poisson regression model was developed for modeling the expected

---

\(^1\) The HSIS is a multi-state highway safety database, developed by the University of North Carolina for the Federal Highway Administration (FHWA), containing crash, roadway inventory, and traffic volume data.
number of truck accidents. For the Poisson model, the number of trucks involved in accidents on each road section was assumed to be Poisson distributed. Models were developed using the explanatory variables horizontal curvature, vertical grade, and shoulder width. The final models selected all failed to pass the chi-square test at a 5% significance level. The researchers hypothesized that overdispersion problems (extra variation), caused by uncertainties in truck exposure data and omitted explanatory variables in the models, were significant in the Poisson regression models failing the chi-square test.

Awad and Janson (1998) used Poisson regression to attempt to improve the fit of a traditional linear regression model for accident data. The authors attempted to explain truck accidents at interchanges in Washington State with various geometric and traffic characteristics. The dependent variable in their study was number of truck accidents or truck accident frequency (TAF). The authors used a square root transformation of TAF to approximate Poisson regression. The $R^2$ value for the transformed model was 0.111. For comparison, the authors constructed a linear regression model using the same independent variables. The $R^2$ was 0.101 for the linear regression model. Many regression models with different combinations of independent variables were tested, without showing any significant changes in the prediction power. The authors concluded that the relationship between TAF and the explanatory variables is complex and cannot be explained by a regression model.

Poisson regression has had mixed results in predicting accidents in the past. Garber and Joshua (1991) indicate a preference for Poisson regression over linear regression models for safety analysis because of better values for goodness-of-fit statistics, the assumption that the occurrence of highway accidents can be described by the Poisson process, and the fact that some variance can be accounted for in the parameter values. However, some studies have found Poisson regression models to not be statistically significant. Poisson regression is more complex than linear regression and the performance of Poisson regression models has not been consistently shown to be significantly better than linear regression models.
2.3.3 NEGATIVE BINOMIAL REGRESSION

One problem that plagues Poisson regression models is overdispersion (Miaou et.al., 1993). Negative binomial regression was developed to be an improvement on the Poisson regression process. The negative binomial regression model allows for overdispersion in the model and can be used to quantify the effect of uncertainty on truck exposure data as well as the effect of the omitted variables on the overdispersion of the Poisson models.

Miaou et.al. (1993) found that the regression parameters for the negative binomial regression model were similar to those obtained from the Poisson regression. The expected number of truck accident involvements, \( E(y_i) \), is of the following form:

\[
E(y_i) = f(x, \beta) = \exp[\beta_0 + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_nx_n]
\]

where: \( x_j \) are geometric variables
\( \beta_j \) are model parameters to be estimated.

The results for this model reaffirmed that the relationships between truck accidents and the examined highway geometric design variables for rural two-lane undivided arterials (developed in the same study using Poisson regression, as previously discussed) were valid even when the overdispersion exists in the model. Based on the AIC values, the negative binomial models for three different classes of roadway were consistently better than the corresponding Poisson models. For rural two-lane undivided arterials, AIC = 5111.6 for Poisson regression while AIC = 4819.7 for negative binomial regression. The authors were encouraged with the preliminary relationships between truck accidents and highway geometric variables and thought that the predictive power of the models could be enhanced by including detailed truck exposure data such as time of day, truck type, and weather conditions. This study recommended broadening the database to include driver, weather, and socioeconomic factors to contribute to the models.

Vogt and Bared (1998) developed negative binomial regression models for modeling accidents at intersections on rural two-lane highways. Accident data in this study were obtained from HSIS files for the states of Minnesota and Washington. Geometric data were obtained from as-built construction plans. The primary variables used in modeling accidents on highway segments included ADT, segment length, lane width, shoulder width, degree of horizontal curvature, grade, posted speed limit, and
commercial vehicle percentage of traffic. Exposure, in million vehicle kilometers, was calculated by the formula:

\[
\text{EXPOSURE} = \text{ADT} \times 365 \times Y \times L_m \times 10^{-6}
\]

where: \( Y \) is the period of study in years

\( L_m \) is the segment length in kilometers.

Vogt and Bared make use of an extended negative binomial regression model, proposed by Miaou and Lum (1993). In this model, roadway segments are divided into subsections and mean number of accidents are estimated by subsection. Then, a weighted sum of these means yields the expected mean for the entire section. Instead of a model of the form \( y_i = (\exp \beta_0)(\exp x_1)^{\beta_1} \ldots (\exp x_n)^{\beta_n} \), each of the expressions \( (\exp x_j)^{\beta_j} \) is replaced by \( \Sigma w_i (\exp x_j)^{\beta_j} \). This extended negative binomial allows a model to address local conditions more precisely than the ordinary negative binomial model. This produces a solution for the problem of roadway segments failing to be homogeneous with respect to some variables – for example, tangents, crest curves, and sag curves all occur on some roadway sections, but not necessarily on the entire length of the roadway section. Rather than averaging variables over different subsegments, the extended negative binomial is an alternative to make a more localized study that acknowledges variability within a roadway segment. The researchers found that the models offer reasonable representations of the effects of highway variables on accidents. Again, it was found that notable limitations to the models were omitted variables, in this case – sight distance and seasonal variation in traffic and weather.

Negative binomial regression appears to have inherent advantages over Poisson regression in describing variation in the data and allowing for overdispersion in the model. However, the conclusions reached by Miaou et.al. (1993) indicate that the examined traffic and highway geometric design variables were valid even when the overdispersion existed in the model.

2.3.4 LOGISTIC REGRESSION

Multiple logistic regression is a commonly used generalized linear model based on the binomial distribution. In logistic regression analysis, the logit is linearly related to the independent variables. Logistic regression obtains maximum likelihood estimates of
the model parameters using an iterative-reweighted least squares algorithm. An
important difference between logistic regression and linear regression is that logistic
regression techniques attempt to predict categorical response variables whereas linear
regression techniques attempt to predict continuous response variables. Logistic
regression can be used to build truck safety models when the dependent variable is
categorical. Categorical variables include binary variables, such as a crash being either
fatal or nonfatal, and ordinal variables such as 0-3 crashes per mile, 3-6 crashes per mile,
and more than 6 crashes per mile.

Garber and Joshua (1991) used logistic regression to describe the observed
probabilities of truck involvement in accidents for a number of independent variables.
Data was obtained from Virginia police accident reports. The model was obtained by
taking the natural parameter as the link function:

\[
\log \left( \frac{\pi_i}{1 - \pi_i} \right) = [X_i]^T
\]

where: \((X_i)^T = \beta_0 + \beta_1x_1 + \beta_2x_2 + \ldots + \beta_nx_n\)

\(x_j\) are geometric variables
\(\beta_j\) are model parameters to be estimated
\(\pi_i = \text{total truck involvements} / (\text{total truck} + \text{nontruck involvements})\).

The value of \(\pi_i\) is the probability of any given accident involving a truck. As with the
linear regression model and the Poisson regression model developed by Garber and
Joshua, logistic regression models were developed for three different roadway types –
interstate highways, divided four-lane highways, and undivided highways. Using logistic
regression, the traffic and geometric factors that best predicted truck accident
involvement depended on roadway type. Logistic regression indicated that lane width
was a factor for undivided 4-lane and 2-lane highways, but not for divided 4-lanes or
interstate highways. Shoulder width and slope change rate were also predictors in the
logistic regression model for undivided highways. The best model was chosen by
selecting the smallest AIC value (AIC = 794.759). Logistic regression was not as
successful at predicting undivided highway accidents as it was at predicting divided
highway accidents according to AIC values. This result is in contrast to results for linear
regression and Poisson regression from the Garber and Joshua study. The AIC value for
logistic regression for undivided highways was higher (meaning the fit was not as good)
than the corresponding AIC values for both linear regression (AIC = 288.46) and Poisson regression (AIC = 62.06).

The Garber and Joshua study indicates that linear regression, Poisson regression, and logistic regression can be used for modeling truck crashes with geometric roadway features. However, the authors suggest that logistic regression models may not be as accurate in describing truck crash involvement as Poisson or linear regression models. Nevertheless, the nature of logistic regression lends itself to using the link function to predict proportions – such as the proportion of crashes that involved trucks or the proportion of truck crashes that are fatal. Also, ordinal logistic regression can be used to evaluate the relative importance of the two severity levels within a certain problem truck point or problem truck mile feature. To accomplish this, the crash rates would be divided into ordinal categories to satisfy the necessity of having categorical response variables.

2.3.5 LOGLINEAR MODELING

Saccomanno and Buyco (1988) considered several methods for calibrating statistical models of truck accident rates. A loglinear approach was chosen to assess the effect of the traffic environment on truck accident rates.

Saccomanno and Buyco quoted a study by Chira-Chavala and Cleveland (1986) in which a loglinear expression was developed combining involvement and exposure to yield an accident rate loglinear expression of the form:

$$\log \left( \frac{m_{ij}}{e_{ij}} \right) = (U - V) + (U_i - V_i) + (U_j - V_j) + (U_{ij} - V_{ij})$$

where: $m_{ij}$ is the expected number of accidents
$e_{ij}$ is the expected volume of truck travel
$U$ and $V$ are parameter estimates of accident and exposure models.

This equation was applied to variables with known, compatible accident and exposure measures. Chira-Chavala and Cleveland found that exposure does not affect the goodness of fit or the selection of the best accident rate model. However, Buyco and Saccomanno (1987) have shown that exposure does play a significant and distinctive role in fitting accident rate models by demonstrating that the use of separate loglinear expressions for accident and exposure data does not provide stable estimates of accident rates. Buyco and Saccomanno demonstrated that the classical weighted least squares
algorithm (WLSA) for calibrating loglinear models, used by Chira-Chavala and Cleveland, produces high residuals for cells that have low cell memberships in the contingency table of causal factors. This is problematic given the nature of accident data.

Saccomanno and Buyco used a generalized linear interactive model (GLIM) to calibrate the loglinear expression of truck accident rates. GLIM uses maximum likelihood techniques for estimating parameters in loglinear expressions. Two loglinear models were developed in this study. Model A is:

\[
\text{LOGNOACC} - k \times \text{LOGEXP} = 1 + R + A + \ldots + \text{RAM}
\]

where: \( \text{LOGNOACC} \) is the logarithm of the number of accidents
\( \text{LOGEXP} \) is the logarithm of truck travel exposure
R, A, …, and RAM are various combinations of factor inputs.

Model B is:

\[
\text{LOGNOACC} - \text{LOGEXP} = 1 + R + A + \ldots + \text{RAM} + (b - 1) \times \text{LOGEXP}
\]

where: all of the terminology is the same as Model A.

Model B is used to check the validity of the assumption that \( k \) in Model A is equal to 1.0. In Model B, exposure is treated as a covariate for an accident frequency expression. Model B can be rearranged to form an accident frequency expression:

\[
\text{LOGNOACC} = 1 + R + A + \ldots + \text{RAM} + b \times \text{LOGEXP}
\]

The best fit model was selected on the basis of the ratio of the likelihoods of the fitted model to the full or saturated model.

The variables used by Saccomanno and Buyco for the truck accident rate models were road type, traffic pattern, traffic volume, truck type, load status, model year, hour of day, and driver age. Using the two loglinear models, no significant relationships were found between the explanatory variables and truck accident involvement rate. The authors cited incompatibility between categorical accident data and continuous measures of exposure as a major problem for calibrating loglinear models of truck accident rates.

The results of the Saccomanno and Buyco study indicate that loglinear models may not provide acceptable prediction capabilities for truck accident rates. The authors did not include specific geometric roadway features in their study. Therefore, there is no way to know if the model would have been statistically significant if specific geometric features had been included.
2.3.6 ARTIFICIAL INTELLIGENCE TECHNIQUES

Artificial intelligence (AI) techniques attempt to advance the understanding of problems by using the mechanisms of intelligent behavior and underlying thought. AI techniques have recently been used to analyze transportation problems such as incident detection.

A literature review by Awad and Janson (1998) revealed no previous studies applying AI techniques to the explanation of truck accidents. The authors used two AI techniques (along with a Poisson – linear regression transformation, as previously discussed) to explain truck accidents at interchanges in Washington State during a 27-month period. The authors used neural networks and a hybrid system of fuzzy logic and neural networks as the two AI techniques. The data input consisted of specific highway geometrics for each ramp in the study (such as ramp width and ramp length), accident information, AADT for the freeways, and ADT for the ramps. The two artificial intelligence approaches showed a high level of performance in identifying different patterns of accidents and presented a better fit when compared to the Poisson – linear regression model.

Awad and Janson concluded that AI techniques are more capable of explaining the complexity of predicting truck accident frequencies than traditional regression procedures. The authors caution that great care should be exercised when using the neural network technique and that much training data are needed to reach reasonable results. Neural networks should be used to provide insights about the strong or weak connections between input variables and the desired output.

2.4 ACCOUNTING FOR EXPOSURE

The Transportation Research Board Special Report 228 (1990) recommends that greater quality control in collecting truck data is required in order to obtain better truck exposure data, which can lead to more meaningful truck crash analyses. Saccomanno and Buyco (1987) found that exposure plays a significant role in fitting accident rate models. The inclusion of exposure data is therefore important to producing efficient truck crash models. This section outlines the many ways that past studies have accounted for exposure in truck safety research.
Garber and Joshua (1991) accounted for exposure using data on vehicle-miles traveled (VMT) for single unit trucks and tractor trailers. Average daily traffic (ADT) figures for 1980 through 1986 were extracted from the police crash files. Using linear regression, accident rates were modeled in terms of accidents per 100 million VMT. Using Poisson regression and logistic regression, exposure data were used as explanatory variables. Wright and Burnham (1985) also obtained exposure data in truck VMT for Georgia roadways used in their study. In their research, exposure was measured in 100 million VMT. This data was used to calculate and model accident rates for trucks. In Vogt and Bared (1998), exposure, in million vehicle kilometers, was calculated by the formula:

\[
\text{EXPOSURE} = \text{ADT} \times 365 \times Y \times L_m \times 10^{-6}
\]

where: \( Y \) is the period of study in years

\( L_m \) is the segment length in kilometers.

The authors used this measure of exposure as an explanatory variable in a negative binomial regression model.

Miaou et al. (1993) used truck exposure data from the Highway Safety Information System (HSIS). Because truck travel data in HSIS were not broken down by truck type, data from the Highway Performance Monitoring System (HPMS\(^1\)) were employed as a supplementary data source whenever exposure data by truck type were needed. Mohamedshah et al. (1993) also used the HSIS to obtain exposure data.

Awad and Janson (1998) used the annual traffic report for Washington State for the years 1992 to 1994. The information included AADT for the main lane of each state route at various major interchanges and truck percentages at selected locations. The authors also made use of a dataset containing traffic counts for approximately 75% of the ramps used in the study and truck percentages for about 10% of the ramps. The exposure data were used as explanatory variables in the AI techniques and in the Poisson and linear regressions.

Saccomanno and Buyco (1988) used a generalized loglinear model for truck accident rates for Ontario for 1983. The algorithm for calibrating loglinear models of

\(^1\) The HPMS is a data collection effort, implemented by the FHWA, designed to provide current data on the mileage and use of highways, and to evaluate highway programs by monitoring changes in highway characteristics and performance.
accident rates permits the inclusion of exposure as a continuous covariate in the expression. Truck travel over the entire road network that was used in this study was estimated from provincial link-specific truck counts, adjusted by weigh station estimates from the 1983 Commercial Vehicle Survey (CVS) for Ontario. The proportion of trucks of a specific type at each weigh station was estimated directly from the CVS counts. These proportions were applied to total truck flows on similar types of roads to yield VMTs for trucks of different types on these roads. Total truck flows on the Ontario highway network were obtained directly from provincial traffic volume data.

The inclusion of exposure data is important in safety analysis. A statement that trucks are involved in only a small number of crashes along a certain roadway is only meaningful if the amount of truck travel, usually expressed in VMT or ADT, is also stated. Exposure has been shown to play a significant role in producing accident rate models (Saccomanno and Buyco, 1987). Based on prior studies, exposure should be accounted for in VMT as either an explanatory variable or as part of a truck crash rate dependent variable. A review of the truck safety literature reinforces the need to include traffic volume, roadway length, and percentage of trucks in developing truck safety models. All are available in some form for this study.

2.5 SUMMARY OF BACKGROUND INFORMATION

This section has provided the findings of a literature review that was conducted in the arena of truck safety research projects that have focused on using geometric roadway features as independent variables. The geometric features that have repeatedly been found to adversely affect truck safety in previous research efforts are lane width, shoulders, and horizontal curvature. Vertical gradient, slope change rate, and vertical curvature have also been found to affect truck safety in certain regression models.

Previous truck safety research efforts have used six different regression techniques to quantify truck safety models. Linear regression has been employed for safety analysis on numerous occasions and the \( R^2 \) goodness of fit statistic has varied widely in different studies. Poisson regression has been used with mixed results. Garber and Joshua (1991) indicate a preference for Poisson regression; however, Miaou et.al. (1993) did not have success using Poisson regression. Negative binomial regression
models have been shown to perform better than Poisson regression using the Akaike’s information criterion (AIC). Logistic regression has been shown less accurate compared to linear regression or Poisson regression in describing truck crash occurrences. One example of a loglinear model found no statistically significant predictors, although geometric roadway features were not used in the model. Artificial intelligence techniques have recently shown promising results in accounting for the complexities involved in modeling truck crash rates.

Obtaining exposure data is important in safety analysis. Exposure can be used in either the dependent variable or as independent variables. Exposure is generally given in vehicle-miles traveled (VMT) when used in truck crash rates as the dependent variable. When exposure data are used as independent variables, quantities such as ADT and percentage of trucks are common measures of exposure. Exposure can be obtained from average daily traffic (ADT) data. Multiplying the ADT by the time span of the study and by the total length of roadway segment being studied is a common method for calculating exposure in VMT. Exposure data is often obtained from the Highway Safety Information System (HSIS), although it can be obtained from traffic reports produced by various state and national agencies.

It is obvious that many of the studies in the arena of predicting truck crash involvement rates with geometric roadway features produce conflicting results. Much of the lack of consistency is probably the result of a combination of factors including lack of data and the omission of key variables such as driver factors and weather. Given the crash rates for trucks as well as the risks they pose to other motorists, it is imperative to continue researching this topic until the factors that affect truck crashes are more readily understood.

3.0 DATABASE DESCRIPTION

The first objective of this research is to refine the truck route evaluation procedure of the 1997-1999 Kentucky Transportation Center (KTC) study “Freight Movement and Intermodal Access in Kentucky” by examining the accuracy of the “adequate” and “less than adequate” weightings that were applied to various problem truck point and problem truck mile features. The second objective is to evaluate the relative influence of different
geometric roadway features on truck crashes to get an aggregate measure of truck route safety. For this research, a total of 81 truck routes (approximately 800 miles) from the KTC freight study will be used.

This section describes the development of the data into a form that allowed the objectives of this research to be accomplished. The first subsection describes the truck facility sites and truck routes used in this research and is followed by a description of the problem truck point and problem truck mile methodology previously developed. The final subsection explains how the database used in this research was built from the crash histories, problem truck points, and problem truck miles collected in the KTC freight study.

3.1 FACILITY SITES AND TRUCK ROUTES

In the KTC freight study, 81 truck routes used to access 46 facility sites or clustered facilities were evaluated. Facility sites were chosen for study based on total trucks per day and distance to the National Highway System (NHS), while ensuring that a variety of modes, commodities, and geographic areas throughout Kentucky were included. The facilities that were evaluated included rail-truck intermodal facilities, riverports, airports, coal tipples, quarries, industrial parks, food suppliers, power plants, and various types of factories and other manufacturing facilities. The sites were located in 37 different counties with at least one site located in each of the 15 Area Development Districts. Figure 1 on the following page presents all of the sites and clusters of sites selected for this study.

The truck routes used to access these sites represented approximately 800 miles of highway. The individual route lengths ranged from 0.4 miles to 54.3 miles with a mean route length of 10.0 miles. The maximum daily truck volume on the routes ranged from 20 to 2283 with a mean of 739. The 81 truck routes represent the routes utilized by trucks to travel between the facility sites and the NHS, except for a few cases in which the truck route evaluated was to the state line or to the actual destination of the freight. Using this approach, the actual route segments could be studied and the actual problems being experienced by truck traffic using the route could be addressed, even when different highway sections fell under the responsibility of different jurisdictions.
More information concerning the facility sites and truck routes can be found in “Final Summary Report on Truck Route Access Evaluation”, produced by the KTC in May 1999, as Research Report KTC-99-48. But the focus in this research is the use of crash histories to develop a crash model based on deficient geometric roadway features to allow the modification of the problem truck point and problem truck mile methodology and to better weight the relative impact of different geometric features and their levels of deficiency into an aggregate measure of the need or priority for truck route improvements.

Figure 1. Kentucky Truck-Generating Sites and Clusters of Sites Selected for Study.
3.2 PROBLEM TRUCK POINT AND PROBLEM TRUCK MILE METHODOLOGY

In the 1999 KTC freight study, problem truck points and problem truck miles were used for both point and continuous quantitatively evaluated truck route features. Problem truck points were used to evaluate point route features: offtracking\(^1\) on horizontal curves, safe speed on horizontal curves, truck stopping sight distance, intersection turning radii, railway crossings, and bridges. Problem truck miles were used to evaluate continuous route features: grade, lane width, and shoulders. Each of the above features at points along the routes or along route sections were graded as “preferred”, “adequate”, or “less than adequate” for truck access using criteria developed in brainstorming sessions and from background sources. The features graded as “preferred” received zero demerit points in the grading scheme. The features graded as “adequate” received one demerit point and the features graded as “less than adequate” received two demerit points. Thus, the “less than adequate” points or miles were weighted twice that of the “adequate” points or miles. For example, a 12 foot lane width was considered “preferred”, an 11 foot lane width was considered “adequate”, and a 10 foot or less lane width was considered “less than adequate”.

In order to obtain an aggregate route measure, the weighted sum of problem truck points along a route with respect to a particular feature was taken. The number of less than “preferred” demerit points (either 1 or 2 for each instance of each feature) was multiplied by the number of trucks per day that passed that particular point. This produced a weighted problem truck point total that accounted for truck volume. For example, for each truck route:

\[
\text{offtracking}_{(PTP)} = \text{offtracking}_{(1)} + (2) \text{offtracking}_{(2)}
\]

where: \(\text{offtracking}_{(PTP)}\) = total number of problem truck points contributed by offtracking for a particular truck route

\(\text{offtracking}_{(1)}\) = points along route rated “adequate” * trucks/day

\(\text{offtracking}_{(2)}\) = points along route rated “less than adequate” * trucks/day.

\(^1\) Offtracking occurs when a portion of a vehicle traversing a horizontal curve is forced either a.) across the centerline of the roadway, or b.) onto the shoulder of the roadway, due to an inadequate lane width given the radius of curvature and the length of the vehicle.
Consider a truck route with four locations of “adequate” offtracking and six locations of “less than adequate” offtracking. This truck route has a truck volume of 200 trucks per day. The number of problem truck points contributed by offtracking for this hypothetical truck route is:

\[
\text{offtracking}^{(PTP)} = \text{offtracking}^{(1)} + (2) \text{offtracking}^{(2)} \\
\text{offtracking}^{(PTP)} = (4)(200) + (2)(6)(200) \\
\text{offtracking}^{(PTP)} = 800 + 2400 \\
\text{offtracking}^{(PTP)} = 3200
\]

The total problem truck points for an individual truck route can be found by summing the problem truck points contributed by each of the six geometric point features. However, in the original KTC freight study, no means to determine the relative value or impact of problem truck points for different geometric features was determined. It is hoped that determining the relative effect of each deficient feature on crash rates in this study can contribute to this goal.

For continuous features, the number of less than “preferred” sections was multiplied by the length (in miles) of the section of the route exhibiting the less than “preferred” characteristics. This number was then multiplied by the number of trucks per day along the route in order to obtain the weighted sum of problem truck miles. For example, for each truck route:

\[
\text{lane width}^{(PTM)} = \text{lane width}^{(1)} + (2) \text{lane width}^{(2)} \\
\text{where: lane width}^{(PTM)} = \text{total number of problem truck miles contributed by lane width for a particular truck route.} \\
\text{lane width}^{(1)} = \text{miles exhibiting “adequate” lane width} \times \text{trucks/day} \\
\text{lane width}^{(2)} = \text{miles exhibiting “less than adequate” lane width} \times \text{trucks/day.}
\]

Consider a truck route with a length of 10 miles. This truck route exhibits three sections of “adequate” lane width of 0.3, 0.8, and 1.1 miles in length. This truck route also exhibits three sections of “less than adequate” lane width of 2.2, 0.5, and 2.9 miles in length. This truck route has a truck volume of 200 trucks per day. The number of problem truck miles contributed by lane width for this hypothetical truck route is:

\[
\text{lane width}^{(PTM)} = \text{lane width}^{(1)} + (2) \text{lane width}^{(2)}
\]
lane width\(_{(PTM)}\) = (0.3+0.8+1.1)(200) + (2)(2.2+0.5+2.9)(200)  
lane width\(_{(PTM)}\) = 440 + 2240  
lane width\(_{(PTM)}\) = 2680

The total problem truck miles for an individual truck route can be found by summing the problem truck miles contributed by each of the three geometric continuous features. Again, in the original KTC freight study, no determination of the relative value of problem truck miles for different geometric features was made.

In this manner, the problem truck points and problem truck miles can be used to compare different truck routes and their relative accommodation of trucks in using the route. Also, the relative urgency of needed route improvements can be examined. Route sections or points that are graded as “preferred” with respect to a particular route feature do not contribute to the sum of problem truck miles or points for that particular truck route. This research will attempt to quantify the actual numerical relationship between the urgency of the “adequate” and “less than adequate” features. In other words, this research will attempt to replace the weight of 2 with a defensible relative weight based on how “adequate” and “less than adequate” features impact truck crash rates. This will provide insight into the validity of weighting the “less than adequate” features as being twice as dangerous as “adequate” features. This research also will attempt to quantify the relative impact of different geometric roadway features on truck crash rates using problem truck points and problem truck miles so one combined overall measure can be used. This will provide insight into the validity of adding the problem truck points for all geometric features for a particular truck route as if all the geometric features were weighted equally in terms of severity of design deficiencies.

3.3 DATABASE CONSTRUCTION

In order to create truck route safety models based on crash histories, it was necessary to create a database containing truck crash data as well as problem truck points and problem truck miles for each geometric roadway characteristic described in the problem truck points and problem truck miles methodology section. The crash data, geometric roadway characteristics, truck volumes, and route lengths were obtained from the individual site reports of the KTC freight study.
Of the 81 truck routes in the KTC freight study, 68 were used in this research. Of the 13 truck routes that were not suitable for use, 7 routes came from 3 facility site reports in which the crash data were only reported by highway name and not by the entire truck route. Due to some truck routes using different sections of the same highway, it was impossible in these 7 cases to accurately state truck crash histories by truck route. Two truck routes had to be eliminated because they came from a facility site report that only listed the crash history for the entire site, not broken down by truck route. Two truck routes out of four at the Hickman-Fulton County Riverport Authority (Site #16) had to be eliminated because the crash data were only reported by highway name and not by the entire truck route. The other two truck routes from this site were usable because the crash histories were reported by truck route. One of the truck routes at the Inland Container Corporation (Site #2632) could not be used since it was along a new road with no reported crash history and little problem truck points or miles. The other truck route that was eliminated was the Edmonson County (Site #2686) truck route, which contained inconsistent crash data.

3.3.1 CRASH DATA

The first task in constructing the database was compiling the crash data for each truck route, which were extracted from the corresponding facility site report. The crash data had been previously extracted from the Kentucky State Police Crash Database for a three year period. For some routes, crash data from 1994 to 1996 was used. For other routes, crash data from 1995 to 1997 was used. The crash data were entered into a Microsoft Excel spreadsheet. The crash data in the reports were divided into non-truck crashes and truck crashes. The data were further sub-divided into fatal crashes and injury crashes. In addition, the number of crashes occurring at intersections was given for the non-truck and truck crashes.

Each row in the spreadsheet represented a different truck route. Descriptions of the truck route were written in the spreadsheet to distinguish truck routes of the same facility site. Columns were created for non-truck and truck crashes. Total crashes could be found by adding the numbers from these two columns for a specific truck route. Truck crashes were sub-divided into fatal crashes, injury crashes, and property-damage
only (PDO) crashes. One column was created for each of these three categories of truck crashes. The fatal truck crashes and the injury truck crashes were found directly from the individual site reports. The PDO truck crashes could be found by subtracting the fatal and injury truck crashes from the total truck crashes. Truck crashes occurring at intersections were recorded in the next column of the spreadsheet.

One major limitation of the crash data available for this research was that the location and characteristics of each individual crash were unknown. The KTC freight study only provided aggregate crash data for each truck route. This proved to be a limitation during the modeling process because it was therefore impossible to divide the truck routes into smaller sections of uniform length or uniform geometric features. The problem truck point and problem truck mile locations were all summarized in the KTC freight reports with detailed descriptions by mileage point along the roadway. It would have been possible to divide the routes into smaller sections, had the crash data been available in the same format. The individual crashes could have been isolated as having occurred in locations where a certain geometric feature or combination of geometric features had been identified along the roadway. It is hypothesized that this division of the truck routes would have been beneficial to the prediction capabilities of the models.

3.3.2 DATA FOR PROBLEM TRUCK POINTS

The next task was to represent the geometric characteristics of the truck routes in the spreadsheet by entering the problem truck points and the problem truck miles. Each of six columns were devoted to one of the geometric characteristics that contribute to problem truck points: offtracking on horizontal curves, safe speed on horizontal curves, truck stopping sight distance, intersection turning radii, railway crossings, and bridges. Data were gathered from each of the individual site reports for the one or more truck routes within a site report. The total number of problem truck points contributed by each of the six geometric characteristics for a given truck route were entered into the spreadsheet in the corresponding column. After the problem truck points for each of the six different geometric characteristics were entered into the spreadsheet, the total number of problem truck points for a particular truck route could be found by summing the values in each of these six columns.
For each truck route, the problem truck points for each feature that were contributed by “adequate” ratings and the problem truck points for each feature that were contributed by “less than adequate” ratings had to be determined in order to accomplish one of the objectives of this research – evaluating the accuracy of assigning one demerit point for “adequate” and two demerit points for “less than adequate”. To extract this data, each facility site report had to be examined individually. Each report contained a table called “Summary of Problem Truck Miles and Points for Truck Route(s)”. For each roadway geometric feature, the number of problem truck points contributed by scores of “adequate” and the number of problem truck points contributed by scores of “less than adequate” had to be determined. These two numbers were entered into separate columns in the database. As a check, these numbers were added to ensure that they summed to the total number of problem truck points contributed by a certain geometric feature. As previously discussed, the “less than adequate” locations were originally deemed twice as undesirable as the “adequate” locations for each geometric roadway feature in the KTC freight study. In that study, the truck volume was multiplied by two (2) for “less than adequate” features when determining the problem truck points contributed by a certain feature. In order to allow comparisons between the “adequate” problem truck points and the “less than adequate” problem truck points, a new column was created in the database in which the problem truck points contributed by the “less than adequate” features were divided by two. This had the effect of taking away the weighting given to “less than adequate” geometric features as being twice as undesirable as “adequate” geometric features. In other words, this new column made the assumption that the “adequate” locations and the “less than adequate” locations were two separate factors. For example, the variable that was used for modeling “less than adequate” offtracking locations was:

\[
\text{offtracking}_{(2)} = \text{points along route rated “less than adequate”} \times \text{trucks/day}.
\]

Notice that this variable does not contain the weighting of 2 that was originally given to it in the KTC freight study. Having the data in offtracking\(_{(1)}\) and offtracking\(_{(2)}\) allows models to be developed that will test what the true ratio is between offtracking\(_{(1)}\) and offtracking\(_{(2)}\) in terms of their truck crash prediction capability. For example, assuming a significant linear regression model is developed for predicting truck crash occurrence, the ratio of the coefficient of offtracking\(_{(2)}\) to the coefficient of offtracking\(_{(1)}\) is the number of
truck crashes contributed by a “less than adequate” offtracking location for every one
crash contributed by an “adequate” offtracking location.

3.3.3 DATA FOR PROBLEM TRUCK MILES

The database construction for problem truck miles proceeded in a similar manner
to the database construction for problem truck points. Data were gathered from each of
the individual site reports in the KTC freight study and the problem truck miles were
entered into the database for each of the three continuous geometric roadway
characteristics. For each truck route, the problem truck miles for each feature that were
contributed by “adequate” ratings and the problem truck miles for each feature that were
contributed by “less than adequate” ratings had to be determined in order to accomplish
one of the objectives of this study – evaluating the accuracy of assigning one demerit
point for “adequate” and two demerit points for “less than adequate”. This data was
extracted in the same manner as described for the problem truck points in the previous
subsection. Just as with the problem truck points, new columns were created in the
database in which the problem truck miles contributed by the “less than adequate”
geometric features were divided by two, in order to allow comparisons between the
“adequate” problem truck miles and the “less than adequate” problem truck miles. This
had the effect of taking away the weighting given to “less than adequate” geometric
features as being twice as undesirable as “adequate” geometric features.

At this point, all of the data that was necessary for the research described in this
paper has been extracted from the truck facility site reports of the 1997-1999 Kentucky
Transportation Center (KTC) study entitled “Freight Movement and Intermodal Access in
Kentucky”.

4.0 ANALYSIS AND RESULTS

The objective of this study is twofold. The first objective is to further develop the
problem truck point and problem truck mile methodology by quantitatively evaluating the
truck crash prediction ability of the two levels of importance (“adequate” versus “less
than adequate”) within a problem truck point or problem truck mile roadway feature.
The second objective is to evaluate the relative influence of different geometric roadway
features on truck crashes to get an aggregate measure of truck route safety. Regression analysis is used to accomplish both objectives using the truck crash histories, problem truck points, and problem truck miles collected in the KTC freight study for non-NHS truck routes in Kentucky. This section provides the results of these research objectives.

4.1 DESCRIPTIVE DATA INFORMATION

Table 2 presents information that summarizes the numbers in the database used for this study. Important characteristics of the geometric features, crash data, and exposure data are provided. Table 2 illustrates that route lengths of the truck routes in the database vary greatly. This indicates that route length should be accounted for in the models by normalizing the crash data to account for the long truck routes experiencing more truck crashes than the short truck routes just because they are longer. A histogram of the resulting truck crash per mile of truck route is shown in Figure 2.

The truck crash per route mile data ranges from a minimum of zero to a maximum of 39.167, yet the 75th percentile truck crash per route mile value is 2.415, only 6% of the maximum. Given this, there are probably some outlier truck crash data points, as can be seen in Figure 2. These unusual observations could prove difficult to model in regression analysis.

The average daily traffic (ADT) variable shown in Table 2 ranges from 818 to 38,589. The standard deviation of the ADT is 8775, indicating that a number of different exposure levels may be present in the database. This emphasizes the need to include ADT in the regression analysis.
Table 2. Descriptive statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>St. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXPOSURE DATA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>10.43</td>
<td>5.20</td>
<td>11.6</td>
<td>0.5</td>
<td>54.4</td>
<td>1.75</td>
<td>14.9</td>
</tr>
<tr>
<td>ADT</td>
<td>11671</td>
<td>9248</td>
<td>8775</td>
<td>818</td>
<td>38589</td>
<td>5227</td>
<td>16935</td>
</tr>
<tr>
<td>Truck ADT</td>
<td>766</td>
<td>639</td>
<td>533.6</td>
<td>20</td>
<td>2283</td>
<td>400</td>
<td>1130</td>
</tr>
<tr>
<td><strong>GEOMETRIC ROADWAY CHARACTERISTICS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off 1</td>
<td>1431</td>
<td>0</td>
<td>3555</td>
<td>0</td>
<td>20262</td>
<td>0</td>
<td>1154</td>
</tr>
<tr>
<td>Off 2</td>
<td>2771</td>
<td>308</td>
<td>5338</td>
<td>0</td>
<td>30800</td>
<td>0</td>
<td>3235</td>
</tr>
<tr>
<td>MSSC 1</td>
<td>465</td>
<td>0</td>
<td>1306</td>
<td>0</td>
<td>6930</td>
<td>0</td>
<td>56</td>
</tr>
<tr>
<td>MSSC 2</td>
<td>1279</td>
<td>0</td>
<td>3873</td>
<td>0</td>
<td>26800</td>
<td>0</td>
<td>796</td>
</tr>
<tr>
<td>Truck SD 1</td>
<td>0.588</td>
<td>0</td>
<td>4.851</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Truck SD 2</td>
<td>4.34</td>
<td>0</td>
<td>35.77</td>
<td>0</td>
<td>295</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T Radii 1</td>
<td>23.8</td>
<td>0</td>
<td>88.8</td>
<td>0</td>
<td>440</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>T Radii 2</td>
<td>157.7</td>
<td>0</td>
<td>582.4</td>
<td>0</td>
<td>4697</td>
<td>0</td>
<td>168</td>
</tr>
<tr>
<td>RR X 1</td>
<td>102.2</td>
<td>0</td>
<td>393.7</td>
<td>0</td>
<td>2576</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RR X 2</td>
<td>28.1</td>
<td>0</td>
<td>151.0</td>
<td>0</td>
<td>1183</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bri 1</td>
<td>967</td>
<td>472</td>
<td>1417</td>
<td>0</td>
<td>7920</td>
<td>0</td>
<td>1326</td>
</tr>
<tr>
<td>Bri 2</td>
<td>124.2</td>
<td>0</td>
<td>486.5</td>
<td>0</td>
<td>3510</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Grade 1</td>
<td>70.9</td>
<td>0</td>
<td>239.3</td>
<td>0</td>
<td>1618.7</td>
<td>0</td>
<td>8.8</td>
</tr>
<tr>
<td>Grade 2</td>
<td>278.4</td>
<td>0</td>
<td>752.7</td>
<td>0</td>
<td>4237.9</td>
<td>0</td>
<td>185.7</td>
</tr>
<tr>
<td>LW 1</td>
<td>1651</td>
<td>128</td>
<td>3788</td>
<td>0</td>
<td>20232</td>
<td>0</td>
<td>1105</td>
</tr>
<tr>
<td>LW 2</td>
<td>1606</td>
<td>347</td>
<td>3347</td>
<td>0</td>
<td>21677</td>
<td>0</td>
<td>1738</td>
</tr>
<tr>
<td>Shou 1</td>
<td>494</td>
<td>0</td>
<td>1793</td>
<td>0</td>
<td>12460</td>
<td>0</td>
<td>217</td>
</tr>
<tr>
<td>Shou 2</td>
<td>3669</td>
<td>1667</td>
<td>4819</td>
<td>0</td>
<td>27100</td>
<td>552</td>
<td>5050</td>
</tr>
<tr>
<td><strong>CRASH DATA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truck Crashes</td>
<td>11.18</td>
<td>8</td>
<td>10.6</td>
<td>0</td>
<td>47</td>
<td>4</td>
<td>15.75</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>107.3</td>
<td>76</td>
<td>93.9</td>
<td>0</td>
<td>356</td>
<td>29</td>
<td>175.3</td>
</tr>
<tr>
<td>Truck Cr. as % of Total Cr.</td>
<td>11.54</td>
<td>10.68</td>
<td>7.3</td>
<td>0</td>
<td>40</td>
<td>5.882</td>
<td>15.248</td>
</tr>
<tr>
<td>Fatal Truck Crashes</td>
<td>0.324</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0.75</td>
</tr>
<tr>
<td>Injury Truck Crashes</td>
<td>3.118</td>
<td>2</td>
<td>3.1</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>4.75</td>
</tr>
<tr>
<td>PDO Truck Crashes</td>
<td>7.740</td>
<td>5</td>
<td>8.3</td>
<td>0</td>
<td>43</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Truck Cr. per Route Mile</td>
<td>2.525</td>
<td>1.127</td>
<td>5.1</td>
<td>0</td>
<td>39.167</td>
<td>0.528</td>
<td>2.415</td>
</tr>
</tbody>
</table>

Length = Route Length, ADT = Average Daily Traffic, Off = Offtracking, MSSC = Maximum Safe Speed on Curves, Truck SD = Truck Sight Distance, T Radii = Truck Turning Radii, RR X = Railroad Crossing, Bri = Bridge, LW = Lane Width, Shou = Shoulder
One glaring characteristic of Table 2 is the number of geometric features that have a median of zero. This indicates that at least half of the data points for these problem truck point or problem truck mile features will be zero, indicating no substandard locations or sections with respect to that geometric feature. Some problem truck point features even have a 75th percentile value equal to zero. It may prove difficult to produce linear relationships describing truck crashes using an explanatory variable with so many zero frequencies. Another noteworthy feature of the geometric characteristics is the large standard deviations. This indicates a wide range of truck route qualities in the database. Given the fact that many of the geometric features have data points of zero and outliers exist in the crash data, care should be exercised in selecting the regression technique.

4.2 DATA CORRELATION

A correlation table (Table 3) of the problem truck point and problem truck mile data was produced to evaluate the viability of producing a truck safety model using all variables in one model to predict truck crashes. From the correlation table, it can be noted that many variables are highly correlated. As a result, obtaining meaningful results
using a composite model of all geometric roadway features is not possible. Of particular interest is the high correlation between the route length and the three continuous geometric roadway features: grade, lane width, and shoulders. Intuitively, one would expect this given that the length of the deficiency is part of the formula for problem truck miles and some routes have a continuous geometric deficiency for their entire length. Other correlations are not unexpected either such as offtracking with maximum safe speed on curves and ADT with truck ADT. Given the existence of correlated variables, individual models for geometric roadway characteristics were developed.

Table 3. Correlation table.

<table>
<thead>
<tr>
<th></th>
<th>Off</th>
<th>MSSC</th>
<th>Tr SD</th>
<th>T Radii</th>
<th>RR X</th>
<th>Bri</th>
<th>Grade</th>
<th>LW</th>
<th>Shou</th>
<th>Length</th>
<th>ADT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSSC</td>
<td>0.727</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truck SD</td>
<td>0.064</td>
<td>-0.045</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Radii</td>
<td>-0.080</td>
<td>-0.048</td>
<td>-0.038</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR X</td>
<td>0.001</td>
<td>0.026</td>
<td>0.543</td>
<td>-0.059</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bri</td>
<td>0.386</td>
<td>0.149</td>
<td>-0.087</td>
<td>-0.065</td>
<td>-0.182</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>0.640</td>
<td>0.434</td>
<td>-0.047</td>
<td>-0.026</td>
<td>-0.117</td>
<td>0.322</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LW</td>
<td>0.679</td>
<td>0.281</td>
<td>-0.047</td>
<td>-0.075</td>
<td>-0.112</td>
<td>0.649</td>
<td>0.732</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shou</td>
<td>0.447</td>
<td>0.126</td>
<td>-0.043</td>
<td>-0.079</td>
<td>-0.144</td>
<td>0.539</td>
<td>0.653</td>
<td>0.834</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>0.569</td>
<td>0.160</td>
<td>-0.096</td>
<td>0.010</td>
<td>0.208</td>
<td>0.538</td>
<td>0.590</td>
<td>0.805</td>
<td>0.766</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADT</td>
<td>-0.024</td>
<td>-0.098</td>
<td>-0.084</td>
<td>0.218</td>
<td>-0.212</td>
<td>0.235</td>
<td>-0.004</td>
<td>-0.031</td>
<td>0.049</td>
<td>-0.062</td>
<td></td>
</tr>
<tr>
<td>Truck ADT</td>
<td>-0.125</td>
<td>-0.170</td>
<td>0.089</td>
<td>0.325</td>
<td>0.103</td>
<td>0.275</td>
<td>-0.114</td>
<td>-0.038</td>
<td>0.145</td>
<td>-0.108</td>
<td>0.489</td>
</tr>
</tbody>
</table>

4.3 REGRESSION RESULTS

To quantitatively evaluate the objectives of this research, individual regression models were developed using the various problem truck point and problem truck mile features. The types of regression used are ordinal logistic regression, linear regression, and negative binomial regression. These regression types appear to be the most applicable given the information gained in the literature review.

Ordinal logistic regression is desirable over binary or nominal logistic regression because it allows the prediction of an ordinal categorical variable. Ordinal variables are categorical response variables that have more than two possible levels with a natural ordering. Truck crashes or truck crash rates can be divided into logical intervals at
natural breakpoints to form the ordinal categorical variable. Linear regression was chosen because it is a relatively simple modeling process, yet has performed as well as or better than many regression analyses in past studies. Negative binomial regression was chosen because it has been shown in previous studies to perform better than Poisson regression using the Akaike’s information criterion (AIC) and can be readily constructed from the truck crash data that is available for this study.

Truck crashes per unit length (truck crashes per route-mile) is the dependent variable in the regression models. The reason for this is to control for the fact that each truck route has a different route length (as discussed in subsection 4.1). Dividing the truck crashes observed on a certain route by the entire length of the route normalizes the measure by providing the number of crashes experienced per mile of truck route. While the actual truck route length could be determined from the data in the KTC freight study, truck ADT and overall ADT data from the KTC freight study are only valid for one section of the truck route – the section exhibiting the maximum ADT. Therefore, using truck crashes per truck-mile was ruled out for the regression models. In some cases, there is a uniform ADT along a truck route; however, in other cases the ADT would vary along the truck route but only the maximum ADT was reported when the original dataset was constructed.

4.3.1 ORDINAL LOGISTIC REGRESSION

An ordinal logistic regression model was one of three approaches produced to meet the objectives of this study. Ordinal variables are categorical response variables that have more than two possible levels with a natural ordering. In order to use ordinal logistic regression for this analysis, the variable truck crashes per route length had to be aggregated into some ordering scheme. To accomplish this, a histogram of truck crashes per mile of route length was produced (Figure 2). From examining the distribution and a list of data points, two breaks in the data were made. Therefore, three categories of truck crashes per route mile were created. Table 4 summarizes the number of routes in each category. The number of observations in each category are reasonably equal given that natural breakpoints in the data were sought when creating the categories in the hopes of producing more meaningful logistic regression models.
Table 4. Frequency distribution for truck crash per truck route length.

<table>
<thead>
<tr>
<th>Truck crash per truck route length</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>y &lt; 0.99</td>
<td>32</td>
</tr>
<tr>
<td>1 &lt; y &lt; 2.99</td>
<td>22</td>
</tr>
<tr>
<td>3 &lt; y</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>68</td>
</tr>
</tbody>
</table>

Nine different logistic regression models were produced, one for each of the nine different geometric roadway features. The dependent variable in all nine models was truck crashes per unit length. There were three explanatory variables: (1) the number of “adequate” problem truck points or miles, (2) the number of “less than adequate” problem truck points or miles, and (3) ADT. The results of the nine distinct ordinal logistic regressions are shown in Table 5. “Adequate” features are represented by 1 and “less than adequate” features are represented by 2.

Table 5. Results of ordinal logistic regression.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off 1</td>
<td>-0.0000316</td>
<td>0.729</td>
<td>T Radii 1</td>
<td>0.0016050</td>
<td>0.586</td>
<td>Grade 1</td>
<td>-0.0025700</td>
<td>0.444</td>
</tr>
<tr>
<td>Off 2</td>
<td>-0.0000300</td>
<td>0.634</td>
<td>T Radii 2</td>
<td>-0.0003071</td>
<td>0.453</td>
<td>Grade 2</td>
<td>-0.0009210</td>
<td>0.215</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0001409</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0001490</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0001444</td>
<td>0.000</td>
</tr>
<tr>
<td>MSSC 1</td>
<td>-0.0008061</td>
<td>0.057</td>
<td>RR X 1</td>
<td>0.0015770</td>
<td>0.017</td>
<td>LW 1</td>
<td>-0.0000243</td>
<td>0.705</td>
</tr>
<tr>
<td>MSSC 2</td>
<td>0.0001053</td>
<td>0.252</td>
<td>RR X 2</td>
<td>0.0043190</td>
<td>0.079</td>
<td>LW 2</td>
<td>-0.0002756</td>
<td>0.035</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0001430</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0001670</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0001503</td>
<td>0.000</td>
</tr>
<tr>
<td>Truck SD 1</td>
<td>-0.6390000</td>
<td>0.999</td>
<td>Bridge 1</td>
<td>-0.0002599</td>
<td>0.209</td>
<td>Shoulder 1</td>
<td>0.0000529</td>
<td>0.712</td>
</tr>
<tr>
<td>Truck SD 2</td>
<td>0.0958800</td>
<td>1.000</td>
<td>Bridge 2</td>
<td>-0.0014840</td>
<td>0.216</td>
<td>Shoulder 2</td>
<td>-0.0001625</td>
<td>0.012</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0001480</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0001568</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0001531</td>
<td>0.000</td>
</tr>
</tbody>
</table>

A summary of the geometric roadway features that were found significant in ordinal logistic regression is given in Table 6. At a significance level of $\alpha = 0.05$, there are three significant geometric predictors: (1) railroad crossing, (2) lane width, and (3) shoulder. There is reasonable evidence to accept two other variables as being practically significant: (1) maximum safe speed on curves with $p=0.057$, and (2)
railroad crossing with $p=0.079$. ADT is a significant predictor in all models. The sign of the ADT coefficient is positive in all models, which indicates that an increase in this explanatory variable leads to an increase in the response variable.

Table 6. Summary of significant geometric predictors – ordinal logistic regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Railroad Crossing 1</td>
<td>0.0015770</td>
<td>0.017</td>
</tr>
<tr>
<td>Lane Width 2</td>
<td>-0.0002756</td>
<td>0.035</td>
</tr>
<tr>
<td>Shoulder 2</td>
<td>-0.0001625</td>
<td>0.012</td>
</tr>
</tbody>
</table>

All other predictors are not significant predictors of truck crashes per route mile given their p-values. This means that the null hypothesis can not be rejected and their slopes can be assumed to be zero. The relevance of this finding is that these geometric variables do not predict truck crashes for the non-NHS routes that are included in this study and therefore may not be useful in the problem truck point and problem truck mile methodology for evaluating truck route improvement needs.

More specifically, these results suggest that most “adequate” rated features are not contributing to truck crashes and should perhaps be excluded from a rating scheme to prioritize needed route improvements. If there are not two severity levels of the same geometric roadway feature, it is unnecessary to obtain a significant ratio of crashes predicted by “adequate” locations to crashes predicted by “less than adequate” locations for any of the geometric features.

Railroad crossings of “adequate” truck access, “less than adequate” lane widths, and “less than adequate” shoulders are significant in predicting truck crashes. The sign of the railroad crossing variable is positive indicating that an increase in the number of “adequate” railroad crossings leads to more truck crashes per mile. The sign of the lane width and shoulder variables are negative indicating that as these problem truck mile values increase, truck crash rates decrease. Perhaps drivers are taking more precautions on the truck routes exhibiting “less than adequate” lane width and shoulders. According to ordinal logistic regression analysis, “adequate” railroad crossing problem truck points need to continue to be collected in future studies in order to evaluate roadways in terms of their truck crash rates. The other features need not be included in problem truck point and problem truck mile analyses that have the goal of predicting truck crashes.
4.3.2 LINEAR REGRESSION

Linear regression was the second analysis technique used to meet the objectives of this study. For linear regression, models were created using truck crashes per unit length (truck crashes per route-mile) as the dependent variable. Again, ADT was used as an explanatory variable. Nine different linear regression models were produced, one for each of the nine different geometric roadway features. There were three explanatory variables: (1) the number of “adequate” problem truck points or miles, (2) the number of “less than adequate” problem truck points or miles, and (3) ADT. The results of the nine distinct linear regressions are shown in Table 7.

Table 7. Results of linear regression.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.652</td>
<td>0.550</td>
<td>Constant</td>
<td>-0.3132</td>
<td>0.726</td>
<td>Constant</td>
<td>0.566</td>
<td>0.592</td>
</tr>
<tr>
<td>Off 1</td>
<td>-0.0001036</td>
<td>0.657</td>
<td>T Radii 1</td>
<td>0.0298200</td>
<td>0.000</td>
<td>Grade 1</td>
<td>-0.0006170</td>
<td>0.872</td>
</tr>
<tr>
<td>Off 2</td>
<td>-0.0000694</td>
<td>0.656</td>
<td>T Radii 2</td>
<td>-0.0009511</td>
<td>0.307</td>
<td>Grade 2</td>
<td>-0.0007900</td>
<td>0.518</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0001962</td>
<td>0.007</td>
<td>ADT</td>
<td>0.0002113</td>
<td>0.001</td>
<td>ADT</td>
<td>0.0001967</td>
<td>0.007</td>
</tr>
<tr>
<td>Constant</td>
<td>0.545</td>
<td>0.616</td>
<td>Constant</td>
<td>-0.468</td>
<td>0.663</td>
<td>Constant</td>
<td>0.983</td>
<td>0.385</td>
</tr>
<tr>
<td>MSSC 1</td>
<td>-0.0004758</td>
<td>0.389</td>
<td>RR X 1</td>
<td>0.0031850</td>
<td>0.043</td>
<td>LW 1</td>
<td>-0.0001381</td>
<td>0.395</td>
</tr>
<tr>
<td>MSSC 2</td>
<td>0.0000223</td>
<td>0.904</td>
<td>RR X 2</td>
<td>0.0019420</td>
<td>0.625</td>
<td>LW 2</td>
<td>-0.0002486</td>
<td>0.177</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0001923</td>
<td>0.009</td>
<td>ADT</td>
<td>0.0002285</td>
<td>0.002</td>
<td>ADT</td>
<td>0.0001931</td>
<td>0.008</td>
</tr>
<tr>
<td>Constant</td>
<td>0.249</td>
<td>0.814</td>
<td>Constant</td>
<td>0.719</td>
<td>0.488</td>
<td>Constant</td>
<td>1.065</td>
<td>0.339</td>
</tr>
<tr>
<td>Truck SD 1</td>
<td>-0.0266000</td>
<td>0.835</td>
<td>Bridge 1</td>
<td>-0.0007506</td>
<td>0.098</td>
<td>Shoulder 1</td>
<td>-0.0000943</td>
<td>0.791</td>
</tr>
<tr>
<td>Truck SD 2</td>
<td>0.0086000</td>
<td>0.618</td>
<td>Bridge 2</td>
<td>-0.0008050</td>
<td>0.519</td>
<td>Shoulder 2</td>
<td>-0.0002131</td>
<td>0.112</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0001988</td>
<td>0.007</td>
<td>ADT</td>
<td>0.0002332</td>
<td>0.002</td>
<td>ADT</td>
<td>0.0002028</td>
<td>0.005</td>
</tr>
</tbody>
</table>

A summary of the significant geometric roadway features that were found using linear regression is given in Table 8. At a significance level of \( \alpha = 0.05 \), there are two significant geometric predictors: turning radii \(_{(1)}\) and railroad crossing \(_{(1)}\). There is reasonable evidence to accept another variable as being practically significant: bridges \(_{(1)}\) with \( p=0.098 \). All other predictors are not significant predictors of truck crashes per route length given their p-values. This means that the null hypothesis can not be rejected and their slopes can be assumed to be zero. The relevance of this finding is that these
geometric variables do not predict truck crashes for the non-NHS routes that are included in this study using linear regression.

Table 8. Summary of significant geometric predictors – linear regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turning Radii 1</td>
<td>0.0298200</td>
<td>0.000</td>
</tr>
<tr>
<td>Railroad Crossing 1</td>
<td>0.0031850</td>
<td>0.043</td>
</tr>
</tbody>
</table>

As with logistic regression, Table 7 indicates that ADT is a significant predictor in all linear regression models. The sign of the ADT coefficient is positive in all models, which in linear regression indicates that an increase in this explanatory variable leads to an increase in the response variable. The coefficient of the ADT variable is consistently around 0.0002. This indicates that each increase of 5000 in ADT leads, on average, to one more truck crash per mile over a three year period.

Using linear regression, the geometric variables that were found significant in predicting truck crashes per mile were both “adequate” locations. This would suggest that the “less than adequate” locations have no effect on truck crashes, an opposite finding from logistic regression. The fact that turning radii(1) and railroad crossing(1) were found to be significant but not turning radii(2) and railroad crossing(2) suggests that drivers may be using more caution in these more severe geometric conditions. As in logistic regression, there are not two severity levels of the same geometric roadway feature that were found to be significant. Therefore, it is unnecessary to obtain a ratio of crashes predicted by “adequate” locations to crashes predicted by “less than adequate” locations for any of the geometric features. The sign of the coefficient for both the turning radii and the railroad crossing variables are positive, indicating that as these problem truck point values increase, truck crash rates also increase. According to linear regression analysis, “adequate” turning radii and “adequate” railroad crossing problem truck points need to continue to be collected in future studies in order to evaluate roadways in terms of their truck crash rates. The other features need not be included in problem truck point and problem truck mile analyses that have the goal of predicting truck crashes. Both linear regression models that contain significant geometric roadway predictors have negative constants that have high p-values. This indicates that the
constants are probably not different from zero. This is acceptable since if there is zero ADT, there will be zero crashes.

4.3.3 NEGATIVE BINOMIAL REGRESSION

Negative binomial regression was the third analysis technique used to meet the objectives of this study. As in logistic regression and linear regression, it was decided for negative binomial regression to model truck crashes per unit length (truck crashes per route-mile) as the dependent variable. Again, ADT was used as an explanatory variable. Nine different negative binomial regression models were produced, one for each of the nine different geometric roadway features. The dependent variable in all nine models was truck crashes per unit length of truck route. There were three explanatory variables: (1) the number of “adequate” problem truck points or miles, (2) the number of “less than adequate” problem truck points or miles, and (3) ADT. The results of the nine distinct negative binomial regressions are shown in Table 9.

Table 9. Results of negative binomial regressions.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
<th>Predictors</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.3301</td>
<td>0.148</td>
<td>Constant</td>
<td>-0.4767</td>
<td>0.028</td>
<td>Constant</td>
<td>-0.3177</td>
<td>0.139</td>
</tr>
<tr>
<td>Off 1</td>
<td>-0.0000361</td>
<td>0.442</td>
<td>T Radii 1</td>
<td>0.0021290</td>
<td>0.155</td>
<td>Grade 1</td>
<td>-0.0005914</td>
<td>0.432</td>
</tr>
<tr>
<td>Off 2</td>
<td>-0.0000135</td>
<td>0.667</td>
<td>T Radii 2</td>
<td>-0.0002593</td>
<td>0.227</td>
<td>Grade 2</td>
<td>-0.0002122</td>
<td>0.377</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0000582</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0000625</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0000582</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.3217</td>
<td>0.145</td>
<td>Constant</td>
<td>-0.6968</td>
<td>0.002</td>
<td>Constant</td>
<td>-0.2076</td>
<td>0.364</td>
</tr>
<tr>
<td>MSSC 1</td>
<td>-0.0002448</td>
<td>0.026</td>
<td>RR X 1</td>
<td>0.0009250</td>
<td>0.003</td>
<td>LW 1</td>
<td>-0.0000317</td>
<td>0.319</td>
</tr>
<tr>
<td>MSSC 2</td>
<td>0.0000286</td>
<td>0.432</td>
<td>RR X 2</td>
<td>0.0011663</td>
<td>0.127</td>
<td>LW 2</td>
<td>-0.0000894</td>
<td>0.015</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0000565</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0000694</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0000574</td>
<td>0.000</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.4576</td>
<td>0.038</td>
<td>Constant</td>
<td>-0.3213</td>
<td>0.135</td>
<td>Constant</td>
<td>-0.1769</td>
<td>0.432</td>
</tr>
<tr>
<td>Truck SD 1</td>
<td>-0.0146600</td>
<td>0.562</td>
<td>Bridge 1</td>
<td>-0.0001345</td>
<td>0.137</td>
<td>Shoulder 1</td>
<td>-0.0000028</td>
<td>0.968</td>
</tr>
<tr>
<td>Truck SD 2</td>
<td>0.0050160</td>
<td>0.147</td>
<td>Bridge 2</td>
<td>-0.0003792</td>
<td>0.130</td>
<td>Shoulder 2</td>
<td>-0.0000672</td>
<td>0.012</td>
</tr>
<tr>
<td>ADT</td>
<td>0.0000597</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0000651</td>
<td>0.000</td>
<td>ADT</td>
<td>0.0000591</td>
<td>0.000</td>
</tr>
</tbody>
</table>

A summary of the significant geometric roadway features that were found using negative binomial regression is given in Table 10. At a significance level of $\alpha = 0.05$, there are four significant geometric predictors: (1) maximum safe speed on curves, (2) railroad crossing, (3) lane width, and (4) shoulder. All other predictors are not
significant predictors of truck crashes per route length given their p-values. This means that the null hypothesis cannot be rejected and their slopes can be assumed to be zero. The relevance of this finding is that these geometric variables do not predict truck crashes for the non-NHS routes that are included in this study using negative binomial regression.

Table 10. Summary of significant geometric predictors – negative binomial regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Safe Speed on Curves 1</td>
<td>-0.0002448</td>
<td>0.026</td>
</tr>
<tr>
<td>RR X 1</td>
<td>0.0009250</td>
<td>0.003</td>
</tr>
<tr>
<td>LW 2</td>
<td>-0.0000894</td>
<td>0.015</td>
</tr>
<tr>
<td>Shoulder 2</td>
<td>-0.0000672</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 9 indicates that ADT is a significant predictor in all models. This is consistent with the ordinal logistic regression models and the linear regression models. The value of the coefficient for ADT is consistently around 0.00006. This indicates that each increase of 16,667 in ADT leads to an average of 2.718 more truck crashes per mile over a three year period, or approximately one more truck crash per mile with each increase of 6000 in ADT. Using linear regression, this increase in ADT was only 5000 for an average of one more truck crash per mile.

The negative binomial regression analysis supports the logistic regression and the linear regression analyses in that there are not two severity levels of the same geometric roadway feature that significantly predict truck crashes per truck route mile. Therefore, the same implications regarding “adequate” versus “less than adequate” drawn previously apply to the negative binomial models also. “Adequate” safe speed on horizontal curve locations and “adequate” railroad crossing locations are significant in predicting truck crashes per mile. “Less than adequate” lane width locations and “less than adequate” shoulders are significant in predicting truck crashes per mile. The sign of the coefficient for the safe speed on horizontal curve variable, the lane width variable, and the shoulder variable is negative, indicating that an increase in problem truck points or miles of these features lead to fewer crashes per mile. The sign of the coefficient for the railroad crossing variable is positive, indicating that an increase in the problem truck points of this feature leads to an increase in truck crashes per mile. According to negative binomial regression analysis, “adequate” railroad crossing problem truck points need to continue to be collected in future studies in order to evaluate roadways in terms of their truck crash
rates. Based on negative binomial regression, the other geometric roadway features need not be included in problem truck point and problem truck mile analyses that have the goal of predicting truck crash rates.

4.4 A CLOSER LOOK AT THE RESULT REGARDING RAILROAD CROSSINGS

Although the variable railroad crossing \( x_{(1)} \) is significant in predicting truck crashes per mile using all three regression techniques, it is not considered a particularly helpful result. Only six truck routes had non-zero railroad crossing \( x_{(1)} \) values. The two largest values of railroad crossing \( x_{(1)} \) came from the two truck routes of the Bells Lane Cluster (Site #7) in Louisville. Both of these truck routes are in a large urban area. The routes exhibit large ADT values, many crashes, and are both only one mile in length. Therefore, the truck crash per mile response variable for these two routes are the second and sixth largest out of the 68 truck routes in the study. Specific crash locations were not available from the KTC freight study report on this facility; however, the report mentioned that many of the crashes occurred at intersections and did not mention the railroad crossings as being a factor in crashes. The third highest total of problem truck points contributed by railroad crossing \( x_{(1)} \) was the Millard Processing Facility (Site #1677) in Pike County. There was only one truck crash at this site over a three year period and it did not occur at a railroad crossing. Table 11 illustrates that the Bells Lane Cluster truck routes may be dominant in determining the significance of railroad crossing \( x_{(1)} \) in this research. It is more likely that the crashes at this site are due to exposure and not railroad crossings.

Table 11. Truck routes exhibiting “adequate” railroad crossings.

<table>
<thead>
<tr>
<th>Truck Route</th>
<th>RR X 1 Prob Truck Points</th>
<th>Truck Crashes per Mile</th>
<th>Truck Crash per Mile Rank (out of 68 with 1 = highest)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 7, Route 2</td>
<td>2576</td>
<td>11.0</td>
<td>2</td>
</tr>
<tr>
<td>Site 7, Route 1</td>
<td>1388</td>
<td>7.0</td>
<td>6</td>
</tr>
<tr>
<td>Site 1677, Route 1</td>
<td>1200</td>
<td>0.909</td>
<td>37</td>
</tr>
<tr>
<td>Site 2031, Route 1</td>
<td>664</td>
<td>1.86</td>
<td>21</td>
</tr>
<tr>
<td>Site 1499, Route 1</td>
<td>660</td>
<td>0.333</td>
<td>59</td>
</tr>
<tr>
<td>Site 33, Route 2</td>
<td>464</td>
<td>0.6</td>
<td>47</td>
</tr>
</tbody>
</table>
To test the validity of concluding that the Bells Lane Cluster site was determining the significance of railroad crossing, an ordinal logistic regression was performed that excluded the Bells Lane Cluster from the analysis, using railroad crossing problem truck points and ADT as explanatory variables. The result was that the sign of the coefficient on railroad crossing changed from positive to negative ($p = 0.071$). This indicates that, excluding the Bells Lane Cluster, an increase in railroad crossing problem truck points actually leads to a decrease in crashes – the opposite result from what was obtained when including the Bells Lane Cluster truck routes. Therefore, the finding of railroad crossing as being statistically significant with a positive coefficient by all three regression techniques is not particularly helpful.

4.5 SUMMARY OF RESULTS

Table 12 summarizes the geometric roadway variables that were found to be significant in the three models. The statistically significant variables are denoted by regression technique with the sign of the coefficient of the variable. The variables lane width and shoulder are significant using both the ordinal logistic regression and the negative binomial regression. The sign of the coefficients of these two variables are negative using both regression types. Therefore, an increase in “less than adequate” problem truck miles for lane width and for shoulders indicate a decrease in truck crashes per mile. This suggests that drivers are possibly more cautious when severe lane width and shoulder conditions are present.

Since no model produced both geometric variables (“adequate” and “less than adequate”) as being significant, it is impossible to calculate the ratio of coefficient two to coefficient one. This would have yielded the relative weight of “less than adequate” locations to “adequate” locations. Instead, the inference is made that all geometric variables that are not statistically significant have coefficients that are not different from zero. Therefore, they do not facilitate the prediction of truck crashes and may not need to be included when evaluating the urgency of route improvements. ADT was found to be statistically significant with a positive coefficient in all models; therefore, ADT should be included when evaluating the urgency of route improvements.
Table 12. Statistically significant predictors of truck crashes per mile of route length.

<table>
<thead>
<tr>
<th>Variable</th>
<th>logistic regression</th>
<th>linear regression</th>
<th>negative binomial regr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offtracking 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offtracking 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSSC 1</td>
<td></td>
<td></td>
<td>negative</td>
</tr>
<tr>
<td>MSSC 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truck SD 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truck SD 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turning Radii 1</td>
<td></td>
<td></td>
<td>positive</td>
</tr>
<tr>
<td>Turning Radii 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RR Crossing 1</td>
<td>positive *</td>
<td>positive *</td>
<td>positive *</td>
</tr>
<tr>
<td>RR Crossing 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bridge 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bridge 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane Width 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lane Width 2</td>
<td>negative</td>
<td></td>
<td>negative</td>
</tr>
<tr>
<td>Shoulder 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder 2</td>
<td>negative</td>
<td></td>
<td>negative</td>
</tr>
<tr>
<td>ADT</td>
<td>positive</td>
<td>positive</td>
<td>positive</td>
</tr>
</tbody>
</table>

* - railroad crossing result is highly suspect

5.0 CONCLUSIONS AND RECOMMENDATIONS

To address the accuracy of assigning “adequate” and “less than adequate” scores for the problem truck point and problem truck mile methodology and to evaluate the relative impact of different geometric roadway features on truck crashes on non-NHS truck routes in Kentucky, this study models truck crashes per mile of truck route using individual problem truck point and problem truck mile components as explanatory variables. Three types of regression models were used based on information collected in the literature review. The data that was used is data that was previously collected during the 1997 to 1999 Kentucky Transportation Center (KTC) study entitled “Freight Movement and Intermodal Access in Kentucky”.

Using available data for truck crash prediction, it is impossible to obtain a significant ratio of crashes predicted by “adequate” problem truck point or problem truck mile locations to crashes predicted by “less than adequate” problem truck point or problem truck mile locations. Most of the coefficients of the geometric roadway features
used to predict truck crashes per unit length are statistically equal to zero. The problem truck point and problem truck mile features that were consistently found to be significant in predicting truck crashes per route length are railroad crossings, lane width, and shoulder. The railroad crossing result is probably not helpful due to the strong influence of one site in which railroad crossings were probably not the cause of many truck crashes. The lane width and shoulder result is counterintuitive because the sign of the coefficient was negative. This indicates “less than adequate” geometric features were causing safer conditions.

From the results of this study, there is insufficient evidence to conclude which if any of the problem truck point and problem truck mile features need to continue to be collected for truck safety analysis or prioritization of route improvements. There was no geometric roadway feature that consistently had positive coefficients using all regression techniques. Most of the statistically significant predictors actually had negative coefficients, seemingly resulting in safer conditions.

It was impossible to evaluate the relative impact of different geometric roadway features due to the majority of the coefficients of the geometric variables not being statistically significant. This would indicate that these variables all had no effect on truck crashes per mile.

The average daily traffic (ADT) variable that was used in this research was the maximum ADT along a particular truck route. This measure of ADT was found to be statistically significant in predicting truck crashes per unit length. It is unknown if a weighted average of ADT for each truck route would have been a less statistically significant explanatory variable. Nevertheless, the sign of the coefficient of ADT was found to be positive and statistically significant in every model produced. Clearly, ADT is an important predictor of truck crashes per mile. As a result, ADT should be used as an urgency measure for truck route improvements.

It is possible that more information could be extracted from the KTC truck route research dataset, but it would be necessary to divide the truck routes into smaller segments (such as one mile in length) and try again to obtain significant results using geometric roadway features. This would provide approximately 800 data points instead of only 68. This research suggests that the total mileage of information is not important.
if the data are not presented in an effective manner. The problem truck point and problem truck mile information is already given by milepost. Using the existing database, the location of all truck crashes used in the study would need to be found by milepost to improve the database to a point that the 68 truck routes could be divided into smaller segments. Geometric characteristics may be sufficient to predict truck crashes using smaller route segments. The advantage of using shorter roadway segments and exact truck crash locations is that the truck crash data will not have to be averaged over the entire length of a truck route. Also, the true cause of a truck crash may be pinpointed more accurately since there is likely to be less geometric deviation over shorter roadway segments. It is unclear if resources will be available for this intensive of an effort.

Future truck route improvement methods developed could use other explanatory variables such as driver characteristics, truck type, and highway type, as well as roadway geometric variables to predict truck crashes. McGee (1986) hypothesized that non-geometric variables such as these would have more influence on truck safety. Omitted explanatory variables are believed to be important in the low prediction capabilities of many of these models. This is consistent with Miaou et.al. (1993) and Vogt and Bared (1998). While not all geometric roadway features were found significant in predicting truck crashes per unit length, it is believed that other variables not available for this research could have possibly performed better.

This study provides useful information for application of the problem truck point and problem truck mile methodology. It is believed that more accurate crash data may be needed and shorter roadway segments should be evaluated for attempting to predict truck crashes using individual problem truck point and problem truck mile components. The research presented in this paper strengthens the priority of obtaining better crash data and analyzing data by smaller route segments. This research also strengthens the priority of obtaining ADT data for route improvement urgency. This study indicates that there could be less geometric data collected in future truck crash prediction studies. Assigning geometric features as being “adequate” or “less than adequate” is not required based on the results of this study. In future data collection efforts, data other than geometric characteristics should be collected for route improvement urgency – such as traffic volume and detailed crash histories.
Using the current form of the database, “adequate” and “less than adequate” distinctions are not able to predict truck crashes per mile. Some “less than adequate” locations behave as if they were “preferred”. These results suggest that using the current form of the database, lane width \(_{2}\) (10 foot wide lanes or less), and shoulder \(_{2}\) (less than 10 feet of usable shoulder) are statistically significant in reducing the number of truck crashes per unit length of non-NHS truck routes.

6.0 REFERENCES


APPENDIX

A DESCRIPTION OF GEOMETRIC FEATURES INCLUDED IN THIS STUDY
PROBLEM TRUCK POINTS

OFFTRACKING

Offtracking is the first of two aspects of trucks traveling around horizontal curves that were evaluated in the Kentucky Transportation Center’s study entitled “Freight Movement and Intermodal Access in Kentucky”. Large trucks require a wider lane width on curves than on tangent sections to accommodate the tracking of the rear wheels outside the path of the front wheels. The following guidelines were used to determine if each curve was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred” : A truck with its front wheels in the center of the lane can negotiate the turn without offtracking.

“adequate” : A truck can keep its back wheels within the lane while placing the front wheels at the outside edge of the travel lane.

“less than adequate” : The lane is narrower than the swept path of the vehicle, and the truck cannot avoid encroaching on the opposing lane or shoulder.

MAXIMUM SAFE SPEED ON CURVES

Maximum safe speed on curves is the second of two aspects of trucks traveling around horizontal curves that were evaluated in the KTC freight study. This variable is related to the speed at which a vehicle can travel around a horizontal curve before centrifugal force causes the vehicle to skid off the road. The ball bank indicator is recommended for evaluating the maximum safe speed. The following guidelines were used to determine if each curve was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred” : The ball bank indicator reading at the speed limit is not over 12.

“adequate” : The ball bank indicator reading is less than or equal to 12 at curve advisory speed.

“less than adequate” : The ball bank indicator reading is greater than 12 at curve advisory speed.

TRUCK SIGHT DISTANCE

Sight distance was one of two geometric characteristics of intersections considered relevant for truck routes. There should be sufficient sight distance at intersections to allow all approaching vehicles to stop when necessary. The following guidelines were used to determine if each intersection was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred” : sight distance > stopping sight distance

“adequate” : sight distance < stopping sight distance, but warning sign is present to indicate stop, signal, or intersection ahead

“less than adequate” : sight distance < stopping sight distance
TRUCK TURNING RADII

Turning radii was the other geometric characteristic of intersections considered relevant for truck routes. All right turns required for the trucks along the route were evaluated, but only those left turns that appeared to have insufficient turning radii were evaluated. Right turns are a concern because if the rear wheels track outside of the lane lines when turning, the wheels may hit the curb at the edge of the lane or track onto the shoulder, potentially hitting objects or pedestrians. The following guidelines were used to determine if each intersection was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred”: Truck turning from right lane can complete maneuver without leaving the travel lanes.
“adequate”: Truck may partially encroach on other lanes in the same direction but does not encroach on opposing traffic.
“less than adequate”: Truck enters opposing traffic lane or must start from a position completely outside the right-most lane.

RAILROAD CROSSINGS

The Kentucky Transportation Cabinet maintains a database containing all of the approximately 2600 railroad crossings in Kentucky. In the KTC study, it was important to consider the appropriateness of particular crossings for trucks rather than overall accident hazard. The following guidelines were used to determine if each crossing was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred”: be close to 90°, have sufficient sight distance, have good pavement/surface quality, and have nearly level approach grades
“adequate”: three of the above four characteristics
“less than adequate”: two or fewer of the desired qualities

BRIDGES

The Kentucky Transportation Cabinet currently maintains an inventory of all bridges in the state as required by FHWA. This database contains detailed geometric and operational information on each bridge, along with a composite sufficiency rating (scale: 1 to 100). This composite rating was a weighted score of three categories: structural adequacy and safety, serviceability and functional obsolescence, and essentiality for public use. The following sufficiency ratings were used to determine if each bridge was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred”: 80.0 – 100.0
“adequate”: 50.0 – 79.9
“less than adequate”: 0 – 49.9
PROBLEM TRUCK MILES

GRADE

Upgrades on highways can cause trucks to lose speed depending on the steepness, length of slope, and the power-to-weight ratio of the vehicle. *A Policy on Geometric Design of Highways and Streets* (AASHTO, 1994) provides speed reduction graphs for heavy trucks given percent and length of upgrade. The following speed reduction criteria were used to determine if each grade was rated as “preferred”, “adequate”, or “less than adequate”.

“preferred” : up to 8 km/hr speed reduction (18 km/hr if preceded by a downgrade)
“adequate” : a maximum of 15 km/hr speed reduction (up to 25 km/hr if preceded by a downgrade)
“less than adequate” : greater than 15 km/hr speed reduction (25 km/hr when preceded by a downgrade)

LANE WIDTH

Lane width is the first of two aspects of cross section to be evaluated in the KTC freight study. Lane width has one of the largest effects on driver safety and comfort of any roadway feature (AASHTO, 1994). Because trucks are significantly wider than passenger cars, the problems resulting from inadequate lane widths are increased for trucks. The following guidelines were used to determine if lane widths were “preferred”, “adequate”, or “less than adequate”.

“preferred” : 12 foot wide lanes
“adequate” : 11 foot wide lanes
“less than adequate” : 10 foot wide lanes or less

SHOULDERs

Shoulders are the second of two cross section elements that were evaluated in the KTC freight study. A shoulder is the usable portion of a roadway adjacent to the travel lanes that can be used for emergencies. The criteria used for evaluating shoulders included width and condition. The following guidelines were used to determine if shoulders were “preferred”, “adequate”, or “less than adequate”.

“preferred” : a minimum of 10 feet of paved shoulder
“adequate” : a minimum of 10 feet of unpaved shoulder
“less than adequate” : less than 10 feet of usable shoulder