The Pennsylvania State University

The Graduate School

College of Engineering

FACTORS AFFECTING DRIVER SPEED CHOICE ALONG TWO-LANE RURAL HIGHWAY TRANSITION ZONES

A Dissertation in

Civil Engineering

by

Ivette Cruzado

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2009

The dissertation of Ivette Cruzado was reviewed and approved* by the following:

Eric T. Donnell
Assistant Professor of Civil Engineering
Dissertation Adviser
Chair of Committee

Paul P. Jovanis
Professor of Civil Engineering

Venky Shankar
Associate Professor of Civil Engineering

Steven F. Arnold
Professor of Statistics

William D. Burgos
Associate Professor of Environmental Engineering
Professor in Charge of Graduate Programs

* Signatures are on file in the Graduate School.

ABSTRACT

Rural highways provide connections between developed areas. In many instances, two-lane rural highways that pass through undeveloped areas provide high levels of mobility that are accompanied by posted speed limits that exceed 45 mph. However, it is common for two-lane rural highways in Pennsylvania to pass through low-speed, developed areas (i.e., rural villages) with posted speed limits that are 35 mph or less. The roadway section between the high- and low-speed environments is referred to as a transition zone. In some cases, transition zone design may be accompanied by changes in roadway geometric features; however, it is hypothesized that drivers fail to adjust their speeds to comply with the change in the regulatory speed at the low-speed end of the transition zone. In other instances, drivers are only informed of the posted speed limit changes by regulatory signs with no corresponding changes in the roadway geometry.

Speed data were collected at 20 two-lane rural highway transition zones in central Pennsylvania. At each study site, speed data were collected at four locations: 500 feet before the transition zone, at the beginning of the transition zone, at the end of the transition zone, and 500 feet after the transition zone. The location of the sensors permitted vehicles to be "tracked," thus the final analysis database included four speed observations collected from 2,859 individual drivers for a total of 11,436 speed observations. Highway characteristic data were also collected at each location, including geometric design features, roadside elements, and access density, among others. The primary objective of this research was to develop speed prediction models to explain the relationship between the roadway features present along a two-lane rural highway transition zone and driver operating speeds. Two general model specifications were considered based on the available speed data. These included point speeds based on the "tracked" vehicles, and speed differentials between successive data collection points in a transition zone.

In the point speed analysis, four repeated speed measurements were collected on each of the 2,859 drivers across 20 different sites. Longitudinal models were used to model these data and compared to the more traditional operating speed modeling approach, ordinary least squares (OLS) regression. Use of OLS regression violates the

assumption of independent observations. The longitudinal models considered in this research were panel data models using both the fixed and random effects estimator, multilevel models, and generalized estimating equations (GEE). From the results of the analyses it was concluded that a three-level model in which speed observations were nested in drivers and drivers were nested in sites is more appropriate in explaining the influence of highway characteristics on driver speeds along two-lane rural highway transition zones. Key relationships between highway features and mean operating speeds in transition zones are as follows:

- When compared to a posted speed limit of 55 mph, a speed limit of 45 mph is associated with a mean operating speed reduction of approximately 3.5 mph. A speed limit of 25 mph is associated with a mean operating speed that is approximately 10.5 mph lower than the baseline of 55 mph. Similarly, a posted speed limit of 35 or 40 mph is associated with a mean operating speed that is approximately 2.4 mph lower than the baseline of 55 mph.
- Wider travel lanes and lateral clearance distances are associated with higher operating speeds along two-lane rural highway transition zones; a mean operating speed increase of 2.4 mph is expected per one-foot of lane width increase while a one-foot increase in lateral clearance is associated with a mean operating speed increase of 0.15 mph.
- The presence of curb is associated with a mean speed reduction of approximately 4 mph while the analysis indicated that a mean speed reduction of 1 mph is associated with a one-unit increase in driveway density.
- The presence of Intersection Ahead and School/Children warning signs were associated with 2 and 1 mph mean speed reductions, respectively, while the presence of a Curve Ahead warning sign was associated with a mean speed increase of almost 1 mph, when compared to the baseline of other warning sign types.
- Finally, the presence of a horizontal curve was associated with a mean speed reduction of 1.5 mph; if the horizontal curve is combined with a warning sign, a mean speed reduction of almost 3 mph is expected when compared to the baseline of a tangent roadway section.

The results from the three-level model also provided the standard deviation associated with each level of the model hierarchy. The standard deviations of the random components from the model developed were: 3.1 mph for highest level (site cluster), 2.1 mph at the second level (driver cluster), and 6.5 mph at the lowest level (speeds).

A second data set was created in which the response variable was change in speed along the transition zone. By considering speed change as the response variable, only one data point per vehicle (driver) was available; however, a site cluster could still be considered in the model specification. Use of the speed differential as the dependent variable in a statistical model eliminated part of the repeated observation issue present in the point speed analysis. As such, two general modeling methods were considered. These included OLS regression and multilevel models in which speeds were nested in sites. The variables that were consistently associated with speed reductions across all models were changes in the posted speed limit, reduction in paved shoulder width (1 mph reduction per one-foot reduction in paved shoulder width), number of driveways (0.36 mph reduction per one-unit increase in driveway density), school/children related warning signs (8 mph mean speed reduction), length of transition zone (0.8 mph average speed reduction per 100 foot increase in transition zone length), and presence of horizontal curve that warrants a warning sign (3.2 mph mean speed reduction is expected with this type of horizontal curve). The presence of a Curve Ahead warning sign and tangent sections were consistently associated with a speed increase along transition zones across all models (3.2 mph average and 2 mph average, respectively).

Several independent variables were not statistically significant in the multilevel speed differential model when compared to the OLS regression model. These included the change in lane width and in lateral clearance, presence of a curb, and Intersection Ahead warning sign. Although the standard errors of the parameter estimates obtained using OLS regression were smaller than those obtained using the multilevel models, the multilevel model is a better representation of the nesting structure of driver speed differential nested within data collection sites.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

AKNOWLEDGEMENTS

I am very grateful to my advisor, Dr. Eric T. Donnell; I do not believe I could have finished this dissertation without his help. I am also grateful for the insight of my dissertation committee members: Dr. Paul P. Jovanis, Dr. Venky Shankar, and Dr. Steven F. Arnold.

I would like to thank all my friends who supported me and helped me with data collection. I would also like to thank the employees at PennDOT as well as the personnel at PTI's Test Track: Rick, Mike (a.k.a. Casanova), and Rae; thanks for making me laugh. Thanks to my friends at PTI's Team Lab, especially to Scott Himes and Vishesh Karwa, for letting me interrupt their work so we could brainstorm about my research. I am very grateful to Miss Terry Reed, who taught me everything I needed to know about networking and etiquette.

Thanks to my best friend, Ingrid Guadalupe, for believing in me. I am most grateful to a certain group of friends, whose help, love, and support during the bad days were vital to arrive to the finish line: Mildred Rodríguez, Marta Ventura, Maria Schmidt, and Nancy Vanessa Vicente. Lastly, I would like to dedicate this dissertation to my mother, Eileen I. Vélez de Cruzado, the most important person in my life.

CHAPTER 1
INTRODUCTION

Rural highways provide connections between developed areas, both residential and commercial. Safety issues may arise when traveling from a high-speed undeveloped to a low-speed developed environment. The roadway section between the high- and low-speed environments is referred to as a transition zone. In some cases, transition zone design may be accompanied by changes in roadway features; however, it is hypothesized that drivers fail to adjust their speeds accordingly. In other instances, drivers are only informed of the required speed changes by traffic signs with no corresponding changes in the roadway geometry. There are currently no geometric design guidelines for transition zones on two-lane rural highways. As such, the objective of this research is to collect operating speed, geometric design, roadside, and land use data along two-lane rural highway transition zones in Pennsylvania. Operating speed models are then estimated in order to obtain information about which roadway, roadside, and land use features are associated with changes in speed along transition zones.

1.1 Background

In 2004, there were more than 4.0 million miles of publicly-owned highways in the United States (U. S.), 77 percent of which are rural roadways (FHWA, 2004).
Two-lane rural highways must balance mobility and access, especially when passing through remote or sparsely developed areas. For the purposes of this research, a "transition zone" is defined as the section of a two-lane rural highway where the regulatory speed changes as the roadway passes through a developed area, either commercial or residential.

Speed limits along high-speed two-lane rural highways typically exceed 40 mph. When passing through a developed area, the posted speed on two-lane rural highways is often reduced. The posted speed limit change is often accompanied by an increase in access density or pedestrian activity in the low-speed section of the two-lane rural highway. Traffic signs are sometimes the only way of communicating to drivers concerning the required change in vehicle operating speeds in transition zones.

The Manual on Uniform and Traffic Control Devices (MUTCD, 2003) contains guidelines on the size, shape, color, and placement of traffic signs. The "Speed Limit Sign" informs drivers about the limit established by law, ordinance, or regulation, and is thus classified as a regulatory sign. The "Reduced Speed Ahead Sign" informs drivers of an upcoming speed limit change; it is classified as a warning sign. Prior to passage of the 2003 edition of the MUTCD, the "Reduce Speed Ahead Sign" was classified as a regulatory sign. Figure 1 shows the evolution of the Reduced Speed Ahead sign, from the 2000 MUTCD edition, R2-5 series, to the 2003 edition, W3-5 series. The pre-2003 speed-zone signs are frequently seen along rural roads in central Pennsylvania.

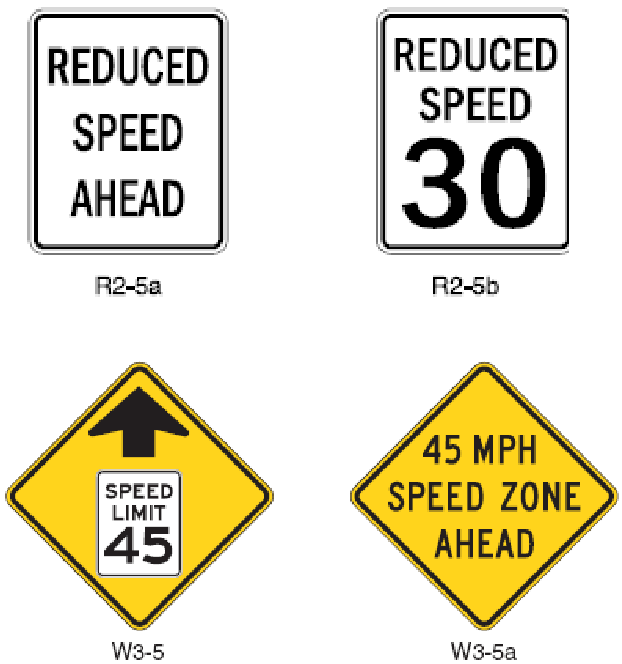

Figure 1 Evolution of Reduce Speed Ahead Sign

Since speed changes should not be abrupt, drivers are warned of speed changes in advance. The Pennsylvania Department of Transportation's (PennDOT) Publication 212 "Official Traffic Control Devices" (2006) indicates that a "Reduced Speed Ahead" or "Speed Reduction" sign must be installed between 500 and 1,000 feet in advance of a speed reduction unless the speed reduction is 10 miles per hour or less.

1.2 Statement of Problem

Rural highways do not serve a vast majority of trips; they often serve traffic volumes less than 100 vehicles per day (McShane, 1998). However, fatal crashes are over-represented on rural highways in the U. S.; it has been estimated that approximately 60 percent of the more than 40,000 annual vehicle-related fatal accidents occurring in the U.S. take place on rural highways (FHWA, 2008). Evans (1991) compared these fatalities by type and functional classification of roads. His research indicated that if all rural and urban non-Interstates had the same fatality rate as the Interstate system, then a 50 percent reduction in fatalities could be achieved. Evans concluded that these statistics demonstrate the influence that roadway characteristics have on traffic safety. Therefore, it has been recommended that highways should be designed in a consistent manner to ensure that driver expectancy is not violated. The Fatal Accident Reporting System (FARS) indicates that nearly 15 percent of fatal crashes in 2005 were attributed to drivers traveling in excess of the posted speed limit (FARS, 2005).

The American Association of State Highway and Transportation Officials' (AASHTO) Policy on Geometric Design of Highways and Streets (2004), commonly referred to as the Green Book, contains a collection of design controls and criteria for all functional classes of highways and streets. The Green Book design criteria intend to provide consistency among design practices nationwide.

Design speed is one of the primary design controls that influence highway design. The design speed is defined as "a selected speed used to determine the various geometric design features of the roadway (AASHTO, 2004)." In highway design, it is desirable to use only a single design speed along a corridor with the anticipation that uniform, consistent operating speeds will result. In the case of transition zones, however, a change in operating speed is required to be in compliance with the associated regulatory speed change, sometimes resulting in speed discord or inconsistencies, particularly in the low-speed operating environment. At the same time, the change in driving environment along transition zones may be accompanied by a change in the roadway or roadside design features. For example, the undeveloped rural area with a clear roadside at the high-speed end of a transition zone may suddenly transform into a developed area with sidewalks, curbs, and a high density of driveways at the low-speed end of a transition zone. While design guidelines are available for both the high- and low-speed environments at either end of a transition zone, there are neither existing guidelines that provide designers with guidelines to link these environments nor are there design guidelines that have been shown to effectively reduce speeds in transition zones.

Safety concerns can arise when drivers fail to appropriately adjust their speeds in transition zones. Since the driving environment changes from high-to-low speed, roadway design features along transition zones represent a challenge to the engineering profession. Furthermore, the low-speed environment presents possible safety concerns due to the presence of pedestrian activity and the increase in turning traffic (TRB, 2007). A recent study sponsored by PennDOT explored the effectiveness of dynamic speed display signs (DSDS) in reducing vehicle operating speeds along 12 two-lane rural highway transition zone sites in central Pennsylvania (Donnell and Cruzado, 2007). The DSDS devices were located 500 feet after the end of the transition zone and speed data were collected before, during, and after implementation of the DSDS. The before data indicated that drivers fail to adjust their speeds along the transition zone; mean operating speeds were 1.4 to 13.9 mph higher than the speed limit at the DSDS location while 85th percentile speeds were 7 to 20 mph higher than the posted speed limit. During DSDS implementation, both mean speeds and 85th percentile speeds next to the DSDS were lower by an average of 6 and 7 mph, respectively. However, after the DSDS was removed, speeds increased to levels similar to the before data collection period suggesting that DSDS were only effective in reducing speeds along transition zones while in place and activated.

Several geometric variables can influence driver behavior as reflected in past research studies (Yagar and Van Aerde, 1983; Poe and Mason, 2000). Therefore, identifying which geometric design elements are associated with operating speeds along transition zones can be the first step in the development of transition zones design guidelines.

1.3 Importance of Research to Engineering

The Transportation Research Board's Committee on Geometric Design (AFB10) and Operational Effects of Geometrics (AHB65) published a strategic research needs document to outline a program to advance geometric design into the 21st century (TRB, 2007). One of the 22 high-priority research needs identified in this long-range plan was to develop design guidelines for high-to-low speed transition zones. The objective of such a research project is to develop treatments and procedures to design high-to-low speed transitions in rural areas. It was recommended that changes in the alignment, vertical profile, and roadway and roadside cross-section be considered as methods to slow vehicle speeds in transition zones. A first step in this process is to estimate speed prediction models along rural highway transition zones to determine the roadway, roadside, and land use characteristics that are associated with driver operating speeds in these areas.

1.4 Research Objectives

Design guidelines are currently not available for the design of transition zones on two-lane rural highways. The development of design criteria for transition zones may produce more uniformity in the roadway and roadside features encountered by motorists along these highway segments. Past research studies have indicated that geometric design, roadside, and land use features influence driver speed choice (Yagar and Van Aerde, 1983; Poe and Mason, 2000; Figueroa and Tarko, 2005), thus changes in these features may influence vehicle operating speeds when high-speed rural highways pass through rural communities. By identifying the highway features that are associated with speed reductions along transition zones, a contribution can be made to the development of design guidelines for high- to low-speed highway sections. As such, the scope of this research is to identify the roadway, roadside, and land use characteristics that are associated with reductions in operating speeds along two-lane rural highway transition zones. Point speed and speed differential models are estimated using a variety of longitudinal and hierarchical modeling methods.

In past operating speed modeling literature, most models have been developed using ordinary least squares regression. Although linear regression models were specified in this research, other analysis methods were also explored and compared in an effort to determine if these alternative methods provide advantages over conventional operating speed modeling methods. The specification of alternative speed prediction models may be helpful in overcoming the limitations of the ordinary least squares regression model in modeling vehicle operating speeds in transition zones.

1.5 Organization of Dissertation

This dissertation is divided into five subsequent chapters. The second chapter discusses previous research studies that are related to the present study and have helped shape the proposed research. Specifically, those studies that have estimated speed prediction models as a function of the roadway environment are critically synthesized for both high-speed, two-lane rural highways and low-speed urban streets. The third chapter describes the site selection process and data collection methods. The fourth chapter discusses the analysis methods used in this research. The results of the analyses and the conclusions from this research are discussed in the fifth and the sixth chapters, respectively.

CHAPTER 2
LITERATURE REVIEW

Rural highways do not serve a vast majority of vehicle trips and often have traffic volumes less than 100 vehicles per day (McShane, 1998). However, approximately 77 percent of publicly-owned highways in the U.S. are classified as rural (FHWA, 2004). More than 50 percent of fatal crashes in the U.S. occur on two-lane rural highways (NHTSA, 2006). Because fatal crashes are overrepresented on two-lane rural roads in the U.S., these roadway types were considered the highest priority research need by the Transportation Research Board's Committee on Geometric Design (Choueiri, et al., 1994). To address this need, the first version of the Federal Highway Administration's (FHWA) Interactive Highway Safety Design Model (IHSDM) contains safety prediction and design consistency modules that can be used to assess the safety and operational performance of current and planned two-lane rural highways (Krammes and Hayden, 2003).

Published literature related to speed prediction along rural highway transition zones between high- and low-speed operating environments is limited. As such, this literature review focuses primarily on speed prediction models that were developed exclusively for both high- and low-speed operating environments. High-speed roadways are considered those with a design speed of 50 mph or greater while low-speed roadways are considered those with a design speed of 45 mph or less (AASHTO, 2004). Much of the high-speed operating speed literature is focused on two-lane rural highways and some of this literature serves as the basis for the IHSDM design consistency module. Most of the low-speed operating speed literature relates to low-speed urban streets. In all cases, speed prediction literature that contains roadway, roadside, and land use characteristics are synthesized in this section of the dissertation.

2.1 High-Speed Rural Highways

Design speed is a fundamental criterion in roadway design as it is used to establish the geometric design features of a highway (AASHTO, 2004). The design speed concept is intended to ensure geometric design consistency. Several operating speed studies have been published on two-lane rural highways that specifically address the relationship between the design speed and operating speed that result from the design process. Operating speeds should be in harmony with the roadway's design speed; discrepancies between design and operating speeds are evidence of a lack of design consistency.

Differences between design and operating speeds led McLean (1979) to develop an alternative concept to the design speed. His research indicated that roadways with design speeds of 70 mph (110 km/hr) or greater had operating speeds that were in accordance with the design speed concept (i.e. operating speeds were uniform and lower than the design speed). McLean showed that operating speeds along horizontal curves on roadways with posted speed limits between 55 and 70 mph (90 and 110 km/hr) were lower than the design speed. On roadways with posted speed limits below 55 mph (90 km/hr), operating speeds exceeded the design speed on horizontal curves. McLean introduced a new concept which indicated that desired operating speeds can be related to the roadway's terrain classification and alignment.

McLean's study considered speed data from 230 sites on two-lane rural highways in Australia, collected on both horizontal curves and the upstream approach tangent. The term "desired speed" was used to identify the speed under free-flow conditions when drivers are not constrained by alignment features, represented by the speed along tangent sections. The data collected indicated that this desired speed was influenced by road function, trip purpose and length, proximity to urban centers, overall design speed, and terrain type. For horizontal curves with design speeds of 60 mph (100 km/hr) and above, results showed that 85th percentile speeds tend to be less than the design speed of a horizontal curve; however, the reverse is true along horizontal curves with lower design speeds. It was determined that available sight distance was correlated with 85th percentile operating speeds, but explained less than one percent of the variability in a statistical model. As such, it was not included in the model specified below:

$$V_C(85) = 53.8 + 0.464V_F - 3.26\left(\frac{1}{R}\right)\times 10^3 + 8.5\left(\frac{1}{R}\right)^2 \times 10^4 \tag{1}$$

where: $V_C(85)$ = 85th percentile curve speed (km/hr);

V_F = desired speed of the 85th percentile car (km/hr); and

R = curve radius (m).

The parameters included in equation (1) were statistically significant at the 99 percent confidence level. The coefficient of determination (R^2) was 0.92. McLean concluded that the horizontal alignment influences vehicle operating speeds on two-lane rural highways.

McLean also indicated that in order to achieve a design that meets driver expectancies, horizontal curves should be designed in a way that will generate speeds which do not differ by more than 5 mph (10 km/hr) along the entire alignment. It was also recommended in the study that changing the speed environment by providing a sequence of carefully designed horizontal curves with each having a predicted speed that is consistent with design guidelines can also promote design consistency. McLean indicated that "when going from a high- to a low-standard, the predicted speed on sequential curves should not differ by more than 10 km/hr (5 mph)."

Yagar and Van Aerde (1983) studied 10 different environmental and geometric design features that were thought to influence operating speeds along two-lane rural highways at 35 locations in Ontario, Canada. The authors theorized that speeds were dependent upon upstream design features along a highway rather than the instantaneous geometric features of the roadway. A speed prediction model was developed using a multiple linear regression model. Five variables were statistically significant in the model. These included: vertical grade, lane width, land use, access, and the posted speed limit. The speed prediction model developed was:

$$Y = 93.3 - 1.8G - 5.7LW - 8.3LU - 8AC - 0.7SL \tag{2}$$

where: Y = mean speed (km/hr);

G = grade (percent);

LW = lane width (m);

LU = land use factor which is set if the adjoining land has access driveways; it represents the fraction of highway on which land use was present upstream (decimal value);

AC = access from other roads; weight value ranging from 0 (no access by any roads) to 4 (controlled intersection); and

SL = posted speed limit (km/hr).

The model explained 85 percent of the variability ($R^2 = 0.85$) in the observed speed data. The radius of curve, presence of an auxiliary lane, available sight distance, and presence of a centerline were not statistically significant in the model and were therefore not included. A variable that represented the distance to lateral obstructions was statistically significant, but it was not included in the final model because its effect was in the opposite direction of what was expected. The results of this study, especially the rejection of curvature as potential predictor variable, are not consistent with the majority of operating speed studies (Andjus and Maletin, 1998; Lamm, et al.; 2002; and Richl and Sayed, 2005).

Andjus and Maletin (1998) studied operating speeds on horizontal curves along two-lane rural highways in Yugoslavia. It was recognized that the main concern in road design is drivers' response to the geometric features present along a roadway. Among all speed parameters considered, the 85th percentile speed from free-flow passenger cars was identified as the speed parameter that best represented driver response to the roadway geometry, particularly along horizontal curves. A total of nine sites were selected for the study. Study sites included horizontal curves with radii ranging from 165 to 2460 feet (50 to 750 meters). To isolate the influence of roadway cross-section elements, sites with speed limit signs and intersections were excluded as well as sites with grades steeper than 4 percent. Speed data from free-flow passenger vehicles, identified as those with time headways greater than 7 seconds, were collected during daylight and dry pavement conditions. Regression models were specified to determine the relationship between horizontal curve radius (*R)* and 50th and 85th percentile operating speeds (V_{50} and V_{85}, respectively). The resulting models are shown in equations (3) and (4) below:

$$V_{50} = 16.92 \ln R - 14.49 \qquad (R^2 = 0.975) \qquad (3)$$

$$V_{85} = 14.75 \ln R - 11.69 \qquad (R^2 = 0.969) \qquad (4)$$

Although horizontal curve radius was the only variable included in the speed prediction model, the authors indicated that there are other factors that influence operating speeds. The authors suggested in their study that vehicle type and driver characteristics should be included in speed prediction models.

Polus, et al. (2000) developed speed prediction models on tangent sections of two-lane rural highways with low volumes in order to determine which geometric design features explain the variability in vehicle operating speeds. A database from 6 states (Minnesota, New York, Pennsylvania, Oregon, Washington, and Texas) with traffic volumes less than 2,000 vehicles per day included speed data from free-flow vehicles (time headway of at least 5 seconds) collected during off-peak hours and during daylight and dry pavement conditions. Speed limits were between 45 and 70 mph (75 and 110 km/hr).

Initially, a single model to predict speed on tangents was developed, which was termed a “geometry measure model.” These models had a low coefficient of determination (R^2), so a family of models was considered in order to obtain better speed predictions; these models were termed “group models.” The primary variables considered in the analysis included tangent length, posted speed limit, enforcement level, curvature before and after the tangent, vehicle deceleration and acceleration characteristics, grade or general terrain, roadway width, roadside slopes, and presence of spiral curves. Secondary variables considered in the analysis were those related to driver workload and speed-choice decisions. Initially it was concluded that, along short tangents, operating speeds are influenced by the geometry of the preceding and succeeding curves; additional factors, such as the posted speed limit and enforcement level appeared to influence operating speeds on long tangent sections.

The database was grouped according to several combinations of tangent length (small, intermediate, and short) and radii (small, intermediate, and reasonable) and different models were developed for these combinations. The resulting regression models are shown in Table 1.

Table 1 Models Developed by Polus, et al. (2000) for Several Radius and Tangent Combinations

Radii (R_1, R_2)	Tangent	Model	R^2	Additional Comments
Less or equal than 250m	Less than 150 m	$SP = 101.11 - \frac{3240}{GM_S}$	0.553	GM_S = geometric measure equal to the average of the radii of previous and following curve (m)
Less or equal than 250 m	Between 150 m and 1000 m	$SP = 94.405 - \frac{3184}{GM_L}$	0.684	$GM_L = \frac{TL \times (R_1 \times R_2)^{1/2}}{100}$
Less or equal than 250 m	Between 150 m and 1000 m	$SP = 105 - \frac{28.107}{e^{(0.00108 GM_L)}}$	0.742	To be used when the maximum 85th percentile speed is established as 65 mph (105 km/hr)
Any reasonable radius	Greater than 1000 m	$SP = 105 - \frac{22.953}{e^{(0.00012 GM_L)}}$	0.838	Radius of horizontal curve does not violate the criterion for design speed
Legend: *SP* = 85th percentile speed (km/hr) *TL* = tangent length (m) R_1, R_2 = previous and following curve radii (m)				

Ottesen and Krammes (2000) evaluated different types of regression models for predicting 85th percentile speed on approach tangents and at the midpoint of a horizontal curve. Data from 138 curves and 78 approach tangents on 29 two-lane rural highways in 5 states were analyzed. Design speeds ranged between 25 to 60 mph (30 and 95 km/hr) and grades were less than 5 percent. The authors evaluated 4 different regression model forms: linear, exponential, inverse, and polynomial. A prediction model for speed on the approach tangent was not successfully developed. For speeds at the midpoint of a horizontal curve, the results of the analyses showed that all regression types had similar values for the coefficient of determination, R^2, ranging from 0.80 to 0.82. Therefore, the authors chose to recommend the following linear regression because of its simplicity and practicality:

$$V85 = 41.62 - 1.29D + 0.0049L - 0.12DL + 0.95V85T \quad (5)$$

where: *V85* = 85th percentile speed at midpoint of curve;

D = degree of curvature, degrees;

L = length of curve; and

V85T = speed of approach tangent.

The model shown in Equation (5) had a coefficient of determination of 0.90. The goodness-of-fit for the model with only the degree of curve as an explanatory variable was 0.80. Adding length of curve and its interaction with degree of curve only increased the R^2 value to 0.81. The authors concluded that a model with only degree of curve is the most appropriate and that Equation (5) is "only useful if approach tangent speeds are actually measured." The authors also concluded that when the degree of curve is less than 4, the operating speeds on the curve are the same as those on long tangents.

Schurr, et al. (2002) studied the relationship between design, operating, and posted speeds along horizontal curves on two-lane rural highways in Nebraska. Various geometric design elements were considered, including length of curve, deflection angle, radius of curve, and superelevation. The designated design speed and posted speed limit for the study sections were also considered in the analysis. In order to isolate the influence of geometric design features on operating speeds, only sites with fair or better pavement surface conditions were considered. Two sensors were placed at each study site; the first was placed along the approach tangent, at least 600 feet (180 m) before the PC, and the second sensor was placed at the midpoint of the horizontal curve. A time headway of at least five seconds was used to identify free-flow vehicles. Only passenger cars during daylight and dry pavement conditions were included in the analyses.

The models developed considered the following operating speeds as dependent variables: mean, 85th percentile, and 95th percentile. The independent variables considered in the analysis were radius of curve, length of curve, length of approach tangent, intersection angle, direction of curve, superelevation, design speed, posted speed, average daily traffic (ADT), roadway width, shoulder width, surfaced shoulder width, percent heavy vehicles, approach grade, departure grade, length of vertical curve, and rate of change of vertical curve.

The results of the analyses showed that the statistically significant variables influencing mean speeds at the midpoint of the curve were intersection angle, length of curve, and posted speed limit. The independent variables significantly influencing 85th percentile operating speeds were approach grade, intersection angle, and length of curve. Finally, the variables identified as significant in the 95th percentile operating speed model were intersection angle, length of curve, and ADT. Table 2 shows the regression equations obtained for each of the response variables considered along with their respective coefficients of determination (R^2).

Table 2 Speed Prediction Models (Schurr, et al., 2002)

Response Variable	Regression Equation	R^2
Mean Speed (km/hr)	$67.4-0.1126\Delta+0.02243L+0.276V_p$	0.55
85th Percentile Speed (km/hr)	$103.3-0.1253\Delta+0.0238L-1.039G$	0.46
95th Percentile Speed (km/hr)	$113.9-0.122\Delta+0.0178L-0.00184ADT$	0.41
Legend: Δ = deflection angle (decimal degrees) L = length of curve (m) V_p = posted speed limit (km/hr) G = approaching grade (percent) ADT = average daily traffic (vpd)		

The design speed, which was inferred from the geometric elements of the roadway, was less than the 95th percentile operating speeds at 17 of the 40 sites considered in the study. This led to the conclusion that 95th percentile operating speeds are somewhat constant when design speed is not considered and that “drivers determine their desired speed on the basis of what they perceive to be reasonable for certain roadway types.”

The study performed by Schurr, et al. (2002) also included binomial proportion tests for comparisons between predicted 85th percentile operating speeds and observed 85th percentile operating speeds at horizontal curve midpoint locations. The results indicated that there is a poor fit between these two parameters. The speeds from curve and tangent sections were compared using paired t-tests. The results showed that few sites had statistically similar operating speed parameters between the two locations at the 95 percent confidence level. This was true for locations with speed limits of 55 and 60 mph (90 and 100 km/hr), thus the authors concluded that drivers choose more uniform speeds at locations where the speed limit is 65 mph (105 km/hr). The authors indicated that this could be attributed to the fact that sharper curves are located on roadways with speed limits of 55 and 60 mph (90 and 100 km/hr). The only factors found to significantly influence operating speed on tangent locations were posted speed (for the three speed parameters) and ADT (for 85th and 95th percentile operating speeds). It was then concluded that large changes in direction cause drivers to slow their speeds; long curves cause drivers to increase their speeds since they have more time to adjust the vehicle to the radius of the curve. The authors recommended that speed models should include the posted speed limit as an explanatory variable.

Lamm, et al. (2002) focused on the parameters influencing the frequency and location of accident clusters by investigating reports from United States, Germany, Greece, and Italy. Since previous studies suggested that abrupt changes in operating speeds, mainly caused by changes in horizontal alignment, are the leading cause of accidents on two-lane rural roads, the authors explored highway geometric design features that influence the consistency of operating speeds. Three safety criteria for two-lane rural highways were used to analyze highway safety; the first two safety criteria were related to speed differentials. The safety criteria were:

1. The difference between design speed and driving behavior. This was defined as variations in observed 85th percentile speeds which are indicators of design consistency.
2. The difference between observed 85th percentile speeds on successive design elements.
3. The difference between side-friction assumed and side-friction demanded for design at 85th percentile speeds on curves.

The goals of the safety criteria were: (1) to select a design speed that it is constant throughout the entire roadway for design consistency and which should be represented by the 85th percentile operating speeds, (2) to achieve constant 85th percentile operating speeds, and (3) to obtain a well-balanced driving dynamic sequence of individual design elements. The parameters considered to evaluate the effects on traffic safety were: curvature change rate of a single curve, length of curve, superelevation rate, lane width, shoulder width, sight distance, percent vertical grade, and traffic volume. Operating speed data were collected on both tangent and curve sections. The study concluded that curvature change rate (CCR_S) was the most successful parameter in explaining most of the variability in operating speeds as well as accident rates. All other parameters were not statistically significant at the 95 percent confidence level.

The authors also developed equations for predicting 85th percentile operating speeds along horizontal curves for two ranges of vertical grades: one equation was developed for roadway sections with vertical grades equal to or less than 6 percent and a second equation for roadway sections with vertical grades greater than 6 percent. The only parameter included in these equations was curvature change rate, CCR_S. The equations developed for these two criteria, along with the coefficients of determination R^2, are shown in Table 3.

Table 3 85th Percentile Speed Prediction Models (Lamm, et al., 2002)

Grade	Equation	R^2
≤ 6%	$V85 = 105.31 + 2 \cdot 10^{-5} \cdot CCR_S^2 - 0.071 \cdot CCR_S$	0.98
> 6%	$V85 = 86 - 3.241^{-9} \cdot CCR_S^3 + 1.61 \cdot 10^{-5} \cdot CCR_S^2 - 4.2610^{-2} \cdot CCR_S$	0.88

Figueroa and Tarko (2005) developed speed prediction models on two-lane rural roadways in Indiana to determine which geometric elements influence vehicle operating speed. The study recognized the difference between the mean speed and speed dispersion factors, justifying the need for developing a speed prediction model that included both. Data were collected at 158 sites during daylight hours under favorable weather conditions. Only free-flow vehicles were considered for the study. Two speed prediction models were developed using ordinary least squares regression: (1) operating speeds along tangent sections and (2) operating speeds along horizontal curves. In the tangent model, the speed limit binary variable explained the greatest amount of variability in the mean speed and speed variance models. Other factors that were included in this model were available sight distance, cross-section dimensions, presence of intersections, truck percentage, and vertical grades. Equation (6) shows the regression model used to estimate operating speeds on tangent roadway sections:

$$\begin{aligned} V_P &= 57.137 - 0.071TR - 3.082PSL_{50} - 0.131GRA - 1.034RES \\ &+ 2.38\times10^{-3}SD - 1.67\times10^{-6}SD^2 - 0.422INT + 0.04PAV \\ &+ 0.394GSW + 0.054USW - 2.233FC + 5.982Z_p + 1.428(Z_pPSL_{50}) \\ &+ 0.061(Z_PGRA) + 0.292(Z_PINT) - 0.038(Z_PPAV) - 0.012(Z_PCLR) \end{aligned} \tag{6}$$

where: V_P = operating speed corresponding to a percentile P (mph);

TR = percentage of trucks (percent);

PSL_{50} = equal to 1 if posted speed limit is 50 mph (80 km/hr); equal to 0 if posted speed limit is 55 mph (90 km/hr);

GRA = highway grade (percent);

RES = equal to 1 if segment has 10 or more residential driveways per mile; 0 otherwise;

SD = sight distance (ft);

INT = equal to 1 if an intersection is located 350 ft (110 m) before or after the spot; 0 otherwise;

PAV = pavement width (ft);

GSW = total gravel shoulder width (ft);

USW = total untreated shoulder width (ft);

CLR = clearance distance including total width of shoulder regardless of type (ft);

FC = equal to 1 if the spot is located on a curve with a radius of 1700 feet (520 m) or more; 0 otherwise; and

Z_p = standardized normal variable corresponding to a selected percentile.

The model developed to predict operating speeds along horizontal curves included the following four explanatory variables: available sight distance, degree of curve, maximum superelevation rate, and presence of residential driveways. Equation (7) shows the regression model used to estimate operating speeds on a horizontal curve:

$$V_p = 47.664 + 3.44 \times 10^{-3} SD - 2.693 RES - 2.541 DC + 7.954 SE$$
$$-0.624 SE^2 + 4.158 Z_p + 0.236 (Z_P DC) - 0.199 (Z_P SE) \qquad (7)$$

where: *DC* = degree of curvature (degrees); and

SE = maximum superelevation rate (percent).

The models for predicting operating speeds along tangent sections and horizontal curves had coefficients of determination, R^2, of 0.844 and 0.932, respectively. The study performed by Figueroa and Tarko (2005) demonstrated that cross-section variables, such as pavement width and lateral clearance distance, influence operating speeds along tangent sections; an increase in any of the lateral dimensions of the cross-section is associated with an increase in operating speeds.

Design consistency is primarily evaluated by calculating the speed differences between tangent and curve sections, thus radius of curve is generally the only variable included in speed prediction models. Recognizing that drivers perceive horizontal curves differently when combined with vertical curves, Richl and Sayed (2005) evaluated 12 already developed speed prediction models in order to incorporate the effects of changes in vertical alignment. The speed prediction models were evaluated with speed data from two sites: an existing alignment with a posted speed limit of 50 mph (80 km/hr) and advisory speed limit signs at some horizontal curve locations ranging from 30 to 45 mph (50 to 70 km/hr), and a proposed alignment with design speeds between 55 and 50 mph (90 and 80 km/hr).

The authors hypothesized that the combination of vertical and horizontal curves may create an optical illusion causing drivers to perceive the radius of a curve differently from its actual radius. Using linear regression, an equation was then developed that explains the relationship between perceived radius (dependent variable) and actual radius, vertical curve, and the combination of both (independent variables). The model developed was:

$$R_P = -51.28 + 0.953R_A + 132.11V + 0.125R_A V \tag{8}$$

where: R_P = perceived radius (m);

R_A = actual radius (m); and

V = indicator variable for vertical crest, equal to 1 for crest vertical curves, 0 for sag vertical curves.

The coefficient of determination (R^2) for the model shown in Equation (8) is 0.996. The speed prediction models were then evaluated for both actual and perceived radius of horizontal curve. The results showed that the majority of the speed prediction models provide similar speed values among each other. Using the value of perceived radius instead of the actual radius resulted in an increase in speed variability on both alignments, the greatest speed differential being the combination of a sharp horizontal curve and a short crest vertical curve. The authors recommended using perceived radius for design consistency evaluation.

Highway designers use the design speed to determine the geometric elements of a roadway. They assume that the design speed will be equal to or exceed the posted speed limit, and that the posted speed should be equal to or greater than the 85th percentile operating speed. Speed harmony or consistency is achieved when the design, operating, and posted speeds are compatible. Achieving operating speed consistency is desired since a "consistent roadway design should ensure that most drivers would be able to operate safely at their desired speed along the entire alignment (Schurr, et al., 2005)." Many studies have focused on speed differentials as a measure of design consistency.

Research conducted by McFadden and Elefteriadou (2000) assessed the implication of using the 85^{th} percentile operating speed for evaluation of design consistency. The research considered speed data from at least 75 vehicles at 21 sites in Pennsylvania (12 sites) and Texas (9 sites). The criteria for site selection included rural highways, in level to rolling terrain, with design and posted speeds of less than 70 mph (110 km/hr), and low-traffic volumes (500 – 4,000 vpd). In order to isolate the effects of horizontal curvature on operating speeds, approach tangents were limited to a minimum of 200 meters (656 ft) and the vertical alignment was limited to an absolute grade of 5 percent. Data collection consisted of information on alignment geometry, cross-section, weather, traffic control devices, light conditions, and terrain and environment. Speed data were collected using a lidar gun, starting 200 m (656 ft) before a horizontal curve and continuing 200 m (656 ft) after the curve. The data only included passenger cars considered to be free-flow vehicles determined using a minimum time headway of five seconds.

Prediction models were developed to estimate 85^{th} percentile speed reduction due to the introduction of a change in alignment (i.e. horizontal curve). Scatter plots and correlation analyses were used to determine if there was a relationship between the speed reduction and the geometric design features of the roadway. The results indicated that there is a statistically significant relationship between speed reduction and the length of approach tangent, radius of curve, deflection angle, pavement width, shoulder width, and posted speed limit. The OLS regression models developed by McFadden and Elefteriadou (2000) are shown in Table 4.

Table 4 85th Percentile Speed Reduction Models Due to Introduction of a Horizontal Curve (McFadden and Elefteriadou, 2000)

Model #	Model	Adj. R^2
1	$V_{85redux}=-14.9+0.144V_{85PC200}+0.0153LAPT+(954.55/R)$	0.712
2	$V_{85redux}=-0.812+(998.19/R)+0.017LAPT$	0.603

Legend:
$V_{85redux}$ = estimated 85th percentile speed reduction (km/hr)
$V_{85PC200}$ = 85th percentile speed 200 meter prior to point of curvature (km/hr)
LAPT = length of approaching tangent (m)
R = radius of curve (m)

The authors concluded that using operating speed profile models at point locations to evaluate design consistency underestimates the actual speed reduction of drivers along a tangent-curve combination. Rather, the use of a single 85th percentile speed reduction measure as a design consistency tool contains more detailed information about driver performance when approaching horizontal curves. The authors also concluded that using only the midpoint location on the approach tangent and midpoint location of the horizontal curve to compute speed reductions does not capture actual minimum and maximum operating speeds and, therefore, collecting operating speed data at several locations approaching and within horizontal curves should be used to determine the speed reduction of drivers.

A study by Park and Saccomanno (2005) considered the difference in 85th percentile speeds between successive highway elements in order to evaluate design consistency. The authors addressed the issue of using aggregate data ("ecological fallacy") from a speed distribution to model operating speeds. The authors recommend use of disaggregate data to model vehicle operating speeds.

Normally, the 85th percentile speed differential, ($\Delta V85$) is calculated as the difference between the 85th percentile speed at a point on the approach tangent and the 85th percentile speed at the midpoint of a horizontal curve (i.e., the difference between point 85th percentile speeds on two successive elements). The authors hypothesized that a better approach is the use of disaggregate data, and that 85th percentile speed differentials should be the 85th percentile of speed differences of individual drivers (the 85th percentile of individual speeds differentials). The authors used linear regression to specify models of operating speed using both aggregate and disaggregate data in order to address this issue. Data from 18 tangent-curve combinations on two-lane rural highway sections were considered for this part of the analysis. When using the speed at the midpoint of the following horizontal curve as the dependent variable, the speed on the approach tangent was not statistically significant when using the aggregate data but was found to be statistically significant when using the disaggregate data. In addition, radius of curve had a higher z-statistic in the disaggregate model, indicating a stronger relationship with operating speeds. Despite these results, the aggregate data model had a higher coefficient of determination, R^2, than the disaggregate model (0.638 vs. 0.275), suggesting that the model using aggregate-level data explained a larger proportion of the variability in operating speeds. The authors concluded that this is evidence that "the presence of summary measures in aggregate data introduces a major source of uncertainty." Additionally, use of aggregate data inflated the coefficient-of-determination and the regression parameter for the radius of curve variable that was included as an explanatory variable in the model.

The authors also specified a multilevel model using the disaggregate data, inferring that this type of model is appropriate for correlated observations. A two-level model was developed: the first level included information about individual vehicle speeds, such as speed on the previous section (tangent), and the second level included the geometric features of the highway segment. The only variable found to be statistically significant in the second level was radius of curve. The results of the two-level model are shown in Table 5.

Table 5 Two-level Model developed by Park and Saccomanno (2005)

Parameter	Estimate	St. Error	Z-value
First level			
Speed at tangent section	0.328	0.023	14.176
Within-group Variance, σ^2	41.023	2.644	15.516
Second level			
1/R	1038.046	241.865	4.292
Between-group Variance, τ_{00}	1.294	0.596	2.173
Fixed Effect Intercept	-18.44	1.742	-10.585
First level R^2	0.242		
Second level R^2	0.755		
Overall R^2	0.283		

The results of the analysis indicated that 75 percent of the variability in the second level is explained by the curvature of the roadway section. Similarly, 24.2 percent of the within section variation was explained by the first level predictor (i.e. approach tangent speed). The authors concluded that the speed differentials of individual vehicles are mostly associated with first level characteristics rather than second level characteristics. The analysis results also indicated that drivers along sharp curves experienced larger speed differentials when compared to mild curves. Lastly, individual driver speed differentials were positively associated with approach speeds, suggesting that faster drivers decrease their speed more so than slower drivers to negotiate a horizontal curve.

Misaghi and Hassan (2005) specified models for both the 85th percentile operating speed at the midpoint of a horizontal curve and the 85th percentile speed differential between the approach tangent and midpoint of a horizontal curve along two-lane rural roads in Canada. Similar to Park and Saccomanno (2005), the 85th-percentile speed difference was calculated based on individual vehicles, thus the authors considered disaggregate data. The objective of the research was to evaluate design consistency by exploring the speed differentials between successive highway elements, specifically from tangent to curve. Speed data were collected at 20 curves, in both directions, along two-lane rural highways with the use of a radar gun. Horizontal curve characteristics (radius, length, etc) varied, but other roadway characteristics that could influence drivers were constant across study sites, such as lane width, traffic signals, and nearby intersections.

Data were then reduced in order to only consider data from free-flow passenger vehicles, during daylight and dry-pavement conditions. In addition, 5 sites were excluded due to the low number of speed observations collected (less than 100).

The predictors considered in the models included the geometric characteristics at each study site. The only variable found to influence operating speeds at the midpoint of a horizontal curve was the radius as shown in Equations (9) and (10) below:

$$V_{85MC} = 91.85 + 9.81 \times 10^{-3} R \tag{9}$$

and

$$V_{85MC} = 94.3 + 8.67 \times 10^{-6} R^2 \tag{10}$$

where: V_{85MC} = 85th percentile speed at middle of curve; and

R = radius of curve (m).

The models shown in Equations (9) and (10) had coefficients of determination of 0.464 and 0.524, respectively. Two additional models were developed to explain the association between geometric characteristics and speed differentials from tangent to curve. The first speed differential model considered data from the 35 sites, while the second model excluded data from nine sites: three sites were excluded because they were considered potential outliers and six sites were excluded for the purpose of model validation. The two models developed are shown in Equations (11) and (12) below:

$$\Delta_{85}V = -83.63 + 0.93V_T + e^{-8.93+3507.1/R} \tag{11}$$

and

$$\begin{aligned} \Delta_{85}V = &-198.74 + 21.42\sqrt{V_T} + 0.11DFC - 4.55SW - 5.36 curve.dir \\ &+ 1.3G + 4.22 drv.flag \end{aligned} \tag{12}$$

where: $\Delta_{85}V$ = 85th percentile speed differential (km/hr);

VT = approach tangent speed (km/hr);

DFC = deflection angle of circular curve (degrees);

SW = shoulder width (m);

curve.dir = indicator variable for direction of curve (1 if right, 0 otherwise); and

drv.flag = driveway flag (1 if intersection on curve, 0 otherwise).

The values for the coefficients of determination, R^2, for Equations (11) and (12) are 0.64 and 0.89, respectively. The authors inferred that, compared to other studies, the relationship between speeds and radius of curve was considered "weak"; they suggested that the use of a radar gun to collect data may cause drivers to slow down due to perceived law enforcement.

Most of the speed prediction models for two-lane rural highways were developed using OLS linear regression; only one study – Park and Saccomanno (2005) – considered multilevel models. Changes in horizontal alignment were related to changes in operating speeds, thus the majority of the equations developed in these studies considered speed along the horizontal curve as the dependent variable; only three studies evaluated prediction models for speeds along tangents (Polus, et al., 2000; Figueroa and Tarko, 2005; and Misaghi and Hassan, 2005). Similarly, only three studies estimated statistical models to predict speed differences due to changes in horizontal alignment (Mc Fadden and Elefteriadou, 2000; Park and Saccomanno, 2005; and Misaghi and Hassan, 2005).

The presence and radius of a horizontal curve is considered the most significant geometric feature influencing operating speeds, therefore the elements of curves, such as deflection angle, radius, and intersection angle, among others, were always found to significantly influence speed parameters (the dependent variable). Radius of curve was sometimes found to be the only significant factor in the models developed (McLean, 1979; Andjus and Maletin, 1998; and Misaghi and Hassan, 2005).

Only one study identified posted speed limit as a factor influencing operating speeds (Schurr, et al., 2002), however the inclusion of speed limit as an explanatory variable has been questioned since the roadway design elements are selected based on speed-related parameters (Wang, et al., 2006). Only two studies identified the presence of roadside geometrics as significant factors influencing operating speeds (Figueroa and Tarko, 2005; and Misaghi and Hassan, 2005). In the study by Figueroa and Tarko (2005), highway grade and driveway density were associated with speed reductions while pavement and unpaved shoulder widths were associated with an increase in operating speeds. The variables of highway density and shoulder width were also found to have the same effect on speed differentials in the study by Misaghi and Hassan (2005). The use of aggregate data is also means for concern since it "introduces a major source of uncertainty"; only two studies (Park and Saccomanno, 2005; and Misaghi and Hassan, 2005) considered disaggregate data for the models developed.

2.2 Low-Speed Urban Streets

Most of the studies along two-lane rural highways have been performed at high-speed locations with posted speed limits of at least 55 mph (90 km/hr) and the literature for this roadway type is vast. There is also a significant body of published literature related to operating speeds on low-speed urban streets.

Poe and Mason (2000) investigated the influence of geometric features on operating speeds at 27 sites located along urban streets in Pennsylvania. The geometric features at the data collection sites varied. The horizontal curve radius ranged from 36 to 679 ft (11 to 207 m) and grades varied from 8.7 to -16.3 percent. The authors inferred that on low-speed highways the geometric features that are associated with operating speeds differ from those on high-speed highways. Speed detectors were placed at several points before, after, and within horizontal curves in order to study roadway, cross-section, roadside, land use, and traffic engineering variables. Posted speed limits were either 25 or 35 mph; only free-flow passenger cars (time headways of at least six seconds) were included in the analyses. A mixed model was used to identify the relationship between operating speeds and roadway geometric elements. A mixed model considers the correlations that may result from multiple observations on the same drivers or observations on drivers at the same site, thus accounting for both random (data collection sites) and fixed (geometric features) effects. The analysis was divided into single-point analysis, where only the detector at the midpoint of a horizontal curve was considered, and multipoint analysis (all data collection points on tangents and curves). Two mixed models were specified: one with a single intercept for all sensors and another with separate intercepts for each sensor. For the single-point models, the analysis showed that the site variable accounted for one third of the residual variance. Three geometric variables were found to be statistically significant at the 95th percentile level: degree of curve, lane width, and roadside hazard rating. The multipoint analysis considered the data from 4 sensors and two models were specified: one with a single intercept and one with separate intercepts for each sensor. The model with a single intercept used a compound symmetry structure for its covariance. In this model only degree of curve and grade were found to be statistically significant at the 95th percentile level. The authors concluded that mixed models were appropriate to model operating speeds on low-speed urban streets, but the variability in operating speed could not be adequately explained by geometric features for multipoint models with a single intercept. The model with separate intercepts used the first-order autoregressive covariance structure and the authors concluded that: (1) vehicles slowed down after entering the curve, (2) as degree of curve increased speed decreased, and (3) as grade increased speed decreased. Other results indicated that upon entering the curve, speeds decreased as lane width increased. This result was attributable to low-speed street design where older urban streets have a wider lane approaching and within the curve. Also, as roadside hazard rating increased speed decreased, except for the sensor located at the endpoint of the curve (PT). Table 6 shows the coefficients of the models that were considered to best explain the relationship between operating speeds and the variables found to be statistically significant for the mixed models with fixed effects developed at 4 data collection locations.

Table 6 Coefficients of the Mixed Models with Fixed Effects by Sensor Location (Poe and Mason, 2000)

Sensor	Intercept	Degree of curvature	Grade	Lane Width	Hazard Rating
PC150	49.59	0.50	-0.35	0.74	-0.74
PC	51.13	-0.10	-0.24	-0.01	-0.57
MID	48.82	-0.14	-0.75	-0.12	-0.12
PT	43.41	-0.11	-0.12	1.07	0.30

Sensor location notes:
PC150 – 150 ft (45 m) before beginning of horizontal curve
PC – beginning of horizontal curve
MID – midpoint of horizontal curve
PT – end of horizontal curve

Tarris, et al. (1996) performed OLS regression and panel data analyses on the same urban street dataset used by Poe and Mason (2000). Since previous studies utilized descriptive statistics obtained through data aggregation, the authors addressed individual drivers and vehicle effects in the study. The analyses included roadway, cross-section, roadside, and land use variables. Other non-highway characteristics were included, such as vehicle type, driver gender and age, and number of passengers in the vehicle. For the panel data models, randomness in the data was attributed to two components: the location of the sensors and the individual vehicles traveling through the data collection site.

Linear regression models were specified using the mean speed (aggregate data) and individual driver speed (disaggregate) data at the midpoint of the curve; only the degree of curve was considered as an explanatory variable in the models. The model obtained using the disaggregate data was:

$$V = 53.8 - 0.272D \qquad (R^2 = 0.63) \tag{13}$$

where: V = mean speed at midpoint of the curve (km/hr); and

D = degree of curvature (degrees).

For the aggregate data, the following model was reported:

$$V = 53.5 - 0.265D \qquad (R^2 = 0.82) \tag{14}$$

The models developed using panel analyses considered data from 4 sensors: 150 ft (45 m) before the curve, at the beginning of the curve (PC), at the midpoint of the curve, and at the end of the curve (PT). Again, the model only included the degree of curve as an explanatory variable. The resulting model is shown in Equation (15) below:

$$V = 52.18 - 0.231D \qquad (R^2 = 0.487) \tag{15}$$

By adding group effects (vehicle) and time variables (sensor location) and looking at the increase in R^2, it was concluded that the group effects variable explained less than 5 percent of the variability in vehicle operating speeds on low-speed urban streets. The authors concluded that, when modeling speed choice, aggregate speed measures should be avoided. The authors also concluded that regression models may explain the influence of geometric features of the roadway, but not the influence of individual driver characteristics on operating speed.

Fitzpatrick, et al. (2005) conducted a study in order to identify the roadway features that influence drivers' speed choice. The study included data from free-flow vehicles collected at 79 tangent sites in suburban/urban areas of six states. The sites considered for the study were mostly flat with straight alignment, good surface conditions, and adequate sight distance. Presence of horizontal curves and traffic control were located far away in study sections in order to obtain data from vehicles not influenced by these features. Free-flow vehicles were identified as those vehicles with time headways of five seconds or greater. Speed data were also collected during the middle of the day (daylight conditions), at times when traffic volumes were low.

Data collected included numerical values of each cross-section feature plus the presence of bike lanes, on-street parking, and median type. Other characteristics, such as pedestrian activity, land development, access density, roadside environment (including type of fixed objects), posted speed limit, number of signals per mile, were also included in the analyses. Speed data were collected using a laser gun connected to a laptop computer. Speed profile plots showed that the variable with the strongest relationship to 85th percentile operating speeds was posted speed limit. This result was expected since 85th percentile operating speeds are commonly used to set posted speed limits (Fitzpatrick and Carlson, 2002).

When examining the speed profile plots, the authors found a negative relationship between access density and pedestrian activity and operating speeds, indicating that drivers tend to select lower speeds along roadways with a higher number of driveways per mile and in the presence of pedestrians. The plots also indicated that operating speeds tend to be lower at sites with no centerline or edge line markings, medians, and at sites where on-street parking is permitted.

Regarding the roadway features, the study findings indicated that roadways with shoulder widths of 6 feet (1.8 m) or greater resulted in operating speeds of 50 mph (80 km/hr) and higher; while shoulder widths between zero and 4 feet (1.2 m) resulted in operating speeds lower than 50 mph (80 km/hr). The presence of curb and gutter produced a range of operating speed values and the research team concluded that there was no evidence that the presence of curb and gutter influenced driver behavior on urban/suburban tangents. Wider pavements resulted in higher speeds but there was no indication of a relationship between lane width and operating speeds. An exploration of the relationship between median width and operating speeds indicated that observed operating speeds increase as the median width increases.

A regression model that shows the relationship between posted speed limit and 85th percentile speeds was developed as shown below in Equation (16). The coefficient of determination was 0.904.

$$FF85 = 12.4 + 0.98SL \tag{16}$$

where: $FF85$ = 85th percentile speed from free-flow vehicles (km/hr); and

SL = posted speed limit (km/hr).

The only variable other than the posted speed limit with a t-statistic greater than 1.0 was access density (t = -1.31). The regression equation that included access density was:

$$FF85 = 25.9 + 0.83SL - 0.054AD \tag{17}$$

where: AD is the access density, defined as the number of access points per 1 mile (1.6 km).

Equation (17) resulted in a coefficient of determination of 0.923. Analyses were also performed by roadway functional classification. The linear regression models developed for different functional classes showed that there is a strong statistical correlation between the posted speed limit and 85th percentile operating speeds on both suburban/urban and rural arterials.

Recognizing that design speed is correlated with the posted speed limit, and subsequently roadway geometrics are correlated with speed limit, Wang, et al. (2006) explored the influence of roadway design features on speeds without including the speed limit variable in any statistical models. The study sites were low-speed urban corridors, with speed limits less than or equal to 40 mph. Since a key characteristic of the urban street environment is the presence of closely spaced intersections, 35 study corridors were selected which had enough distance between intersections in which drivers could accelerate to a desired free-flow speed. Speed data were collected using in-vehicle Global Positioning Systems (GPS) during daylight, dry pavement, and non-peak hours. Acceleration and deceleration data were removed from the data in order to consider only uninterrupted trips (trips not influenced by pedestrians or turning movements).

Since the data included observed speeds from the same driver, the authors considered the development of a mixed-effects models in order to allow for the correlation between observations (within-subject correlation), thus adding a “variable (that reflects) the influence from each driver.” Speed prediction models were developed for both the 85th and 95th percentile speeds along the corridors, which are considered representative of drivers’ speed choice. The model developed by Wang, et al. (2006) for the prediction of 85th percentile speeds was:

$$V85 = 31.6 + 6.5 lane.num - 0.1 roadside - 0.05 driveway - 0.082 INT \\ + 3 curb - 4.26 sidewalk - 3.2 parking + 3.3 landuseI + 3.27 landuseII \qquad (18)$$

where: $V85$ = 85th percentile cruising speed (mph);

lane.num = number of lanes;

roadside = density of roadside objects divided by their average offsets from roadside (ft);

driveway = number of driveways per mile;

INT = number of T-intersections per mile;

curb = indicator variable for presence of curb;

sidewalk = indicator variable for presence of sidewalk;

parking = indicator variable for the presence of on-street parking;

landuseI = 1 if land use is residential, 0 otherwise; and

landuseII = 1 if land use in non-commercial and non-residential, i.e. "other", 0 otherwise.

The model developed had an intra-class correlation (ICC) value of 0.35 which meant that 35 percent of the unexplained variance of speeds is caused by driver or vehicle characteristics. In addition, the authors also developed a model for the 95th percentile speeds; the model was almost identical to the one shown in Equation (18). The authors inferred that adding a speed limit variable would result in many of the explanatory variables not being statistically significant. Although the variable for the presence of a curb was found to be associated with higher speeds (positive coefficient), it was not until subsequent work (Wang, 2006) that it was suggested that drivers select higher speeds because the curb represents a barrier between the through travel lane(s) and roadside objects.

Although linear regression was considered for many of the studies, other analysis methods, such as panel data (Tarris, et al., 1996) and mixed effects (Poe and Mason, 2000, and Wang, et al., 2006), were applied to model speed relationships on low-speed urban streets. Contrary to high-speed, two-lane rural highways, speed prediction models along low-speed urban streets consider a variety of additional roadway features, such as driveway density and parking, among others. Vertical grade and lane width were found to be statistically significant in some models (Yagar and Van Aerde, 1983; and Poe and

Mason, 2000) while driveway density was also included in several models (Yagar and Van Aerde, 1983; Fitzpatrick, et al., 2005; and Wang, et al., 2006). Similarly to high-speed rural roads, changes in horizontal alignment were associated with changes in operating speeds along urban streets. Degree of curvature was included in speed prediction models in two studies (Poe and Mason, 2000; and Tarris, et al., 1996).

In one study speed limit was found to be the only significant predictor variable at the 95 percent confidence level (Fitzpatrick, et al., 2005). However, concerns over including the posted speed limit as an explanatory variable in speed prediction models was addressed by Wang, et al. (2006) which may explain the results from Fitzpatrick, et al. (2005).

Changes in the driving environment are associated with changes in operating speeds. However, speed differentials have been negatively associated with safety (Garber and Gadiraju, 1989, and Lamm, et al., 2002). To achieve design consistency, it has been recommended that along entire sections of rural highways operating speeds should not differ by more than 5 mph (10 km/hr [McLean, 1979]). In the case of transition zones, a speed difference is indeed desired.

2.3 Rural to Urban Transition Zone Highways

Based on the design consistency literature, minimum speed differentials are desired along sections of roadway with a single design speed. In the case of transition zones, however, speed differentials are desired. In Germany, design guidelines indicate the use of sometimes unnecessary alignment changes in order to obtain desired speed differentials (Wooldridge, 1994). Only one study defined a transition zone as intended in this research – a location where changes in operating speeds are required and communicated to drivers by the presence of speed limit signs as a result of traveling from a high-speed to a low-speed area.

A two phase study by Rowan and Keese (1962) investigated possible factors that influence operating speeds along rural-to-urban transition areas in order to develop new criteria for the establishment of speed zones. More than 150 sites were studied, which included several combinations of roadway functional classification, traffic volumes, and land use development. In the first phase of the research, before-and-after studies were conducted in which posted speed limits were either reduced from 60 to 30 mph (95 to 50 km/hr) or increased from 30 to 55 mph (50 to 90 km/hr) in increments of 5 mph (10 km/hr). The results indicated that posted speed limits have little effect on operating speeds. The second phase of the study aimed to identify geometric features that influence drivers' choice of speed in rural-to-urban transition areas by use of two study methods: individual vehicle speed (IVS) and the test car methods. The IVS method, which measured individual speeds throughout various sections of the study sites with the use of an event recorder combined with road tubes and air switches, concluded that horizontal and vertical curves are the two most common elements that influence operating speeds, mainly due to sight distance restrictions. It was also concluded that changes in the cross-section resulted in traffic speed variations, but these factors could not be isolated in order to quantify them. The study results indicated that vehicles reduced their operating speeds when traveling from rural to developed areas. The study also showed that commercial developments have more influence on operating speeds than residential areas, and that residential areas with good lateral clearances have less influence than those with trees and shrubs near the curbline as indicated by lower operating speeds. These results suggest that appearance, and not density, of developed areas is a factor that influences driver behavior. The test car method was able to study the influence of sight distance on operating speeds when sight distances were less than 1000 feet (305 meters) and up to 1200 feet (365 meters). The results of this part of the study indicated that research participants decelerated more rapidly each time the sight distance became more restrictive.

The relationship between design consistency and driver error was studied by Wooldridge (1994). One objective of design consistency is to meet driver expectations in order to increase safety on highways. Driver expectancy is defined as those observable and measurable roadway features that are able to increase driver awareness for a particular task. One way to examine driver expectancy is to measure the speed differential along a section of a roadway. If driver expectancy is met, then there should not be any abrupt changes in operating speeds.

Wooldridge studied driver workload on two-lane rural highways with a speed limit of 55 mph (90 km/hr) that had a lower advisory speed sign on some sections.

Driver workload was measured using Messer's procedure (1979), a model based on "the presumption that the roadway itself provides most of the information that the driver uses to control (the) vehicle; hence the roadway imposes a workload on the driver." Messer's procedure consists of assigning ratings to roadway features based on their severity and, consequently, their contribution to driver workload—lane width reductions and crossroad overpasses are considered "more severe" than bridges and lane drops. Wooldridge's study (1994) included other factors, such as sight distance and driver expectation, in the analysis. A workload value was assigned to each geometric feature along the roadway segment being considered. The conclusions of the study indicated that large changes in driver workload over a short distance of roadway are strongly correlated with high accident rates. Roadway segments with high workload values are also correlated with high accident rates. Wooldridge recommended that future studies focus on the combination between driver workload and speed variation along a series of roadways, and to analyze this relationship using the Messer procedure.

Rural roads require adjustment in both cross-section elements and operating speeds when passing through a rural community in order to adapt to the upcoming developed area. Therefore, such projects may require the use of flexibility in design: by using design values not recommended by the Green Book, an alternative solution may be obtained for those situations that would normally be addressed with the conventional design philosophy. One study addressed the need to develop geometric design criteria for transition zones, and to provide information about projects where flexibility in design is often employed (Stamatiadis et al., 2004 and 2006). The authors inferred that rural roads, when passing through communities, should contain different cross-sections and posted speeds, thus there is a need to design transition zones to effectively influence driver behavior and to assist drivers in adjusting their speeds accordingly. The research included identification of appropriate case study sites to demonstrate flexibility in design. Curb and gutter design in transition zones, instead of a full cross-section with clear zones, was identified as one of the possible applications where design flexibility could be used along roadways passing through rural communities.

A total of 22 sites in 11 states were considered for a before and after study. Three transition zone scenarios were studied: (1) Type A, physical transition from rural area to a built-up section where the transition is a point location; (2) Type B, roadway passing through a rural community where the transition is a point location, and (3) Type C, a longitudinal transition zone was present, as identified by changes in the posted speed limit. The posted speed limit at both type A and B scenarios remained constant, thus changes in roadway environment specified the location of the transition point. Figure 2 illustrates the three categories.

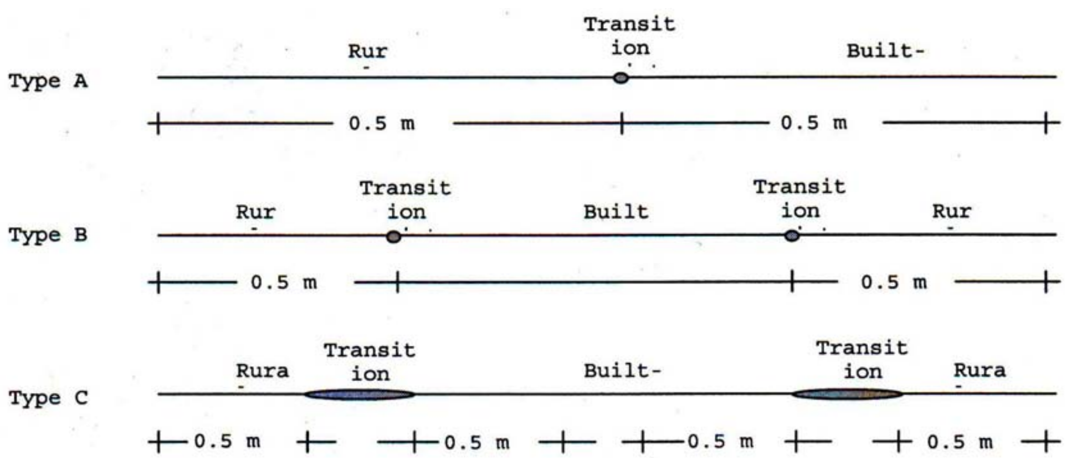

Figure 2 Study Sites Classification (Stamatiadis et al., 2004)

The design elements considered in the study were: design speed, horizontal alignment, vertical alignment, superelevation, lane width, clear zone, sight distance, median/two-way left-turn lane (TWLTL), side slopes, intersection design, and shoulder width. The before period represented the existing roadway condition while the after period consisted of a physical change in the roadway. Examples of design flexibility application in the after period included the following: reconstruction of a highway with right of way constraints, use of curb and gutter design instead of a full cross-section with clear zones in transition zones to a rural community, use of innovative approaches for intersection design, modification of design elements to address pedestrian/bicycle access, altering (lowering) design elements, altering (lowering) design speed, application of traffic calming devices, reduce/retain the footprint of the roadway, and shielding roadside obstacles with barriers rather than eliminating obstacles.

For each of the 22 sites, simple observational before-after safety analyses were performed. The total number of crashes, crash rate, crash severity, speed, and an overall roadway score were computed and compared in the before and after periods. Surveys indicated that the design element most frequently introduced along transition zones in the after period was the conversion of a median to a two-way left-turn lane (18 out of 22 cases), followed by a change in shoulder width (narrower or no shoulder in 17 out of 22 cases). By performing an expert panel analysis, the potential contributing factors (driver, vehicle, environmental, and roadway) on crash occurrence were determined. In rural sections, although the driver was identified as the major contributing factor, the roadway was a contributing factor in all but two cases.

The direct safety consequences in the design elements were not able to be isolated because more than one element changed. For almost all of the cases, the operating speed was higher than the design speed and posted speed limit, indicating that design speed had little influence on operating speeds. Drivers also disregarded posted speed limits because the geometric design elements did were not restrictive. The presence of curb and gutter had a small effect on operating speeds, but these were brief encounters in the study. The use of speed limit signs was found not to be an adequate means for attracting the attention of drivers. The authors concluded that there is a need to focus research on the design of transition areas to properly inform the driver of the upcoming posted speed limit changes. The authors also concluded that eliminating or reducing the shoulder width reduction did not pose major safety consequences, as observed in the after period of sites. Several sites studied had posted speed limits higher than their corresponding design speeds.

The research performed by Stamatiadis et al. (2004 and 2006) is evidence that there is a lack of design guidance for transition zones between rural and developed areas. Several of the study sections had a curb and gutter design in the built-up section, which was often the only visual indication of changes in the driving environment. However, curb and gutter appeared to have little effect on operating speeds. There is a need for other forms of design flexibility applications to properly reduce vehicle speeds. These include increased signage, introduction of landscaping features, or more limiting design features such as a lower design speed or the introduction of smaller radius or successive horizontal curves. The authors also noted that there is a need to study the relationship between design and operating speeds in transition zones. A recommendation was made to add transition zone design guidance to the AASHTO Green Book.

2.4 Summary

Most of the speed prediction models along high-speed two-lane rural highways have been developed using data collected at the midpoint of a horizontal curve. Only two speed models were developed to predict operating speeds along tangent sections approaching a horizontal curve (Polus, et al., 2000; and Figueroa and Tarko, 2005). Horizontal curve data, such as radius, degree of curvature, or the deflection angle, have been included in most operating speed prediction models for two-lane rural highways. A general consensus among past two-lane rural highway operating speed research is that sharper horizontal curves (i.e., smaller radius or higher degree of curve) reduce vehicle operating speeds. Other variables that have been shown to be negatively correlated with vehicle operating speeds on two-lane rural highways are:

- Presence of horizontal curve to the left, as compared to a horizontal curve to the right
- Length of approaching tangent before entering a horizontal curve
- Highway grade
- Average daily traffic
- Truck percentage
- Driveway density
- Presence of nearby intersections

The predictor variables that have been shown to be positively correlated with vehicle operating speeds on two-lane rural highways are:

- Length of horizontal curve
- Posted speed limit
- Sight distance
- Pavement width
- Shoulder width, either paved or unpaved
- Maximum superelevation rate

Along low-speed urban streets, published operating speed models have generally reached consensus that the degree or radius of a horizontal curve is strongly correlated with the operating speed. Increases in the degree of curve have been shown to reduce vehicle operating speeds. Other variables that have been found to be negatively correlated with operating speeds on low-speed urban streets include:

- Grade
- Driveways
- Presence of sidewalk
- Pedestrian activity
- On-street parking
- Density of roadside objects
- Number of intersections

The predictor variables that have been shown to be positively correlated with vehicle operating speeds on two-lane rural highways are:

- Shoulder width
- Posted speed limit
- Number of lanes
- Presence of curb
- Presence of centerline and edge line pavement markings

In the present research, rural highway transition zones include both a high-speed and a low-speed segment. The two-lane rural highway and urban street operating speed research provides some important insights regarding the geometric design, roadside, and land use characteristics that may be associated with operating speeds along transition zones; however, operating speed models for transition zones do not currently exist. Since these highway sections require changes in operating speeds to comply with the change in the regulatory speed limit, research is needed to quantify the effects of geometric design, roadside, and land use characteristics on operating speeds.

The most common method of data analysis as presented in the literature review is ordinary least squares (OLS) linear regression. However, using OLS regression to develop speed prediction models along transition zones may result in the violation of the independent observations assumption. When collecting speed data at several point locations along a study sites, correlated speed data is expected; the speed at a downstream location is dependent on the speed at an upstream location. Alternatives methods that are able to model correlated data have been explored in past studies; a two-level model for speed differences was estimated by Park and Saccomanno (2005) while panel data models were explored by Tarris et al. (1996) to model speeds along horizontal curves on urban streets. As such, one of the purposes of this research is to explore longitudinal data methods for developing speed prediction models along two-lane rural highway transition zones.

The inclusion of speed limit as a potential explanatory variable in operating speed models is questionable since it may be endogenous with highway design features. It has been recognized that design elements, such as lane and shoulder widths, are selected based on a design speed. Speed limit values are typically posted at levels equal to or less than the designated design speed, thus it has been suggested that speed limit should not be included when exploring the highway characteristics that are influential on operating speeds (Wang, et al., 2005). However, speed prediction models that have included this variable have been on sections of highway with a constant speed limit. Since transition zones have posted speed limit changes and only a single designated design speed, the inclusion of speed limit as a potential explanatory variable would not necessarily pose concerns related to endogeneity.

CHAPTER 3
DESCRIPTION OF DATA

This chapter details the data collection methodology used in the present research. Operating speed, roadway and roadside design features, and land use characteristics were collected along 20 transition zones in central Pennsylvania to identify which highway characteristics are associated with operating speeds. Study site selection and data collection is described in this chapter, followed by summary statistics for all operating speed and highway characteristics measured at each study site.

3.1 Site Selection

As previously noted, transition zones are defined as highway sections in which a change in operating speed is required based on changes in the regulatory speed limit. The study focuses on high-to-low speed transition zones on two-lane rural highways.

An essential requirement of the study sites was the presence of both a Reduced Speed Ahead sign followed by a Speed Limit sign. The limits of the transition zone were then defined by the location of these two signs--the Reduced Speed Ahead sign indicated the beginning of the transition zone and the downstream Speed Limit sign indicated the end of the transition zone (i.e., beginning of low-speed environment). Figure 3 shows the limits of the transition zone in relation to the static speed signs.

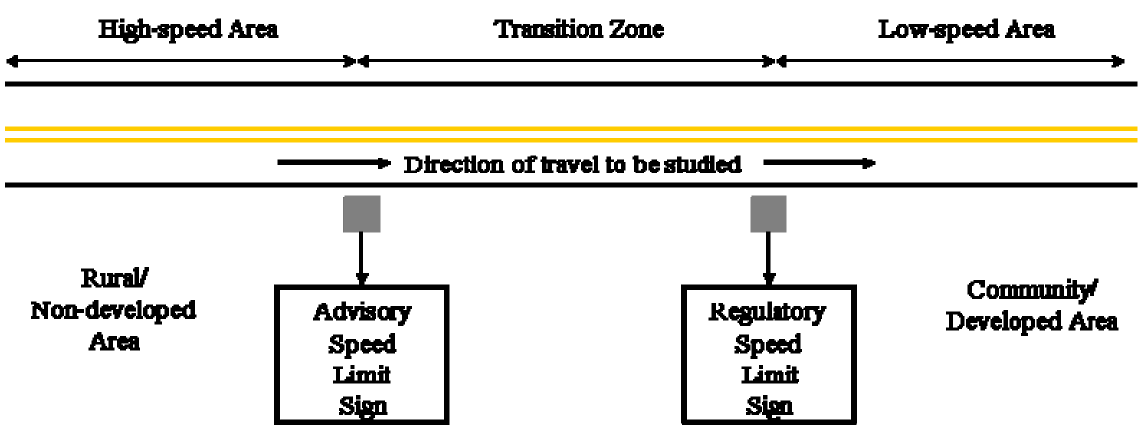

Figure 3 Transition Zone Illustration

Additionally, only sites with the version of the Reduced Speed Ahead sign specified in the 2000 edition of the MUTCD were considered. As noted previously, the Reduced Speed Ahead sign has changed in both size and color in the new edition of the MUTCD (see Figure 1). However, the 2003 MUTCD version of this sign is not frequently encountered along two-lane rural highways in central Pennsylvania. Figure 4 shows an example of a transition zone with a Reduced Speed Ahead sign.

Figure 4 Example of a Transition Zone with a Reduced Speed Ahead Sign

PennDOT's online video photolog system was used as a tool to identify potential study sites. Field visits confirmed if the locations were indeed appropriate for the research. In order to minimize the probability that driver behavior will be influenced by factors other than the geometric features, sites with the following characteristics were then identified as candidates for the present research:

1. Free of signalized or stop-controlled intersections along the major road in the direction of travel within the transition zone.
2. Less than 10 percent heavy vehicles since trucks and other heavy vehicles may influence drivers' speed choice.
3. Low-volume highways in order to maximize the probability of collecting free-flow vehicles. Past research has identified low-volume highways as those

with an ADT less than 4,000 vehicles per day (McFadden and Elefteriadou, 2000).

4. Smooth pavement surfaces and visible pavement markings.

Additionally, the study sites included a range of roadway, roadside, and land use characteristics, such as variable lane and shoulder widths, and vertical grades. Details on the highway features present at each site are discussed later in this chapter. Using the site selection criteria described previously, 20 sites in central Pennsylvania were selected for the present study. Table 7 provides a description of the study sites.

Table 7 Description of Study Sites

Site ID	Town	County	Route	Segment(s)	Speed Limit Reduction (mph)	Transition Zone Length (ft)
1	Alverda	Indiana	553WB	0160-0170	55 → 35	535
2	Brush Valley	Indiana	56WB	0420	55 → 35	690
3	Corsica	Jefferson	322 WB	0020-0030	55 → 35	725
4	Cross Keys	Juniata	35 NB	0050-0060	55 → 40	540
5	Cross Keys	Juniata	35 SB	0070-0080	55 → 40	375
6	Curwensville	Clearfield	453 NB	0390-0410	45 → 25	750
7	Curwensville	Clearfield	879 EB	0100-0110	45 → 25	500
8	Ernest	Indiana	110 EB	0070-0080	55 → 35	1065
9	Fousetown	Huntingdon	655 SB	0540-0550	55 → 35	750
10	Freeport	Butler	356 NB	0110	55 → 40	690
11	Homer City	Indiana	3035 NB	0010-0030	55 → 35	1020
12	Madisonburg	Centre	192 EB	0210-0220	55 → 40	800
13	Mifflintown	Juniata	35 NB	0510-0520	45 → 25	570
14	Orbisonia	Huntingdon	522 NB	0210	40 → 25	925
15	Osceola Mills	Clearfield	53 NB	0480-0490	45 → 25	460
16	Shirleysburg	Huntingdon	522 SB	0310-0320	55 → 35	700
17	Spruce Creek	Huntingdon	45WB	0080-0090	55 → 35	675
18	Unionville	Centre	3040 NB	0360-0370	45 → 35	665
19	Warriors Mark	Huntingdon	550SB	0110-0120	55 → 35	410
20	Zion	Centre	550 NB	0520-0540	55 → 40	860

3.2 Data Collection

The data collected for this research consist of speed data (response variable) and the roadway, roadside, and land use characteristics (explanatory variables) at each site. The methods used to collect these data are described in this section of the dissertation.

3.2.1 Speed Data

Speed data were collected using Nu-metrics Hi-Star sensors, which use vehicle magnetic imaging technology. The Hi-Star sensors are non-intrusive, thus eliminating the possibility of drivers adjusting their speeds due to visible equipment and data collection personnel. The dimensions of the sensors are 6.5 inches by 5.5 inches with a profile of 0.625 inches--they were placed in the center of the travel lane. A rubber cover was used to protect them and to reduce their conspicuity. In addition to speed data, Hi-Star sensors time stamp the data and also provide information related to the pavement temperature, pavement condition (dry or wet), and vehicle length. The time stamp can be used to identify free-flow vehicles.

As previously noted, the limits of the transition zone were defined as related to the position of the traffic signs that inform drivers of changes in the regulatory speed. The position of the Reduced Speed Ahead sign marked the beginning of the transition zone. The end of the transition zone is marked by the Speed Limit sign that indicates the lower posted speed limit. Since drivers may be influenced by upstream geometric design features (Yagar and Van Aerde, 1983), speed data were collected in advance of the transition zone. Additionally, since it was hypothesized that drivers are influenced by the highway features instead of the traffic signs, speed data were also collected downstream of the transition zone. Therefore, the Hi-Star sensors were placed at four points along each study site in order to collect speed data before, within, and after the transition zone. The following four points correspond to the sensor locations where point speed data were collected: (1) 500 feet before the beginning of the transition zone; (2) at the beginning of the transition zone; (3) at the end of the transition zone; and (4) 500 feet after the end of the transition zone. Figure 5 shows the four locations where the Hi-Star sensors were placed at each study site.

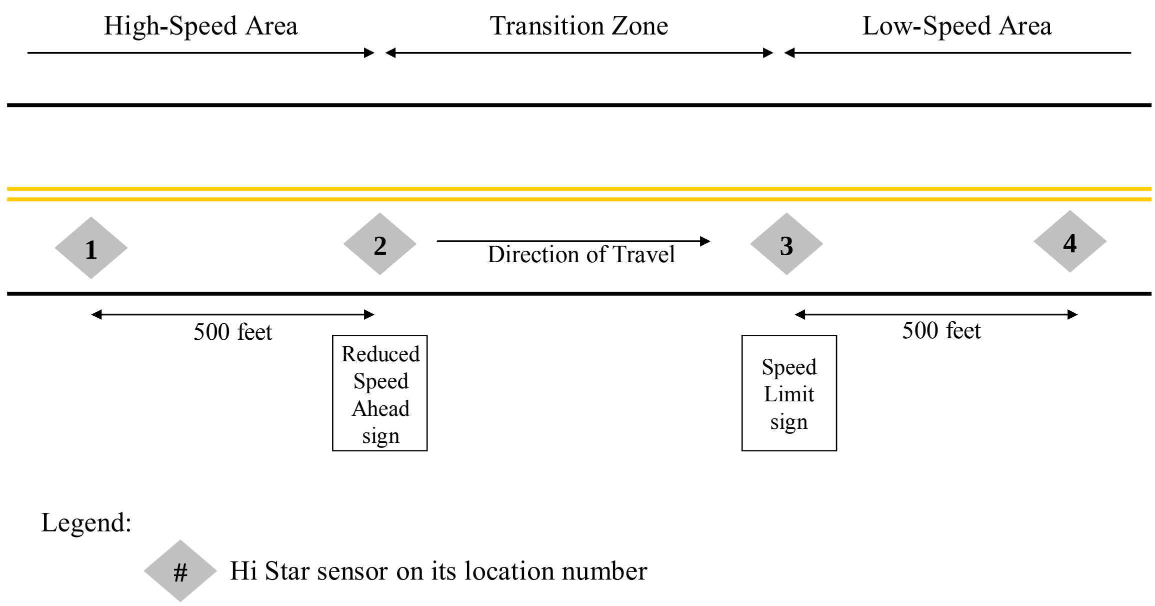

Figure 5 Sensor Layout

Information collected by the Hi-Star sensors was downloaded into a Microsoft Excel spreadsheet using the Highway Data Management (HDM) software. In order to isolate the effects of highway features on operating speeds, only data from free-flow vehicles were considered for analyses. Past research has indicated that free-flow vehicles should have a minimum time headway of five seconds (McFadden and Elefteriadou, 2000). Vehicles with time headways less than five seconds were discarded from the database. In addition, data were collected during daylight, under favorable weather conditions (no rain or snow and dry pavement). Data were also collected during non-peak travel hours in order to maximize the probability of observing free-flow vehicles.

The layout of the sensors permitted a vehicle at a site to be "tracked." Only vehicles in which speed information was collected at all four sensor locations were included in the analysis database. As such, the number of observations for each sensor location at a specific site across all four sensors was the same.

A minimum sample size of 100 free-flow passenger vehicles per site was desired for data analysis. This sample size was obtained using the following equation (Institute of Transportation Engineers [ed. Robertson], 1994):

$$N = (S\frac{K}{E})^2 \tag{19}$$

where: N = minimum number of measured speeds;

S = estimated sample standard deviation (mph);

K = constant corresponding to the desired confidence level; and

E = permitted error in the average speed estimate (mph).

A value of 5.3 is representative of the sample standard deviation, S, for two-lane rural highways (Robertson, 1994). By substituting several values for the confidence level constant, K, a range of sample sizes can be obtained for a specific value of permitted error, E. Table 8 shows the computed sample sizes for 90, 95, and 99 percent confidence levels with a permitted error, E, of ±1 mph and a standard deviation, S, of 5.3.

Table 8 Sample Sizes for Different Levels of Confidence

K	Confidence Level	N
1.64	90%	76
1.96	95%	108
2.58	99%	187

Although a minimum sample size of 100 free-flow vehicles at each site was desired, in some instances there were fewer than 100 speeds collected at a study site. This was primarily due to low traffic volumes during the four- to six-hour data collection period. There were 2,859 free-flow passenger vehicles included in the analysis database for a total of 11,436 individual vehicle point speeds. Table 9 summarizes the speed data collected at each study site, including the sample size, mean speed, and sample speed standard deviation at each sensor location for all 20 sites. Figure 6 shows a graph of mean speed at each sensor location for each study site.

Table 9 Mean Speed and Speed Deviation at each Study Site

Site ID	Sample Size	Mean Speeds per Sensor, mph (Speed Deviation, mph)			
		1	2	3	4
1	124	47.9 (7.24)	49.6 (7.20)	50.3 (6.07)	47.6 (7.12)
2	68	52.8 (9.50)	52.4 (7.71)	44.2 (8.47)	43.1 (8.04)
3	98	51.3 (5.44)	49.9 (5.72)	46.3 (6.17)	43.1 (6.15)
4	104	57.6 (7.97)	53.9 (7.69)	52.6 (6.67)	48.2 (6.75)
5	231	58.2 (6.78)	52.3 (7.16)	49.6 (6.79)	45.5 (6.81)
6	99	42.6 (7.18)	41.6 (6.00)	35.8 (7.37)	28.7 (6.26)
7	159	52.0 (6.11)	47.0 (5.62)	44.4 (6.91)	37.4 (6.19)
8	149	57.1 (6.20)	53.0 (7.02)	49.5 (7.50)	46.7 (6.48)
9	478	58.4 (6.40)	53.1 (5.92)	48.3 (7.02)	47.4 (7.65)
10	148	51.7 (6.08)	51.0 (5.82)	49.6 (5.90)	49.0 (6.14)
11	141	43.3 (6.88)	41.4 (6.11)	36.6 (5.30)	36.9 (5.96)
12	73	54.5 (6.20)	52.6 (5.96)	48.8 (7.09)	38.9 (9.57)
13	130	43.8 (6.89)	41.7 (5.02)	28.4 (4.08)	30.1 (4.42)
14	112	53.4 (7.03)	49.2 (6.17)	39.2 (5.81)	36.1 (5.97)
15	81	46.7 (5.98)	41.7 (5.02)	41.7 (5.42)	36.6 (5.89)
16	122	54.0 (6.90)	50.8 (5.51)	45.7 (6.20)	36.7 (5.66)
17	164	58.2 (6.25)	55.5 (6.08)	50.4 (6.04)	46.3 (6.02)
18	52	58.1 (7.54)	53.3 (7.06)	52.0 (6.31)	50.8 (6.08)
19	178	50.2 (5.75)	45.5 (5.31)	49.5 (6.44)	42.5 (6.01)
20	148	53.3 (5.56)	52.0 (6.01)	47.8 (5.58)	43.9 (6.26)
Total:	2,859				

[a] High speed zone is located between sensors 1 and 2
[b] Low speed zone is located between sensors 3 and 4

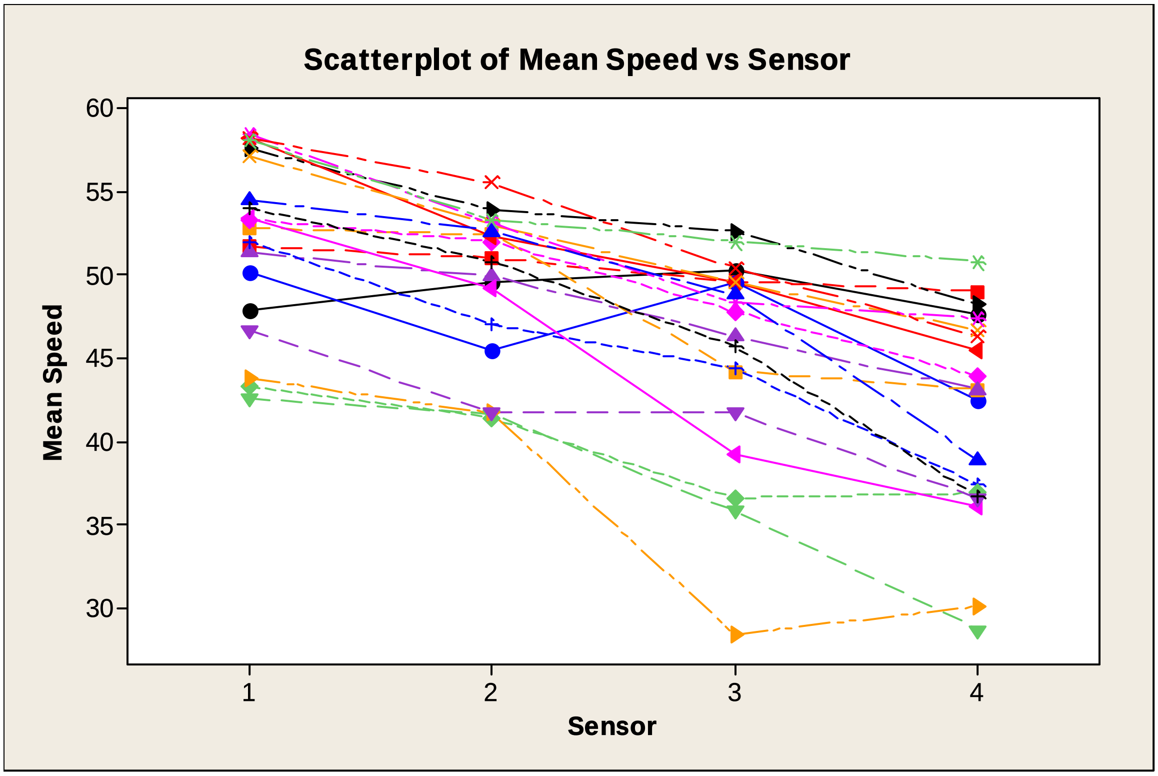

Figure 6 Mean Speed Plot for each Data Collection Point at each Study Site

As shown in Figure 6, it can be seen that operating speeds and speed differentials vary across sites. A steep slope is related to a greater speed change, while horizontal lines are indicative of no change in mean operating speed between two successive data collection points. The variability in the speed reductions observed at each study site location suggests that roadway, roadside, or land use characteristics may be influencing observed operating speeds. Several interesting observations can be noted from both Figure 6 and Table 9:

Before the transition zone (between sensors 1 and 2):

1. Mean speed increase only at Site 1 (by 1.7 mph).
2. Mean speeds reduced by less than 2 mph at 7 sites (Sites 2, 3, 6, 10, 11, 12, and 20).
3. Mean speeds decreased between 2 and 5 mph at 8 sites (Sites 4, 8, 13, 14, 16, 17, 18, and 19).
4. Mean speeds reductions of at least 5 mph at 4 sites (Sites 5, 7, 9, and 15).

Within the transition zone (between sensors 2 and 3):

1. Two sites experienced an increase in mean speed; Site 19 by approximately 4 mph and Site 1 by 0.7 mph.
2. Four sites experienced less than a 2 mph mean speed reduction (Sites 4, 10, 15, and 18).
3. Sites 3, 5, 7, 8, 9, 11, 12, and 20 experience speed reductions between 2 and 5 mph.
4. Sites 2, 6, 16, and 17 experience speed reductions between 5 and 10 mph.
5. Two sites (Sites 13 and 14) experience speed reductions of at least 10 mph.

Beyond the transition zone (between sensors 3 and 4):

1. Sites 11 and 13 experienced mean speed increases of 0.3 and 1.7 mph, respectively.
2. Four sites experience mean speed reductions of 2 mph or less (Sites 2, 9, 10, and 18)
3. Sites 1, 3, 4, 5, 8, 14, 17, and 20 experienced mean speed reductions between 2 and 5 mph.
4. Six sites (Sites 6, 7, 12, 15, 16, and 19) experienced mean speed reductions of at least 5 mph.

From the variability in speed patterns observed in Figure 6, specifically the speed changes that take place beyond the limits of the transition zone (sensor 3 in Figure 6), it can be inferred that drivers are traveling at speeds in excess of the posted speed limit at the low-speed end of the transition zone. Since several study sites have the same posted speed limit changes, it can also be inferred from the variability in these speed patterns that changes in operating speeds are influenced by the different roadway, roadside, or land use characteristics.

As noted in Chapter 1, safety issues may arise when drivers do not adjust their operating speeds along a transition zone. One method to set posted speed limits is via an engineering study. The 85th percentile operating speed is the most common measure to set posted speed limits. Posted speed limits may also be set based on local statutes which apply across specific roadway functional classes or geographic areas (TRB, 1998). The

85th percentile operating speeds were calculated at each sensor location at each study site; these data are shown in Table 10 along with the regulatory speed limit.

Table 10 85th Percentile Speeds

Site ID	Speed Limit (mph)		85th Percentile Speeds per Sensor (mph)			
	High Speed Zone[a]	Low Speed Zone[b]	1	2	3	4
1	55	35	55.40	57.03	56.62	54.95
2	55	35	62.60	60.43	52.94	51.49
3	55	35	56.97	55.83	52.70	49.48
4	55	40	65.86	61.88	59.54	55.21
5	55	40	65.22	59.73	56.59	52.58
6	45	25	50.06	47.84	43.40	35.19
7	45	25	58.30	52.82	51.55	43.77
8	55	35	63.52	60.29	57.28	53.39
9	55	35	65.03	59.18	55.54	55.29
10	55	40	58.01	57.08	55.72	55.36
11	55	35	50.38	47.72	42.05	43.08
12	55	40	60.96	58.82	56.14	48.78
13	45	25	50.92	47.63	32.65	34.64
14	40	25	60.72	55.63	45.24	42.25
15	45	25	52.90	46.86	47.31	42.74
16	55	35	61.20	56.48	52.12	42.57
17	55	35	64.68	61.77	56.70	52.58
18	45	35	65.93	60.66	58.54	57.14
19	55	35	56.13	51.04	56.13	48.77
20	55	40	59.10	58.20	53.59	50.36

[a] High speed zone is located between sensors 1 and 2
[b] Low speed zone is located between sensors 3 and 4

By comparing 85th percentile operating speeds to the regulatory speed limit at each of the speed sensor locations, the magnitude of speeding vehicles can be determined. At the high-speed zone (sensor 1):

1. Only one site (Site 11) had observed 85th percentile speeds lower than the posted speed limit.
2. Observed 85th percentile speeds at two sites were 20 mph higher than the posted speed limit (Sites 14 and 18).

3. Observed 85th percentile speeds at Sites 4, 5, 7, and 9 were between 10 and 20 mph higher than the posted speed limit.
4. At the remaining 12 sites (Sites 1, 2, 3, 6, 8, 10, 12, 13, 15, 16, 17, 19, and 20), the observed 85th percentile speeds were higher than the posted speed limit by less than 10 mph.

At the beginning of the transition zone (sensor 2):

1. The observed 85th percentile speeds at two sites were at least 10 mph higher than the posted speed limit (Sites 14 and 18).
2. The observed 85th percentile speeds at five sites were between 5 and 10 mph higher than the posted speed limit (Sites, 2, 4, 7, 8, and 17)
3. Sites 11 and 19 had 85th percentile speeds lower than the posted speed limit.
4. The remaining 11 sites had 85th percentile speeds higher than the posted speed limit by an amount of 5 mph or less (Sites 1, 3, 5, 6, 9, 10, 12, 13, 15, 16, and 20).

At the end of the transition zone (sensor 3), the following trends were found:

1. The observed 85th percentile speeds were 20 mph higher than the posted speed limit at nine sites (Sites 1, 7, 8, 9, 14, 15, 17, 18, and 19).
2. The observed 85th percentile speeds were between 10 and 20 mph higher than the posted speed limit at nine sites (Sites 2, 3, 4, 5, 6, 10, 12, 16, and 20).
3. At the remaining two sites (Sites 11 and 13), 85th percentile speeds were higher than the posted speed limit by less than 10 mph (approximately 7 mph at both sites).

At the low-speed zone (sensor 4):

1. Two sites had 85th percentile speeds 20 mph higher than the reduced posted speed limit (Sites 9 and 18).
2. Fourteen sites experienced 85th percentile speeds between 10 and 20 mph higher than the speed limit (Sites, 1, 2, 3, 4, 5, 6, 7, 8, 10, 14, 15, 17, 19 and 20).
3. Four sites (Sites 11, 12, 13, and 16) had 85th percentile speeds between 5 and 10 mph higher than the posted speed limit.

Table 10 shows that 85^{th} percentile speeds are lower at sensor location 4 when compared to those speeds at sensor location 3, thus drivers do not appear to fully adjust their speeds within the transition zone. The observed 85^{th}-percentile speeds in the low-speed zone provide evidence that, although drivers keep decelerating after the end of the transition zone, operating speeds exceed the posted speed limit.

The study sites included in this research have different speed limit reductions; the posted speed limit changes from 55 to 35 mph at nine sites; from 55 to 40 mph at five sites; from 45 to 25 mph at four sites; from 40 to 25 mph at one site; and from 45 to 35 mph at one site. The speed changes observed at each site do not provide any consistent pattern in relation to the posted speed limit changes. This underscores the need to determine which roadway, roadside, and land use characteristics are associated with speed reductions along two-lane rural highway transition zones.

3.2.2 Highway Characteristics

It has been hypothesized that various roadway, roadside, and land use characteristics are associated with drivers' speed choice along transition zones. Since speed data were collected at four points at each study site, the highway characteristics at each point were also collected. The roadway, roadside, and land use characteristics that were collected at each study site include the following:

- Changes in the posted speed limit
- Lane width
- Paved shoulder width
- Stabilized shoulder width;
- Paved roadway width
- Lateral clearance distance
- Presence of guide rail
- Vertical curve and grade data
- Presence of a horizontal curve
- Type of centerline marking
- Type and number of both regulatory and warning signs
- Number of driveways

- Presence and/or introduction of curb and gutter

Some of the geometric roadway features, such as lane width, shoulder width, paved roadway width, and grade, were measured at each of the sensor locations. Data for other highway features, such as type and number of traffic signs and number of driveways, were collected and assigned to a sensor location according to their proximity to each sensor (i.e., influence zone). Figure 7 illustrates how some of these features were assigned to each sensor (color coded).

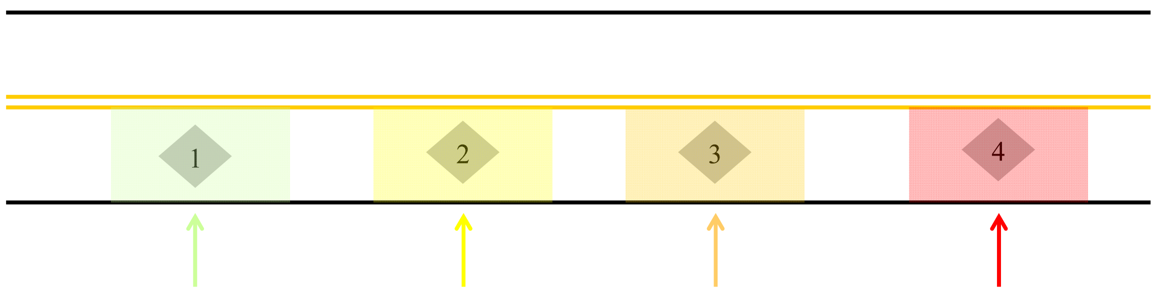

Figure 7 Area Assigned at each Sensor Location

The highway characteristics collected to be included in the data analysis as predictors for operating speeds can be categorized into groups: continuous (those that were measured), count (those that were counted), and categorical (used as indicator variables in the analysis). Table 11 shows the summary statistics for those highway characteristics that were either measured or counted at each study site.

The summary statistics for speed limit are not shown in Table 11 since this information is provided in Table 7. However, the summary statistics for speed limit reduction are shown in Table 11. In addition to these highway characteristics, other variables were created. For example, a variable for "rounded lane width" was created since it was hypothesized that, although a lane width of 9.8 feet was measured, it was intended to be a 10-foot lane. Indicator variables were also created for many of the highway features shown in Table 11, such as lateral clearance (less than 10 feet, between 10 and 20 feet, greater than 20 feet), vertical alignment (flat, downgrade, upgrade, sag vertical curve, and crest vertical curve), speed limit reduction, and number of driveways.

Table 11 Summary Statistics for Quantitative Highway Features

Measured Highway Feature	Mean	St Dev	Minimum	Maximum
Lane Width	10.65	0.570	9.7	13
Paved Shoulder	3.43	1.911	0	8.6
Stabilized Shoulder	1.27	1.807	0	12
Paved Roadway	28.33	4.493	23	41
Lateral Clearance	8.88	5.244	0	30
Grade	-0.50	2.926	-9.2	5.6
Speed Limit Reduction	18.39	2.523	10	20
Counted Highway Feature			**Minimum**	**Maximum**
Number of Driveways - Adjacent Side			0	5
Number of Driveways - Opposite side			0	5
Total Number of Driveways			0	7
Number of Warning Signs			0	3
Number of Regulatory Signs*			0	1
Number of Junction Signs			0	2

* In addition to the speed limit signs that specify the beginning and end of the transition zone.

The following highway features were coded as categorical variables in the analysis database:

- Information on horizontal alignment (tangent, presence of curve and direction, locations of PC, MC, and PT)
- Presence of curb
- Presence of non-residential buildings (i.e. post office, school, fire station)
- Type of warning sign (intersection ahead, passing and non-passing zones, school zone, etc.)
- Presence of guide rail
- Type of centerline marking (no passing, passing on both sides, passing on opposite side, passing on adjacent side)

The final set of explanatory variables considered for data analyses consisted of approximately 50 potential predictors, including indicator variables. Tables 12, 13, and 14 lists the frequency, percent, and cumulative percent for the indicator variables included in the dataset. Lastly, different warning signs were included, thus Table 15 lists the number of warning signs per study site.

Table 12 Summary Statistics for Indicator Variables for Change in Roadway Alignment

Presence of Horizontal Curve Regardless of Direction				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
HC (0 value)	5,543	48.47	48.47	No Horizontal Curve
HC (1 value)	5,893	51.53	100	Presence of Horizontal Curve
Total	11,436	100		
Presence of Horizontal curve Considering Direction				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
HCRight	3,920	34.28	34.28	Curve to the Right
HCLeft	1,727	15.10	49.38	Curve to the Left
HCReverse	246	2.15	51.53	Reverse Curve
Subtotal	*5,893*			
Tangent	5,543	48.47	100	No Horizontal Curve
Total	11,436			
Interaction between Horizontal Curve and Curve Ahead Warning Sign				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
Curve_w_ws	2,312	20.22	20.22	Curve with Warning Sign
Curve_wo_ws	3,335	29.16	49.38	Curve without Warning Sign
HCReverse	246	2.15	51.53	Reverse Curve (no warning sign)
Subtotal	*5,893*			
Tangent	5,543	48.47	100	No Horizontal Curve
Total	11,436			
Presence of Vertical Grade regardless of direction				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
Grade (0 value)	7,899	69.1	69.1	Grade is less or equal than | 3% |
Grade (1 value)	3537	30.9	100	Grade is greater than | 3% |
Total	11,436			
Presence of Vertical Curve Considering Direction				
Variable	**Freq.**	**Percent**	**Cum.**	**Description Comments**
G_UP	1,518	13.3	13.3	Grade is greater than + 3%
G_DOWN	2,019	17.7	30.9	Greater is less than - 3%
Subtotal	3,537			
G_FLAT	7899	69.1	100	Grade is less or equal than +/- 3%
Total	11,436			

Table 13 Summary Statistics for Indicator Variables for Speed Limit, Total Number of Driveways, Warning Signs, and Centerline

Speed Limit				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
sl25	1,162	10.2	10.2	Speed Limit 25 mph
sl35	3,148	27.5	37.7	Speed Limit 35 mph
sl40	1,632	14.3	52.0	Speed Limit 40 mph
sl45	1,042	9.1	61.1	Speed Limit 45 mph
sl55	4,452	38.9	100	Speed Limit 55 mph
Total	11,436	100		
Total Driveways				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
td0	2,234	19.5	19.5	no driveways
td1	3,562	31.2	50.7	1 driveway
td2	1,776	15.5	66.2	2 driveways
td3	1,632	14.3	80.5	3 driveways
td4	1,154	10.1	90.6	4 driveways
td5	271	2.4	92.9	5 driveways
td6	536	4.7	97.6	6 driveways
td7	271	2.4	100	7 driveways
Total	11,436	100		
Warning Signs				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
Intersection	1069	9.3	9.3	Intersection Ahead
School/Children	1172	10.2	19.6	School Zone / Presence of Children
Curve	1307	11.4	31.0	Curve Ahead
Other	831	7.3	38.3	Other Type of Warning Sign
None	7057	61.7	100	No Presence of Warning Sign
Total	11436			
Centerline				
Variable	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
C0	8,978	78.51	78.51	No passing
C1	648	5.67	84.17	Passing allowed both sides
C2	228	1.99	86.17	Passing allowed this side
C3	1,582	13.83	100	Passing allowed other side
Total	11,436	100		

Table 14 Summary Statistics for Indicator Variables for Lateral Clearance, Guiderail, Curb, Building, and Regulatory Signs

Lateral Clearance				
Value	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
0	8,485	74.2	74.2	Lateral Clearance less or equal to 10 ft
1	2,951	25.8	100	Lateral Clearance greater than 10 ft
Total	11,436	100		
Guiderail to the Right				
Value	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
0	8,950	78.3	78.3	No Guiderail to the Right
1	2,486	21.7	100	Presence of Guiderail to the Right
Total	11,436	100		
Curb				
Value	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
0	9,394	82.1	82.1	No Curb
1	2,042	17.9	100	Presence of Curb
Total	11,436	100		
Building				
Value	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
0	10,145	88.7	88.7	No building
1	1,291	11.3	100	Presence of School, Post Office, etc
Total	11,436	100		
Regulatory Sign (in addition to the transition zone indicators)				
Value	**Freq.**	**Percent**	**Cum.**	**Description / Comments**
0	11,001	96.2	96.2	No Regulatory Sign
1	435	3.8	100	Presence of Regulatory Sign
Total	11,436	100		

3.3 Summary

Twenty study sites were identified in central Pennsylvania to explore the relationship between operating speeds and highway characteristics along transition zones. All study sites required both Reduced Speed Ahead and Speed Limit signs to identify the beginning and the end of the transition zone, respectively.

Speed data were collected using Hi-Star sensors, which are considered to be non-intrusive. Speed data were collected during daylight and dry pavement conditions. Only data from free-flow passenger vehicles (those with headways of at least five seconds) and those vehicles for which speed information was available at all four sensors were included in the data analyses in order to track individual driver speeds.

All potential highway features that were considered to influence drivers' speed choice were included in the database as potential explanatory variables. The final data set included 11,436 individual speed observations from 2,859 vehicles and more than 50 potential explanatory variables.

CHAPTER 4
ANALYSIS METHODOLOGY

This chapter describes the methodology used to determine the association between various roadway characteristics and operating speeds along two-lane rural highway transition zones. Statistical models of mean vehicle operating speed were estimated using a variety of methods. The explanatory variables considered in the analysis include roadway and roadside design features, traffic control characteristics, and the surrounding land use.

The statistical analyses can be categorized into two sections: point speed models and speed difference models. In the point speed models, an initial OLS regression model was developed so that the results of this traditional speed modeling method could be made to the following three longitudinal models considered in this research: panel data analysis, multilevel models, and generalized estimating equations (GEE). In the speed difference models, in which the change in speed along the transition zone length was used as the response variable, both OLS regression and multilevel models were considered. Use of only a single observation per driver (speed difference) in this dataset removed the issue of correlation among observations, thus panel data and GEE models were not considered with this dataset. Figure 8 shows a flowchart of the model development process for both point speeds and speed difference scenarios.

This section of the dissertation is organized into two sections. The first describes the point speed modeling methodology, and the second describes the speed difference modeling methodology. In both point speed and speed difference models, the general functional form of the model specification is provided in the following sections, along with a discussion of the key assumptions of the model and model estimation procedures.

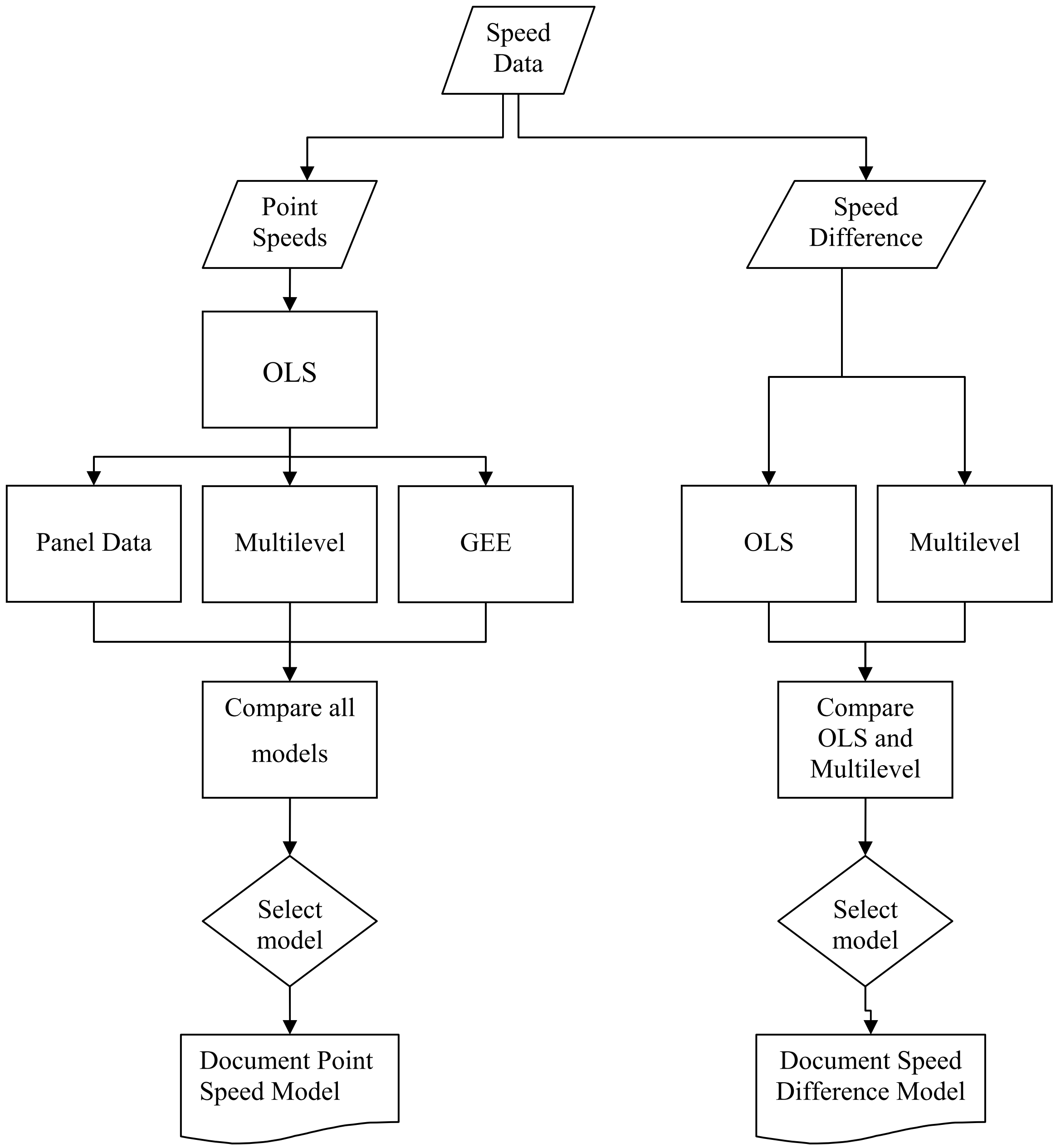

Figure 8 Flowchart of Model Development and Identification

4.1 Point Speed Analysis

As described in Chapter 3 of this dissertation, speed and roadway data were collected at four sensor locations at each study site. In addition to the operating speed data, the data collection equipment provided information concerning the time headway of each vehicle passing through the transition zone. This information permitted each vehicle to be "tracked" through the transition zone. As a result, driver-specific information contained in the analysis dataset could be explored. In the present study, panel data analyses, multilevel models, and generalized estimating equations (GEE), an extension of the general linear model (GLM), were applied to account for the correlation between observations due to driver-specific speed information. The results obtained from these models were compared to the traditional operating speed modeling approach of OLS regression.

4.1.1 Ordinary Least Squares

The method of OLS linear regression is perhaps the most common statistical method used to obtain parameter estimates of vehicle operating speeds as described previously in Chapter 2. In this method, it is assumed that a linear relationship exists between the dependent variable and the independent variables. Let y_i be the i^{th} observation of the response variable (i = 1, 2, …, n), the linear relationship is commonly expressed in the following equation:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1} \qquad (20)$$

where: Y = the column vector for dependent variable (speed);

X = referred to as the design matrix, containing the set of independent variables (highway features);

β = column vector of regression parameters to be estimated; and

ε = column vector that contains the random errors.

The relationship between three terms explains the methodology behind OLS. These three terms are: observations (y_i), overall mean ($\bar{y}$), and predictors ($\hat{y}_i$), also viewed as group-specific sample means.

The purpose of OLS is to minimize the total sum of squares, defined as the difference between the predicted values and the observed data. These are explained below along with their respective equations:

1. Total Sum of Squares, *TSS*, is defined as the sum of squared deviations of each observation from their mean, given by:

$$TSS = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 \tag{21}$$

2. Sum of Squared Errors, *SSE*, also known as residual sum of squares, is the sum of squared deviation of observations from their respective sample means (i.e. predictors), given by:

$$SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 \tag{22}$$

3. Mean Sum of Squares, *MSS*, also referred to as the regression sum of squares, is the sum of squared deviations of the sample means (predictors) from the overall mean, given by:

$$MSS = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2 \tag{23}$$

The terms *MSS* and *SSE* can be interpreted as the between-group sum of squares and the within-group sum of squares, respectively. By minimizing *TSS*, estimates are obtained for the explanatory variables that better explain the response variable. This can be obtained by either minimizing *MSS* or *SSE*, as seen in the following equation:

$$TSS = MSS + SSE \tag{24}$$

There are five assumptions associated with the OLS estimator, which were previously listed in Section 4.1.1. Each assumption is described in more detail below, along with methods to assess each:

1. *The dependent variable is linearly associated with the independent variable(s) plus an error term.* Violations of this assumption include having the wrong regressors (either by being omitted or by being irrelevant), nonlinearity, and parameter estimates not being constant. Scatter plots showing the relationship between speed (dependent variable) and each independent variable were used to detect any possible non-linear relationships in the data. The t-test is used to

identify if a variable is associated with the response variable; to identify which variables should be included in the model, a p-value of 0.05 or less was used. In addition, the F-test and the coefficient of determination R^2 are used to provide information about the fit of the model.

2. *The error term has a zero expected value and is normally distributed and uncorrelated with the independent variables*. Plots of residuals against fitted values, normal probability plots, and time sequence residual plots are useful in determining of this assumption is met. The graph of residuals should be randomly scattered and centered around zero and should not show any patterns. A remedial measure to address the violation of this assumption is to transform the variables.
3. *The error terms have equal variances and are not correlated with one another.* Two problems are associated with the violation of this assumption: heteroskedasticity (non-equal variances) and autocorrelated errors. The plot of residuals against the response variable can be used to check for heteroskedasticity; the absolute magnitudes of the residuals should be on average the same for all values of the response variable. In addition, the Breusch-Pagan test is used to check for unequal variances. The null hypothesis is that the error term has a constant variance across all observations. The Durbin-Watson statistic, d, can be used to test for lack of randomness in least squares residuals. The null hypothesis is that no autocorrelation is present among the residuals: when there is no autocorrelation, the d-statistic is approximately 2.0.
4. *The observations on the independent variables are fixed in repeated samples*. Errors in measuring the independent variables and autoregression are problems associated with the violation of this assumption. To confirm if this assumption is met, the independent variables should not be correlated with the error term. The Hausman test is used to test for the equality of the estimates produced by the null and alternative estimators. A p-value of 0.05 or less results in rejecting the null hypothesis of no correlation between the error and the independent variables.

5. *The number of observations is greater than the number of independent variables and the independent variables are not correlated.* Violating this assumption results in multicollinearity. When this happens, the OLS estimates cannot be computed. A correlation matrix between the independent variables can be used to identify which variables have high correlation coefficients. In addition, the diagonal elements of the inverse of the correlation matrix are known as the variance inflation factors (VIF). Variables with VIF higher than 10 indicate harmful collinearity and should not be included in the model.

4.1.2 Panel Data

Panel data analysis has been used previously to investigate speed relationships (Tarris, et al., 1996). In the present context, panel models offer advantages over traditional ordinary least squares (OLS) linear regression models because observations are correlated for the same driver along a transition zone. Past operating speed models that have used a panel data analysis approach are limited to roadways classified as low-speed urban streets and high-speed, two-lane rural highways. No operating speed models currently exist for two-lane rural highway transition zones.

Panel data is a form of longitudinal data in which observations in a sample are collected at two or more points in time. The sample is viewed as a cross-section of drivers where the speed observations are repeated measurements on each driver over time. In this study, driver speeds were observed sequentially at the following four points in time: (1) before the beginning of the transition zone (high-speed area), (2) at the beginning of the transition zone, (3) at the end of the transition zone, and (4) after the end of the transition zone (low-speed area). Figure 9 illustrates the general two-level cluster that represents the panel data structure in this research. In Figure 9, each speed measurement is clustered within an individual driver j. The number of drivers observed varies per data collection site. For the purposes of this dissertation, a disaggregate analysis refers to the instance where all of the individual speed observations are used in the model specification.

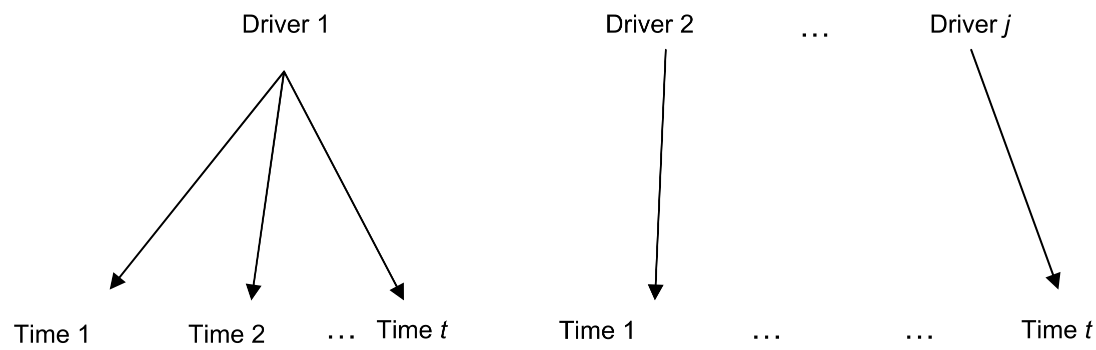

Figure 9 Panel Data Illustration

As illustrated in Figure 9, information on each driver *j* is collected at several time periods *t*. In this particular study, $t = 1, 2, \ldots, T$ where $T = 4$ corresponding to the four sensor locations. When there are no missing observations, the panel is balanced. Since drivers are the clusters, it is expected that the observations within clusters will be correlated (vehicles are tracked, thus the speed data from a specific driver is assumed to be dependent on the previous speed). The advantages of using panel data are as follows (Brüderl, 2005):

- There is more variability, less collinearity, and more degrees of freedom, therefore panel data analysis is considered more informative than other modeling methods when the data contain both cross-section and time elements.
- The estimates are more efficient than the OLS estimator.
- Panel data analysis allows one to study individual driver dynamics by considering unit-specific clusters (i.e., characteristics on individual drivers).
- The time-ordering of individual speed observations are explicitly taken into consideration.
- Individual unobserved heterogeneity (the variation of observations due to variables not included in the model) is accounted for in the model.

It is important to note that, in this study, the study sites *k* produce a third-level cluster as opposed to the two-level cluster represented in Figure 9. The three-level cluster that shows the information on driver *j* is nested in site *k* is shown in Figure 10.

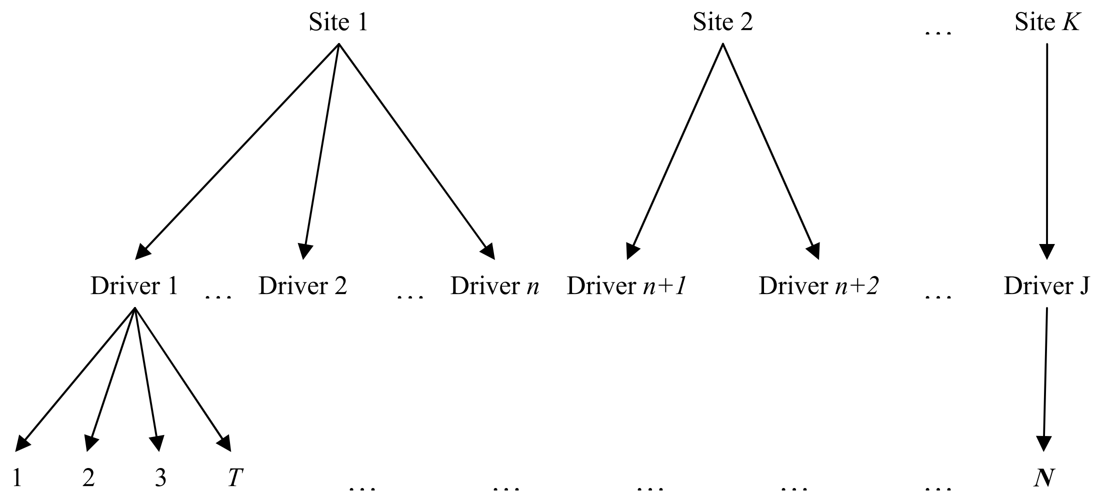

Figure 10 Three-Level Hierarchical Data Structure

In Figure 10, speed observations *i* collected at time *t* are nested within drivers *j*, which are then nested within the study sites *k*. Figure 10 also shows that driver speed observations are site-specific, meaning that drivers only traveled through one of the study sites therefore no speed data are available for other sites for the same driver. The variable for study site cannot be included in the panel model specification because the matrix of regressors (explanatory variables) would produce perfect collinearity with the study site variable, and would thus eliminate the possibility of exploring the association of roadway, roadside, land use, and traffic control characteristics on operating speeds.

In addition to the disaggregate-level analysis referred to previously, this dissertation also considers an aggregate-level panel data model in order to compare the coefficient estimates between the two datasets as well as measures of model efficiency (coefficient of determination, t-statistics). Aggregation is done by computing a mean operating speed for all drivers at each sensor location within a study site. Although past operating speed research has used aggregate data to determine the statistical association between vehicle operating speed and roadway design features, there are several limitations associated with aggregating data, thus recent research has considered disaggregate-level data (Park and Saccomanno, 2005; and Misaghi and Hassan, 2005). First, aggregating speed data may result in an "ecologic fallacy", a term that is used to imply that, although conclusions are developed for a group, they may not apply to an individual (Park and Saccomanno, 2005). By aggregating data, some information belonging to individual drivers is lost. Although using aggregate data may improve the goodness-of-fit of a statistical model (e.g., coefficient of determination), doing so may introduce a major source of uncertainty. Data aggregation may also bias the parameter estimates of a statistical model (Garrett, 2003). Nevertheless, both disaggregate and aggregate models of point speed for panel data models were specified in this dissertation to compare the results since the ecologic fallacy issue has only be addressed for OLS models.

Before introducing fixed- and random-effects panel models, consider first the following pooled linear regression model:

$$y_{jt} = \alpha + \beta x'_{jt} + \varepsilon_{jt} \tag{25}$$

where: y_{jt} = speed of driver j at time t, $j = 1, 2, \ldots, n$; $t = 1, 2, \ldots, T$;

α, β = vector of estimable parameters;

x'_{jt} = vector of explanatory variables corresponding to driver j at time t;

ε_{jt} = disturbance term corresponding to driver j at time t.

The ordinary least square (OLS) estimator is appropriate only if the assumptions of the classical linear regression model are met. These include (Greene, 2008):

1. Relationship between set of explanatory variables and dependent variable is linear;
2. Independence across observations i;
3. Conditional mean of the disturbances is zero ($E[\varepsilon_{jt}]=0$);
4. Homoskedastic disturbances ($Var[\varepsilon_{jt}] = \sigma^2$); and
5. Strict exogeneity of x_{jt} ($Cov[\varepsilon_{jt}, \varepsilon_{ls}]$) = 0 if $j \neq k$ or $t \neq s$.

These assumptions are discussed in detail below (see Section 4.2). When estimating a pooled regression model, the regression parameters are constant across drivers and time. The disturbance term (ε_{it}) accounts for the variation that is not explained by the independent variables in the model. In panel data analysis, the effects of omitted variables are collectively significant (Washington et al. 2003). These effects can be absorbed by the intercept for individual drivers, time periods, or both. In the present research, individual driver heterogeneity is accounted for in the model, but time heterogeneity is not considered. The error term in such a model therefore includes a term for the unobserved driver-specific effects.

To account for individual driver effects in the model, the intercept can differ while the slope coefficients (β) are constant across drivers. Two methods can be used to estimate a different intercept for each driver. The first is to use a dummy variable for each driver and use OLS to estimate the model. In the context of the present study, such an approach would require the use of 2,858 dummy variables, which leads to a significant loss in degrees of freedom. Another method is by applying a fixed-effects model to the data set which uses the fixed-effects estimator, also known as the within estimator, and amounts to using OLS in order to estimate the slope coefficients (β) thus being treated as fixed and estimable (STATA Press, 2005). Furthermore, statistical inference can only be made on the drivers included in the sample. The fixed-effects model takes the following general form:

$$y_{jt} = \alpha_j + \beta x'_{jt} + \varepsilon_{jt} \tag{26}$$

where: y_{jt} = speed of driver j at time t, $i = 1, 2, \ldots, n$; $t = 1, 2, \ldots, T$;

α_j = driver-specific constant;

β = vector of estimable parameters;

x'_{jt} = vector of explanatory variables corresponding driver j at time t; and

ε_{jt} = disturbance term corresponding to driver j at time t.

An F-test can be used to test the hypothesis that the individual driver-specific effects (α_j) are equal. The null hypothesis is that that pooled model is the efficient estimator. The fixed-effects model allows the unobserved driver-specific effects to be correlated with the explanatory variables included in the model specification (Greene 2008). If the driver-specific effects are not correlated with the explanatory variables included in the model, it is possible to model the individual driver intercepts as randomly-distributed from a pool of possible intercepts. The random-effects model takes the following general form:

$$y_{jt} = (a + u_j) + Bx'_{jt} + e_{jt} \tag{27}$$

where: y_{jt} = speed of driver j at time t, $i = 1, 2, \ldots, n$; $t = 1, 2, \ldots, T$;

a = constant;

u_j = random heterogeneity specific to j^{th} driver that is constant over time;

B = vector of estimable parameters;

x'_{jt} = vector of explanatory variables corresponding driver j at time t; and

e_{jt} = disturbance term corresponding to driver j at time t.

The following assumptions are associated with the strict exogeneity assumption in the random-effects model (Greene, 2008):

$$E[e_{jt} \mid x] = E[u_j \mid x] = 0 \tag{28}$$

$$E[e^2{}_{jt} \mid x] = \sigma_\varepsilon^2 \tag{29}$$

$$E[u^2{}_j \mid x] = \sigma_u^2 \tag{30}$$

$$E[e_{jt} u_j \mid x] = 0 \text{ for all } j \text{ and } t \tag{31}$$

Let $n_{jt} = e_{jt} + u_j$, so the error components in the random-effects model are:

$$E[n_{jt} \mid x] = \sigma_\varepsilon^2 + \sigma_u^2 \tag{32}$$

Feasible generalized least squares (FGLS) was used to estimate the regression parameters in the random-effects model. A Breusch-Pagan Lagrange multiplier test can be used to test the appropriateness of the random-effects model. The null hypothesis is that the variance component for the driver (σ_u^2) is zero. The test is chi-squared-distributed with one degree of freedom. Rejecting the null hypothesis suggests that the random-effects model is more appropriate than the classical linear regression model.

The Hausman test is used to determine which model, the random-effects or the fixed-effects, is more appropriate. Under the null hypothesis, both OLS in the fixed-effects model and GLS in the random-effects model are consistent, but OLS is not efficient. The Hausman test is used to test the assumption that there is no correlation between the individual driver effects (α_j) and the vector of explanatory variables. The null and alternative hypotheses for the Hausman test, in terms of the covariance for the between-subject residual for the fixed-effects model, α_j, are defined as follows:

H_0 : Cov(x_{jt}, α_j) = 0

H_1 : Cov(x_{jt}, α_j) ≠ 0

Failure to reject the null hypothesis indicates that the random-effects model is favored over the fixed-effects model. Rejecting the null hypothesis favors the fixed-effects model.

The STATA software provides in its output the values for the between-subject and within-subject standard deviations, $\sqrt{\psi}$ and $\sqrt{\theta}$ respectively (STATA Press, 2005). STATA also provides a value for the within-subjects correlation, ρ, given by the equation:

$$\rho = \frac{\psi}{\psi + \theta} \tag{33}$$

where ψ is the between-subject variance and θ is the within-subject variance (σ_ε in the output). If the value of ρ is close to 1, then there are no differences between observations for an individual driver.

4.1.2 Multilevel Models

Similar to panel data analyses, multilevel models are also used in longitudinal studies where the response from an individual are correlated and the data has a clustered structure. In multilevel models, several levels of clusters can be recognized, thus a cluster level may be nested in another cluster level, creating a "super cluster." Multilevel models are able to recognize the data hierarchy while allowing a residual component at each level. The benefits of multilevel models are:

1. The ability of recognizing the hierarchy of the data structure, therefore the estimates and standard errors are more efficient. Underestimating the standard errors can lead to incorrect statistical inferences of the parameters.
2. They are able to provide information about the level variables (i.e., group variables).
3. They allow for prediction of both group effects and the group variable itself simultaneously by adding a dummy variable (i.e. can include the characteristics at each sensor plus a dummy variable for sensor).
4. Each cluster (i.e. group variable) can be treated as a random sample from a population.

5. Multilevel models can allow for non-nested models; they allow for several levels to be "crossed." An example could be drivers nested in sites, where drivers and county of residence are crossed.

Panel data models are only able to accommodate two-level data structures; by setting a panel variable (driver) and a time variable (sensor) it is specified that longitudinal data (speed observations) are nested in drivers. In multilevel analysis, this structure is represented by two levels: speed observations at the lower level which are nested in driver clusters, the higher level. The two-level model has the following functional form:

$$y_{ij} = \beta_0 + \sum_{p=1}^{P} \beta_p X_{pij} + \varsigma_j^{(2)} + \varepsilon_{ij} \tag{34}$$

where: y_{ij} = observation i for driver j;

β_0 = fixed intercept (slope);

$\sum_{p=1}^{P} \beta_p X_{pij}$ = sum of the explanatory variables (X) and their parameter estimates (β);

$\varsigma_j^{(2)}$ = random intercept for level 2 (drivers), with variance $\psi^{(2)}$; and

ε_{ij} = random error term (residual) with variance θ.

The maximum likelihood estimator is used to estimate the parameters in multilevel models. The maximum likelihood method is the joint probability density of all the observed responses (speeds) as a function of the model parameters β, ψ, and θ.

The maximum likelihood estimators are expressed in terms of the model sum of squares, *MSS*, and the sum of squared errors, *SSE*. For a two-level model, expressed in Equation (34), the MSS is the sum of squared deviations of cluster means (drivers) from the overall mean, and is given by:

$$MSS = \sum_{j=1}^{J} \sum_{i=1}^{n} (y_{\cdot j} - \overline{y_{\cdot\cdot}})^2 \tag{35}$$

where $\overline{y_{\cdot\cdot}} = \frac{1}{Jn} \sum_{j=1}^{J} \sum_{i=1}^{n} y_{ij}$ which is the population mean, β.

The *SSE* is the sum of squared deviations of responses from their cluster means, and is given by the following equation:

$$SSE = \sum_{j=1}^{J} \sum_{i=1}^{n} (y_{ij} - \overline{y_{.j}})^2 \tag{36}$$

where $\overline{y_{.j}} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$ which is the mean for a specific cluster.

The maximum likelihood estimators of the within- and between-cluster variances, θ and ψ, are then computed in terms of *MSS* and *SSE*:

$$\hat{\theta} = \frac{1}{J(n-1)} SSE \tag{37}$$

and

$$\hat{\psi} = \frac{MSS}{Jn} - \frac{\hat{\theta}}{n} \tag{38}$$

If the model is true, then the estimators for β and θ are unbiased. The estimator for ψ, however, has downward bias. The unbiased moment of estimator (or ANOVA) of ψ is:

$$\hat{\psi}^{M} = \frac{MSS}{n(J-1)} - \frac{\hat{\theta}}{n} \tag{39}$$

As shown in Figure 10, the dataset created for this research specifies that its structure consists of three levels. The higher level, which corresponds to the site variable, cannot be taken into account in panel data analyses. The class diagram that illustrates the three-level model, which corresponds to the unit diagram shown in Figure 10, is shown in Figure 11.

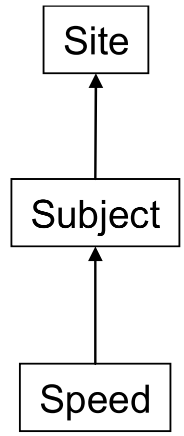

Figure 11 Class Diagram for Multilevel Model Dataset

The equation that describes the three-level unconditional model is:

$$y_{ijk} = \beta_1 + \varsigma_{jk}^{(2)} + \varsigma_{k}^{(3)} + \varepsilon_{ijk} \tag{40}$$

where: y_{ijk} = observation i for driver j at site k;

β_1 = fixed intercept (slope);

$\varsigma_{jk}^{(2)}$ = random intercept for level 2 (drivers), with variance $\psi^{(2)}$;

$\varsigma_{k}^{(3)}$ = random intercept for level 3 (site), with variance $\psi^{(3)}$; and

ε_{ijk} = random error term (residual) with variance θ.

As shown in Equation (40), multilevel models are able to add a random intercept at each level of the data structure. The random part of the three-level model included in Equation (40) is shown in the following equation:

$$y_{ijk} = \varsigma_{jk}^{(2)} + \varsigma_{k}^{(3)} + \varepsilon_{ijk} \tag{41}$$

Equation (41) can be represented by the path diagram shown in Figure 12.

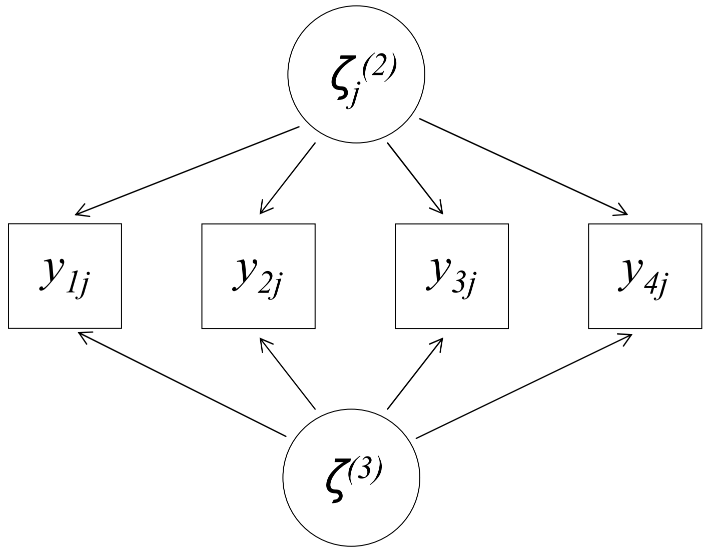

Figure 12 Random Path Diagram for Unconditional Three-Level Model

In the dataset created for this research, for a driver *j* (level 2), nested in site *k* (level 3), there are 4 observations (y_{ij}), which are the observed variables (the responses) inside the rectangular elements shown in Figure 12. The presence of clusters creates possible correlation within the clusters. The variance component terms can be explored by fitting unconditional models, (i.e. models without explanatory variables). The likelihood-ratio test is used to determine if a specific level of the data hierarchy is indeed necessary by fitting the unconditional models, with and without the random intercept for the level in question. A p-value of 0.05 or less indicates that the models fitted are significantly different at the 95 percent confidence level and that the level specified is indeed necessary.

Multilevel models can be classified according to the characteristics of the model components, such as type of response variable and type of structure, among others. The data set for this research is then classified as follow:

1. *Type of response variable*. The response variable is speed, which is a continuous variable with a normal distribution.
2. *Type of data structure.* Speed observations were collected at four sensor locations at twenty study sites. In addition, vehicles were tracked, thus specifying speed observations to specific drivers. The data structure for this

research is then hierarchical in nature with three levels: speeds (level 1) are nested in driver clusters (level 2) and drivers are nested in sites (level 3).

3. *Type of variance structure.* For this research, the model is assumed to be a variance components model, also known as a random intercept model, since only the intercept is assumed to vary randomly across higher levels (drivers and sites). In this model, there is a residual intercept at each level as described previously in relation to Equation (40).
4. *Other*. In this classification the options include models with measurement error, missing data, and spatial models. For this research, a spatial model seems appropriate since this type of model is able to account for driver clusters to be crossed with sites (i.e., there is speed information for a specific driver at more than one site). Besides speed information, no additional driver-specific data were collected, thus it was assumed that drivers were site-specific.

4.1.3 Generalized Estimating Equations (GEE)

Another analysis method used in this dissertation to estimate speeds of individual drivers over time is the generalized estimating equations. Generalized estimating equations (GEE) are used in longitudinal models when there is correlation among the sample data. GEE is an extension of the generalized linear model (GLM), but instead of using maximum likelihood theory for independent observations, GEE is based on quasi-likelihood estimation which allows for overdispersion of data (greater variability). Zeger and Liang (1986) described the GEE method for discrete and continuous outcomes. The method has been used in transportation research primarily to model crash occurrence (Abdel-Aty and Wang, 2006; and Lord and Persaud, 2000); however, it has not been applied to speed data which are continuous, normally-distributed data.

GEE models are population-averaged (marginal) models rather than conditional (cluster-specific) models such as the panel models described previously (Zorn, 2001). In the former, the regression parameters represent the average effect of the explanatory variables across the population on the dependent variable. Alternatively, the regression parameters in a conditional model represent the effect of a change in the explanatory variables on the dependent variable for an individual driver. In GEE, few subpopulations are thought to exist and they can be identified as having shared values for the independent variables (Ghisletta and Spini, 2004). This is partially true of drivers on two-lane rural highways in central Pennsylvania – there are likely few subpopulations and the independent variables across many of the study sites in the present research are alike (e.g., lane width, posted speed limits, regulatory speed limit changes, land use characteristics, etc.).

In generalized linear models (GLM), the probability density of the response *Y*, which is assumed to have exponential form, is expressed as:

$$f(y) = \exp\left\{\frac{y\theta - b(\theta)}{\alpha(\phi)} + c(y,\phi)\right\} \tag{42}$$

for some functions *a*, *b*, and *c* that determine the specific distribution. The mean and the variance of *Y* are:

$$E(Y) = b'(\theta) \tag{43}$$

$$Var(Y) = \frac{b''(\theta)\phi}{\omega} \tag{44}$$

In GLM, the probability distributions of the response *Y* are parameterized in terms of the mean μ and dispersion parameter ϕ as opposed to the natural parameter *θ*. Several combinations of family and link options are available. A GEE model with Gaussian family and identity link is the basic GLM model. The probability function for the normal (Gaussian) family can be expressed as:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \tag{45}$$

for $-\infty < y < \infty$.

The variance of the response *Y* is:

$$Var(Y) = \phi = \sigma^2 \tag{46}$$

A link function (*g*) is used to relate the mean of the i^{th} observation to a linear predictor ($x_i'\beta$) as follows:

$$g(\mu_i) = x_i'\beta \tag{47}$$

where: x_i = vector of explanatory variables; and

β = vector of estimable regression parameters

Log-likelihood functions for the distributions are parameterized in terms of means μ_i and the dispersion parameter ϕ and are of the form:

$$L(y,\mu,\phi)=\sum_i \log\left(f\left(y_i,\mu_i,\phi\right)\right) \tag{48}$$

where the sum is over the observations; each individual contribution is:

$$l_i=\log\left(f\left(y_i,\mu_i,\phi\right)\right) \tag{49}$$

For the normal (Gaussian) family, the individual contributions l_i, which are expressed in terms of the mean and dispersion parameters, are:

$$l_i=-\frac{1}{2}\left[\frac{\omega_i\left(y_i-\mu_i\right)^2}{\phi}+\log\left(\frac{\phi}{\omega_i}\right)+\log\left(2\pi\right)\right] \tag{50}$$

In the generalized estimating equations framework (GEE), there are repeated observations made on the same subject. As such, let Y_{jt} be the response variable (speed) on subject (driver) j during time period t, which corresponds to the sensor locations (where j = 1, 2, …, J and t = 1, 2, …, T). Because the data are correlated, the covariance structure of the data is modeled in GEE. The link function and the linear predictor shown in equations (47) and (50) are the same in the GEE framework except that the vector of explanatory variables includes both the driver j and time t. To estimate the vector of regression parameters, the following equation is used:

$$S(\beta)=\sum_{i=1}^{K}\frac{\partial\mu_i^{'}}{\partial\beta}V_i^{-1}[Y_i-\mu_i(\beta)]=0 \tag{51}$$

The primary benefit of GEE models is that they can account for the correlation within clusters. In the GEE framework, $R_i(\alpha)$ is a working correlation matrix with n_i by n_i dimensions. The covariance matrix of the response variable is modeled as:

$$V_i=\phi A_i^{1/2}R(\alpha)A_i^{1/2} \tag{52}$$

where A_i = n_i by n_i diagonal matrix with $v(\mu_{it})$ as the t^{th} diagonal element.

Four working correlation structures can be considered for GEE models:

1. *Independent*. The observations for a cluster (driver) are independent of each other, therefore the GEE estimates are the same as the regular GLM but with different standard errors. The working correlation matrix is diagonal in this

case. Letting y_{jt} be the t^{th} observation on the j^{th} driver, the correlation between two observations is: $Corr(y_{jt}, y_{jk}) = \begin{cases} 1 & t = k \\ 0 & t \neq k \end{cases}$ and the correlation matrix, V_i, for t=4 is $V_i(4 \times 4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

2. *Exchangeable*. All measurements are equally correlated (the correlations are constant within a driver). In this case the correlation between two observations is $Corr(y_{jt}, y_{jk}) = \begin{cases} 1 & t = k \\ \alpha & t \neq k \end{cases}$ and the correlation matrix for a given cluster, assuming four observations per subject, is

   $$V_i(4 \times 4) = \begin{bmatrix} 1 & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & 1 \end{bmatrix}.$$

3. *Autoregressive*. The correlations between observations for each subject depend on the distance between measurements; as the distance/time increases between the time periods, the correlation decreases. The correlation for any two observations is $Corr(y_{jt}, y_{j,t+n}) = \alpha^n, n = 0,1,2,\ldots,n_j - t$ and the correlation matrix for a given cluster is $V_i(4 \times 4) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha & 1 & \alpha & \alpha^2 \\ \alpha^2 & \alpha & 1 & \alpha \\ \alpha^3 & \alpha^2 & \alpha & 1 \end{bmatrix}$.

4. *Unstructured*. No assumptions about the correlations, thus the correlation between any two observations for a driver are different. The correlation between two observations can be identified as $Corr(y_{jt}, y_{jk}) = \begin{cases} 1 & t = k \\ \alpha_{tk} & t \neq k \end{cases}$

while the correlation matrix can be viewed as

$$V_i(4\times 4)=\begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & 1 & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & 1 & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 1 \end{bmatrix}.$$

Any multicollinearity issues with the data do not violate any assumptions of the GEE models and do not cause biased, inefficient, or inconsistent estimators; only the standard errors are affected. For model verification, a plot of the residuals should not show any patterns if the model is specified correctly. For model (and working correlation matrix) selection, the quasi-likelihood under independence model criterion (QIC) proposed by Pan (2001) and the marginal coefficient of determination (marginal R-square or $R^2{}_m$) values are used.

In GLM, the AIC criterion is used to choose the best model. The AIC criterion cannot be applied to GEE models due to the (possible) violation of normally-distributed data and independency. Instead, the QIC criterion, an alternate method to the AIC criterion, can be used in GEE models. The QIC is a function of the working correlation matrix; it is used to identify which is the best correlation structure. The model with the smallest value for QIC is the best GEE model.

Another model selection method is the marginal R-square, R^2m, which follows the theory of coefficient of determination for OLS regression, and can also be applied for selection of the best working correlation matrix in GEE models. In addition to the QIC criterion, the marginal R-square, R^2m, can provide additional information about the fitness of the model (James Cui, 2007). The marginal R-square, R^2m, is given by the following equation:

$$R^2m = 1-\frac{\sum_{j=1}^{K}\sum_{t=1}^{n_i}(Y_{jt}-\hat{Y}_{jt})^2}{\sum_{j=1}^{K}\sum_{t=1}^{n_i}(Y_{jt}-\overline{Y}_{jt})^2} \tag{53}$$

where Y_{jt} is the observation for subject j at time period t, $\hat{Y}_{jt}$ is the predicted value (obtained from the model developed under consideration), and $\overline{Y}_{jt}$ is the marginal mean across all time periods and given by the following equation:

$$\overline{Y}_{jt} = \frac{1}{nT} \sum_{t=1}^{T} \sum_{j=1}^{n} Y_{jt} \tag{54}$$

The R^2m value measures the fitness of the model being tested by comparing it to the null model. The marginal R-square then is the defined as the amount of variance in the response variable that is explained by the fitted model.

4.2 Speed Differential Analysis

An alternative method to explore the relationship between operating speeds and roadway, roadside, and land use characteristics along transition zones is to consider the change in speed between sensors as the dependent variable. In the following analyses it was hypothesized that changes in the driving environment are responsible for changes in driving behavior, thus differences in roadway characteristics were included as explanatory variables in several operating speed differential models. Ordinary least squares (OLS) linear regression and multilevel model analyses were used to estimate speed reductions in transition zones along two-lane rural highways.

By developing a speed differential prediction model, the response variable is identified as the change in speed between the limits of the transition zone. The new dataset created consists of one observation per vehicle, eliminating driver-specific characteristics and correlated observations. Since the assumption of independent observations is no longer expected to be violated, OLS analysis can be applied to the dataset. The methodology for OLS analyses was previously discussed in Section 4.1.1.

One observation per vehicle (driver) suggests that longitudinal models such as panel data and GEE models are not longer appropriate. However, the site cluster is still present, thus a two-level model in which speed observations are nested in sites can also be considered in order to explore the highway characteristics that influence changes in operating speeds along transition zones. The general form of a two-level model was previously expressed in Equation (34) in section 4.1.2 of this chapter. Equation (34), however, specified that point speed observations (level 1) were nested in drivers (level 2). When considering speed differentials along the transition zone, only one observation per driver is available, thus the driver cluster no longer exists. The two-level model for predicting speed differentials along transition zones considers changes in operating speeds for each driver at the lowest level (level 1), which are nested in sites (level 2). The theory of the maximum likelihood estimator for two-level models is also included in section 4.1.2 in this chapter (see Equations [35] to [39]).

CHAPTER 5
DATA ANALYSIS RESULTS

The data collected from the Hi-Star sensors were carefully examined in order to include only information from free-flow passenger vehicles. Those vehicles with headways less than five seconds were excluded from the data set as well as those with vehicle lengths greater than 20 feet. Only vehicles whose speed information was available at all four sensor locations were considered for the analyses. The final dataset consisted of 11,436 point speed observations from 2,859 identified drivers distributed across 20 study sites. In addition, a separate dataset was created that considered only speed changes between the limits of the transition zone. This chapter discusses the results from the data analyses performed for both point speeds and speed differentials.

5.1 Point Speed Analysis Results

This section of the chapter discusses the development of speed prediction models that considered point speed observations as the response variable. The results from OLS regression, panel data analysis, multilevel models, and generalized estimating equations (GEE) are discussed in this section.

5.1.1 Correlation Analyses

Correlation analyses were undertaken in order to initially indentify the highway characteristics that were associated with speeds in transition zones. The variable most highly correlated with speed observations was posted speed limit (correlation value of 0.51). Other variables identified as potential variables in statistical model building, along with their correlations values with the response variable (operating speed), were:

- Total number of driveways (-0.29)
- Presence of curb (-0.26)
- Number of warning signs (-0.23)
- Presence of Intersection Ahead warning sign (-0.26)
- Presence of school/children related warning sign (-0.19)

All other variables had correlation values less than an absolute value of 0.2.

In addition, correlations between explanatory variables were explored in order to assess potential multicollinearity problems due to the inclusion of two correlated explanatory variables. Any issues related to collinear explanatory variables included in the model-building process are described in subsequent sections of this dissertation.

5.1.2 Ordinary Least Squares

Linear regression has been the most common method used to estimate speed prediction models, as previously discussed in Chapter 2. Since the dataset created for point speed observations consists of correlated observations (four observations per driver), the independency assumption for linear regression models is expected to be violated. Nevertheless, an OLS regression model was estimated in this section for the following two reasons: (1) to obtain initial insights regarding which highway characteristics influence operating speeds along transition zones, and (2) to compare the longitudinal models estimated in this research to the more traditional OLS regression model. An initial OLS regression model was estimated using a backward elimination procedure. A correlation matrix was computed to verify that any two independent variables were not significantly correlated. In addition, variance inflation factors (VIF) were calculated to detect multicollinearity. The correlation matrix indicated low correlation levels among the independent variables included in the OLS regression model (less than an absolute value of 0.4) and the VIF values were all less than 10. Both methods indicated that no collinear variables were present in the model specification. The results of the OLS regression model are summarized in Table 15.

Table 15 Linear Regression Model Results

Parameter	Estimate	SE	t	p-value	VIF
Speed Limit 25 mph	-12.62	0.327	-38.62	<0.001	2.1
Speed Limit 35/40 mph	-2.71	0.210	-12.93	<0.001	2.3
Speed Limit 45 mph	-5.76	0.283	-20.37	<0.001	1.4
Lane Width Addition	2.03	0.138	14.68	<0.001	1.3
Lateral Clearance	0.02	0.014	1.43	0.152	1.1
Total Driveways	-1.10	0.044	-24.87	<0.001	1.3
Curb	-4.48	0.238	-18.79	<0.001	1.8
Intersection WS	-1.76	0.272	-6.48	<0.001	1.3
School/Children WS	-2.82	0.266	-10.61	<0.001	1.4
Curve WS	2.38	0.237	10.00	<0.001	1.2
Curve with WS	-0.73	0.224	-3.27	0.001	1.7
Curve without WS	-0.67	0.186	-3.63	<0.001	1.5
Constant	50.91	0.269	189.59	<0.001	-
Analysis of Variance					
Source	**df**	**SS**	**MS**	**F**	
Model	12	351936.4	29328.0	540.62	
Residual	11423	619683.9	54.2		
Total	11435	971620.4	85.0		

The coefficient of determination, R^2, for the OLS model shown in Table 15 is 0.3622, indicating that 36 percent of the variance in speed observations can be explained by the model. The result of the F-test shown in the Analysis of Variance table indicates that the null hypothesis that the parameter estimates, including the constant, are zero is rejected, thus there is an association between the independent variables and the response variable. All except one highway feature are statistically significant at the 95 percent confidence level; lateral clearance is statistically significant at the 80 percent confidence level as indicated by its p-value. As shown in Table 15, the variables that are associated with higher operating speeds are:

- *Lane Width Addition*: a one-foot increase in lane width is associated with a mean operating increase of 2 mph.
- *Lateral Clearance*: for each one-foot increase in lateral clearance, a 0.02 mph mean operating speed increase is expected.
- *Curve Ahead Warning Sign*: the presence of this warning sign is associated with a mean speed increase of 2.4 mph when compared to the baseline of no

warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.

The parameter estimate for the presence of a Curve Ahead warning sign ("Curve WS" variable in Table 15) is not consistent with engineering intuition since it is associated with a mean speed increase. However, it was observed that these warning signs are located before the beginning of a horizontal curve (i.e., along the approach tangent) where vehicle operating speeds tend to be higher relative to speeds within a horizontal curve.

The highway features associated with mean speed reductions are:

- *Speed Limit 25 mph*: a posted speed limit of 25 mph is associated with a mean speed reduction of 12.6 mph when compared to the baseline of 55 mph.
- *Speed Limit 35/40 mph*: a posted speed limit of either 35 or 40 mph reduces the mean operating speed by 2.7 mph when compared to the baseline of 55 mph.
- *Speed Limit 45 mph*: a posted speed limit of 45 mph is associated with a mean speed decrease of 5.8 mph when compared to the baseline of 55 mph.
- *Total Driveways:* a mean speed reduction of 1.1 mph is expected per one-unit increase in driveway density.
- *Curb* : the presence of a curve is associated with a mean speed reduction of 4.5 mph when compared to the baseline of no curb.
- *Intersection Ahead Warning Sign*: the presence of this sign is associated with a mean speed reduction of 1.8 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.
- *School/Children Warning Sign:* the presence of this sign is associated with a mean speed reduction of 2.8 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.
- *Curve with Warning Sign:* the presence of a horizontal curve that warrants a warning sign is associated with a mean speed reduction of 0.7 mph when compared to the baseline of a tangent roadway section.

- *Curve without Warning Sign*: the presence of a horizontal curve that does not warrant a warning sign is associated with a mean speed reduction of 0.7 mph when compared to the baseline of a tangent roadway section.

The parameter estimates for the speed limit variables may appear to be counterintuitive since a speed limit of 35 or 40 mph should be associated with lower operating speeds when compared to a speed limit of 45 mph. However, it is important to note that speed limits of 35 and 40 mph were, with exception of one site (Site ID 14 in Table 10), encountered in the low-speed section (sensor locations 3 and 4). Highway features that were only encountered in the low speed zone, such as the presence of a curb, may be associated with the lower operating speeds estimated by the 35 and 40 mph posted speed limit variable in the model. This may be an indication that the interaction between highway features and the posted speed limit variables should be explored. However, the purpose of this investigation was to explore the main effects that are influential on operating speeds along transition zones.

The inclusion of indicator variables for speed limit was preferred over the speed limit as a continuous variable. Use of the indicator variables resulted in a higher value for the coefficient of determination (0.3622 as compared to 0.3168 when including speed limit as a continuous variable). Also, the root mean square error is lower when considering indicator variables for speed limit when compared to the root mean square error when considering speed limit as a continuous variable (7. 36 as opposed to 7.62). These are indications that the use of indicator variables for speed limit results in a better model fit.

Linear regression assumes that the speed observations are independent; since driver-specific data are included in the dataset, this assumption is violated. It is expected that the speed from a particular driver at a downstream location is dependent on the speed at an upstream location for the same driver. Therefore, speed prediction models that are able to account for correlation among observations are preferred.

To test for the assumption of equal variances among the errors in the OLS regression model, the Breusch-Pagan test was used. A χ^2 value of 1.09, corresponding to a p-value of 0.2968, was obtained. Therefore, the null hypothesis of equal variances is not rejected and it can be concluded that the data are of homoskedastic nature and no transformations are necessary.

To test for the assumption of no autocorrelation among the residuals, the Durbin-Watson test was performed resulting in a value, *d*, of 1.077. This was indicative that positive autocorrelation was present in the model. In order to correct for this problem, the Prais-Winsten approach was performed. In the Prais-Winsten procedure, the error term for a particular period is assumed to be linearly associated with the error term at a previous period. However, the lag variable cannot be calculated for the first observation, resulting in loss of observations. Prais-Winsten regression generates values for the lost observations and recalculates the Durbin-Watson statistic. The model developed using Prais-Winsten regression is shown in Table 16.

Table 16 Prais-Winsten Speed Prediction Model

Parameter	Estimate	SE	t	p-value
Speed Limit 25 mph	-11.48	0.368	-31.22	<0.001
Speed Limit 35/40 mph	-2.21	0.177	-12.49	<0.001
Speed Limit 45 mph	-4.62	0.325	-14.23	<0.001
Lane Width Addition	2.33	0.198	11.77	<0.001
Lateral Clearance	0.15	0.010	14.95	<0.001
Total Driveways	-1.07	0.034	-31.48	<0.001
Curb	-4.00	0.227	-17.60	<0.001
Intersection WS	-2.40	0.226	-10.63	<0.001
School/Children WS	-1.31	0.199	-6.56	<0.001
Curve WS	1.28	0.180	7.11	<0.001
Curve with WS	-2.64	0.196	-13.46	<0.001
Curve without WS	-1.25	0.155	-8.08	<0.001
Constant	49.22	0.358	137.34	<0.001
Analysis of Variance				
Source	**df**	**SS**	**MS**	**F**
Model	12	304530.6	25377.6	610.74
Residual	11423	474650.6	41.6	
Total	11435	779181.2	68.1	

The model developed using the Prais-Winsten procedure had a coefficient of determination, R^2, of 0.3908 indicating that almost 40 percent of the variation in speeds is explained by the model. The Durbin-Watson statistic, *d*, for the corrected model was

2.20, a value close to 2.0 which is indicative of no autocorrelation. The values for the parameter estimates were very similar to those obtained using the OLS estimator as indicated by the consistency in signs and by the small differences in magnitudes. The interpretations of the parameters obtained using the Prais-Winsten regression procedure are:

- *Speed Limit 25 mph*: a posted speed limit of 25 mph is associated with a mean speed decrease of 11.5 mph when compared to the baseline of 55 mph.
- *Speed Limit 35/40 mph*: a posted speed limit of either 35 or 40 mph reduces mean speed by 2.2 mph when compared to the baseline of 55 mph.
- *Speed Limit 45 mph*: a posted speed limit of 45 mph is associated with a mean speed decrease of 4.6 mph when compared to the baseline of 55 mph.
- *Lane Width Addition*: a one-foot increase in the lane width at a study segment is associated with a 2.3 mph increase in the mean operating speed.
- *Lateral Clearance*: the mean speed increases by 0.15 mph for each one-foot increase in lateral clearance.
- *Total Driveways*: the mean speed decreases by 1.1 mph for each one-unit increase in the number of driveways within a study segment.
- *Curb*: the presence of a curb is associated with a mean speed reduction of 4 mph when compared to the baseline of no curb.
- *Intersection Ahead Warning Sign*: the presence of this sign is associated with a mean speed reduction of 2.4 mph when compared to the baseline. The baseline in this case is no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.
- *School/Children Warning Sign*: the presence of a sign related to the presence of a school or children is associated with a mean speed reduction of 1.3 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.
- *Curve Ahead Warning Sign*: the presence of this sign is associated with a mean speed increase of 1.3 mph when compared to the baseline of no warning

sign or presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.

- *Curve with Warning Sign*: a horizontal curve that warrants a Curve Ahead warning sign is associated with a mean speed reduction of 2.6 mph when compared to the baseline of a tangent section.
- *Curve without Warning Sign*: a horizontal curve without a warning sign is associated with a mean speed reduction of approximately 1.3 mph when compared to the baseline of a tangent section.

The Prais-Winsten procedure is able to produce a regression model that corrects for autocorrelated error terms. Because individual driver speeds were tracked through each data collection site, the OLS regression assumption of independent observations is violated. Longitudinal models consider this issue and are the focus of the remainder of this section on point speeds.

5.1.3 Panel Data Analysis Results

As previously noted, panel data are repeated measures on one or more subjects. The sensor locations permitted vehicles to be "tracked", thus speed data were available for each driver at each of the sensor locations. Therefore, the variable "driver," which corresponds to subject *j* mentioned in the analysis methodology, was set as the panel variable while the time variable was the "sensor" data collection point.

An initial investigation of the explanatory variables that were correlated with vehicle operating speeds was performed in order to examine the variables that should be considered in the model when performing panel data analysis. An iterative process in which various predictors (explanatory variables) were considered was performed, while examining the consistency of their coefficients in estimating various panel models. The variables found to be statistically significant were: speed limit (indicator), lateral clearance (continuous) , total number of driveways (continuous), presence of curb (indicator), presence of intersection ahead warning sign (indicator), presence of school/children related warning sign (indicator), presence of curve ahead warning sign (indicator), and presence of horizontal curve with and without a warning sign (indicator).

A fixed-effects model was initially used in the present analysis. For this analysis, drivers were nested in sites. The STATA software was used to estimate the fixed-effects panel data model. In addition to the parameter estimates, STATA also provides the results of an F-test that can be used to test the null hypothesis that the constant terms are equal across units, as well as information on the between- and within-subject variances. The results of the fixed-effects panel data model and the comparison between this model and the OLS linear regression model developed previously with the Prais-Winsten approach are shown in Table 17.

Table 17 Fixed-Effects Panel Data Model

Parameter	Fixed-Effects Panel Data			OLS Model	
	Estimate	St. Error	t	Estimate	St. Error
Speed Limit 25 mph	-10.46	0.537	-19.49	-11.48	0.368
Speed Limit 35/40 mph	-2.20	0.173	-12.71	-2.21	0.177
Speed Limit 45 mph	-3.41	0.481	-7.09	-4.62	0.325
Lane Width Addition	3.49	0.354	9.85	2.33	0.198
Lateral Clearance	0.16	0.011	15.33	0.15	0.010
Total Driveways	-0.95	0.034	-27.69	-1.07	0.034
Curb	-4.01	0.235	-17.09	-4.00	0.227
Intersection WS	-1.91	0.228	-8.36	-2.40	0.226
School/Children WS	-1.08	0.199	-5.43	-1.31	0.199
Curve WS	0.84	0.186	4.51	1.28	0.180
Curve with WS	-3.46	0.197	-17.51	-2.64	0.196
Curve without WS	-1.68	0.164	-10.25	-1.25	0.155
Constant	47.05	0.604	77.95	49.22	0.358
Sigma_u ($\sqrt{\psi}$)	6.2022			-	
Sigma_e ($\sqrt{\theta}$)	5.007			-	
Rho (ρ)	0.6054			-	
R^2 within	0.4723			-	
R^2 between	0.2220			-	
R^2 overall	0.3266			0.3908	
F-test	$F(2858, 8565) = 5.65$			$F(12, 11423) = 360.68$	

All of the parameter estimates for the fixed-effects panel data model shown in Table 17 have p-values less than 0.05, indicating that each explanatory variable is statistically significant at the 95-percent confidence level. The signs for these estimates are consistent across both models. The variables with the highest differences in their magnitudes as well as the values for the standard error were the indicators variables for a posted speed limit of 25 mph, a posted speed of 45 mph, and lane width addition (absolute differences of 1.02, 1.21, and 1.16, respectively). The variables for both speed limits (25 mph and 45 mph) indicate greater speed reductions for the Prais-Winsten regression approach while the variable for lane width addition indicates a greater speed increase when estimating the fixed-effects panel data model. All other parameter estimates differed by an absolute value less than 0.82 and had almost identical values for the standard errors. Interpretation of the parameter estimates for the fixed-effects panel data model are:

- *Speed Limit 25 mph*: a posted speed limit of 25 mph is associated with a mean speed decrease of 10.5 mph when compared to the baseline of 55 mph.
- *Speed Limit 35/40 mph*: a posted speed limit of either 35 or 40 mph reduces mean speed by 2.2 mph when compared to the baseline of 55 mph.
- *Speed Limit 45 mph*: a posted speed limit of 45 mph is associated with a mean speed decrease of 3.4 mph when compared to the baseline of 55 mph.
- *Lane Width Addition*: a one-foot increase in the lane width at a study segment is associated with a 3.5 mph increase in the mean operating speed.
- *Lateral Clearance*: mean speed increases by 0.16 mph for each one-foot increase in lateral clearance.
- *Total Driveways*: mean speed decreases by nearly 1 mph for a one-unit increase in the number of driveways within a study segment.
- *Curb*: the presence of curb is associated with a mean speed reduction of 4 mph when compared to the baseline of no curb.
- *Intersection Ahead Warning Sign*: the presence of this sign is associated with a mean speed reduction of 1.9 mph when compared to the baseline. The baseline in this case is no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.
- *School/Children Warning Sign*: the presence of a sign related to the presence of a school or children is associated with a mean speed reduction of 1 mph

when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or presence of children.

- *Curve Ahead Warning Sign*: the presence of this sign is associated with a mean speed increase of 0.84 mph when compared to the baseline of no warning sign or presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Curve with Warning Sign*: a horizontal curve that warrants a Curve Ahead warning sign is associated with a mean speed reduction of 3.4 mph when compared to the baseline of a tangent section.
- *Curve without Warning Sign*: a horizontal curve without a warning sign is associated with a mean speed reduction of approximately 1.7 mph when compared to the baseline of a tangent section.

The overall coefficient of determination, R^2, for the fixed-effects panel data model is lower than the selected OLS linear regression model (0.33 as compared to 0.39). This was expected since the OLS linear regression model had smaller values for the standard errors, thus suggesting a better fit which is reflected in the R^2 value. The panel data model also produces R^2 values for the variance within and between drivers: 0.47 and 0.22, respectively. This indicates that the panel data model developed explains 47 percent of the variance associated with the driver cluster while explaining 22 percent of the variance associated with different drivers (from driver to driver).

For the fixed-effects panel data model, the F-test statistic results in a value of 5.65, thus the null hypothesis is rejected indicating that there are differences between individuals (drivers *j*) and there is individual-specific heterogeneity. Therefore, a pooled model would produce inconsistent estimates suggesting that use of a panel data model is favored over a pooled model.

The output from STATA for panel data models also provides the value of *Sigma_u* and *Sigma_e*, which correspond to between-subject standard deviation ($\sqrt{\psi}$) and the within-subject standard deviation ($\sqrt{\theta}$), respectively. A standard deviation of 6.2 mph is associated with different drivers while a standard deviation of 5 mph is associated with the presence of the driver cluster. The intraclass correlation, ρ (*rho* in the output) represents the within-cluster correlation. If the intraclass correlation is close to 1, it indicates that there are no differences between observations for each subject (i.e., speed at sensor locations are the same). The value for the intraclass correlation of the fixed-effects panel data model was 0.6054, indicating that observations for a specific driver are not similar, which it was expected since it was hypothesized that speeds vary along the studied transition zones. The intraclass correlation value is then interpreted as 60 percent of the variance in speed that is not explained by the covariates is due to time-invariant driver-specific characteristics.

In order to confirm if the fixed-effects model was indeed appropriate, the random-effects model was also estimated using the same variables as the fixed-effects specification and a Hausman test was performed. A comparison between both fixed- and random-effects models and the selected OLS model is shown in Table 18.

Table 18 Fixed-Effects and Random-Effects Comparison

Parameter	Random-effects		Fixed-effects		OLS	
	Estimate	St. Error	Estimate	St. Error	Estimate	St. Error
Speed Limit 25 mph	-12.04	0.333	-10.46	0.537	-11.48	0.368
Speed Limit 35/40 mph	-2.52	0.164	-2.2	0.173	-2.21	0.177
Speed Limit 45 mph	-4.85	0.290	-3.41	0.481	-4.62	0.325
Lane Width Addition	2.14	0.178	3.49	0.354	2.33	0.198
Lateral Clearance	0.12	0.010	0.16	0.011	0.15	0.010
Total Driveways	-0.97	0.034	-0.95	0.034	-1.07	0.034
Curb	-3.79	0.211	-4.01	0.235	-4.00	0.227
Intersection WS	-2.05	0.218	-1.91	0.228	-2.40	0.226
School/Children WS	-1.49	0.200	-1.08	0.199	-1.31	0.199
Curve WS	1.42	0.179	0.84	0.186	1.28	0.180
Curve with WS	-2.41	0.184	-3.46	0.197	-2.64	0.196
Curve without WS	-1.41	0.153	-1.68	0.164	-1.25	0.155
Constant	49.77	0.325	47.05	0.604	49.22	0.358
Sigma_u ($\sqrt{\psi}$)	4.8348		6.2022		-	
Sigma_e ($\sqrt{\theta}$)	5.007		5.007		-	
Rho (ρ)	0.4825		0.6054		-	
R^2 within	0.4684		0.4723		-	
R^2 between	0.2692		0.2220		-	
R^2 overall	0.3522		0.3266		0.3908	

The standard errors for the random-effects panel data model are smaller when compared to those obtained using the fixed effects and OLS regression models. The variables for speed limit indicate greater speed reductions associated with this highway characteristic while the variable for lane width addition is associated with a lower speed increase for the random-effects panel data model.

When comparing the panel data models, the differences in standard errors of the explanatory variables for both models are, for the most part, less than 0.03 mph. For three of the explanatory variables – speed limit 25, speed limit 45, and lane width addition – the difference in standard errors range from 0.18 to 0.2 mph. Similarly, for all except four variables, the parameter estimates between fixed and random effects panel data models are similar, differing by less than a value of 0.6. The variables of speed limit 25, speed limit 45, lane width addition, and presence of horizontal curve with warning sign, differ between the models by a value of 1.58, 1.44, 1.35, and 1.05, respectively. These differences may be evidence that the posted speed limit and lane width addition variables in the random-effects model are picking-up site-specific effects that were not detected using the fixed-effects estimator.

The values of the within-subject standard deviation ($\sqrt{\theta}$) are exactly the same for both the random-effects and the fixed-effects models; a standard deviation of approximately 5 mph is attributed to the residual term. This was expected since the residual term includes the variation not explained by the explanatory variables and both models have the same variables included in the model. The between-subject standard deviation ($\sqrt{\psi}$), however, is higher for the fixed-effects model: a between-driver standard deviation of 6.2 mph was estimated in the fixed-effects models as compared to 4.8 mph variation between drivers in the random-effects model, thus the random-effects model is associated with less variability between drivers. The standard deviation values for the variance components in the between- coefficient of determination for the random-effects model is higher than the one for the fixed-effects model (0.27 as compared to 0.22). Additionally, the overall coefficient of determination in the random-effects models is higher than in the fixed-effects model. This suggests that the random-effects model provides a better fit to the operating speed data collected along the 20 rural highway transition zones.

The random-effects model assumes that the correlation between the predictors and the between-subject error term is zero. In the fixed-effects model, this correlation was found to be 0.0187, indicating very little correlation between the explanatory variables and the variance between drivers. The Hausman test was then performed in order to determine which model specification is preferred. The test resulted in a chi-square statistic (χ^2) of 10,211.31 with a p-value less than 0.0001. As such, the null hypothesis that the random-effects model estimator is consistent is rejected, favoring the fixed-effects model. Because the independent observations assumption of the OLS regression model is violated, and the Hausman test suggests that the parameter estimates from random-effects panel data model are inconsistent, it is recommended that a fixed-effects panel data model is more appropriate to represent the point speed data in the present research.

Some researchers have addressed the implications of including the effects of speed limit when modeling operating speeds that consider the effects of highway geometrics (Wang et al, 2006). In highway design, one of the primary design controls is design speed. Highway design criteria are selected based on the design speed while the posted speed limit may be set at a level that is equal to or less than the designated design speed. Therefore it is expected that highway geometrics may be correlated with the speed limit. Past researchers (Wang et al., 2006) found that including the posted speed limit variable in a regression model significantly changes the statistical inferences that can be made on other explanatory variables in the model. A panel data model was performed without considering the speed limit variable. The results for the fixed-effects model and the comparison with the fixed-effects model including the speed limit variable are shown in Table 19.

Table 19 Fixed-Effects Panel Data Models with and without Speed Limit

Parameter	Without Speed Limit		With Speed Limit	
	Estimate	St. Error	Estimate	St. Error
Speed Limit 25 mph	-	-	-10.64	0.543
Speed Limit 35/40 mph	-	-	-2.19	0.173
Speed Limit 45 mph	-	-	-3.47	0.481
Lane Width Addition	4.08	0.366	3.42	0.355
Paved Shoulder	0.13	0.046	-0.10	0.046
Lateral Clearance	0.16	0.011	0.17	0.011
Total Driveways	-1.13	0.032	-0.96	0.034
Curb	-7.80	0.18	-3.99	0.235
Intersection WS	-3.71	0.228	-2.00	0.232
School/Children WS	-2.71	0.194	-1.05	0.199
Curve WS	-0.39	0.184	0.85	0.186
Curve with WS	-4.01	0.185	-3.37	0.202
Curve without WS	-0.68	0.164	-1.66	0.164
Constant	44.70	0.636	47.43	0.630
Sigma_u ($\sqrt{\psi}$)	7.0329		6.1515	
Sigma_e ($\sqrt{\theta}$)	5.1979		5.0060	
Rho (ρ)	0.6467		0.6016	
R^2 within	0.4312		0.4726	
R^2 between	0.0329		0.2354	
R^2 overall	0.1858		0.3342	
F Test	$F(2858, 8567) = 6.08$		$F(2858, 8564) = 5.52$	

All variables shown in Table 19 are statistically significant at the 95 percent confidence level (p-values less than 0.05). When excluding the posted speed limit variables from the model, the variable for paved shoulder width is statistically significant with a positive coefficient, indicating that the mean speed is expected to increase by 0.14 mph for each additional foot of paved shoulder width provided along the transition zone segment included in the sample dataset. The magnitude of the paved shoulder width variable in the model that includes the posted speed limit indicator variables suggests that a one-foot increase in the paved shoulder width is associated with a 0.1 mph speed decrease, which is not consistent with engineering intuition.

The difference between the parameter estimates for the lane width addition variable is less than 0.7 mph; the model without the speed limit variable indicates a speed increase of 4 mph per foot of lane width addition within the transition zone as compared to 3.4 mph with the model that includes the speed limit indicator variables. When comparing the estimates for the lateral clearance variable, the parameter estimate was almost the same for both models (0.16 versus 0.17). The influence of number of driveways on operating speeds is also very similar between the two models: a 1.1 mph speed reduction per driveway in the model without the posted speed limit as compared to a speed reduction of 1 mph in the model with speed limit variables included.

The indicator variable for the presence of curb had the highest difference between the parameter estimates when comparing the models with and without the posted speed limit variables. When not considering speed limit, the speed reduction associated with this variable is almost 8 mph as compared to a 4 mph speed reduction indicated by the model with the speed limit indicator variables. The speed reduction associated with the presence of an Intersection Ahead warning sign increased from 2 mph (model with speed limit) to 3.7 mph (model without speed limit). Similarly, the speed reduction indicated by the School/Children warning sign increased from 1 mph (model with speed limit) to 2.7 mph (model without speed limit).

Similar to the paved shoulder variable, the variable for presence of Curve Ahead warning sign also resulted in a contradictory interpretation when compared to the previous model developed. In the model without the posted speed limit indicator variables, an expected mean speed reduction of 0.4 mph was estimated for the presence of this sign while in the previous model with the posted speed limit this variable was associated with an expected mean speed increase of approximately 0.8 mph.

The estimate for the presence of curve without a warning sign variable indicated an expected mean speed reduction of 0.6 mph, compared to a mean speed reduction of 1.7 mph for the model that included speed limit indicator variables. For a curve that warranted a warning sign, the mean speed reductions associated with this variable increased from 3.4 mph (model with speed limit) to 4 mph (new model without speed limit variable).

Although most of the parameter estimates are similar when comparing the models with and without the posted speed limit shown in Table 19, the values for the coefficient of determination as well as the between- and within- subject standard deviations differ between the two models. Both between- and within- subject standard deviations are greater for the model that does not include the speed limit (7.03 and 5.2 as compared to 6.2 and 5.0, respectively), indicating that the within- and between-standard deviations are higher in the model without the posted speed limit indicator variables. The higher values for the random component standard deviations for the model without speed limit result in lower values for both between- and within- coefficients of determination (0.03 and 0.43 as compared to 0.22 and 0.47, respectively) and consequently a lower overall coefficient of determination (0.19 as compared to 0.33). This is an indication that the speed limit variable is able to explain more of the variance in the observed speeds, thus the model that includes this variable provides a better fit to the data. However, the model without the posted speed limit indicators produces parameter estimates for all independent variables that are consistent with engineering intuition (i.e., paved shoulder width and Curve Ahead warning sign variables).

It is important to note that, in this study, the study sites create a three-level cluster dataset as opposed to the general two-level cluster in which speed observations per sensor location are nested within subjects. Since a random sample was collected at the study sites, it is assumed that drivers are site-specific (a driver only drives through one of the study sites) thus creating a higher level cluster in which individual drivers are nested within sites. This hierarchy was previously shown in Figure 10 of Chapter 4.

A variable for site cannot be included in any model because then the matrix of the predictors, $\Sigma\beta X_{jt}$, would create perfect collinearity with the study site variable. By eliminating the subject variable (driver) and calculating a mean speed for each of the sensors at each study site, the data can be aggregated into a two-level cluster with observations nested in sites, as shown in Figure 13. The driver-specific information is aggregated and the panel variable is site k with values 1 to 20; the time variable would still be occasion (sensor) t with values 1 to 4. However, several authors have explained the importance of considering disaggregate data and the problems associated with modeling aggregate data (Park and Saccomanno, 2005; Misaghi and Hassan, 2005).

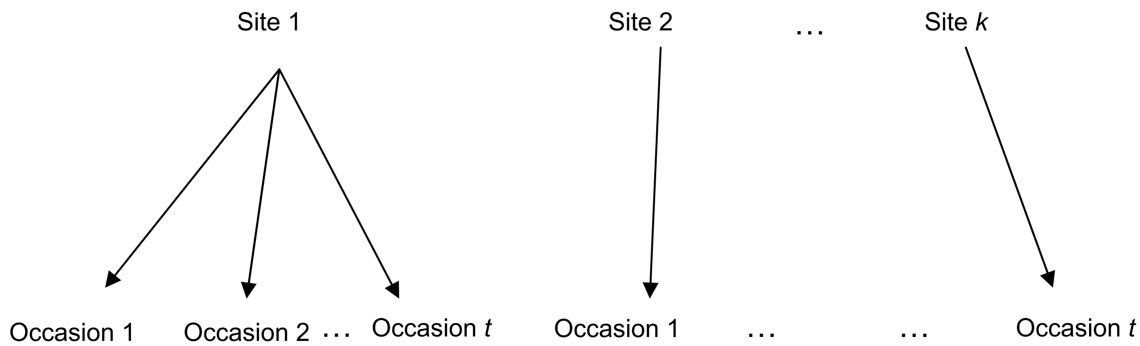

Figure 13 Model hierarchy for Aggregate Data

Regression analyses performed using aggregate data may result in an "ecologic fallacy," a term that is used to indicate that, although conclusions are developed for a group, they may not apply to an individual (Park and Saccomanno, 2005). By aggregating data, some information belonging to the individuals is lost. Although using aggregate data may produce higher values for the coefficient of determination, the use of summarized data introduces a major source of uncertainty. It also may cause higher values of the parameter estimates when, in reality, they should be lower.

While the use of aggregate data is not recommended, the studies that have addressed this issue are related only to OLS regression models; the ecologic fallacy issue has not been explored in longitudinal data analysis, although it is expected that the same recommendation will result. As such, the complications that may arise from the use of aggregate data are further explored using a panel data analysis framework. The dataset was aggregated by calculating mean speeds at each sensor location for each study site. This aggregate dataset consisted of a total of 80 observations corresponding to the four mean speeds collected at each of the four sensor locations for each of the 20 study sites.

Correlation analyses were performed to determine the association between the explanatory variables and the response. These correlation values are shown in Table 20 for both the aggregate and disaggregate data.

Table 20 Correlation Values with Response Variable Mean Speed

Explanatory Variables	Aggregate Data	Disaggregate Data
Speed Limit	0.6960	0.5050
Lateral Clearance	0.2351	0.1334
Total Driveways	-0.4831	-0.2936
Curb	-0.4075	-0.2684
Intersection Warning Sign	-0.3836	-0.2636
School/Children Warning Sign	-0.2592	-0.1922
Curve Ahead Warning Sign	-0.0101	0.0819
Curve with Warning Sign	-0.1096	-0.1059
Curve without Warning Sign	-0.0733	-0.0752

As shown in Table 20, the estimates for the correlations are higher for the aggregate data when compared to the disaggregate data indicating that there is a stronger correlation between the explanatory variables and response in the aggregate dataset. The sign of the coefficients are consistent between the aggregate and disaggregate datasets with the exception of the sign for the Curve Ahead warning sign variable. As noted earlier in this chapter, the parameter estimate for the Curve Ahead warning sign (see Table 17) was positive and counterintuitive. This finding resulted from the use of disaggregate data. Based on the correlations shown in Table 20, the correlation between the Curve Ahead warning sign and operating speed is negative, which is consistent with engineering intuition. A fixed-effects panel data was specified using the aggregate data and compared to the results from the disaggregate data. For comparative purposes, both models are shown in Table 21.

Table 21 Fixed-Effects Panel Data Models for Aggregate and Disaggregate Data

Parameter	Aggregate		Disaggregate	
	Estimate	St. Error	Estimate	St. Error
Speed Limit 25 mph	-10.29	2.635	-10.46	0.537
Speed Limit 35/40 mph	-2.52	0.958	-2.20	0.173
Speed Limit 45 mph	-3.47 *	2.166	-3.41	0.481
Lane Width Addition	4.12	1.873	3.49	0.354
Lateral Clearance	0.17	0.059	0.16	0.011
Total Driveways	-0.89	0.195	-0.95	0.034
Curb	-3.96	1.456	-4.01	0.235
Intersection WS	-2.14 *	1.253	-1.91	0.228
School/Children WS	-0.14 *	1.125	-1.08	0.199
Curve Ahead WS	-0.20 *	1.174	0.84	0.186
Curve with WS	-3.19	1.130	-3.46	0.197
Curve without WS	-1.53 *	0.937	-1.68	0.164
Constant	45.21	3.231	47.05	0.604
Total number of observations, N	80		11436	
Sigma_u ($\sqrt{\psi}$)	3.8196		6.2022	
Sigma_e ($\sqrt{\theta}$)	2.4295		5.0070	
Rho (ρ)	0.7120		0.6054	
R^2 within	0.8263		0.4723	
R^2 between	0.4556		0.2220	
R^2 overall	0.6204		0.3266	
F Test	F(19, 48) = 6.96		F(2858, 8565) = 5.65	
*Not statistically significant at the 0.05 alpha level				

As shown in Table 21, the coefficients for the variables are very similar for the aggregate and disaggregate models indicating that the predictors are associated with operating speeds in a similar manner. The only variable that has an opposite parameter estimate when comparing the aggregate to disaggregate data is the Curve Ahead warning sign variable. As noted previously, the change in sign resulting from the aggregate-level data analysis produces results that are more consistent with engineering intuition. However, three of the explanatory variables (School/Children warning sign, Curve Ahead warning sign, and presence of horizontal curve without a warning sign) were found not to be statistically significant in the aggregate-level model when compared to the disaggregate-level model.

As expected, the coefficients of determination (overall, within- and between-), are greater for the aggregate model when compared to the disaggregate-level model. The values for the between- and within-subject standard deviations (*Sigma_u* and *Sigma_e* in the output, respectively) are greater for the disaggregate model than for the aggregate model, indicating that the aggregate model explains is associated with less variability. Therefore the values for the coefficient of determination are greater, an indication that the aggregate model is a better fit to the data even if this model includes variables that were not statistically significant (speed limit 45 mph, Intersection and Curve Ahead warning signs as well as a School/Children warning sign, and presence of horizontal curve that does not warrant a warning sign).

The majority of the parameter estimates for the explanatory variables are very similar when comparing the two models; however, the standard errors of these estimates are greater for the aggregate model. Therefore, although the coefficient of determination indicates that the aggregate model is a better fit, the estimates obtained are more precise for the disaggregate model as indicted by the small values of standard errors.

Since relying solely on the coefficient of determination for selection of a model that best fits the data may result in an ecologic fallacy, three additional measures of model validation were explored in order to provide a better comparison between the aggregate and disaggregate speed prediction models. These model validation measures are: mean prediction bias (*MPB*), mean square error (*MSE*), and mean absolute deviation (*MAD*). The results of the model validation measures are shown in Table 22.

Table 22 Measures of Fit for the Aggregate and Disaggregate Fixed-effects Panel Data Models

Measure of Fit	Equation	Aggregate Model	Disaggregate Model
Mean Prediction Bias	$MPB = \frac{1}{n}\sum\left(\hat{Y}_i - Y_i\right)$	-0.667	-0.019
Mean Squared Error	$MSE = \frac{1}{n}\sum\left(\hat{Y}_i - Y_i\right)^2$	59.260	57.221
Mean Absolute Deviation	$MAD = \frac{1}{n}\sum\left\lvert\hat{Y}_i - Y_i\right\rvert$	6.116	6.007
where: $\hat{Y}_i$ = predicted value; Y_i = observed value; and n = total number of observations.			

As shown in Table 22, all values for the model validation measures are greater for the aggregate model when compared to the disaggregate model. A mean prediction bias of approximately -0.7 mph is associated with the aggregate model while the disaggregate model is associated with a mean prediction bias of -0.02 mph. The variance associated with the aggregate model is 59.3 mph^2 as compared to 57.2 mph^2 for the disaggregate model. Finally, a mean absolute deviation of 6.1 mph is expected with the use of the aggregate model as compared to a mean absolute deviation of 6 mph when using the disaggregate data. Although the coefficient of determination indicated that the aggregate model was a better fit than the disaggregate model, the additional model validation measures indicate that there is less variability associated with the disaggregate model, thus contradicting the implications given by the coefficient of determination. Based on the findings from the panel data analysis, it is therefore recommended that operating speeds along two-lane rural highway transition zones be modeled using a fixed-effects estimator with disaggregate-level data.

5.1.4 Multilevel Model Analysis Results

Panel data analyses are only able to accommodate two-level data structures while multilevel models can recognize additional hierarchical levels. The two-level variance components model in which speed observations i are nested in drivers j was previously shown in Equation (34). The two-level variance components model was estimated with the multilevel model (xtmixed) and the maximum likelihood (mle) options in Stata, including only the variables found to be statistically significant from the panel data analysis. A comparison between the two-level model and both fixed- and random-effects panel data models is shown in Table 23.

Table 23 Comparison between Two-Level and Panel Data Models

Parameter	**Two-Level Models (MLE)**		**Panel Data**			
			Fixed-effects		**Random-effects**	
	Estimate	**St. Error**	**Estimate**	**St. Error**	**Estimate**	**St. Error**
Speed Limit 25 mph	-11.95	0.340	-10.46	0.537	-12.04	0.333
Speed Limit 35/40 mph	-2.49	0.161	-2.20	0.173	-2.52	0.164
Speed Limit 45 mph	-4.76	0.296	-3.41	0.481	-4.85	0.290
Lane Width Addition	2.19	0.186	3.49	0.354	2.14	0.178
Lateral Clearance	0.13	0.010	0.16	0.011	0.12	0.010
Total driveways	-0.97	0.033	-0.95	0.034	-0.97	0.034
Curb	-3.77	0.208	-4.01	0.235	-3.79	0.211
Intersection WS	-2.05	0.214	-1.91	0.228	-2.05	0.218
School/Children WS	-1.41	0.195	-1.08	0.199	-1.49	0.200
Curve Ahead WS	1.35	0.176	0.84	0.186	1.42	0.179
Curve with WS	-2.54	0.182	-3.46	0.197	-2.41	0.184
Curve without WS	-1.45	0.151	-1.68	0.164	-1.41	0.153
Constant	49.63	0.339	47.05	0.604	49.77	0.325
Random Components						
Driver ($\sqrt{\psi}$)	5.4774		6.2022		4.8348	
Residual ($\sqrt{\theta}$)	5.018		5.007		5.0071	

As shown in Table 23 the estimates obtained using maximum likelihood for the two-level model, as well as their respective standard errors, are very similar to the ones obtained with the random-effects panel data model. When comparing the two-level model with the fixed-effects panel data model, the differences in magnitude of the parameter estimates are greater because the fixed-effects model estimator was used for panel data while the maximum likelihood estimator was used for the multilevel model. However, these estimates are similar for both options and all the parameter estimates are similar in sign and magnitude between the two models. The coefficients for the multilevel model can be interpreted as:

- *Speed Limit 25 mph*: a posted speed limit of 25 mph is associated with a mean speed decrease of approximately 12 mph when compared to the baseline of 55 mph.
- *Speed Limit 35/40 mph*: a posted speed limit of either 35 or 40 mph reduces mean speed by 2.5 mph when compared to the baseline of 55 mph.
- *Speed Limit 45 mph*: a posted speed limit of 45 mph is associated with a mean speed decrease of almost 5 mph when compared to the baseline of 55 mph.
- *Lane Width Addition*: for a minimum lane width of 9 feet, a mean speed increase of 2.2 mph is associated with per every one-foot increase in the lane width.
- *Lateral Clearance*: mean speed increases by 0.13 mph per every 1 ft increase in lateral clearance.
- *Total Driveways*: mean speed decreases by nearly 1 mph for each additional driveway in a transition zone.
- *Curb*: the presence of curb is associated with mean speed reduction of almost 4 mph when compared to the baseline of no curb.
- *Intersection Ahead Warning Sign*: the presence of this sign is associated with a mean speed reduction of approximately 2 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *School/Children Warning Sign*: the presence of a sign related to the presence of school or children is associated with a mean speed reduction of 1.4 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Curve Ahead Warning Sign*: the presence of this sign is associated with a mean speed increase of 1.4 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.

- *Curve with Warning Sign*: a horizontal curve that warrants a Curve Ahead warning sign is associated with a mean speed reduction of 2.5 mph when compared to the baseline of a tangent section.
- *Curve without Warning Sign*: a horizontal curve without a warning sign is associated with a mean speed reduction of approximately 1.4 mph when compared to the baseline of a tangent section.

The values for the within-subject standard deviation ($\sqrt{\theta}$) are also similar across all models. The between-subject standard deviation ($\sqrt{\psi}$) is lower for the two-level model (5.5 mph) when compared to the between-subject standard deviation in the fixed-effects panel model (6.2 mph) but higher when compared to the random-effects panel data model (4.83). This indicates that, when comparing the two-level model that uses the maximum likelihood estimator with the fixed-effects panel data model, the two-level model explains more of the variability between drivers, an indication of a better fit to the observed data.

In addition to the two-level hierarchy presented in Table 23, a three-level multilevel model was developed in order to include one additional hierarchy that corresponds to the site variable. The unit and class diagrams that represent the three-level data structure in which the higher level for sites is accounted for are shown in Figures 10 and 11, respectively (see Chapter 4). The equation that describes the three-level unconditional model, which includes a random intercept at each level of the data structure, was previously shown in Equation (40).

A likelihood-ratio test is used to determine if a specific level of the data hierarchy is indeed necessary. The unconditional models with and without the random intercept for the second-level group, driver ($\varsigma_{jk}^{(2)}$), were estimated. These models are:

$$y_{ijk} = \beta_1 + \varsigma_{jk}^{(2)} + \varsigma_{k}^{(3)} + \varepsilon_{ijk} \tag{55}$$

$$y_{ijk} = \beta_1 + \varsigma_{k}^{(3)} + \varepsilon_{ijk} \tag{56}$$

Equation (55) shows the three-level unconditional model in which speed data are nested in drivers which are nested in sites, while Equation (56) shows the two-level unconditional model in which speed data are nested in sites. The likelihood-ratio test was used to determine if the combination of site and driver would produce better estimates than the model that does not include the driver cluster (i.e., if the driver level in the hierarchy is necessary). The results of the likelihood-ratio test between the models shown in Equations (55) and (56) resulted in a test-statistic of χ^2 = 745.20 which corresponds to a p-value of zero. Therefore, the null hypothesis that the variance component for drivers is zero is rejected, thus the thee-level model that includes the driver cluster will produce more accurate estimates than the two-level model that does not take into account the driver level.

Similarly, in order to verify that the higher level component (random intercept for sites) is needed, a likelihood-ratio test was performed between the unconditional models, with and without the random effect for site ($\varsigma_k^{(3)}$). The two-level unconditional model in which speed data are nested in drivers is shown in Equation (57):

$$y_{ijk} = \beta_1 + \varsigma_j^{(2)} + \varepsilon_{ijk} \tag{57}$$

The value of the likelihood-ratio test (χ^2) between the models shown in Equations (57) and (55) was 1724.01 corresponding to a p-value less than 0.0001, indicating that the three-level model in which speed observations are nested in drivers which are nested in sites is favored when compared to a two-level model in which the site level is not taken into account.

In order to investigate if the class variable "sensor" should be added as an additional level, the unconditional models with and without this level term were fitted. A four-level unconditional model in which speed observations are nested in sensors, sensors are nested in drivers, and drivers are nested in sites, was compared to the three-level unconditional model shown in Equation (55). The likelihood-ratio test revealed a value of zero, corresponding to a p-value of 1.0. This indicates that a random intercept for sensor is not required. The p-value of 1.0 indicates perfect multicollinearity – this is because there are no repeated measurements; there are 4 speed observations which correspond to the four sensors. However, the group variable sensor may replace the driver level; this option will later be explored in this section.

Table 24 shows the estimates for three unconditional models developed:

- *Model 1*: two-level model in which speed observations are nested in sites (Equation [56])

- *Model 2*: two-level model with speed observations nested in drivers (Equation [57])
- *Model 3*: three-level model with speed observations nested in drivers which are nested in sites (Equation [55])

Table 24 Maximum Likelihood Estimates for Multilevel Unconditional Models Fitted

Parameter	Two-Level		Three-Level
	Model 1	Model 2	Model 3
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed Part			
β1	47.2 (1.13)	47.9 (0.13)	47.2 (1.13)
Random Part			
Site ($\sqrt{\psi}$)	5.032 (0.80)	-	5.022 (0.80)
Driver ($\sqrt{\psi}$)	-	6.125 (0.11)	3.777 (0.09)
Residual ($\sqrt{\theta}$)	7.851 (0.05)	6.888 (0.05)	6.888 (0.05)
Log Likelihood	-39,845.2	-40,334.6	-39,472.6

The random terms shown in Table 24 can be interpreted as follow:

- *Model 1*: only considers the sites and ignores the fact that there are drivers nested within sites. A standard deviation of approximately 5 mph is associated with the presence of a site cluster while a standard deviation of 7.85 mph is associated with the residual term, θ, i.e. the overall variance that cannot be explained by the explanatory variables.
- *Model 2*: does not take into account the higher level (sites); speed data are nested in drivers. The standard deviation for the residual term, $\sqrt{\theta}$, is lower when compared to Model 1 (6.89 as compared to 7.85) indicating less variability with this model. However, the standard deviation for the random intercept for the second level (driver) is higher for Model 2 when compared to Model 1 (6.125 as compared to 5.032, respectively). This is an indication that, when considering only two levels, the site cluster is able to explain more of the variance than the driver cluster.
- *Model 3*: by specifying the variance component term for the combination of driver and site, another level to the hierarchy is added and between-driver

within-site heterogeneity is accommodated. The random effect for subject is nested within sites in the sense that it does not take on the same value for a given subject across all sites, but takes on a different value for each combination of site and driver. The standard deviation for the site component term is basically the same as Model 1 above (only decreased from 5.03 to 5.02). A standard deviation of approximately 3.8 mph is associated with the combination of driver and site. By adding higher levels that better represent the data structure, the variability associated with these levels can be specified.

The results of the likelihood ratio tests, performed for Models 1, 2, and 3, indicate that a three-level model was appropriate for the dataset developed in this research. Therefore, an initial model was developed in order to find the relationship between the driving environment (explanatory variables) and the operating speed (response variable) along the transition zones. The hierarchy of this model is as follows:

- Level 3 – Sites (20)
- Level 2 – Subjects (i.e., drivers) [2859 total]
- Level 1 – Speeds (response variable) [4 per subject]

A three-level model was estimated with the same variables found to be statistically significant in previous models. Table 25 shows the comparison of the estimates between the three-level model including the site level, the two-level model without the site cluster, and both the fixed- and random-effects panel data models.

Table 25 Comparison between Three-level, Two-level and Fixed-Effects Panel Data Models

Parameter	Multilevel Models (MLE)				Panel Data			
	Three-Level		Two-Level		Fixed-effects		Random-effects	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Speed Limit 25 mph	-10.54	0.524	-11.95	0.340	-10.46	0.537	-12.04	0.333
Speed Limit 35/40 mph	-2.21	0.173	-2.49	0.161	-2.20	0.173	-2.52	0.164
Speed Limit 45 mph	-3.48	0.469	-4.76	0.296	-3.41	0.481	-4.85	0.290
Lane width addition	3.34	0.342	2.19	0.186	3.49	0.354	2.14	0.178
Lateral Clearance	0.16	0.011	0.13	0.010	0.16	0.011	0.12	0.010
Total driveways	-0.95	0.034	-0.97	0.033	-0.95	0.034	-0.97	0.034
Curb	-4.00	0.233	-3.77	0.208	-4.01	0.235	-3.79	0.211
Intersection WS	-1.93	0.227	-2.05	0.214	-1.91	0.228	-2.05	0.218
School/Children WS	-1.09	0.199	-1.41	0.195	-1.08	0.199	-1.49	0.200
Curve Ahead WS	0.85	0.186	1.35	0.176	0.84	0.186	1.42	0.179
Curve with WS	-3.42	0.197	-2.54	0.182	-3.46	0.197	-2.41	0.184
Curve without WS	-1.67	0.163	-1.45	0.151	-1.68	0.164	-1.41	0.153
Constant	46.70	0.976	49.63	0.339	47.05	0.604	49.77	0.325
Random Components								
Site	3.4316		N/A		N/A		N/A	
Driver	4.457		5.4774		6.2022		4.8348	
Residual	5.0036		5.018		5.007		5.0071	

A comparison across all models shows that the three-level model and the fixed-effects panel data model produce both parameter estimates and standard errors (*SE* in the Table 25) that are almost identical to each other. Similarly, the parameter estimates and their standard errors for the two-level and the random-effects panel data models are also almost identical.

The residual terms between the four models compared in Table 25 are also very similar, which was expected since the models all include the same explanatory variables. The three-level model indicates that a standard deviation of 3.4 mph is associated with the site cluster, information that cannot be obtained with the two-level and the fixed-effects panel data models. When comparing the standard deviation values associated with the driver cluster, the three-level model indicates that this model explains more of the driver variance (lowest standard deviation value [4.46 mph]).

Although the estimates obtained with both the fixed-effects panel data model and the three-level model using the maximum likelihood estimator (*mle*) are almost identical, and the standard errors of these estimates are higher when compared to the two-level and random-effects panel data models, the three-level model provides additional information when compared to all other models. The three-level model provides information about the variance that is not explained by the explanatory variables included in the model specification when compared to either the two-level or the panel data models. By including random components at each level of the hierarchy, the variance associated with each level can be obtained. Consequently, although the three-level model produces estimates with higher standard errors, this model is a better representation of the data that includes the site cluster.

As previously indicated, the data structure can be altered by replacing the level variable *driver* with the level variable *sensor*, since speed data were collected at four sensor locations at each site. Therefore the data hierarchy is observations (speed data) nested in sensors which are nested in sites. Figure 14 represents the class diagram for the alternative hierarchy and the unit diagram is presented in Figure 15.

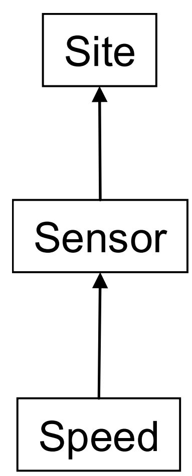

Figure 14 Class Diagram for Alternative Hierarchy

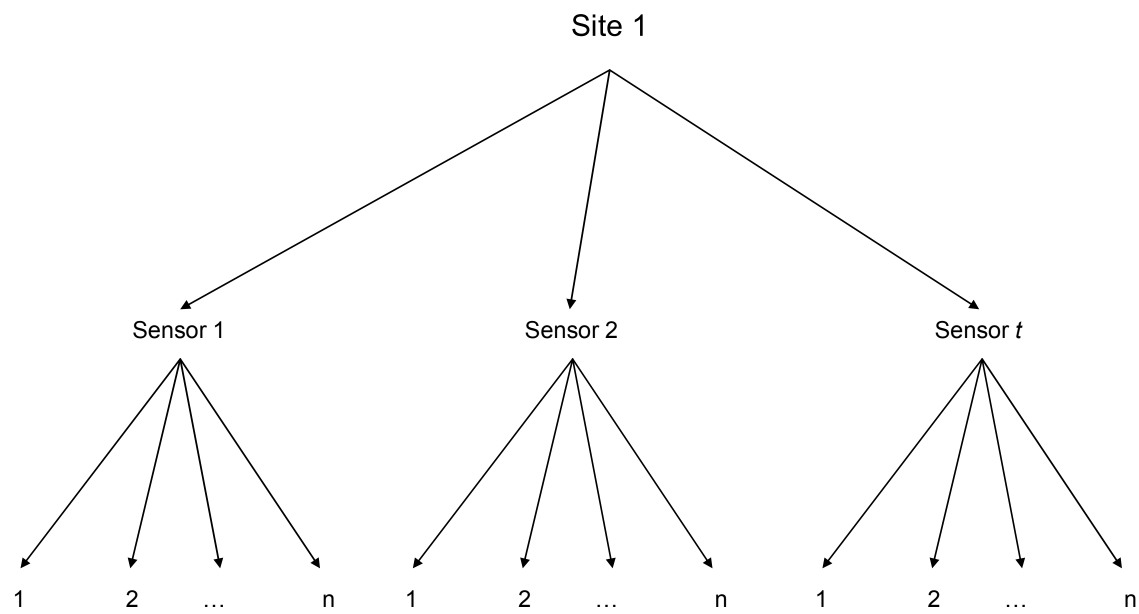

Figure 15 Unit Diagram for the Alternative Data Hierarchy

The same methodology for the initial data hierarchy was applied to the alternative hierarchy by replacing the level variable driver with the level variable sensor. For the model that specifies that operating speeds (response variable) are a function of the driving environment (explanatory variables), the hierarchy description is as follows:

- Level 3 – sites $k = 1, 2, \ldots, 20$
- Level 2 – sensors $j = 1, 2, 3, 4$
- Level 1 – speed $i = 1, 2, \ldots, 2859$

The total number of observations is the same as the initial three-level model with the different hierarchy: 2859 vehicles each passing through 4 sensors for a total of 11,436 observations. For this alternative hierarchy, instead of having 4 speed measurements per driver that are nested in the driver cluster, there is one observation per driver in the sensor cluster.

Similarly to the previous hierarchy, in which the variance component for driver was explored, the presence of the sensor cluster was explored by comparing the unconditional models with and without the variance component for sensor and performing a likelihood-ratio test. The value for the likelihood-ratio test was 4155.43 (p-value < 0.0001), rejecting the null hypothesis that the variance component for sensor is zero. This indicates that the three-level model that includes the sensor level is preferred over the two-level model that only considers speed observations nested in sites. However, there is one disadvantage when considering the sensor cluster instead of the driver cluster: there is no driver-specific information (driver behavior).

A three-level model with the alternative data hierarchy was developed for predicting speeds along transition zone sections. In addition, a two-level model that did not take into account the level for sensors (sensor cluster) was also developed in order to investigate any changes in parameter estimates due to the sensor cluster. A comparison between the two- and three-level models is shown in Table 26.

Table 26 Two- and Three-Level Variance Components Models for the Alternative Hierarchy

Variable	Three-level			Two-level		
	Estimate	St. Error	Z	Estimate	St. Error	Z
Speed Limit 25 mph	-10.95	1.701	-6.44	-10.44	0.639	-16.33
Speed Limit 35/40 mph	-2.45	0.811	-3.02	-2.29	0.220	-10.42
Speed Limit 45 mph	-3.85	1.457	-2.64	-3.11	0.600	-5.19
Lane Width Addition	2.43	1.051	2.31	3.36	0.447	7.51
Lateral Clearance	0.15	0.051	3.01	0.16	0.014	11.17
Total Driveways	-0.95	0.171	-5.56	-0.98	0.045	-21.69
Curb	-3.87	1.108	-3.49	-3.99	0.277	-14.43
Intersection WS	-2.44	1.075	-2.27	-1.99	0.304	-6.55
Curve with WS	-2.80	0.968	-2.89	-3.60	0.259	-13.88
Curve without WS	-1.46	0.767	-1.91 *	-1.90	0.211	-8.99
Constant	48.33	1.968	24.56	46.83	1.105	42.38
Random Components						
Site	3.1473	0.580		3.5046	0.566	
Sensor	2.1118	0.211		N/A	N/A	
Residual	6.4688	0.043		6.7002	0.044	
* Not significant at the 0.05 alpha value (p-value = 0.056)						

A likelihood-ratio test was performed between these two models; the value of the test was 643.62 (p-value less than 0.001). The result from the likelihood-ratio test provides evidence that the specification of a cluster for sensors (a three-level model) is preferred over the two-level model. However, the standard errors of the parameter estimates obtained with the three-level model are considerably higher than those obtained with the two-level model (more than twice their values).

In contrast to previous models, the variables for presence of both Curve Ahead and School/Children warning signs were found not to be statistically significant for the three-level model, thus they were not included in the model and are not shown in Table 26. The estimates for the random effects indicate that there is less variability in the three-level model when compared to the two-level model. When adding the sensor cluster, the standard deviation associated with the site cluster reduces from 3.5 to 3.1 mph. Similarly, the standard deviation of the residual term decreases from 6.7 to 6.5 mph. Although the differences between the standard deviations for both site and residual random components when comparing the three-level and two-level models are not of great magnitude (less than 0.5 mph), the advantage in adding the sensor variance component is that information about the variability in the response variable due to the extra level is gained. The standard deviation attributed to the sensor cluster is approximately 2 mph.

In addition, the estimates obtained with the three-level model that represents the site-sensor-speed hierarchy (specified in Figures 14 and 15) were compared to the estimates obtained with the original hierarchy of site-driver-speed (specified in Figures 10 and 11). The comparison between the three-level models is shown in Table 27.

Table 27 Comparison between Three-level Models Hierarchies

Parameter	Alternative Hierarchy		Original Hierarchy	
	Estimate	St. Error	Estimate	St. Error
Speed Limit 25 mph	-10.85	1.821	-10.54	0.524
Speed Limit 35/40 mph	-2.38	0.838	-2.21	0.173
Speed Limit 45 mph	-3.92	1.483	-3.48	0.469
Lane Width Addition	2.37	1.051	3.34	0.342
Lateral Clearance	0.15	0.051	0.16	0.011
Total Driveways	-0.95	0.171	-0.95	0.034
Curb	-3.95	1.202	-4.00	0.233
Intersection WS	-2.43	1.075	-1.93	0.227
School/Children WS	-0.47*	1.001	-1.09	0.199
Curve Ahead WS	0.28*	1.004	0.85	0.186
Curve with WS	-2.78	0.970	-3.42	0.197
Curve without WS	-1.37†	0.789	-1.67	0.163
Constant	48.40	1.961	46.70	0.976
Random Components				
Level 3 - Site	3.1160		3.4316	
Level 2 - Sensor/Driver	2.1124		4.4570	
Level 1 - Residual	6.4689		5.0036	

* Not significant (p-values greater than 0.6)
† Not significant at the 0.05 alpha value (p-value = 0.083)

As shown in Table 27, the signs for the parameter estimates are generally similar between the two models. For the posted speed limit variables, the estimates for the alternative hierarchy indicate greater speed reductions when compared to the original hierarchy, but these differences between estimates are modest (additional speed reduction of approximately 0.5 mph or less). The lane width addition variable is associated with speed increase of 2.4 mph for the hierarchy that considers the sensor cluster as opposed to an increase in speed of 3.3 mph for the original hierarchy. The parameter estimates for the variables of lateral clearance and presence of curb are almost identical when comparing the two models, differing by an absolute value of 0.01 and 0.05, respectively, while the estimates for number of driveways are identical between the two models. The presence of an Intersection Ahead warning sign is associated with speed reduction of 2.4 mph for the alternative hierarchy, 0.5 mph greater than the speed reduction predicted by the original hierarchy. The presence of a horizontal curve that warrants a warning sign is associated with a speed reduction of 2.8 mph when considering the sensor cluster; however, the original hierarchy indicated a greater speed reduction (3.4 mph) associated with this variable. The variables for the presence of School/Children and Curve Ahead warning signs were not statistically significant for the alternative hierarchy (p-values of 0.64 and 0.78, respectively).

The standard errors of the estimates obtained with the alternative hierarchy are greater when compared to the original hierarchy, indicating that the estimates obtained with the model that consider the driver cluster are more consistent. The variance associated with the site cluster is less for the alternative hierarchy when compared to the original hierarchy (standard deviation of 3.12 mph as compared to 3.43 mph, respectively) indicating that including the sensor cluster reduces the variance associated with the site cluster. At the second level, specifying a sensor cluster indicates a standard deviation of 2.1 mph while the variability when considering a driver cluster at this level is greater (4.5 mph). At the lower level, however, the original hierarchy explains more of the variance as indicated by the standard deviation of the residual term (5 mph as compared to 6.5 mph). The values for the standard errors of the estimates and the variance associated with the residual term indicate that the original hierarchy is a better fit for modeling speeds along transition zones.

Specifying a three-level model in which the sensor cluster replaces the driver cluster results in the loss of driver-specific information. And although this hierarchy results in greater standard errors for the estimates as well as greater variance at the lower level (speeds), an advantage is that additional driver-specific information can be included in the model. By considering the previous speed of each driver j, this information can be included in the three-level model with the alternative hierarchy.

The information on previous speed can be included in the three-level model as a random intercept which is independent across subjects. This model is referred to as the random intercept model and is shown below:

$$y_{ijk} = \beta_1 + \sum \beta X_{jk} + \varsigma_{jk}^{(2)} + \varsigma_k^{(3)} + \varsigma_j + \varepsilon_{ijk} \qquad (58)$$

where: ς_j = random intercept for previous speed of driver j.

Besides the random intercept model, a random coefficient model can be specified in which a random coefficient is included for any of the time-varying variables. By including previous speed and its random coefficient allows for the effect of this variable to vary between-subjects. The model for the random coefficient model is shown in Equation (59).

$$y_{ijk} = \beta_1 + \sum \beta X_{jk} + \beta L_{ijk} + \varsigma_{jk}^{(2)} + \varsigma_k^{(3)} + \varsigma_j L_j + \varepsilon_{ijk} \tag{59}$$

where: L_j = is the time-varying variable (previous speed) for driver j.

In order to include the speed at the previous sensor location, a new data set was created since only the previous speed was known at sensor locations 2, 3, and 4 (previous speed information was not available for sensor location 1). The higher levels remained the same with the lower level having fewer observations in each cluster, for a total of 8,577 observations in the dataset. The hierarchy of this model is provided below:

- Level 3 – site k = 1, 2, …., 20
- Level 2 – sensor j = 1, 2, 3
- Level 1 – speed, i = 1, 2, …, 2859

The new dataset consisted of a total of 8,577 speed observations for the response variable. Both random intercept and random coefficient models were developed for the new data set. Table 28 shows the estimates obtained with both options.

Table 28 Three-Level Models with Previous Speed for Alternative Hierarchy

Model	***Random Intercept***			***Random Coefficient***		
Parameter	**Estimate**	**St. Error**	**Z-value**	**Estimate**	**St. Error**	**Z-value**
Fixed Components						
Total Driveways	-0.44	0.411	-1.07*	-0.64	0.281	-2.28
Intersection WS	-4.69	2.499	-1.88†	-4.97	1.704	-2.92
Previous Speed	-	-	-	0.59	0.014	41.89
Constant	18.56	1.444	12.85	18.77	1.048	17.91
Random Components						
Site	2.83	1.250		2.10	0.818	
Sensor	4.93	0.799		3.18	0.621	
Previous Speed	0.59	0.055		0.08	0.012	
Residual	5.16	0.040		5.17	0.040	

* Not significant at the 0.05 alpha-level (p-value of 0.286)
† Not significant at the 0.05 alpha level (p-value of 0.060)

A likelihood-ratio test was performed between these two models; a χ^2 value of 206.76 (p-value of zero) resulted, which indicates that the random coefficient model is preferred over the random intercept model (previous speed should be included as an explanatory variable and the model should include its random coefficient). This can also be seen in the Z-statistic values for the coefficient estimates; these indicate significance at the 0.05 alpha level only for the random coefficient model.

As shown in Table 28, only the variables for total number of driveways, Intersection Ahead warning sign, and previous speed were found to be significant for the model selected. This can be interpreted as:

- *Total Driveways*: mean speed decreases by nearly 0.6 mph per unit increase in the total number of driveways in a study segment.
- *Intersection Ahead Warning Sign*: the presence of this sign is associated with a mean speed reduction of approximately 5 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Previous Speed*: for every 1 mph speed increase at a sensor location, there is a 0.6 mph speed increase at the next sensor location.

The standard deviation values for the random components are for the most part lower for the random coefficient model, indicating less variability associated with this model when compared to the random intercept model. However, including previous speed as an explanatory variable results in loss of information on highway characteristics that influence operating speeds. It is assumed that the previous speed of a vehicle depends on highway characteristics, therefore including previous speed in the model results in several highway-related variables not being statistically significant. In addition, the presence of a sensor cluster also results in some explanatory variables not being statistically significant, even without considering the previous speed variable.

Based on the findings from the multilevel models analyses, it is recommended that a three-level model, in which speed observations are nested in drivers and drivers are nested in sites, should be used when modeling speeds along transition zones. This hierarchy uses information about the highway characteristics that influence operating speeds along transition zones as well as information on driver-specific variability.

5.1.5 Generalized Estimating Equations (GEE) Analysis Results

The transition zone speed dataset was also analyzed using generalized estimated equations (GEE). As discussed in Chapter 4, GEE are longitudinal models that allow for correlation within clusters in the sample data. The hierarchy of the dataset indicates the presence of driver clusters; therefore, the speed data are correlated for the same driver. The correlation within clusters is accounted for by exploring different correlation matrix structures.

GEE models for each of the four working correlation matrices (independent, exchangeable, autoregressive, and unstructured) were applied to the disaggregate data in this study, exploring the same variables that were initially identified as significantly influential on operating speeds along transition zones. The QIC criterion and the marginal coefficient of determination (R^2m) values were also evaluated in order to identify which model, and consequently which correlation matrix, best fit the data. Table 29 shows the estimates for the coefficients for each of the variables considered as well as their standard errors and the QIC criterion and the marginal R-square values for each of the working correlation matrices.

Table 29 Variable Coefficients for each of the GEE Models According to Working Correlation Structures

Variable	Coefficients (Semi-Robust Standard Error)			
	Independent	Exchangeable	AR 1	Unstructured
Speed Limit 25 mph	-12.62 (0.388)	-11.95 (0.355)	-11.17 (0.366)	-12.49 (0.352)
Speed Limit 35/40 mph	-2.71 (0.196)	-2.49 (0.166)	-1.87 (0.167)	-2.58 (0.168)
Speed Limit 45 mph	-5.76 (0.350)	-4.76 (0.308)	-5.39 (0.305)	-4.94 (0.305)
Lane Width Addition	2.03 (0.210)	2.19 (0.180)	2.38 (0.188)	2.02 (0.177)
Lateral Clearance	0.02* (0.013)	0.13 (0.009)	0.15 (0.009)	0.15 (0.009)
Total Driveways	-1.10 (0.040)	-0.97 (0.031)	-1.03 (0.031)	-1.03 (0.031)
Curb	-4.48 (0.296)	-3.77 (0.222)	-3.56 (0.224)	-3.41 (0.222)
Intersection WS	-1.76 (0.274)	-2.05 (0.215)	-2.11 (0.216)	-1.59 (0.209)
School/Children WS	-2.82 (0.238)	-1.41 (0.178)	-0.57 (0.174)	-0.96 (0.171)
Curve Ahead WS	2.38 (0.188)	1.35 (0.153)	1.03 (0.152)	1.24 (0.145)
Curve with WS	-0.73 (0.236)	-2.54 (0.181)	-2.41 (0.184)	-2.54 (0.180)
Curve without WS	-0.67 (0.201)	-1.45 (0.141)	-0.96 (0.146)	-1.20 (0.137)
Constant	50.91 (0.383)	49.63 (0.331)	48.81 (0.341)	49.67 (0.328)
R^2m	0.3622	0.3505	0.2876	0.3488
QIC	619718.1	631086.0	637993.9	632716.9
* Not significant at the 0.05 alpha level (p-value = 0.138)				

The signs of the parameters for all of the explanatory variables were consistent across the correlation matrices considered. This is one of the advantages of the GEE modeling procedure; the estimates of the parameters are consistent even if the correlation structure chosen is incorrect (i.e. the estimates are robust to misspecifications of correlations [Zorn, 2001]). The magnitude of these coefficients does, however, vary across GEE models since the correlation matrix is included in the variance term of the speed prediction model (see Equations [46] and [47]). However, the differences in the majority of the parameter estimates is less than one in magnitude, indicating that the efficiency gains in estimates obtained by selecting the appropriate correlation matrix is modest (Zorn, 2001).

For speed limit variables, the estimates obtained across the GEE models differed by a value of 1.45 or less. The estimates obtained for the 25 mph speed limit indicator variable ranged from -12.6 for the independent model to -11.2 for the autoregressive (AR 1) model. For the speed limit 35 and 40 mph indicator variable, these parameter estimates ranged from -2.71 (Independent model) to -1.87 (AR 1 model). The range of the estimates obtained for 45 mph speed limit indicator variable varied from -5.76 (Independent model) to -4.76 (Exchangeable model).

For the lane width addition and lateral clearance explanatory variables, the difference between parameter estimates across GEE models was less than 0.4. The parameter estimates for the lane width variable ranged from 2.38 (AR 1 model) to 2.02 (Unstructured model). The lateral clearance variable was not statistically significant in the Independent model (an estimate value of 0.02); across the other GEE models it ranged from 0.15 (both AR 1 and Unstructured models) to 0.13 (Exchangeable model).

The estimates obtained for the number of driveways variable indicated that this variable is associated with an approximate 1 mph speed reduction for each additional driveway present along a two-lane rural highway transition zone; the parameter estimates ranged from -1.1 (Independent model) to -0.97 (Exchangeable model). The highest speed reduction associated with the presence of curb was for the Independent model (-4.5 mph); the lowest speed reduction associated with the curb present indicator variable was for the Unstructured model (-3.4 mph).

Both the Intersection Ahead and School/Children warning signs were associated with lower operating speeds along two-lane rural highway transition zones as indicated by the negative signs of the parameter estimates. The Intersection Ahead warning sign was associated with speed reductions between 2.1 and 1.6 mph (AR 1 and Unstructured models, respectively). The estimates obtained for the School/Children warning sign variable differed by the greatest magnitude across models when compared to the rest of explanatory variables. These parameter estimates ranged from -2.82 (Independent model) to -0.57 (AR 1 model). The parameter estimates obtained for the variable for Curve Ahead warning sign were positive; these ranged from 2.4 for the Independent model to 1.0 for the AR 1 model.

Finally, changes in horizontal alignment, with or without a sign that warns drivers of upcoming curves, were associated with lower operating speeds. The variable for a horizontal curve sign that warrants a warning sign had parameter estimates between -2.54 and -0.73 (Unstructured and Independent models, respectively). For curves that are not combined with a related warning sign, the speed reductions ranged from -1.45 (Exchangeable model) to -0.67 (Independent model).

All z-values for the parameter estimates in all GEE models indicated that the explanatory variables were statistically significant at the 0.05 alpha-level, with the exception of the variable for lateral clearance in the independent correlation structure (z-value of 1.48 corresponding to a p-value of 0.138). The statistical software package STATA also produces the working correlation matrix for each GEE model. Since all working correlation matrices are symmetric, and the correlations between individuals have a value of zero for the independent GEE model, the lower triangles of the correlation matrix for the exchangeable, autoregressive, and unstructured GEE models are:

1. Exchangeable: $V_i(4\times 4)=\begin{bmatrix} 1 & & & \\ 0.544 & 1 & & \\ 0.544 & 0.544 & 1 & \\ 0.544 & 0.544 & 0.544 & 1 \end{bmatrix}$

2. Autoregressive: $V_i(4\times 4)=\begin{bmatrix} 1 & & & \\ 0.598 & 1 & & \\ 0.358 & 0.598 & 1 & \\ 0.214 & 0.358 & 0.598 & 1 \end{bmatrix}$

3. Unstructured: $V_i(4\times 4)=\begin{bmatrix} 1 & & & \\ 0.699 & 1 & & \\ 0.556 & 0.503 & 1 & \\ 0.473 & 0.467 & 0.563 & 1 \end{bmatrix}$

The working correlation matrix for the Exchangeable model indicates that the correlation between speed observations is 0.544, regardless of the time (or in this case, distance) between sensor locations. The working correlation matrix for the AR 1 model specifies that between any two consecutive sensors, the correlation between speed observations is approximately 0.6. For speed observations between sensors 1 and 3 and sensors 2 and 4, the correlation is 0.358, while between sensors 1 and 4 this correlation is 0.214, indicating that speed observations between sensor locations are less correlated as sensor locations become further apart. For the Unstructured model, the working correlation matrix indicates different correlation values, regardless of the position of the sensors. For adjacent sensors, the correlations between speed observations are 0.7, 0.5, and 0.56 for sensors 1 and 2, 2 and 3, and 3 and 4, respectively. The speeds at sensors 3 and 4 are correlated to speeds at sensor 1 by a value of 0.56 and 0.47, respectively. Finally, the correlation in speed observations between sensors 2 and 4 is 0.467.

In order to identify which GEE model best describes the data, the QIC criterion and the marginal coefficient of determination, R^2m, were used. Both the QIC criterion and the R^2m favored the independent correlation structure (highest R^2_m and lowest QIC). However, Ballinger (2004) recommends selecting the model with the correlation structure that makes more sense theoretically and to use the QIC criterion when undecided between two correlation structures. The identification of the independent correlation matrix as the best model is counterintuitive since it specifies that speed observations within the driver cluster are independent. The autoregressive structure is selected when data within a cluster is correlated over time; however, the location of the sensors is not identical for all study sites since transition zone lengths vary over study sites. Based on these statements, the exchangeable and the unstructured correlation matrices are the most representative of the nature of how the data were collected. Both the QIC criterion and the marginal coefficient of determination values favor the exchangeable GEE model, thus this model was selected as the most appropriate to model operating speeds along transition zone among all GEE models. The parameter estimates for the explanatory variables of the exchangeable model can be interpreted as:

- *Speed Limit 25 mph*: a posted speed limit of 25 mph is associated with a mean speed decrease of almost 12 mph when compared to the baseline of a 55 mph posted speed limit.
- *Speed Limit 35/40 mph*: a posted speed limit of either 35 or 40 mph reduces mean speed by 2.5 mph when compared to the baseline of a 55 mph posted speed limit.
- *Speed Limit 45 mph*: a posted speed limit of 45 mph is associated with a mean speed decrease of almost 5 mph when compared to the baseline of a 55 mph posted speed limit.

- *Lane Width Addition*: for a minimum lane width of 9 feet, a mean speed increase of approximately 2.2 mph is associated per one-foot increase in the lane width addition variable.
- *Lateral Clearance*: a mean speed increase of 0.13 mph is associated with a one-foot increase in lateral clearance.
- *Total Driveways*: mean speed decreases by 1 mph per one-unit increase in the total number of driveways present in a two-lane rural highway transition zone.
- *Curb*: presence of curb is associated with mean speed reductions of 3.8 mph when compared to the baseline of no curb presence.
- *Intersection Ahead Warning Sign*: the presence of this sign is associated with a mean speed reduction of approximately 2 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *School / Children Warning Sign*: the presence of a sign related to the presence of school or children is associated with an additional mean speed reduction of 1.4 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Curve Ahead Warning Sign*: the presence of this sign is associated with a mean speed increase of 1.4 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Curve with Warning Sign*: a horizontal curve that warrants a Curve Ahead warning sign is associated with an additional mean speed reduction of approximately 2.5 mph when compared to the baseline of no change in horizontal alignment.
- *Curve without Warning Sign*: a horizontal curve without a warning sign is associated with a mean speed reduction of 1.5 mph when compared to the baseline of no change in horizontal alignment.

5.1.6 Point Speed Analyses Summary

The speed data collected at each of the four sensor locations permitted the vehicles to be "tracked;" therefore, individual driver-speed information was available. Panel data and GEE methodologies were applied to the data since they are longitudinal models that allow for correlation among the observations, specifying that speed observations were nested in drivers. In order to incorporate a higher level, several three-level models were developed which specified a site cluster at the highest level.

For panel data analyses, a fixed-effects model that includes the posted speed limit with disaggregate data produced a better goodness-of-fit to the data when compared to the model without the posted speed limit indicators. As noted previously, however, the fixed-effects panel model without the posted speed limit indicators resulted in the signs for two parameters changing and becoming more consistent with engineering intuition (i.e., paved shoulder width and Curve Ahead warning sign). For multilevel models, the model that best fit the data was a three-level model with a data structure that specified speeds nested in drivers and drivers nested in sites. Among the GEE models, the Exchangeable working correlation matrix was selected as the most appropriate to model the data. A comparison of the recommended models selected as appropriate among the point speed analysis methods described in this section of the dissertation is shown in Table 30.

Table 30 Comparison Between all Model Selected as Appropriate

Parameter	Panel Data (FE)		Three Level (MLE)		GEE (Exchangeable)	
	Estimate	St. Error	Estimate	St. Error	Estimate	St. Error
SL 25 mph	-10.46	0.537	-10.54	0.524	-11.95	0.355
SL 35/40 mph	-2.20	0.173	-2.21	0.173	-2.49	0.167
SL 45 mph	-3.41	0.481	-3.48	0.469	-4.76	0.308
Lane Width Addition	3.49	0.354	3.34	0.342	2.19	0.180
Lateral Clearance	0.16	0.011	0.16	0.011	0.13	0.009
Total Driveways	-0.95	0.034	-0.95	0.034	-0.97	0.031
Curb	-4.01	0.235	-4.00	0.233	-3.77	0.222
Intersection WS	-1.91	0.228	-1.93	0.227	-2.05	0.215
School/Children WS	-1.08	0.199	-1.09	0.199	-1.41	0.178
Curve Ahead WS	0.84	0.186	0.85	0.186	1.35	0.154
Curve with WS	-3.46	0.197	-3.42	0.197	-2.54	0.181
Curve without WS	-1.68	0.164	-1.67	0.163	-1.45	0.141
Constant	47.05	0.604	46.70	0.976	49.63	0.331
Random Components						
Site	-		3.4316			
Driver	6.2022		4.457		-	
Residual	5.007		5.0036		-	
R^2	R^2overall = 0.3266		-		R^2marginal = 0.3505	

The variables found to significantly influence operating speeds were consistent across all models specified in this section of the dissertation. The highway characteristics found to be associated with speed reductions, regardless of the modeling methodology, were changes in posted speed limit, number of driveways, presence of a curb, presence of Intersection Ahead and School/Children warning signs, and presence of a horizontal curve, either with or without a Curve Ahead warning sign. Only the variables related to lane width and lateral clearance distance were found to be associated with higher operating speeds. The magnitude of the parameter estimates for all the variables were generally consistent across all models.

The parameter estimates obtained from the fixed-effects panel data model and the three-level model are nearly identical; they all differed by an absolute value of 0.15 or less. When compared to the GEE parameter estimates obtained with the Exchangeable correlation matrix, the magnitudes of these differences are higher. The posted speed limit

25 mph indicator and the posted speed limit 45 mph indicator were associated with approximately 1.5 mph greater speed reductions for the GEE exchangeable model when compared to the panel and multilevel models. The variable for presence of a horizontal curve that warrants a warning sign was associated with a speed reduction of 2.5 mph for the GEE model when compared to a speed reduction of approximately 3.4 mph for both the panel data and multilevel models. The estimates obtained for the lane width addition variable indicated only a speed increase of 2.2 mph per one-foot increase in lane width for the GEE model, a lower value than those obtained with the panel data and multilevel models (3.5 and 3.3 mph, respectively). All other variables had similar estimates across all models, differing by approximately an absolute value of 0.5 or less.

When comparing the values for the standard errors of the estimates, these were also almost identical for the panel data and the multilevel models (differences less than an absolute value of 0.02). The standard errors obtained from the Exchangeable GEE model are smaller when compared to the ones obtained using the panel data and multilevel models; however, the differences in standard errors for most of the variables were less than an absolute value of 0.04. The greatest differences between standard errors were for the indicator variables for both speed limits of 25 and 45 mph and for the continuous variable of lane width addition; still, these differences were less than an absolute value of 0.2.

The coefficient of determination, R^2, is available for panel data and GEE model analyses; multilevel models do not provide this information. The overall coefficient of determination for the fixed-effects panel data model was approximately 0.33. However, the marginal coefficient of determination computed for the Exchangeable GEE model was 0.35, indicating that the exchangeable GEE model provides a better fit to the data. Although the GEE model indicated less variance, as indicated by the smaller standard errors and the coefficient of determination, the advantage of the panel data and multilevel models is that estimates for the random components can be obtained, which is not available when specifying GEE models.

The parameter estimates obtained with the panel data and the multilevel models were almost identical, but the three-level model is able to provide additional information about the variability associated with the site cluster. The use of multilevel models enabled the addition of study sites to the data structure hierarchy. Therefore, the variability associated with the higher level (sites) can be obtained with the use of a three-level model, information that panel data analyses fail to provide. In addition, the results of the multilevel model analyses indicated that a three-level model was preferred over a two-level model. Therefore a three-level model that specifies a data hierarchy in which speed observations are nested in drivers and drivers are nested in sites is preferred over the other models considered in this dissertation.

5.2 Speed Differential Analysis Results

The previous models developed – panel data, multilevel models, and GEE – considered the study section in its entirety by treating each data collection location as a point speed. Additional models were developed using ordinary least squares regression (OLS) and multilevel models that only considered the transition zone sections.

As discussed in Chapter 3, speed data were collected at 4 sensor locations along 20 two-lane rural highway transition zones in Central Pennsylvania. The sensor locations permitted vehicles to be tracked along each of the study areas; therefore, it was possible to obtain driver-specific speed differentials. Several authors have indicated the importance of considering individual speed differentials when developing prediction models (Park and Saccomanno, 2005; and Misaghi and Hassan, 2005). Since sensor locations 2 and 3 defined the beginning and the end of the transition zone, respectively, the response variable for the speed differential model was defined as:

$$\Delta Y_{2\text{-}3} = Y_2 - Y_3 \qquad (60)$$

where: $\Delta Y_{2\text{-}3}$ = change in speed between limits of the transition zone (mph);

Y_2 = speed at sensor location 2 (mph); and

Y_3 = speed at sensor location 3 (mph).

Table 31 shows the summary statistics for the response variable for each study site included in this research.

Table 31 Speed Differential along Transition Zone Descriptive Statistics

Site ID	N	Response Variable: $\Delta Y_{2\text{-}3}$ (mph)			
		Mean	St Dev	Minimum	Maximum
1	124	-0.758	5.808	-20	19
2	68	8.28	8.49	-11	31
3	98	3.592	5.447	-13	15
4	104	1.288	5.298	-12	14
5	231	2.762	5.006	-13	26
6	99	5.859	7.387	-15	21
7	159	2.616	5.77	-19	19
8	149	3.503	6.372	-12	25
9	478	4.793	7.495	-17	29
10	148	1.439	5.506	-16	19
11	141	4.83	5.877	-9	23
12	73	3.849	5.338	-12	18
13	130	13.315	5.452	-3	33
14	112	10.018	6.416	-10	27
15	81	-0.0247	4.552	-10	11
16	122	5.074	5.702	-12	25
17	164	5.03	5.069	-9	19
18	52	1.346	6.426	-11	18
19	178	-3.916	4.608	-17	10
20	148	4.169	5.62	-10	22

In Table 31, a positive value of $\Delta Y_{2\text{-}3}$ indicates a speed reduction, while a negative value for $\Delta Y_{2\text{-}3}$ indicates a speed increase. This relationship was helpful in associating speed reductions with a desirable outcome (speed reductions are desired along each of the transition zones) while associating speed increases with an undesirable outcome. A linear regression model was identified to model mean speed difference; the OLS model is then:

$$\Delta Y_{2-3} = a + \Sigma bX + \varepsilon \tag{61}$$

In the mean speed differential OLS model, the amount of variation not explained by the independent variables is contained in the error term, ε.

As previously noted, information on roadway characteristics were also collected at each sensor location in order to include them as potential predictors. In order to develop a speed differential prediction model along transition zones, only roadway characteristics between the limits of the transition zone (i.e. sensor locations 2 and 3) were included in the set of explanatory variables (see Chapter 3). However, since it was hypothesized that changes in operating speeds were a result of changes in the driving environment, additional explanatory variables were created in order to include the geometric design feature changes within the limits of the transition zone. To perform the speed differential analysis, the following variables were created in the database:

1. *Change in cross-sectional roadway characteristics*. Variables for the change in lane width, shoulder width, paved roadway width and lateral clearance were created by subtracting these measurements collected at sensor location 3 from their respective measurements collected at sensor location 2. Therefore a positive value for the variable Delta Lane Width would indicate a lane width reduction between the beginning and end of the transition zone. The descriptive statistics of these changes in lateral dimensions (cross-sectional) are shown in Table 32.
2. *Average of cross sectional roadway characteristics*. Similarly, for the variables of lane width, shoulder width, paved roadway width, and lateral clearance, the average value of their respective measurements at both sensor locations 2 and 3 was calculated and included as potential predictors.
3. *Change in driveways*. Different variables were included that described the change in driveway density. By identifying an area for each sensor, as shown in Figure 7 (see Chapter 3), driveways were assigned to that area, thus *delta driveways* was calculated by subtracting the number of driveways assigned to sensor location 3 from the number of driveways assigned to sensor location 2. A positive value indicated a reduction in driveway density while a negative value indicated an increase in driveway density. The descriptive statistics for change in driveways are also shown in Table 32. In addition, indicator variables were created to indicate either a driveway density increase or a decrease in driveway density.
4. *Transition zone length*. This variable was not considered in the point speed models since it was specific to each study site. Transition zone lengths for

each study site are shown in Table 7 in Chapter 3. The descriptive statistics for length of transition zone are shown in Table 32.

5. *Introduction of Curb*. An indicator variable was created to indicate if a curb was introduced within the limits of the transition zone.
6. *Horizontal Alignment*. Three indicator variables were created following the procedure in the analyses for point speeds; horizontal curve that warrants a warning sign, horizontal curve that does not warrant a warning sign, and a tangent roadway section.

Table 32 Descriptive Statistics for Continuous and Indicator Variables

Continuous Variables	Mean	St Dev	Minimum	Maximum
Delta Speed Limit	18.39	2.523	10	20
Delta Lane Width, ft	-0.09	0.274	-1	0.4
Delta Paved Shoulder, ft	-0.41	2.101	-4.5	6
Delta Stabilized Shoulder, ft	1.23	3.011	-5.5	12
Delta Paved Roadway, ft	-0.61	3.731	-10.4	11.5
Delta Lateral Clearance, ft	-0.45	3.897	-9	6.4
Delta No. of Driveways - Next Side	0.24	1.414	-3	3
Delta No. of Driveways - Other side	-0.37	1.418	-3	2
Delta Total No. of Driveways	-0.13	2.489	-5	4
Transition Zone Length (ft)	681.45	190.590	375	1065
No. Warning Signs	0.27	0.686	0	3
Indicator Variables	**Mean**	**St Dev**	**Minimum**	**Maximum**
Curb	0.31	0.464	0	1
Curve with Warning Sign	0.34	0.485	0	1
Curve without Warning Sign	0.24	0.429	0	1
Tangent Section	0.38	0.485	0	1

In addition, several variables for the reduction in speed limit were created in order to be considered as potential explanatory variables, including both continuous and indicator variables. As presented in Chapter 3 in Table 7, the distribution of speed limit changes per study site was as followed:

- Nine sites indicated a posted speed limit reduction from 55 to 35 mph
- Four sites indicated a posted speed limit reduction from 45 to 25 mph
- The reduction in speed limit at five sites was from 55 to 40 mph

- One site indicated a speed limit reduction from 40 to 25 mph (Site 14)
- One site indicated a reduction in speed limit from 45 to 35 mph (Site 18)

Since the changes in speed limit at two of the study sites were not observed at any of the other sites (sites 14 and 18), these two sites were either combined with other sites. The categories for the speed limit reduction indicator variables, including the inclusion of how sites 14 and 18 were considered, were as follow:

- *Speed Limit 55-35mph.* Speed limit reduction from 55 to 35 mph. Nine sites observed this speed limit reduction. Since site ID 18 indicated a speed limit reduction from 45 to 35 mph, and it was observed that speed limit upstream of the study site was 55 mph, this site was included in this category.
- *Speed Limit 55-40 mph*. Speed limit reduction from 55 to 40 mph; this speed limit reduction was observed at five sites.
- *Speed Limit 45/40-25 mph.* Speed limit reduction from either 45 or 40 mph to 25 mph. Four sites indicated a speed limit reduction from 45 to 25 mph. Site ID 14 was the only site which indicated a speed limit reduction from 40 to 25 mph, therefore it was included in this category.

Data for other roadway characteristics between sensor locations 2 and 3 were combined for some of the variables, such as type of warning signs and grade. Interaction terms were not included in the OLS analyses in order to identify the influence of individual roadway features on vehicle operating speed changes in a transition zone. However, similar to the point speed analyses presented earlier, an indicator variable for the presence of a horizontal curve that required a warning sign was included in the dataset of potential explanatory variables.

The regression model focused only on driver behavior along the transition zone and, since driver-specific speed data were available, disaggregate data were used for the analysis. The dataset considered for this analysis included 2,959 observations which correspond to the number of vehicles included in the data sample. The procedure followed to develop the OLS model for change in speed along transition zones was:

1. Determine correlations between response variable and explanatory variables to identify potential predictors for the model specification.

2. Correlations between explanatory variables to identify presence of possible collinearity.
3. Centralization of continuous variables: Centering a continuous variable entails subtracting each value by the mean and is done to reduce the correlation with other variables and avoid multicollinearity. Two continuous variables have been identified for centralization: speed difference along the transition zone (response variable) and speed before the transition zone at sensor location one (explanatory variable).
4. One-Way Analysis of Variance (ANOVA) for each potential explanatory variable to exclude any variables that do not have any statistical influence on the response variable.

Each step for the modeling procedure is described in detail in the following sections.

5.2.1 Correlation Analyses

Correlation analyses were initially performed to investigate which explanatory variables were strongly correlated with the response variable. Additionally, these correlation analyses were also helpful in investigating if there were some explanatory variables that were correlated with each other.

The correlation analysis showed that reductions in speed limit to 25 mph, regardless of the initial speed limit (speed limit at the high-speed zone) were highly correlated with the response variable. When considering the cross-sectional features of the roadway, the variables related to lane width and both paved and stabilized shoulder width were strongly correlated. In addition, explanatory variables related to lateral clearance distances were also correlated to the response variable.

As related to changes in alignment, either horizontally or vertically, the presence of horizontal curve, regardless of direction, had the strongest correlation with the response variable. As related to driveway density related variables, the strongest correlation to the response variable was total number of driveways. The presence of warning signs, as well as the type of warning sign located along the transition zone, was also correlated to the speed reductions. The length of the transition zone was also found to have a correlation with the response variable. These correlation values are shown in Table 33.

Table 33 Correlations between Potential Explanatory Variables and Response Variable

Potential Explanatory Variable	Correlation
Speed Limit change from 45/40 to 25 mph	0.204
Average Lane Width	0.159
Change in Lane Width	-0.134
Average Paved Shoulder	0.114
Average Stabilized Shoulder	0.181
Change in Stabilized Shoulder	0.170
Average Lateral Clearance	0.242
Change in Lateral Clearance	0.134
Total Number of Driveways	0.268
Presence of Horizontal Curve	0.122
Number of Warning Signs	0.277
Presence of Intersection Ahead Warning Sign	0.225
Presence of School/Children Warning Sign	0.236
Presence of Curve Ahead Warning Sign	0.201
Length of Transition Zone	0.208

The majority of the variables shown in Table 33 are positively correlated with the response variable, indicating that these are associated with speed reductions. The variable of change in lane width has a correlation of -0.134, indicating that an increase in lane width is associated with a decrease in speed reduction (i.e., speed increase). The variables for average paved and stabilized shoulders and average lateral clearance indicated that high values for these lateral distances beyond the travel lane are associated with speed reductions. The presence and number of warning signs were also associated with speed reductions along the transition zone.

The indictor variables related to speed limit reductions that were correlated to the response variable were only those in which the posted speed limit in the low-speed area was 25 mph. Although variables were available for changes in driveway density, the only variable related to this highway characteristic that had a high correlation value with the response variable was total number of driveways along the transition zone. The correlation for the transition zone length variable indicated that longer distances between the speed limit signs that mark the limits of the transition zone are associated with greater speed reductions. The only variable that had an unexpected correlation value with the response variable was average lane width (0.159). The results indicated that high values for lane width are associated with greater speed reductions, which contradicts the results obtained previously which suggested that as the travel lane narrows, the speed reductions along the transition zone increase.

In addition, correlations between explanatory variables were explored in order to identify which of these variables had high correlation values among each other. High correlation values between explanatory variables indicate the potential for multicollinearity if these are included in the same model. The variables related to paved roadway width were strongly correlated to those related to lane width, and both paved and stabilized shoulder widths, therefore paved roadway was not considered in the model. Similarly, the variables for the same cross sectional characteristic (i.e. average lane width and change in lane width) were strongly correlated, thus indicating that only one variable specific to the cross sectional characteristic under consideration should be included in the model.

5.2.2 Centralization of Continuous Variables

The second step in the in the speed differential analysis in the transition zone included the centralization of continuous variables in order to reduce multicollinearity. Multicollinearity exists when one of the independent variables is highly correlated to one or more of the other explanatory variables in a multiple regression model. It has been suggested that, in order to reduce the impact of multicollinearity, to increase the sample size or to "center" the variables (Motulsky, 1995). Centering variables involves subtracting the mean from each individual observation. By subtracting each observation by a scalar number (the overall mean), the histogram is "shifted", and the range between the variable values remains the same. Centering a variable is useful when the variable is continuous, and one advantage is that the results are easier to interpret. The independent variable for speed at sensor location one ($SpeedS1$) was then selected to be centralized and the new variable is given by Equation (62) below:

$$SpeedS1ctr_j = SpeedS1_1 - \frac{1}{n}\sum_{j-1}^{n} SpeedS1_j \tag{62}$$

where $SpeedS1ctr_j$ = is the speed at sensor 1 centralized for driver j and n=2859.

Figure 16 shows the histogram for both SpeedS1 and SpeedS1ctr in which it can be seen that the histogram's shape remains the same.

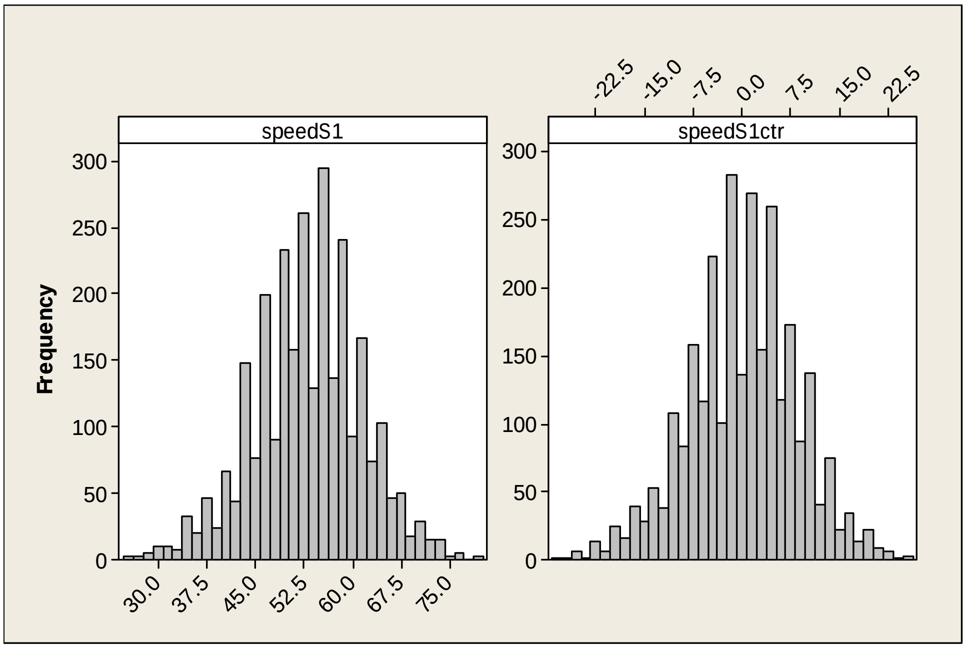

Figure 16 Histograms for Speed at Sensor 1 (Original and Centralized)

5.2.3 One-Way ANOVA

The third step in the speed differential analysis consisted of applying one-way Analysis of Variance (ANOVA) for each of the individual explanatory (categorical) variables in order to narrow down the potential predictors to be included in the final model (i.e. in order to further explore which variables could be included in the regression model and which variables can be excluded). The following variables were found to have a statistically significant influence on the response variable (which indicates that they could be in the regression model):

- All speed limit related variables: speed limit in the low-speed zone, speed limit in the high-speed zone, all speed limit indicators, and change in speed limit

- All cross-sectional characteristics (lane width, paved and stabilized shoulder, lateral clearance)
- Presence of a steep positive grade (greater than 3 percent)
- All driveway related variables
- All horizontal alignment variables, including their interaction with a curve-ahead warning sign.
- Introduction of curb
- All warning sign-related variables
- Transition zone length
- Speed at sensor 1(centered)

Although the correlation analysis did not indicate that the presence of a curb, posted speed limits of 35 and 40 mph, and a steep positive grade were associated with speed reductions in the transition zone, the ANOVA results did indicate that each was statistically significant. As such, each of these variables was included in the list of potential explanatory variables for the speed reduction OLS regression model specified in the following section.

5.2.4 Linear Regression Model and Variance Inflation Factors

Regression models were fitted with those explanatory variables that were identified as having an influence on the response variable (those identified either by correlation analyses, one-way ANOVA, or a combination of both). Best subsets analyses were performed and both the Mallows Cp and the AIC criterion values were explored: a low value for both the Cp and the AIC scores are used to identify the best regression model. In addition, the variance inflation factor (VIF) was used to detect multicollinearity. Those variables with a VIF value of 10 or higher were excluded since this was an indication that the particular variable is associated with an increase in the variance of the estimated coefficients. Based on the coefficient of determination and the VIF values, as well as the Cp and AIC criterion values, a model was developed. The model developed had a coefficient of determination, R^2, of 0.248 indicating that approximately 25 percent of the variation in speed differentials is explained by the explanatory variables. It also had the lowest Cp value (15.0) as well as the lowest AIC criterion score (18,444.2) and the highest coefficient of determination (R^2) during the best subsets procedure. The results of the OLS model, including the results for the Analysis of Variance (ANOVA) are shown in Table 34.

Table 34 Speed Differential OLS Results

Predictor	**Estimate**	**St. Error**	**t**	**p-value**	**VIF**
Speed1 Centered	0.16	0.017	9.20	<0.001	1.6
Speed Limit 55-40 mph	2.98	0.394	7.56	<0.001	2.2
Speed Limit 45/40-25	2.94	0.399	7.37	<0.001	2.0
Delta Lane Width	2.41	0.978	2.47	0.014	5.6
Delta Paved Shoulder	1.06	0.121	8.71	<0.001	5.0
Delta Lateral Clearance	0.09	0.040	2.28	0.023	1.9
Total Driveways	0.38	0.081	4.65	<0.001	1.7
Curb Intro	1.21	0.547	2.22	0.026	5.0
Intersection WS	3.11	0.615	5.06	<0.001	2.2
School/Children WS	7.33	0.644	11.39	<0.001	2.8
Curve WS	-3.60	0.614	-5.85	<0.001	2.6
Transition Zone Length	0.68	0.091	7.44	<0.001	2.4
Curve with WS	4.27	0.448	9.54	<0.001	3.7
Tangent	-1.31	0.348	-3.78	<0.001	2.2
Constant	-4.95	0.648	-7.64	<0.001	-
Source	**df**	**SS**	**MS**	**F**	**P**
Regression	14	34628.5	2473.5	67.03	<0.001
Residual Error	2844	104938.3	36.9		
Total	2858	139566.8			

The influence of each of the explanatory variables on speed differentials along transition zones can be interpreted as follows:

- *Speed1 Centered*: a mean speed reduction of 0.16 mph is expected per unit increase of the speed 500 ft before the beginning of the transition zone.
- *Speed Limit 55-40*: a mean speed reduction of approximately 3 mph is expected when the speed limit decreases from 55 to 40 mph, as compared to the baseline of a change in posted speed limit from 55 to 35 mph.

- *Speed Limit 45/40-25*: a mean speed reduction of 2.9 mph is expected if the speed limit changes from 45 or 40 mph to 25 mph when compared to the baseline of a posted speed limit reduction from 55 to 35 mph.
- *Delta Lane Width*: for each one-foot change in the lane width reduction, a mean speed reduction of 2.4 mph is expected.
- *Delta Paved Shoulder*: a one-foot paved shoulder width reduction is associated with a mean speed reduction of 1mph.
- *Delta Lateral Clearance*: a one-foot lateral clearance reduction is associated with a mean speed reduction of 0.1 mph.
- *Total Driveways*: a mean speed reduction of 0.4 mph is expected for each unit increase in the total number of driveways.
- *Curb Intro*: the introduction of curb is associated with a mean speed reduction of approximately 1.2 mph when compared to the baseline of no curb.
- *Intersection Ahead Warning Sign*: the presence of an Intersection Ahead warning sign is associated with a mean speed reduction of 3.1 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *School/Children Warning Sign*: the presence of a warning sign related to school or presence of children is associated with a mean speed reduction of 7.3 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Curve Ahead Warning Sign*: the presence of a Curve Ahead warning sign is associated with a mean speed increase of 3.6 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Transition Zone Length*: a mean speed reduction of 0.7 mph is associated with each 100-foot increase in the transition zone length.

- *Curve with Warning Sign*: the presence of a horizontal curve that warrants a warning sign is associated with a mean speed reduction of 4.3 mph when compared to the baseline of presence of a curve without a Curve Ahead warning sign.
- *Tangent*: the presence of a tangent along the transition zone is associated with a mean speed increase of 1.3 mph when compared to the baseline of presence of a curve without a Curve Ahead warning sign.

5.2.5 Additional Remedial Measures and Linear Regression Assumptions

The assumptions of OLS were previously discussed in Chapter 4. The VIF values have ensured the absence of multicollinearity. Several assumptions can be checked using residual plots. Scatterplots of the standardized residual versus the fitted values can be used to check the assumption of normality, linearity, and equal variances (homoskedasticity). If the scatterplot is randomly scattered about zero it is a good indication that the assumptions of regression are met. Histograms were also used to check for normality. Figure 17 shows the scatterplot of residuals versus fitted values for the linear regression model developed. A histogram for the residuals is shown in Figure 18.

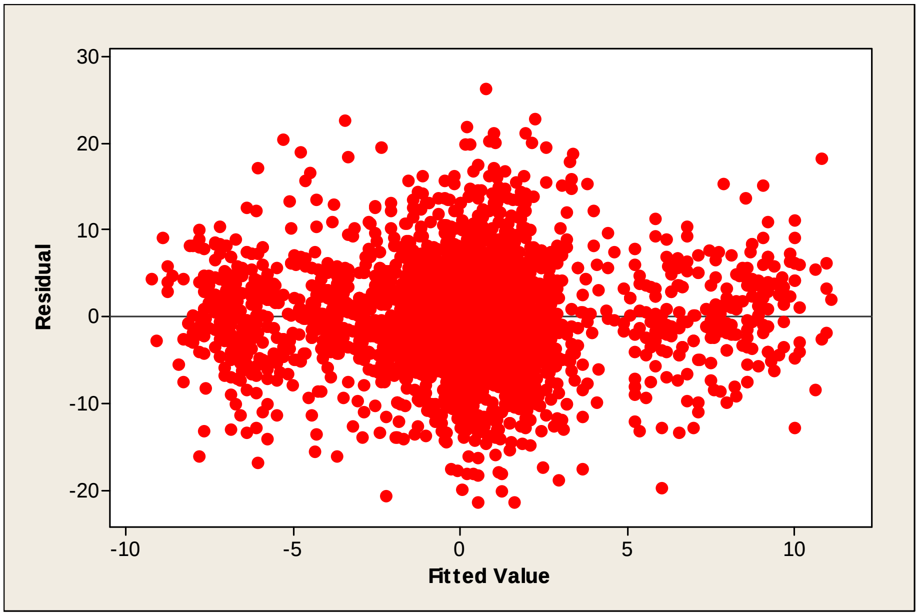

Figure 17 Scatterplot of Residuals versus Fitted Values

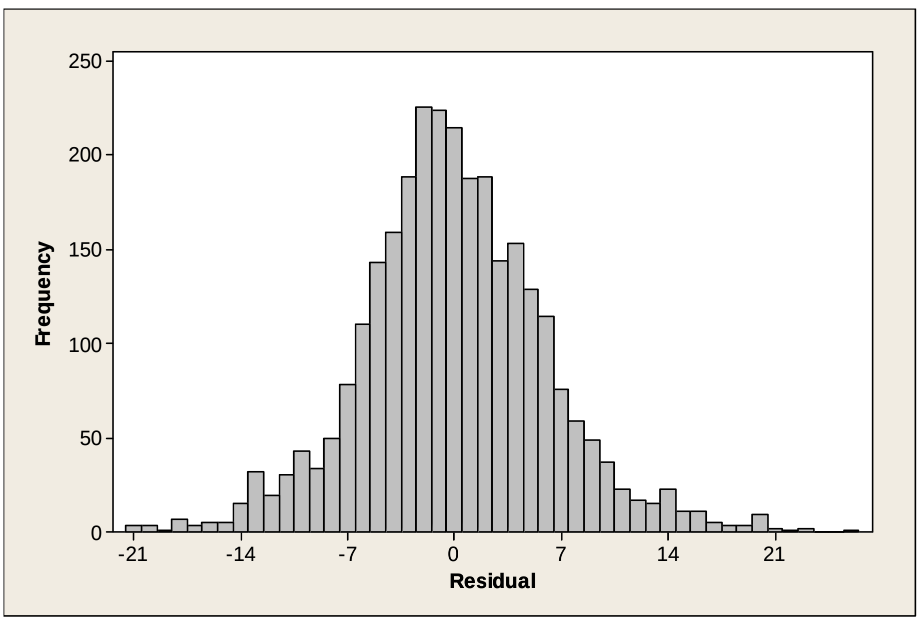

Figure 18 Histogram of Residuals

The plot of residuals versus fitted values does not show any pattern and it is scattered around zero, thus the assumptions of normality, linearity, and homoskedasticity were met. The histogram of residuals is bell-shaped centering on zero, confirming that the normality assumption is indeed met.

In addition, the plot of residuals versus the order of the data can be used to check the assumption of homoskedasticity and check for autoregression problems. This plot is shown in Figure 19.

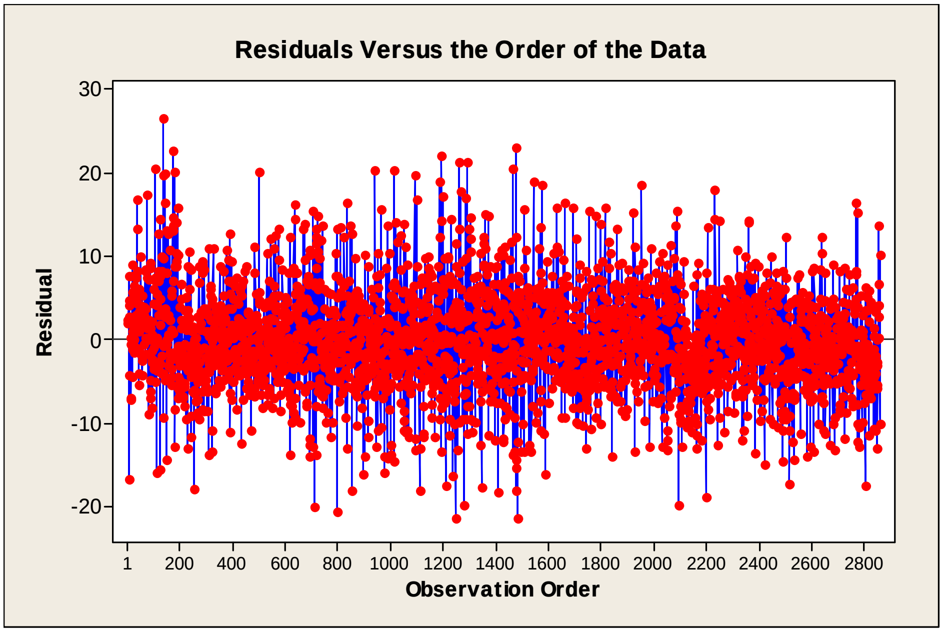

Figure 19 Residuals versus the Order of the Data

The plot shown in Figure 19 is also scattered around zero and does not indicate any patterns in the residual data, thus the assumption of equal variances was met. In addition, the Durbin-Watson test was performed resulting in a value, *d*, of 1.96. As previously indicated in Chapter 4, if the d-statistic is close to a value of two, there is little evidence that autocorrelation exists among the residuals.

5.2.6 Multilevel Model for Speed Differential

In addition to OLS regression, a multilevel model in which vehicle speed differences (level 1) were nested in sites (level 2) was developed. The response variable remained the same: change in speed along the transition zone as expressed in Equation (60) in Section 5.2. The two-level model with change in speed as the response variable can be expressed as:

$$\Delta Y_{2-3\,jk} = \beta_0 + \sum \beta_1 X_{jk} + \varsigma_k^{(2)} + \varepsilon_{jk} \qquad (63)$$

where: $\Delta Y_{2\text{-}3jk}$ = speed difference for driver *j* at site *k*;

β_0 = fixed intercept (slope);

$\Sigma\beta_1 X_{jk}$ = matrix of independent variables (X) and their coefficients (β);

$\varsigma_k^{(2)}$ = random intercept for level 2 (sites), with variance $\psi^{(2)}$; and

ε_{jk} = random error term (residual) with variance θ.

An initial two-level model was estimated with the explanatory variables that were included in the OLS regression model estimated previously. Table 35 shows the results from both models for comparison purposes.

Table 35 Comparison between Two-Level and OLS Models

Parameter	Two-Level		OLS	
	Estimate	St Error	Estimate	St Error
Speed Centered	0.17	0.017	0.16	0.017
Speed Limit 55-40 mph	2.87	1.088	2.98	0.394
Speed Limit 45/40-25	2.98	1.098	2.94	0.399
Delta Lane Width	2.22*	2.517	2.41	0.978
Delta Paved Shoulder	1.09	0.311	1.06	0.121
Delta Lateral Clearance	0.10*	0.116	0.09	0.040
Total Driveways	0.38	0.224	0.38	0.081
Curb Introduction	0.67*	1.673	1.21	0.547
Intersection WS	2.47	1.728	3.11	0.615
School/Children WS	7.64	1.900	7.33	0.644
Curve WS	-2.91	1.793	-3.60	0.614
Transition Zone Length	0.75	0.252	0.68	0.091
Curve with WS	3.03	1.258	4.27	0.448
Tangent	-1.88	1.026	-1.31	0.348
Constant	-4.75	1.905	-4.95	0.648

* p-value greater than 0.020
† p-value less than 0.020 and greater than 0.05

When compared to the OLS regression model developed previously, several variables were not statistically significant at the 80 percent confidence level (p-value greater than 0.20) for the multilevel model. These variables were: change in lane width (p-value = 0.377), change in lateral clearance (p-value = 0.410), and introduction of a curb (p-value = 0.689). In addition, four variables that were previously found statistically significant at the 95 percent confidence level (p-values less than 0.05) for the OLS model, were significant between the 80 and 95 percent confidence levels (p-values between 0.20

and 0.05). These variables were: total number of driveways (p-value = 0.091), Intersection Ahead warning sign (p-value = 0.153), Curve Ahead warning sign (p-value = 0.104), and presence of a tangent roadway section (p-value = 0.067).

Those variables that were consistently significant at the 95 percent confidence level for both the multilevel and OLS regression models had estimates that were similar in both sign and magnitude, differing by an absolute value less than 0.6. The standard errors obtained with the OLS regression model were smaller when compared to those obtained with the multilevel model. However, similar as the analyses for point speeds, multilevel models have two advantages: they better represent the data hierarchy and they provide information of the variance at each level of the data hierarchy. Therefore, an iterative process was used to develop a multilevel model which included the variables found to significantly influence changes in operating speeds. The results of the multilevel model are shown in Table 36 below.

Table 36 Two-Level Model for Speed Differential

Parameter	Estimate	SE	Z	p-value
Speed1 Centered	0.17	0.017	9.85	<0.001
Speed Limit 55-40 mph	2.91	1.056	2.75	0.006
Speed Limit 45/40-25 mph	3.52	0.973	3.62	<0.001
Delta Paved Shoulder	0.98	0.262	3.76	<0.001
Total Driveways	0.35*	0.227	1.54	0.123
Intersection WS	2.03*	1.551	1.31	0.191
School/Children	7.65	1.817	4.21	<0.001
Curve WS	-3.26*	1.718	-1.90	0.058
Transition Zone Length	0.71	0.245	2.92	0.004
Curve with WS	2.91	1.263	2.31	0.021
Tangent	-1.90*	1.017	-1.87	0.061
Constant	-4.45	1.730	-2.57	0.010
Random-effects				
Site	1.4876	0.2818	-	-
Residual	5.9502	0.0790	-	-
* Not significant the 95 percent confidence level				

Three of the variables shown in Table 36 were not significant at the 0.05 alpha-level: total number of driveways and both presence of Intersection Ahead and Curve

Ahead warning signs. The results indicated that the variable of Intersection Ahead warning sign was significant at the 80 percent confidence level. Similarly, the variables of number of driveways and Curve Ahead warning sign were significant at the 85 and 90 percent confidence levels, respectively. The influence on speed reductions along the transition zone of the variables shown in Table 36 based on their estimates can be interpreted as follow:

- *Speed1 Centered*: a mean speed reduction of 0.17 mph is expected for each unit increase in vehicle speed 500 ft before the beginning of the transition zone.
- *Speed Limit 55-40*: a mean speed reduction of 2.9 mph is expected when the speed limit decreases from 55 to 40 mph, as compared to the baseline of a change in posted speed limit from 55 to 35 mph.
- *Speed Limit 45/40-25*: a mean speed reduction of 3.5 mph is expected when the speed limit decreases from either 45 or 40 mph to 25 mph when compared to the baseline of a posted speed limit reduction from 55 to 35 mph.
- *Delta Paved Shoulder*: a one-foot paved shoulder width reduction is associated with a mean speed reduction of 1mph.
- *Total Driveways*: a mean speed reduction of 0.35 mph is expected for each unit increase in the total number of driveways in a transition zone.
- *Intersection Ahead Warning Sign*: the presence of an Intersection Ahead warning sign is associated with a mean speed reduction of 2 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *School/Children Warning Sign*: the presence of a warning sign related to school or presence of children is associated with a mean speed reduction of 7.7 mph when compared to the baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.
- *Curve Ahead Warning Sign*: the presence of a Curve Ahead warning sign is associated with a mean speed increase of 3.3 mph when compared to the

baseline of no warning sign or the presence of a warning sign that does not indicate a change in highway alignment, a change in access density, or indicates presence of children.

- *Transition Zone Length*: a mean speed reduction of 0.7 mph is associated with every 100 ft increase in transition zone length.
- *Curve with Warning Sign*: the presence of a horizontal curve that warrants a warning sign is associated with a mean speed reduction of 2.9 mph when compared to the baseline of presence of a curve without a Curve Ahead warning sign.
- *Tangent*: the presence of a tangent section is associated with a mean speed increase of 1.9 mph when compared to the baseline of presence of a curve without a Curve Ahead warning sign.

The values for the standard between- and within- standard deviations indicate a variability of 1.5 mph associated with the site level and almost 6 mph variability for the residual term which cannot be explained by the variables included in the model.

In addition, the output provided by Stata includes the results of the likelihood-ratio test that tests the null hypothesis that the estimates obtained by linear regression are more efficient. The test results indicated a value of $\chi^2 = 66.68$ (p-value = <0.001) thus rejecting the null hypothesis and favoring the two-level model.

Similar to the multilevel models developed for the alternative hierarchy that consider the sensor cluster in the point speed analyses (see Section 5.1.2), random intercept and random coefficient models were developed for the two-level model that considered the difference in speed as the response variable. Once again, the speed at sensor 1 was included in these models as a random intercept and as a time-varying variable with a random coefficient. The random intercept and random coefficient models are shown below in Equations (64) and (65), respectively:

$$\Delta Y_{2-3\,jk} = \beta_1 + \sum \beta X_k + \varsigma_k^{(2)} + \varsigma_j + \varepsilon_{ijk} \qquad (64)$$

$$\Delta Y_{2-3\,jk} = \beta_1 + \sum \beta X_k + \beta L_{jk} + \varsigma_k^{(2)} + \varsigma_j L_j + \varepsilon_{jk} \qquad (65)$$

where: ς_j = random intercept for speed at sensor 1 of driver *j*; and

L_j = is the time-varying variable (speed at sensor 1) for driver *j*.

The models developed that correspond to Equations (64) and (65) are shown in Table 37 below.

Table 37 Random Intercept and Random Coefficient Models for Two-Level Speed Differential Prediction Model

Model:	*Random Intercept*		*Random Coefficient*	
Parameter	**Estimate**	**St. Error**	**Estimate**	**St. Error**
Speed1 Centered	-	-	0.14	0.029
Speed Limit 50-40 mph	3.59	1.121	3.39	1.132
Speed Limit 45/40-25 mph	3.80	0.989	3.95	0.995
Delta Paved Shoulder	1.19	0.266	1.17	0.269
School/Children WS	10.20	1.837	9.95	1.844
Curve WS	-2.70*	1.779	-2.61†	1.793
Transition Zone Length	0.90	0.230	0.89	0.231
Curve with WS	2.97	1.344	2.78	1.355
Tangent	-2.88	1.033	-2.73	1.043
Constant	-4.52	1.783	-4.39	1.799
Random Components				
Site	0.1692	0.0314	0.1009	0.0237
Speed1 Centered	1.5721	0.3068	1.5941	0.3046
Residual	5.9064	0.0787	5.9052	0.0786

* p-value = 0.130
† p-value = 0.145

The parameter estimates of the variables obtained from both the random intercept and random coefficient models were very similar to each other, differing by an absolute value of 0.3 or less. When compared to the previous model that only considered previous speed as an explanatory variable, some of these estimates differed by more than a value of one. In addition, the variables for number of driveways and Intersection Ahead warning sign were not significant at the 80 percent confidence level for the random models when compared to the two-level models that included speed at sensor 1 as an explanatory variable. The interpretations of the parameter estimates obtained with the random intercept and random coefficient models are as follow:

- *Speed1 Centered*: the random coefficient model indicated a mean speed reduction of 0.14 mph per unit increase in the vehicle speed at sensor 1 compared to an initial speed reduction of 0.17 mph in the two-level model shown in Table 36.

- *Speed Limit 55-40*: an average speed reduction of 3.5 mph was associated with this variable in both random models when compared to a speed reduction of 2.9 mph indicated by the initial two-level model shown in Table 36.
- *Speed Limit 45/40-25*: a mean speed reduction of 3.9 mph is associated with this variable for the random models when compared to an initial speed reduction of 3.5 mph indicated in the two-level model shown in Table 36.
- *Delta Paved Shoulder*: the random models indicated an average speed reduction of 1.2 mph per every one-foot of paved shoulder width reduction when compared to an initial speed reduction of 1 mph in the two-level model shown in Table 36.
- *School/Children Warning Sign*: the initial two-level model shown in Table 36 indicated a mean speed reduction of 7.7 mph associated with this sign while the random models estimated an average speed reduction of 10.1 mph for the same variable.
- *Curve Ahead Warning Sign*: the random models indicated an average speed increase of 2.7 mph for the presence of a Curve Ahead warning sign while the initial two-level model shown in Table 36 estimated a mean speed increase of 3.3 mph for this variable. This variable was significant at the 85 percent confidence level for both random intercept and random coefficient models.
- *Transition Zone Length*: the initial two-level model shown in Table 36 estimated a mean speed reduction of 0.7 mph for each 100 ft increase in the transition zone length while the random models estimated an average speed reduction of 0.9 mph.
- *Curve with Warning Sign*: the initial two-level model shown in Table 36 indicated a mean speed reduction of 2.9 mph due to the presence of a horizontal curve that warrants a warning sign and this same value was estimated by the random models.

- *Tangent*: the random models estimated an average speed increase of 2.8 mph associated with a tangent section when compared to a speed increase of 1.9 mph for the initial two-level models shown in Table 36.

The standard errors for the estimates were also almost identical for the random models, with the random intercept model having standard errors slightly greater than the ones for the random coefficient model (between 0.001 and 0.014 greater). The values for the between- and within- standard deviations were also similar between the random intercept and the random coefficient models. When adding the time-varying variable of speed at sensor 1 (*speedS1ctr*) and including a random coefficient for it (random coefficient model), the variability associated with this decreases from 0.17 to 0.1 mph as seen by the values of its standard deviation.

The likelihood-ratio tests resulted in χ^2 values of 188.69 and 129.84 for the random intercept and the random coefficient models, respectively, thus rejecting the null hypothesis that estimates obtained with linear regression are more efficient. Therefore a multilevel model for speed differential is preferred over simple linear regression. Contrary to multilevel models developed before (See section 5.1.2), since a driver cluster is not specified, adding the variable for previous speed (speed at sensor 1) does not significantly change the parameter estimates of other variables, thus highway characteristics have been found to significantly influence operating speeds in an consistent manner. The random coefficient model explains more of the variability as seen in the standard deviation values for the random terms. This model also provides more information due to including a random coefficient for the variable for speed at sensor 1.

5.2.7 Speed Differential Analyses Summary

Contrary to speed point analyses, a new response variable was computed by subtracting the speed at the end of the transition zone (sensor 2) from the speed at the beginning of the transition zone (sensor 2). The new dataset included 2859 speed differential observations from each of the 2859 vehicles collected in the field. Two statistical methods were explored: ordinary least squares (OLS) linear regression and multilevel models.

The assumptions of OLS were met, thus indicating that a linear regression model was appropriate to model mean speed reduction between the limits of the transition zone. Correlation analyses and one-way ANOVA were performed to initially select the explanatory variables that were associated with the response variable. Besides changes in speed limit, lane width, paved shoulder, and lateral clearance differences were associated with changes in the expected mean speed along transition zones. The presence of warning signs for the presence of intersection and related to school and children as well as an increase in driveway density were associated with mean speed reductions, as was the presence of curb. The presence of a Curve Ahead warning sign was associated with mean speed increases. The results of the linear regression analysis indicated that the length of a transition zone was positively associated with mean speed reductions. Finally, changes in horizontal alignment are associated with mean speed reductions. These speed reductions are greater than if the horizontal curve warrants a warning sign.

Two-level models were explored in which speed differential observations were nested in sites. Three multilevel models were developed based on the manner in which speed at sensor 1 was incorporated into the model: as an explanatory variable only, as a random term only (random intercept model), and as a time-varying variable with a random coefficient (random coefficient model). The highway characteristics found to be significant were, for the most part consistent and similar to the ones originally obtained with the linear regression model. Table 38 shows the estimates and their standard errors obtained with all models developed for predicting speed differentials along transition zones. Only statistically significant variables at the 80-percent confidence level are included in Table 38.

Table 38 Speed Differential Models Comparison

Parameter	Estimates (Standard Error)			
	Linear Regression	Two-Level Models		
		Explanatory Variable	Random Intercept	Random Coefficient
Speed1 Centered	0.16 (0.017)	0.17 (0.017)	-	0.14 (0.029)
SL 55-40 mph	2.98 (0.394)	2.91 (1.056)	3.59 (1.121)	3.39 (1.132)
SL 45/40-25 mph	2.94 (0.399)	3.52 (0.973)	3.80 (0.989)	3.95 (0.995)
Delta Lane Width	2.41 (0.978)	-	-	-
Delta Paved Shoulder	1.06 (0.121)	0.98 (0.262)	1.19 (0.266)	1.17 (0.269)
Delta Lateral Clearance	0.09 (0.040)	-	-	-
Total Driveways	0.38 (0.081)	0.35† (0.227)	-	-
Curb Introduction	1.21 (0.547)	-	-	-
Intersection WS	3.11 (0.615)	2.03† (1.551)	-	-
School/Children WS	7.33 (0.644)	7.65 (1.817)	10.20 (1.837)	9.95 (1.844)
Curve WS	-3.60 (0.614)	-3.26* (1.718)	-2.70† (1.779)	-2.61† (1.793)
Transition Zone Length	0.68 (0.091)	0.71 (0.245)	0.90 (0.230)	0.89 (0.231)
Curve with WS	4.27 (0.448)	2.91 (1.263)	2.97 (1.344)	2.78 (1.355)
Tangent	-1.31 (0.348)	-1.90 (1.017)	-2.88 (1.033)	-2.73 (1.043)
Constant	-4.95 (0.648)	-4.45 (1.730)	3.59 (1.121)	0.14 (0.029)

* p-value between 0.05 and 0.1
† p-value between 0.1 and 0.20

Including the speed at sensor 1 as an explanatory variable in the multilevel models resulted in three variables not being statistically significant predictors of mean operating speed reductions along two-lane rural highway transition zones (change in lane width, change in lateral clearance, and presence of curb). In addition, when including speed at sensor 1 only as a fixed explanatory variable, the variables for total number of driveways and both indicator variables for presence of a Curve Ahead and Intersection Ahead warnings signs were not statistically significant (p-value greater than 0.05 but less than 0.20). When adding speed sensor 1 as either a random intercept or as a time-varying variable with random coefficient (random intercept and random coefficient models, respectively) the variables for number of driveways and Intersection Ahead warning sign were not statistically significant (p-value greater than 0.20). The variable for Curve Ahead warning sign was significant at the 85 percent confidence level for the random models (p-value less than 0.15). All other geometric design, roadside, and traffic control, were found to be statistically significant in the multi-level models.

As shown in Table 38, a speed limit reduction from 55 to 40 mph was associated with speed reductions ranging from 2.9 to 3.6 mph, while a speed limit reduction from either 45 or 40 mph to 25 mph was associated with speed reductions ranging from 2.9 to 4 mph. A one-foot reduction in paved shoulder width was associated with speed increases along the transition zones of approximately 1 mph, regardless of which speed differential model was applied. The range of the estimates obtained for the presence of a School/Children warning sign indicated speed reductions from 7.3 to 10.2 mph. The presence of a horizontal curve that warrants a warning sign was associated with speed reductions from 2.8 to 4.3 mph, while the presence of a tangent section was associated with speed increases ranging from 1.3 to 2.9 mph. Finally, all speed differential models indicated that longer transition zones were associated with greater speed reductions; per every 100 feet of transition zone length, a mean speed reduction ranging from 0.7 to 0.9 mph is expected.

The standard errors of the estimates obtained by linear regression are smaller than those obtained from the multilevel models as shown in Table 38. This indicates that the standard errors obtained by linear regression may be underestimated when compared to other modeling methods that account for the hierarchical nature of the data. In addition, the OLS model identified a greater number of highway characteristics as statistically significant when compared to those in the multilevel models. However, when developing the multilevel models, the output in Stata provides the result from a likelihood-ratio test that tests the efficiency of the estimates as compared to linear regression estimates; these always favored the use of multilevel models. In addition, multilevel models are able assign the variability in speed differentials associated with each level, information that linear regression models fail to provide. Thus, it is recommended that a random coefficient two-level model is more appropriate to predict speed differentials along transition zones.

CHAPTER 6
CONCLUSIONS AND RECOMMENDATIONS

Speed data were collected at several transition zone sites in Central Pennsylvania, in which a Reduced Speed Ahead sign was present to indicate a regulatory speed reduction to drivers. In order to develop speed prediction models, highway characteristics at each site were collected and included in the data analyses as potential explanatory variables. Several data analysis methods were explored and the coefficients of the explanatory variables found to significantly influence operating speeds were described. This chapter contains conclusions from the research as well as a discussion of how to apply the recommended speed prediction models in highway engineering practice. Lastly, recommendations for future research are provided.

6.1 Conclusions

A total of 11,436 speed observations were included in the dataset which corresponded to 2859 vehicles as they traveled along the four sensors at each of the 20 study sites. The data were organized and analyzed according to two response variables: (1) point speeds at pre-defined data collection locations before, within, and after the transition zone; and (2) speed differences between the two sensor locations that defined the limits of the transition zone.

The point speed analyses considered panel data models, multilevel models, and generalized estimating equations (GEE), an extension of the generalized linear model (GLM) for continuous and discrete outcomes. Each method can be used to model longitudinal data and are able to account for the correlation between observations due to driver-specific information, which cannot be addressed with OLS regression. Several variables were consistently found to significantly influence operating speeds along transition zones, regardless of the data analysis method. A summary of the findings for each model is listed below:

- *Panel Data*. Both fixed-effects and random-effects models were explored and, although the results from the fixed-effects model indicated low values of correlation between the explanatory variables and the between-subject error term, the Hausman test rejected the null hypothesis that this correlation was zero, thus favoring the fixed-effects model. An examination of aggregate versus disaggregate data confirmed that using aggregate data results in an ecologic fallacy: the estimates obtained for a group do not necessarily apply to an individual.

- *Multilevel Model*. Panel data models are only able to accommodate two levels of the data, therefore they fail to account for site variability within the data structure hierarchy represented by the data collected in the present study. A three-level model was specified in which speeds were nested in drivers which were nested in sites. The parameter estimates were obtained using the maximum likelihood estimator and the variance component term for the combination of site and driver was obtained from the results. Multilevel models also permitted the addition of a previous speed variable as a random component. An alternative hierarchy was explored in which the driver cluster was replaced by the sensor cluster, thus specifying that speeds were nested in sensor locations which were nested in sites. Since the alternative data hierarchy did not include driver-specific information, the variable for previous speed could be added in the model as an explanatory variable. In addition, this variable could also be considered as both a random term (random intercept model) and as a time-varying variable with a random coefficient (random coefficient model).
- *GEE Models*. One of the advantages of performing GEE analysis was to explore which working correlation matrix best represented the data in the present research. Based on the marginal coefficient of determination (R^2m) and the quasi-likelihood under independence model criterion (QIC), the independent working correlation matrix, which specified that speed observations for the same driver are independent from each other, provided the best fit for the data. However, alternative correlation structures, such as the exchangeable, unstructured, and autoregressive, are more intuitive based on the data collection protocol. When considering only these three correlation

structures, the exchangeable matrix provided the best fit to the data based on the R^2m and QIC values.

The comparison between the random- and fixed-effects panel data models led to the notion that the speed limit and lane width variables were picking up site-specific effects, as indicated by the differences between the standard errors and parameter estimates for these variables. For some roads, the speed limits are set by local jurisdiction while for other roads, the speed limits, as well as the highway geometrics, are set according to design guidelines. The latter can lead to endogeneity issues and future studies should explore this matter further.

Although panel data is a two-level model, the findings of this study indicated that a three-level model in which speeds were nested in drivers and drivers were nested in sites was more appropriate to model the data. An advantage of the alternative site-sensor-speed data is that it allows for the previous speed variable to be included in the model, either as an explanatory variable, a random intercept, or a time-varying variable with random coefficient. However, adding previous speed resulted in some variables being excluded from the model because they were not statistically significant; these variables were: speed limit reduction from 45 to 25 mph, presence of Curve Ahead warning sign, and presence of a tangent section (only when considering previous speed as an explanatory variable). This was expected since it is theorized that speeds are influenced by highway characteristics, therefore, including previous speed as an explanatory variable in the model results in multicollinearity. Thus the original hierarchy – site-driver-speed – is most appropriate for representing the data structure as compared to the hierarchy that considers a sensor cluster.

When examining the results from the GEE analyses, these also produced coefficient estimates similar to the panel data and multilevel models. However, concerns arose due to the selection of the Independent working correlation matrix as the best representative of the correlation within driver clusters (for the same driver). The Independent correlation matrix specifies that, for a specific driver, the correlation between the speed observations is zero, which is counterintuitive. This led to the selection of the Exchangeable working correlation matrix across all GEE model although the selection criteria did not originally favor this model. When comparing all longitudinal data models, it was concluded that a three-level model with data hierarchy site-driver-speed best represented the data in this research.

According to the three-level model selected, in which speeds were nested in drivers and drivers were nested in sites, the following variables were associated with lower speeds along transition zones: posted speed limit, number of driveways, presence of curb, warning signs related to presence of intersection and presence of children and school, and changes in horizontal alignment. On the contrary, the highway characteristics that were associated with higher speeds were wider lane widths, wider lateral clearance distances, and presence of a Curve Ahead warning sign. The development of a three-level model provides an advantage over panel data and GEE models: the variance associated at each level of the data hierarchy can be obtained. The results of the three-level model indicated that approximately 3.4 mph of standard deviation is associated with the combination of site and drivers while a standard deviation of 4.5 mph is associated with the driver cluster (second level). The standard deviation associated with the residual term was 5 mph; this is a measure of the variance that cannot be explained by the explanatory variables included in the model.

In the second part of the speed analysis, the response variable was change in operating speeds along the transition zone (between sensors 2 and 3) as opposed to point speeds collected at all four sensor locations. The speed differences were modeled using both linear regression and multilevel models. By considering speed differential as the response variable, only one observation per driver is available. Therefore for multilevel models, only a two-level model in which speeds were nested in sites was applied. The number of variables significant in the OLS model was greater than those found in the multilevel models. However, likelihood-ratio tests always favored the estimates obtained with multilevel models when compared to those obtained by linear regression. The variables found to be associated with speed reductions in the multilevel models were posted speed limit reductions, reduction in paved shoulder width, presence of warning signs related to school and children, longer transition zone lengths, and presence of a horizontal curve that may be perceived as “sharp” since it is combined with a Curve Ahead warning sign.

The speed at the upstream location (sensor 1, upstream of the transition zone) was found to be statistically significant regardless of the methodology. The results indicated that the higher the speeds before the transition zone, the greater the speed reductions (drivers that were traveling at low speeds do not reduce their speeds as much as those traveling at higher speeds prior to the transition zone). The only two variables associated with speed increases within a transition zone were presence of a Curve Ahead warning sign (statistically significant at the 80 percent confidence level for multilevel models) and presence of a tangent highway section. Tangent sections have consistently been found to be associated with higher speeds than curved roadway sections, and as a result, speed differentials in transition zones that do not have a horizontal alignment change are lower than speed differentials in transition zones with horizontal alignment changes (McLean, 1979; Andjus and Maletin, 1998; and Misaghi and Hassan, 2005). For the presence of a Curve Ahead warning sign, the finding in the present research may be inconsistent with engineering intuition. However, the Curve Ahead warning sign is placed on the approach tangent in advance of a horizontal curve. Based on the findings of this research, tangent sections are associated with speed increases (a correlation analysis showed that the presence of this warning sign and the tangent section indicator were not strongly correlated).

The point speed and speed differential models estimated in this research both provided consistent results related to the association between mean speed and various explanatory variables present along two-lane rural highway transition zones. Tables 30 and 38 in Chapter 5 contain a comparison of the parameter estimates obtained using the various statistical models estimated in this dissertation. A brief summary of these findings are provided below.

In the point speed models, the variables that were associated with speed reductions along transition zones, regardless of the statistical analysis methodology used, were:

- Reductions in the posted speed limit
- Increase in number of driveways
- Presence of a curb
- Presence of warning signs related to intersection and school/children

- Presence of a horizontal curve, with or without a warning sign

The variables that were consistently found to be associated with speed increases along two-lane rural highway transition zones in the point speed models were:

- Increase in lane width
- Presence of a Curve Ahead warning sign

In the speed differential models, the highway characteristics that were associated with a speed reduction along two-lane rural highway transition zones were:

- Posted speed limit reductions from 55 to 40 mph and from either 45 or 40 mph to 25 mph when compared to the baseline of speed limit reduction from 55 to 35 mph
- Decrease in paved shoulder width
- Presence of warning signs related to intersection and school/children
- Presence of horizontal curve that warrants a warning sign

There are two explanatory variables that were not statistically significant in the panel data and GEE model specifications but that should be carefully examined in future studies (paved shoulder width and lateral clearance). In the fixed-effects panel data model, when speed limit was not considered in the point speed analysis, paved shoulder width was statistically significant (positively correlated with speed) which is consistent with the results for the speed differential analysis. In the GEE model specification, lateral clearance was not statistically significant when using the independent working correlation matrix. However, lateral clearance was statistically significant when using all other GEE working correlation matrix specifications. As noted previously, the marginal coefficient of determination (R^2_m) and QIC criterion indicated that the independent working correlation matrix produced the best fit to the data collected in the present research; however, the alternative working correlation matrices are intuitively more representative of the data.

Although the three-level model in which speeds are nested in drivers and drivers are nested in sites provides estimates with standard errors higher than other models, these standard errors are modestly higher. In addition, the three-level model is a better representation of the data hierarchy, thus it is selected as the most appropriate model for point speed predictions along transition zones. Similarly, the two-level model for predicting speed changes between the limits of the transition zone is also selected as the model that best represents the data.

6.2 Application and Relevance to Transportation Engineering

The mission of the American Association of State Highway and Transportation Officials (AASHTO) is to “advocate transportation-related policies”; its Green Book contains geometric design criteria for highways and streets in the U.S. The Green Book contains geometric design criteria for all functional class highways, including high-speed two-lane rural highways as well as low-speed urban streets. However, design criteria are not available for the transition from a high-speed zone to a low-speed zone and vice versa.

Extensive literature exists that focuses on the development of speed prediction models as a function of the driving environment for both high- and low-speed roads. Limited literature is available for highways in which changes in operating speeds are required as indicated by changes in the regulatory speed. Future studies may benefit from the data analysis methodology presented in this research as well as from the results described herein. Although the goal of this research was not to develop design criteria for transition zones, the results from this study may be considered as an initial step in the process of guidelines development. By knowing which geometric design, roadside, traffic control, and land use variables influence vehicle operating speeds in transition zones, future research can be focused on validating the results, expanding the number of possible explanatory variables included in statistical model specifications, and then developing guidelines for creating “self-enforcing” transition zones.

The majority of past operating speed studies used OLS regression to determine the statistical association between speed and various explanatory variables. Recent research studies have considered the use of different models to explain the variability in operating speeds. Although linear regression was considered for speed differential prediction models, a two-level model was also applied for predicting speed changes along transition zones. For point speeds, several longitudinal data analysis methods were considered, including panel data, multilevel models, and GEE models. The appropriateness of a three-level model was established for the point speed analysis, thus demonstrating the importance of selecting a model that best represents the hierarchy of the data structure.

The data analyses performed in this research included several highway characteristics that have not been considered in past studies as potential factors that influence operating speeds. The results indicated that presence of both Intersection Ahead and School/Children warning signs are associated with speed reductions; these are traffic control devices that have not been explored in previous studies. In addition, although past studies include the value of horizontal curve radii, these studies have not considered the effects of a horizontal curve that warrants a warning sign. The exploration of warning signs and the results confirming their effect on speed parameters may inspire future researchers to collect information related to the benefits of traffic control devices in reducing vehicle operating speeds.

As discussed earlier, multilevel models, for both point speed and speed differential analyses, were selected as the modeling methodology to best describe the data in this research. In addition, regardless of the methodology, several variables were consistently found to influence operating speeds along transition zones. However, there are advantages and disadvantages related to the use of either point speed or speed differential models developed in this study. For the point speed prediction model developed, highway site characteristics should be collected at four locations along the study site: at the beginning and end of the transition zone (identified by the location of the posted speed limit signs) and 500 ft before and after the transition zone. On the contrary, the speed differential model developed requires the collection of highway characteristics only at two locations: at both the beginning and at the end of the transition zone.

Although one advantage of the speed differential model is that it requires less highway characteristic data, and consequently less data to be input in the model, this model also requires that speed data should be collected 500 ft before the beginning of the transition zone (the operating speed at this upstream location is included as an explanatory variable in the speed differential models). The collection of speed data is associated with several model application disadvantages: the selection of a non-intrusive data collection device so that drivers do not perceive the data collection equipment as enforcement, the possibility of stopping traffic in order to install the data collection device, and the need to screen the data in order to identify free-flow passenger cars. To overcome these disadvantages, the following suggestions are recommended:

- Input the mean speed at sensor 1 obtained in this study for the upstream speed explanatory variable in the speed differential models. The value of the mean speed at this location was 53.2 mph (standard deviation = 8.32 mph). In addition, the mean speed value, plus or minus one standard deviation, can also be input in the model to explore the changes in speed differentials for the majority of the driving population.
- Input the posted speed limit as the speed at sensor 1. The posted speed limit in the high-speed zone at 14 study sites in the research was 55 mph and the posted speed limit at five sites was 45 mph; the remaining site had a posted speed limit of 40 mph 500 ft before the beginning of the transition zone.
- Use of a simulation program to obtain the expected mean speed before the beginning of the transition zone and input this value as the speed at sensor 1. The FHWA Interactive Highway Safety Design Model's (IHSDM) Traffic Analysis Module contains a two-lane rural highway simulation model that can be used for this purpose. In the program, the highway alignment, vertical profile, and cross-section can be entered and vehicle speed at various points along the alignment can be output.

While the use of the point speed prediction model requires the collection of various geometric design, roadside, traffic control, and land use data, such models do not require the collection of vehicle operating speed data. Additionally, the point speed models are able to predict mean operating speeds beyond the limits of the transition zone, which could be advantageous in determining how drivers adopt their operating speeds before entering a two-lane highway transition zone or downstream of the transition zone segment.

6.3 Recommendations

The final dataset for this research consisted of 11,436 speed observations from 2859 vehicles, across 20 sites. In some cases, the variability in the highway site characteristic data was limited. Future research should include a larger sample of two-lane rural highway transition zones with more variability in the explanatory variables considered in this research, particularly the horizontal alignment, vertical profile, and cross-section elements.

The speed prediction models developed in the present study indicate that the presence of a horizontal curve was associated with operating speed reductions in transition zones along two-lane rural highways. Although an indicator variable was used to define the presence of horizontal curve, the radius of curve was not available. This was because as-built roadway construction plans were not available for most of the selected study sites. Future operating speed models for two-lane rural highway transition zones should include the as-built radius as an explanatory variable rather than an indicator variable for the presence of a horizontal curve.

Similarly, drivers may perceive the presence of a Curve Ahead warning sign as an indication of an upcoming "sharp" curve that requires significant operating speed reductions when compared to horizontal curves that are not accompanied by an advance warning sign. Although the advance curve warning sign was found to influence operating speeds (positive correlation), the presence of this warning sign does not necessarily indicate that the curve requires a significant operating speed change. This underscores the need to include the as-built horizontal curve radius in future operating speed prediction models along two-lane rural highway transition zones.

Lastly, the statistical models estimated in the present study were for operating speeds and not speed variance. The existing literature indicates that speed variance can be used as a surrogate measure of safety; therefore, future operating speed prediction models for two-lane rural highway transition zones should consider both mean speed and speed variance. Design consistency is usually measured in terms of speed changes between adjacent roadway segments (e.g., tangent-to-curve): low values for speed differentials are associated with a good and consistent design (Glennon and Harwood, 1978; McLean, 1979; McFadden and Elefteriadou, 2000; and Fitzpatrick and Carlson, 2002). As such, the principle of design consistency would indicate that large speed differentials along a highway alignment are not desirable. In the case of transition zones, however, a speed differential is desired. Several studies have suggested that large speed differentials are associated with an increase in accident frequency (Garber and Gadiraju, 1989; and Lamm, et al., 2002). As such, the goal of transition zone design should be to create an alignment, profile, cross-section, and roadside that produces a gradual change in the speed profile of drivers when traveling from a high- to low-speed section of highway. The following is a list of future research recommendations that should be considered to address the issue of speed differentials in transition zones on two-lane rural highways:

1. Consider using a simulation model as a tool to efficiently develop a variety of "test cases" or design scenarios to further explore the association between highway design features and vehicle operating speeds along transition zones of two-lane rural highways. This would permit researchers to isolate the effects of various geometric elements on vehicle operating speeds and also to evaluate the effects of a combination of highway elements on operating speeds (e.g., overlapping horizontal/vertical curves, consecutive horizontal curves with increased radii, narrowing lane/shoulder widths, etc.). The goals of such an approach would be to create design scenarios that produce a gradual decrease in vehicle operating speeds over a pre-determined transition zone length. The IHSDM is an example of a tool that could be used for this purpose. As noted previously, it contains a traffic simulation program in the Traffic Analysis Module (TWOPAS) as well as a design consistency algorithm.
2. Perform crash-based safety studies along transition zone highway sections. Data on crash frequency and severity of crashes should be collected along transition zones and compared to other two-lane rural highways in the absence of transition zones. Although it is perceived that operating speeds in excess of the posted speed limit is a safety concern, research is required to investigate this claim. The influence of highway and roadside design features on crash parameters can then be explored in a similar manner to the operating speeds in the present research. Crash prediction models could then be utilized to identify the highway characteristics that are associated with crash frequency and crash severity and compared to the same highway features used in the operating speed prediction models. Together, these models could be used to

design transition zones that not only achieve the desired operating speeds, but also produce desirable safety outcomes.

3. Determine a threshold speed differential value over some specified transition zone length that can be used to determine if these zones are substantively "safe" (no negative safety implications) or "unsafe" (associated with an increase in vehicle accidents). It is then recommended that, when exploring the highway characteristics that are associated with speed reductions along transition zones, to be certain that these are associated with gradual (and safe) speed reductions that will not compromise highway safety.
4. Future studies should also focus on two-stage speed limit reduction transition zones, in which speed limit changes take place along two sections of the highway as opposed to a one-stage speed limit reduction. An example of a two-stage speed reduction could be when a speed limit of 55 mph is reduced over two adjacent sections to 25 mph; the first section of the highway indicates a speed limit reduction from 55 to 40 mph while the second section indicates a speed limit reduction from 40 to 25 mph.

REFERENCES

Abdel-Aty, M., and X. Wang. Crash Estimation at Signalized Intersections Along Corridors: Analyzing Spatial Effect and Identifying Significant Factors. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1953*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 98-111.

Andjus, V., and M. Maletin. Speeds of Cars on Horizontal Curves. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1612*, Transportation Research Board of the National Academies, Washington, D.C., 1998, pp. 42-47.

A Policy on Geometric Design of Highway and Streets. AASHTO, Washington, D.C., 2004.

Ballinger, G. A. Using Generalized Estimating Equations for Longitudinal Data Analysis. In *Organization Research Methods*, Sage Publications, 2004.

Brüderl, J. Panel Data Analysis. Manheim; Baden-Württemberg: University of Manheim, 2005.

Choueiri, E. M., R. Lamm, J. H. Kloeckner, and T. Mailaender. Safety Aspects of Individual Design Elements and Their Interactions on Two-Lane Highways: International Perspective. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1445*, Transportation Research Board of the National Academies, Washington, D.C., 1994, pp. 34-46.

Cui, J. QIC Program and Model Selection in GEE Analyses. *The Stata Journal*, Vol. 7, No. 2, 2007, pp. 209-220.

Donnell, E. T. and I. Cruzado. *Effectiveness of Speed Minders in Reducing Driving Speeds on Rural Highways in Pennsylvania*. Final Report prepared for the Pennsylvania Department of Transportation, 2007.

Evans, L. *Traffic Safety and the Driver*. Van Nostrand Reinhold, a division of International Thomson Publishing, Inc., New York, 1991.

Fatality Analysis Reporting System (FARS). *2005 National Highway Traffic Safety Administration*. www.-fars.nhtsa.gov/. Accessed Nov. 23, 2008.

FHWA, *The U. S. Department of Transportation Rural Safety Initiative*. February 2008. www.dot.gov/affairs/ruralsafety/ruralsafetyinitiativeplan.htm. Accessed Nov. 22, 2008.

Figueroa, A. M., and A. P. Tarko. Speed Factors on Two-Lane Rural Highways in Free-Flow Conditions. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D,C, 2005.

Fitzpatrick, K., and P. Carlson. Selection of Design Speed Values. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1796*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 3-11.

Fitzpatrick, K., S. Miaou, M. Brewer, P. Carlson, and M. D. Wooldridge. Exploration of the Relationships between Operating Speed and Roadway Features on Tangent Sections. *Journal of Transportation Engineering*, Vol. 131, No. 4, 2005, pp. 261-269.

Garber, N., and R. Gadiraju. Factors Affecting Speed Variance and its Influence on Accidents. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1213*, Transportation Research Board of the National Academies, Washington, D.C., 1989, pp. 64-71.

Garrett, T. A. Aggregated versus Disaggregated Data in Regression Analysis: Implications for Inference. *Economics Letters*, Vol. 81, No. 1, 2003, pp. 61-65.

Geometric Design Strategic Research, Transportation Research Circular E-C110, Transportation Research Board of the National Academies, Washington D.C., 2007.

Ghisletta, P. and D. Spini. An Introduction to Generalized Estimating Equations and an Application to Assess Selectivity Effects in a Longitudinal Study on Very Old Individuals. *Journal of Educational and Behavioral Statistics*, Vol. 29, No. 4, 2004, pp. 421-437.

Glennon, J. C. and D. W. Harwood. Highway Design Consistency and Systematic Design Related to Highway Safety. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 681*, Transportation Research Board of the National Academies, Washington, D.C., 1978, pp. 77-88.

Greene, W.H. *Econometric Analysis*, Prentice Hall, New Jersey, 2008.

Krammes, R.A., and C. Hayden. Making Two-Lane Roads Safer. *Public Roads*, Vol. 66, No. 4, 2003, pp. 16-21.

Lamm, R., B. Psarianos, and S. Cafiso. Safety Evaluation Process for Two-Lane Rural Roads: A 10-Year Review. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1796*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 51-59.

Lord, D., and B. N. Persaud. Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1717*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 102-108.

Manual of Transportation Engineering Studies (ed. H. D. Robertson). Institute of Transportation Engineers. Prentice Hall, New Jersey, 1994.

Manual on Uniform Traffic Control Devices, FHWA, Washington, DC, 2000 and 2003.

McFadden, J. and L. Elefteriadou. Evaluating Horizontal Alignment Design Consistency of Two-Lane Rural Highways. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1737*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 9-17.

McLean, J. An Alternative to the Design Speed Concept for Low Speed Alignment Design. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 702*, Transportation Research Board of the National Academies, Washington, D.C., 1979, pp. 55–63.

McShane, W. R., R. P. Roess, and E. S. Prassas. *Traffic Engineering*. Prentice Hall, Inc, New Jersey, 1998.

Messer, C. J., Mounce, J. M., and Brackett R.Q. Highway Geometric Design Consistency Related to Driver Expectancy. *Report FHWA-RD-79-35,* Federal Highway Administration, Washington, D.C. (1979).

Misaghi, P., and Y. Hassan. Modeling Operating Speed and Speed Differential on Two-Lane Rural Roads. *Journal of Transportation Engineering*, Vol. 131, No. 6, 2005, pp. 408-418.

Motulsky, D. Multicollinearity in Multiple Regression. www.graphpad.com/articles/Multicollinearity.htm. Accessed Jan. 17, 2009.

NHTSA, *Traffic Safety Facts 2005: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*, Report US DOT HS 810 631, Washington DC, 2006.

Official Traffic Control Devices, Publication 212, Commonwealth of Pennsylvania, Department of Transportation, Pennsylvania, 2006.

Ottesen, J. L., and R. A. Krammes, Speed-Profile Model for a Design-Consistency Evaluation Procedure in the United States. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1701*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 76-85.

Pan, W. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics*, Vol. 57, No. 1, 2001, pp. 120-125.

Park, Y-J., and F. Saccomanno. Evaluating speed consistency between successive elements of a two-lane rural highway. *Transportation Research Part A*, Vol. 40, 2005, pp. 375-385.

Poe, C. M., and J. M. Mason. Analyzing Influence of Geometric Design on Operating Speeds Along Low-Speed Urban Streets. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1737*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 18-25.

Polus, A., K. Fitzpatrick, and D. Fambro. Predicting Operating Speeds on Tangent Sections of Two-Lane Rural Highways. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board,* *No. 1737*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 50-57.

Rabe-Hesketh, S., and A. Skrondal. *Multilevel and Longitudinal Modeling Using Stata*. Stata Press, Texas, 2005.

Richl, L., and T. Sayed. Effect of Speed Prediction Models and Perceived Radius on Design Consistency. *Canadian Journal of Civil Engineering*, Vol. 32, No. 2, 2005, pp. 388-399.

Rowan, N. J., and C. J. and Keese. A Study of Factors Influencing Traffic Speeds. *Accident HRB Bulletin 341*, Highway Research Board, Washington, D.C., 1962, pp. 30-76.

Schurr, K., P. T. McCoy, G. Pesti, and R. Huff. Relationship of Design, Operating, and Posted Speeds on Horizontal Curves of Rural Two-Lane Highways in Nebraska. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1796*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 60-71.

Schurr, K. S., B.W. Spargo, R. R. Huff, and G. Pesti. Predicted 95th Percentile Speeds on Curved Alignments Approaching a Stop. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.

Special Report 254, Managing Speed: Review of Current Practice for Setting and Enforcing Speed Limits, Transportation Research Board, National Academy Press, Washington, D.C., 1998.

Stamatiadis, N., J. G. Pigman, and D. Hartman. *Safety Consequences of Flexibility in Highway Design for Rural Communities*. Draft of Final Report prepared for National Cooperative Highway Research Program, University of Kentucky, 2004.

Stamatiadis, N., J. G. Pigman, and D. Hartman. Safety Consequences from Design Flexibility in Rural to Urban Transitions. Presented at 85th Annual Meeting of the Transportation Research Board, Washington D.C., 2006.

Tarris, J., C. Poe, J. M. Mason, and K. Goulias. Predicting Operating Speeds on Low-Speed Urban Streets: Regression and Panel Analysis Approaches. In *Transportation Research Record: Journal of the Transportation Research Board,* *Transportation Research Board, No. 1523*, Transportation Research Board of the National Academies, Washington, D.C., 1996, pp. 46-54.

Wang, J., K. K. Dixon, H. Li, and M. Hunter. Operating-Speed Model for Low-Speed Urban Tangent Streets Based on In-Vehicle Global Positioning System Data. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1961*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 24-33.

Wang, J. Operating Speed Models for Low Speed Urban Environments Based on In-Vehicle GPS Data: A Dissertation Presented to the Academic Faculty, Georgia Institute of Technology, 2006.

Washington, S. P., M. G. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC Press, Florida, 2003.

Wooldridge, M. D. Design Consistency and Driver Error. In *Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board, No. 1445*, Transportation Research Board of the National Academies, Washington, D.C., 1994, pp. 148-155.

Yagar, S. and M. Van Aerde. Geometric and Environmental Effects on Speeds of Two-Lane Highways. *Transportation Research Part A*, Vol. 17A, 1983, pp. 315-325.

Zeger, S. L. and K. Y. Liang. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, Vol. 42, No.1, 1986, pp. 121-130.

Zorn, C. W. Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *American Journal of Political Science*, Vol. 45, No. 2, 2001, pp. 470-490.

Curriculum Vitae
Ivette Cruzado

Education
Doctor of Philosophy in Civil Engineering, Pennsylvania State University, 2009.
Master of Science in Civil Engineering, Michigan State University, 2001.
Bachelor of Science in Civil Engineering, University of Puerto Rico at Mayagüez, 1999.

Work Experience
Researcher, Pennsylvania State University and the Pennsylvania Department of Transportation: *Study of Bead Gun Angle when Applying Glass Beads on Waterborne Paint*, 2008-2009.
Project Manager, Pennsylvania State University and the United State Sign Council: *The Effects of On-premise Sign Lighting Level on Nighttime Sign Legibility and Traffic Safety*, 2008.
Researcher, Pennsylvania State University and the Pennsylvania Department of Transportation: *Effectiveness of Speed Minders on Rural Highways in Pennsylvania*, 2007-2008.
Researcher, Pennsylvania State University and the Pennsylvania Department of Transportation: *Evaluation of Wide Edge Lines on Horizontal Curves on Two-Lane Rural Highways*, 2006.
Instructor, University of Puerto Rico at Mayagüez, *Highway Design*, *Statistics Applied to Civil Engineering*, and *Civil Engineering Seminar* courses, 2001-2004
Instructor, Polytechnic University of Puerto Rico, *Transportation Engineering* and *Highway Design* courses, 2001.
Research and Teaching Assistant, Michigan State University, 1999-2001.

Professional and Student Associations
Engineering Graduate Student Council (EGSC), 2005-2009.
Civil and Environmental Engineering Graduate Student Association (CEEGSA), 2006-2008.
Colegio de Ingenieros y Agrimensores de Puerto Rico (CIAPR), Active member since August, 2001
Institute of Transportation Engineers (ITE), Active member since August, 1998

Publications
The Effects of Internally Illuminated On-Premise Sign Brightness on Nighttime Sign Visibility and Traffic Safety; M. T. Pietrucha, P. M. Garvey, and I. Cruzado, prepared for the United States Sign Council Foundation, 2009.
Effectiveness of Speed Minders in Reducing Driving Speeds on Rural Highways in Pennsylvania; E.T. Donnell and I. Cruzado, Final Report, prepared for the Pennsylvania Department of Transportation, June, 2008.
Operational Effects of Wide Edge Lines Applied to Horizontal Curves on Two-Lane Rural Highways; E. T. Donnell, M. D. Gemar, and I. Cruzado, prepared for the Pennsylvania Department of Transportation. November, 2006.
Sustainable Transportation Systems; I. Cruzado, Urban Transport XI: Urban Transport and the Environment in the 21st Century, WIT Press 2005.
Safe Ways to School; V. Sisiopiku and I. Cruzado, The Sustainable City II: Urban Regeneration and Sustainability, WIT Press 2002.
Parking on the State Trunkline System; Final Report, prepared for the Michigan Department of Transportation, August 2000.

Fellowships
International Road Federation (IRF) Executive Leadership Fellowship Grant, 2008
Sloan Fellowship Recipient, Pennsylvania State University, 2006-2009.
Carmen E. Turner Graduate Scholarship, WTS Philadelphia, 2005
Sloan Fellowship Recipient, Michigan State University, 1999-2001.

Licenses
Engineer in Training (EIT), license #18008, Puerto Rico, 1999.