Evaluating Effect of Operating Speed on Crashes of Rural Two-Lane Highways

Fahida Rahman
Rowan University

Xu Zhang

Mei Chen

WILEY | Hindawi

Research Article

Evaluating Effect of Operating Speed on Crashes of Rural Two-Lane Highways

Fahmida Rahman [iD],[1] Xu Zhang [iD],[2] and Mei Chen [iD][3]

[1]*Department of Civil & Environmental Engineering, Rowan University, Glassboro, NJ 08028, USA*
[2]*Kentucky Transportation Center, 266 Raymond Bldg. Lexington, KY 40506, USA*
[3]*Department of Civil Engineering, University of Kentucky, Lexington, KY 40506, USA*

Correspondence should be addressed to Fahmida Rahman; rahmanf@rowan.edu

Crashes on a roadway are influenced by various factors, including but not limited to road geometries, traffic volume, and environmental conditions. Among these factors, traffic volume and segment length are commonly used to predict crashes. Recently, the role of speed in crashes has been recognized as a significant factor, prompting its incorporation as a variable in crash modeling. Nevertheless, previous research studies that examined speed-related factors are mostly concentrated on higher functional class roads where speed data are abundant. Lack of actual speed data has limited the scope of such a study on rural two-lane highways. Due to recent advancements in data collection methodologies, there has been a significant increase in the accessibility of speed data pertaining to these roads. This study aims to assess the significance of speed as a predictor of crashes on rural two-lane highways, utilizing actual speed data. The results of this study showed a negative correlation between speed and crash frequency on rural two-lane roadways. In addition, it was observed that the impact of speed in the crash model becomes more pronounced at higher operating speed conditions of these roads. The aforementioned observation prompted us to consider a categorizer based on speed and, afterwards, separating crash prediction models for various speed ranges. This approach ultimately resulted in enhanced accuracy in crash prediction. Based on our analysis, developing separate models at different speed levels is recommended to better evaluate the safety performance of these roads under various conditions. Such models can also be useful for transportation planners and policymakers to identify high-risk segments and allocate resources to improve the safety of these roads.

1. Background

According to a study conducted in the United States, rural two-lane highways account for a significant proportion, specifically 76% of the overall paved road mileages [1]. In Kentucky, a substantial proportion of roadway crashes are attributed to these specific roadways. In particular, they are responsible for forty percent of all crashes, forty-seven percent of crashes that result in injuries, and sixty-six percent of fatal crashes, occurring on roads maintained by the state [2, 3]. There are various factors that lead to roadway crashes, encompassing attributes such as roadway geometric conditions, traffic volume, environmental conditions, and speed characteristics. Of these factors, speed is often cited as a primary cause of crashes [4].

The traditional approach in the Highway Safety Manual (HSM) incorporates annual average daily traffic (AADT) and segment length as the base conditions for crash prediction, which can be further adjusted for different road geometric attributes [5]. Multiple studies have provided empirical evidence supporting the relationship between speed and traffic crashes [4, 6–11]. Furthermore, these studies have recommended the incorporation of speed as one of the variables in crash prediction models [11–17]. The speed considered in the analysis may represent either individual driver speed [4, 6, 7, 9, 10] or aggregated roadway

speed, depending on the purpose of the analysis [8, 11–13, 15]. However, such analyses are predominantly carried out on routes with high levels of traffic, such as interstates and arterials. The examination of the correlation between speed and safety on rural two-lane highways has primarily been conducted within the framework of geometric design consistency [18–22]. Geometric design consistency refers to the uniformity and predictability of road features, such as curves, slopes, and intersections, which can affect driver behavior and safety. Especially, the 85th percentile speed serves as a metric for assessing design consistency throughout different segments. Oftentimes, due to limited available data, speed is calculated through models.

In recent times, there has been a notable increase in the availability of measured speed data, especially on higher functional class roads. This has led to many studies examining the association between measured speed and crash frequency relationship on these roads [23–34]. Some studies observed that considering speed variables in crash prediction models can lead to improved performance when compared to traditional approaches [26]. One recent study utilized measured speeds from Ohio and Washington while developing crash prediction models for a wide range of roadways and found certain operating speed-related measures to be significant when modeling total crashes, fatal and injury crashes, and property damage crashes [24]. Another study by Das et al. developed crash modification factors (CMFs) using several speed metrics for evaluating the safety effectiveness of a countermeasure specific to speed. They considered different levels of data aggregation in their analysis [25]. Further studies were carried out to examine how the speed and crash relationship varies depending on the data aggregation approach used [33, 34].

Certainly, previous literature has extensively investigated the role of speed among the contributing factors of crash occurrence. This highlights the importance of considering speed when assessing the crashes of a particular location. Nevertheless, the significance of speed is yet to be systematically explored in relation to crashes that occur on rural two-lane highways, which are less traveled with limited availability of speed data while constituting a significant portion of the nation's roadways. Neglecting the importance of speed in rural two-lane crash studies may result in incorrect decision-making during the selection of safety countermeasures and roadway design processes. This, in turn, can have a significant impact on the investment made by the Department of Transportations (DOTs) in highway projects aimed at reducing crashes.

The objective of this study is to investigate the significance of speed in relation to crashes of rural two-lane highways. To achieve this, the authors develop a model to predict crashes on these roads. This model incorporates speed as a factor, utilizing aggregated speed metrics at the segment-level, including average speed and the 85th percentile speed. The significance of speed is evaluated across different speed ranges. Such analyses offer insights into ways to enhance the model's performance. The subsequent sections of the paper are structured in the following manner: In

Section 2, an overview of the data sources is provided. Section 3 presents a zero-inflated negative binomial model to estimate the crash frequency as a function of AADT, length, and speed. In Section 4, the performance of the model is analyzed and ways to enhance its performance are discussed. In Section 5, a summary of findings and future research direction concludes the paper.

2. Data Collection and Preparation

The study particularly utilized rural two-lane highway segments in Kentucky. Datasets on roadway, speed, and crashes were collected for these roads. The crash datasets used in this study were obtained from the Kentucky State Police collision database, covering the time frame from 2013 to 2017. In addition, the roadway geometry data and traffic counts were collected from the Highway Information System (HIS) maintained by the Kentucky Transportation Cabinet (KYTC). The crashes were further linked to the homogenous segments of roads based on the attributes such as traffic counts, functional classes, horizontal curves, shoulders, and grades [35].

Following the study by Ng, crashes that occurred within a distance of one hundred feet of intersections were classified as intersection crashes [22]. These crashes were excluded from the dataset since it is more likely that they were caused by a different combination of contributing factors. While HSM recommends 250 ft for intersection-related crashes, this value can be too restrictive for this study considering the low-volume condition on most of the segments. Furthermore, as suggested by Hauer and Bamfo, we also excluded the segments that were shorter than 0.1 miles [36].

Speed data from GPS-based probes were collected for the years 2015–2017. These data were obtained from a third-party data vendor known as HERE Technologies [37]. The data were available in 5-minute epochs for each day and in both directions of study segments, whenever probes were observed. These speeds were referenced to the HERE road network, which was then conflated with the homogeneous segments to create a spatial linkage among speed, roadway attributes, and crash dataset. Details on the conflation process are documented by Zhang and Chen [38]. Subsequently, a screening process was conducted to assess the adequacy of the speed data, ensuring that only segments containing enough data were included in the analysis. To identify the minimum required sample size of the speed data for each segment, this study used equation (1) by Li et al. [39]. Such a method is commonly used to estimate a reasonable sample size for collected traffic data to be within an allowable error range by incorporating data dispersion [39, 40]:

$$n_{\min} = \left(\frac{Z \times \sigma}{\varepsilon}\right)^2, \quad (1)$$

where the value of Z is 1.96 for a 95% confidence interval, σ is the standard deviation from the speed data, and the allowable error value, ε, is used as 5 units.

The estimated minimum sample sizes for each segment were compared with available speed data, and only the

segments meeting the minimum sample sizes of speed data were included in this study. Note that daytime speed data from 6 am to 8 pm were used, as nighttime data could be sparse in some rural areas. For each segment, we calculated aggregated speed metrics, especially the average speed and the 85th percentile speed, by utilizing the 5-minute epoch speed data available during the daytime period of 2015–2017. After all preprocesses, the final dataset contained 44,008 segments with 93,820 crashes recorded over a 5-year period in both directions of the road. The segments collectively encompass 21,240 centerline miles of rural two-lane segments in Kentucky, as depicted in Figure 1.

3. Methodology

This section outlines the methodology employed in the development of the model for predicting crashes of rural two-lane highways in this study. Multiple models were explored with separate speed measures to come up with the most reasonable measure to properly explain how speed affects the crashes on these roads.

3.1. Zero-Inflated Negative Binomial Model. Since crashes are infrequent, it is likely that a significant proportion of instances in the dataset contain zero-observed crashes. The threshold for determining the percentage of zero observations that warrants the use of zero-inflated (ZI) models remains debatable [41–45]. Existing literature has employed such models with zero observations ranging from 11% to 62% [41–45]. In our dataset, approximately, 40% of rural two-lane segments had no observed crashes, making it necessary to address the overdispersion issue caused by excess zeros. To tackle this, we utilized the zero-inflated negative binomial (ZINB) model, a statistical approach that has demonstrated a good statistical fit in previous studies [46]. It is important to note that certain studies argue against the use of zero-inflated models, claiming that the high percentage of zero-crash sites is not due to inherently safe and unsafe sites but rather results from specific conditions such as a mix of low exposure, high heterogeneity, and high-risk crash sites [47–49]. In addition, issues such as short time or small spatial scales of analysis, missing or misreported crash data, or omitted key variables in the model are cited as potential factors contributing to the high percentage of zero crashes [48]. However, there are studies that advocate for considering ZI models for crash count modeling. These studies suggest

that ZI models do not make assumptions about roads being inherently safe or unsafe but instead take into account the possibility of observing zero crashes [46, 50]. Furthermore, it is important to highlight that the main goal of model selection is to determine a model that effectively fulfills the research objectives, rather than seeking the ultimate "true" model [46]. Given the objectives of our study as well as the long-time period and large spatial scale of the data collected, the ZINB model is considered a reasonable choice to effectively model crashes in this study.

ZINB is formed by integrating a logit model and a negative binomial (NB) model [51]. The logit model is associated with excess zero crash occurrences, whereas the NB model generates the crash frequency in a segment, including instances of zero crash occurrences, based on a binomial process. If we indicate the likelihood of a crash frequency generated by the logit model as p_i, then the likelihood of the crash frequency produced by the NB model can be represented as $(1 - p_i)$. In ZINB, the parameter p_i is commonly estimated by employing a logistic regression model that incorporates explanatory variables [52]. In this study, we considered AADT and length of the segment (L) as the independent variables in addition to the speed measure (V), following existing practices [52, 53]. Here is the equation showing the logistic regression model:

$$\text{Ln}\left(\frac{p_i}{1 - p_i}\right) = \gamma_0 + \gamma_1 \text{AADT} + \gamma_2 L + \gamma_3 V. \tag{2}$$

In equation (2), the term $p_i/1 - p_i$ denotes the odds associated with the crash frequency resulting from the logit model. In particular, it represents the ratio between the likelihood of the crash frequency from the logit model and the likelihood of the crash frequency from the NB model. The equation also includes an intercept term, γ_0, along with regression coefficients γ_1, γ_2, and γ_3. The calculation of the likelihood of the zero crash frequency from the logit model can be adjusted as follows (equation (3)). A p_i value that is somewhat close to 1 indicates that segment i is unlikely to experience any crashes and is hence considered as a safe segment:

$$p_i = \frac{e^{(\gamma_0 + \gamma_1 \text{AADT} + \gamma_2 L + \gamma_3 V)}}{1 + e^{(\gamma_0 + \gamma_1 \text{AADT} + \gamma_2 L + \gamma_3 V)}}. \tag{3}$$

Now, the distribution of ZINB can be used to express the likelihood of the crash frequency, Y_i, on segment i [54]:

$$\Pr(Y_i = 0) = p_i + (1 - p_i)\left(\frac{1}{1 + \alpha \mu_i}\right)^{\alpha^{-1}},$$

$$\Pr(Y_i = y_i) = (1 - p_i)\frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(1 + y_i)\Gamma(\alpha^{-1})}\left(\frac{\alpha \mu_i}{1 + \alpha \mu_i}\right)^{y_i}\left(\frac{1}{1 + \alpha \mu_i}\right)^{\alpha^{-1}} \quad y_i = 1, 2, 3, \cdots\cdots,$$
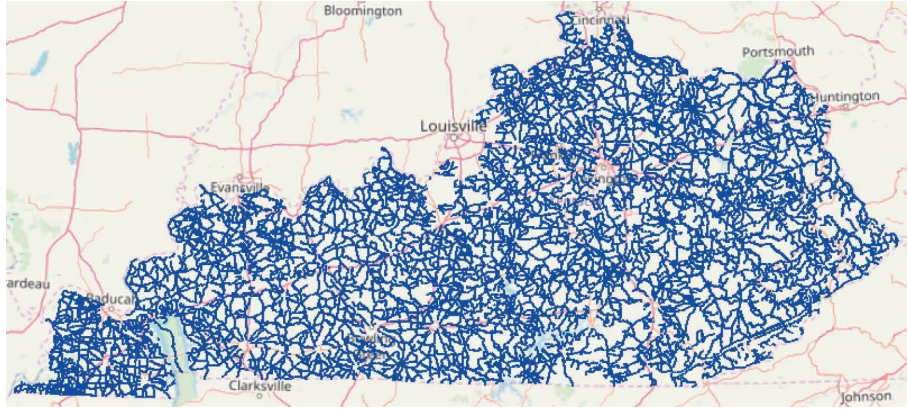
$$\tag{4}$$

where Γ is the gamma function, α is the overdispersion parameter estimated using equation (5), and μ_i refers to the mean of the underlying distribution of NB, which can be expressed as a function of the independent variables, as shown in equation (6):

$$\alpha = \frac{\text{Var}\,(Y_i) - \mu_i}{\mu_i^2}. \tag{5}$$

Here, $\text{Var}\,(Y_i)$ is the variance of Y_i, and μ_i is calculated from the following equation:

$$\mu_i = (1 - p_i)e^{(\varepsilon_i + \beta_1 \text{AADT} + \beta_2 L + \beta_3 V)}. \tag{6}$$

Here, μ_i represents the expected crash frequency in 5 years. Besides speed measures, we included AADT and length as the independent variables, similar to previous studies [19, 26, 27, 32, 55, 56]. The equation also includes an error term ε_i that follows a gamma distribution, as well as regression coefficients β_1, β_2, and β_3, which are to be estimated.

3.2. Variable Selection. By utilizing the 5-minute epoch speed data collected over a span of 3 years during the daytime, several speed metrics were computed for each direction of the segments. These include average speed (V_{avg}), the 85th percentile speed (V_{85}), the difference between average speed and speed limit ($V_{\text{avg}} - V_{\text{sl}}$), and the difference between the 85th percentile speed and speed limit ($V_{85} - V_{\text{sl}}$). The metrics were aggregated from both directions of a segment; and crashes were summed up. The ZINB model was utilized to examine each of the speed variables, together with AADT and length, as provided in equation (7) However, it should be noted that the model did not include geometric attributes such as lane width and shoulder width, as these variables indicated a high correlation with AADT based on the Pearson correlation coefficient:

$$\mu = e^{\varepsilon} . \text{AADT}^{\beta_1} . L^{\beta_2} . e^{\beta_3 V}. \tag{7}$$

In equation (7), we applied a natural logarithm transformation to AADT and L, as they exhibited a skewed distribution. No transformations were considered necessary for the speed measures due to their normal distribution.

Table 1 displays the descriptive information for the independent variables (i.e., AADT, length, and speed metrics) and the dependent variable, which is the crash frequency observed over a period of 5 years, considered in this study. It is noteworthy to mention that the dataset includes segments with low average speeds, which can be attributed to highly restrictive geometric conditions, such as narrow lanes and sharp curvature. Furthermore, the study data contain 14 segments with very low-speed limits, such as 10 mph, primarily located in mountainous areas. Moreover, many study segments exhibited average speeds or the 85th percentile speeds well below the default speed limit of 55 mph for rural two-lane roads in Kentucky. This is largely due to the limiting geometrics of these roads.

To assess the relative performance of models employing alternative speed metrics, we utilized the Akaike information criterion (AIC) and Bayesian information criteria (BIC), which were computed using equations (8) and (9), respectively:

$$\text{AIC} = -2\log Q + 2K, \tag{8}$$

and

$$\text{BIC} = -2\text{Log}Q + (\ln{(i)})K, \tag{9}$$

where Q represents the maximized likelihood function for the model, K denotes the number of parameters included in the model, and i is the total number of observations. According to previous research, models with lower values of AIC and BIC are considered to be more favorable [57].

To evaluate the prediction accuracy of the models, we examined various metrics of goodness-of-fit using data that were not previously observed by the model. These metrics include the root mean squared error (RMSE), mean absolute percentage error (MAPE), mean absolute deviation (MAD), and generalized R^2 value. RMSE is calculated by taking the square root of the mean squared error (MSE), which is obtained by averaging the squared errors of predicted crash frequencies across all segments. MAPE calculates the absolute error by comparing it to the actual crash frequency while excluding segments with no crash [58]. MAD quantifies the average absolute difference between the predicted crash frequency by the model and the actual crash frequency.

TABLE 1: Descriptive information of the study segments.

Variables	Unit	Statistics			
		Min.	Max.	Mean	Standard deviation
AADT	Vehicles	2	19619	1456	1895
Segment length (L)	Mile	0.10	2.97	0.48	0.30
Average speed (V_{avg})	mph	5.36	69.67	38.94	10.37
Speed limit (V_{sl})	mph	10	55		
The 85th percentile speed (V_{85})	mph	9.10	70	47.90	8.77
Difference between average speed and speed limit ($V_{avg} - V_{sl}$)	mph	−49.64	20.66	−14.07	10.98
Difference between the 85th percentile speed and speed limit ($V_{85} - V_{sl}$)	mph	−45.87	32.78	−5.11	9.41
Crash frequency in 5 years		0	173	2	4

Generalized R^2 is derived from the likelihood function Q, wherein an upper limit of 1 is applied to the scale. This approach offers a simplified version of the traditional R^2 metric, eliminating the need for assumptions regarding the distribution of the dependent variable, such as a normal distribution. Generalized R^2 is estimated with the following equation:

$$R^2 = 1 - \exp\left[-\frac{2}{i} \left\{ \mathrm{Log}\, Q(\widehat{\beta}) - \mathrm{Log}\, Q(0) \right\} \right], \qquad (10)$$

where $\mathrm{Log}\, Q(\widehat{\beta})$ and $\mathrm{Log}\, Q(0)$ indicate the log-likelihoods of the fitted and null models with only the intercept, respectively.

We evaluated five models presented below with rural two-lane segments in this study. The conventional model form, which consists of only AADT and length of the segment, was used as a benchmark for evaluating the performance of other models, each of which contained at least one of the speed metrics. The goal was to assess the impact of incorporating speed as a variable in the crash prediction model and determine the extent to which it improved the accuracy of predictions:

(1) AADT and length-only model

(2) AADT, length, and V_{avg}-based model

(3) AADT, length, and V_{85}-based model

(4) AADT, length, and $(V_{avg} - V_{sl})$-based model

(5) AADT, length, and $(V_{85} - V_{sp})$-based model

For model development, we utilized 75% of the dataset for training and the remaining 25% for testing the model. Table 2 provides a summary of all the tested models, including coefficients, AIC, BIC, generalized R^2, RMSE, MAPE, and MAD values. It is interesting to note that models that include speed metrics tend to match the data better than the conventional model, as evidenced by lower values of AIC and BIC. In addition, each model shows that all of the speed metrics are significant at a significance level of 5%. Among all the models, the one utilizing the 85th percentile speed appears to exhibit the least amount of error, closely followed by the average speed model. Given that the 85th percentile speed is frequently employed in highway planning to evaluate safety [57], it is plausible that this model would be more appropriate for such purposes. However, it is necessary to collect a substantial amount of data to achieve an accurate estimate of the 85th percentile speed. Since average speed

provides a better representation of actual operating conditions, the model with AADT, length, and average speed, as shown in equation (11), was ultimately selected for further analysis.

$$\mu = e^{-4.09} . \mathrm{AADT}^{0.89} . L^{1.02} . e^{-0.01V}. \qquad (11)$$

4. Integration of Speed for Better Performance

We have observed that speed is certainly a significant contributor to crashes. In this section, we discuss how speed and other independent variables are correlated with crashes using the average speed-based model shown in equation (11). We also evaluated how well the model fits the data, which further helped us adopt a refined approach of incorporating speed and ultimately improving model performance.

In equation (11), it is observed that both AADT and length exhibit a significant positive association with the crash frequency, as anticipated. The model also reveals a negative correlation between average speed and crash frequency, which suggests that more crashes tend to take place at lower speeds. This observation aligns with a recent investigation conducted by Dutta and Fontaine, which specifically examined interstates [26]. The negative relationship can also be noticed through marginal model plots, which illustrate how responses align with an independent variable while setting all other variables constant at their average values [59]. The obtained marginal model plots in Figure 2 illustrate that segments with lower average speeds tend to have a higher crash frequency, while the crash frequency increases with AADT and length.

We further justified the negative relationship between the average speed and crash frequency by normalizing the crash data in proportion to the vehicle miles traveled (VMT), utilizing AADT and length. A clear decreasing trend was noticed on the normalized crash frequency with a higher average speed. To be more specific, when other factors, such as AADT and length, remain constant, the crash frequency in the region with a higher average speed is actually lower, despite the fact that the total crash frequency may be higher due to high traffic volume.

Further analysis of the performance of the model was carried out utilizing cumulative residual (CURE) plots. The construction of CURE plots followed the methodology

TABLE 2: Model parameter estimates and goodness-of-fit.

Variables	Traditional model		Speed-based models							
			V_a		V_{85}		$V_a - V_{sp}$		$V_{85} - V_{sp}$	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Intercept (ε)	−4.18	0.042	−4.09	0.041	−3.69	0.046	−4.62	0.059	−4.52	0.050
Ln (AADT)	0.81	0.006	0.89	0.007	0.88	0.007	0.86	0.008	0.85	0.007
Ln (L)	0.98	0.010	1.02	0.010	1.03	0.010	0.99	0.010	0.99	0.010
V			−0.01	0.001	−0.02	0.001	−0.01	0.001	−0.01	0.001
AIC	106960		106617		106487		106850		106803	
BIC	107002		106668		106538		106901		106853	
R^2	0.446		0.452		0.454		0.448		0.449	
RMSE	3.71		3.66		3.65		3.70		3.70	
MAPE (%)	68.48		66.86		66.78		67.98		68.08	
MAD	1.61		1.59		1.58		1.61		1.61	

*Note: p value <0.0001 for all the variables.

outlined by Hauer and Bamfo [36]. These plots display the cumulative residual, which represents the difference between the observed crash frequency and the predicted crash frequency derived from the model. The independent variables are ordered in ascending order in the plot. The purpose of such a plot was to get a visual representation of how well the model matched the dataset. An acceptable cumulative residual curve is defined as one that remains within a range of two standard deviations ($\pm 2\sigma$) [23].

Figure 3 presents the CURE plots for the three independent variables employed in the average speed-based model. Evidently, the model exhibits inadequate fit to the data as a substantial proportion of the CURE extends beyond the $\pm 2\sigma$ limit, considering all independent variables. Furthermore, it is apparent that the model consistently overestimates or underestimates the crash frequency where the speed and AADT are higher. The average speed plot shows that the model constantly overestimates or underestimates at three speed intervals, deviating from the expected ranges. These observations prompted us to explore a different approach, outlined in the following section, which involved utilizing speed as a categorizer.

4.1. Speed as a Categorizer for Model Development. In this section, we attempted to investigate the most effective means by which speed can be incorporated into crash models. Based on Figure 3, it is clear that the current model exhibits a steady tendency to overestimate the crash frequency as the average speed increases up to approximately 30 mph. Subsequently, there is a shift towards underestimation until the average speed reaches roughly 50 mph. After this point, the model goes back to overestimating the crash frequency.

Considering these transitions in the CURE plot in terms of average speed, the study dataset was divided into three speed ranges based on average speed, and three distinct models were developed. The three speed ranges were categorized as follows: low speed, which encompassed speeds below 30 mph, medium speed, which included speeds ranging from 30 mph to 50 mph, and high speed, which referred to speeds over 50 mph. The respective proportions of total segments were approximately 21%, 61%, and 18%.

For each individual speed range, we developed crash prediction models with the ZINB form. Similar to the overall model, 75% of the segments within each speed range were used to train the model, and the remaining 25% were used for testing after model calibration. The influence of speed was analyzed across all speed levels. In the next subsections, we explain the importance of including speed as the variable in the model, in addition to how the crash frequency is affected by speed in different speed ranges.

4.1.1. Low-Speed Roads. The dataset for low-speed roads had 9,371 individual segments, all of which had an average speed of less than 30 mph. These segments had a total of 8,158 crashes in 5 years. Of the three independent variables considered, AADT and length exhibited statistical significance (p value <0.0001) at a significance level of 5%. However, the average speed was found to be insignificant and, therefore, not included in the model. The final model specification is presented in Table 3.

Quantifying the variables is one method for determining the relative significance of each independent variable in the model. Equation (12) provides a method for quantifying the significance of an independent variable:

$$\text{Variable importance} = \frac{\text{Var}(E(y/X))}{\text{Var}(y)}. \tag{12}$$

Here, the variance of the crash frequency, y, and given independent variable, X, denoted as $\text{Var}(E(y/X))$, is calculated by taking into account the predicted crash frequency, y, in relation to the conditional distribution of the variables under consideration. The variance is subsequently calculated throughout the probability distribution of variable X. $\text{Var}(y)$ is calculated as the variance of y. Based on the results, the relative importance of AADT and length on low-speed roads is 68% and 32%, respectively.

4.1.2. Medium-Speed Roads. Within the medium-speed group, a total of 27,075 distinct segments were identified, each characterized by an average speed ranging from 30 to
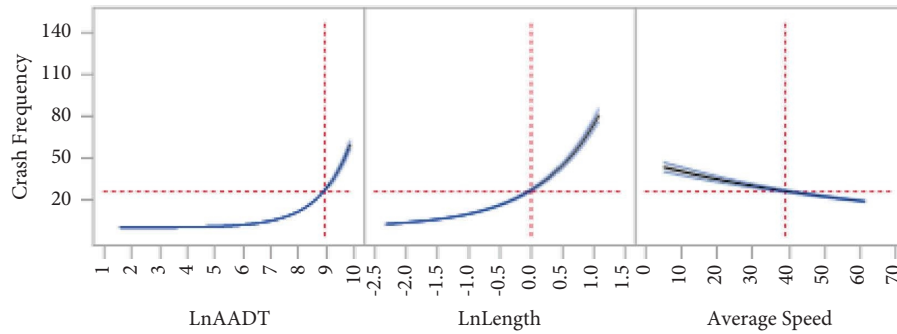
FIGURE 2: Marginal model plots for the average speed-based model.

50 mph. This category had a total of 58,104 crashes in 5 years. Based on the calibrated model, all three variables, i.e., AADT, length, and average speed, exhibited statistical significance (p value <0.0001) at a significance level of 5%. For comparison purposes, a traditional model with only AADT and length was also fitted with the same dataset. Table 4 presents the specifications and performance of the two models.

While the statistical significance of average speed is observed within the medium-speed group, its relative importance is only about 1%. In contrast, AADT and L exhibit significantly higher levels of importance, accounting for 59% and 40%, respectively. It would appear that the influence of speed is quite insignificant for this group, which is supported by the marginal model plots in Figure 4. Based on the figure, the line remains relatively flat, suggesting that there is no significant change in the crash frequency with average speed. However, the plot does indicate that other factors are playing an important role in influencing the crash frequency. Based on this finding, it appears that taking the average speed out of the model does not change the accuracy of the model very much.

Based on the above finding, we proceeded with the conventional model form and developed CURE plots for AADT and length, as illustrated in Figure 5. The plots indicate the possibility of further partitioning the data to enhance the accuracy of the model. Clearly, the plot suggests that there is a noticeable pattern of consistently underpredicting values, which then shifts to consistently overpredicting values when Ln (AADT) reaches a value of approximately 8, which corresponds to an AADT value of around 3000. The medium-speed dataset was further separated into low-volume and high-volume subsets using this value as a cutoff.

In order to assess the potential improvement in prediction accuracy, we conducted calibration and testing on two separate submodels: one developed for low-volume roads and another for high-volume roads. The purpose was to determine if incorporating AADT as an additional categorizer could enhance the predictive capabilities of the models. The ZINB formulation was utilized in both submodels, and AADT and length were used as the independent variables. Table 5 shows the specifications and prediction performance of these models. We then combined the predicted crash frequency from the two submodels to compare their overall performance with that of the single model.

From the table, it can be observed that the performance of the two submodels, when combined, shows a marginal improvement compared to the performance of the single model. Furthermore, Figure 6 shows that the corresponding CURE plots for both submodels fit better, demonstrating the effectiveness of considering AADT as an additional categorizer for medium-speed roads.

4.1.3. High-Speed Roads. High-speed roads included a total of 7,561 segments, each of which had an average speed of higher than 50 mph. These segments had a total of 27,648 crashes in 5 years. Upon calibration, it is evident that average speed is statistically significant (p value <0.0001) for crashes on high-speed roads. As expected, AADT and length are also significant. Table 6 shows variable coefficients and error metrics for the speed-based model. The estimated coefficient of average speed indicates a negative correlation between the crash frequency and speed of these roads. Further investigation revealed that these roads are characterized by high geometric standards. Compared to low and medium-speed roads, lanes and shoulders are wider with the presence of straighter sections. Within this particular category, the model gives 8% weight to average speed, while AADT and length account for 52% and 40%, respectively. This indicates that, as compared to its effect on other roads, speed has a greater impact on crash predictions on high-speed roads.

In addition, the traditional model was developed and is included in Table 6 for comparison purposes. It should come as no surprise that integrating speed in the crash frequency prediction model results in an enhanced performance over the traditional approach. The inclusion of average speed in the model leads to improved performance measures, as displayed in the table.

Further evaluation of CURE plots for the speed-based model showed that overprediction occurs after an AADT of nearly 5,000. However, due to the relatively small number of samples available in the high-speed range, we decided not to further subdivide the dataset on the basis of AADT. As more data become accessible in subsequent periods, it will be possible to reexamine this analysis.

4.2. Overall Performance Result. We evaluated the combined performance of the models that were based on speed and AADT categorizers with the performance of the initial model
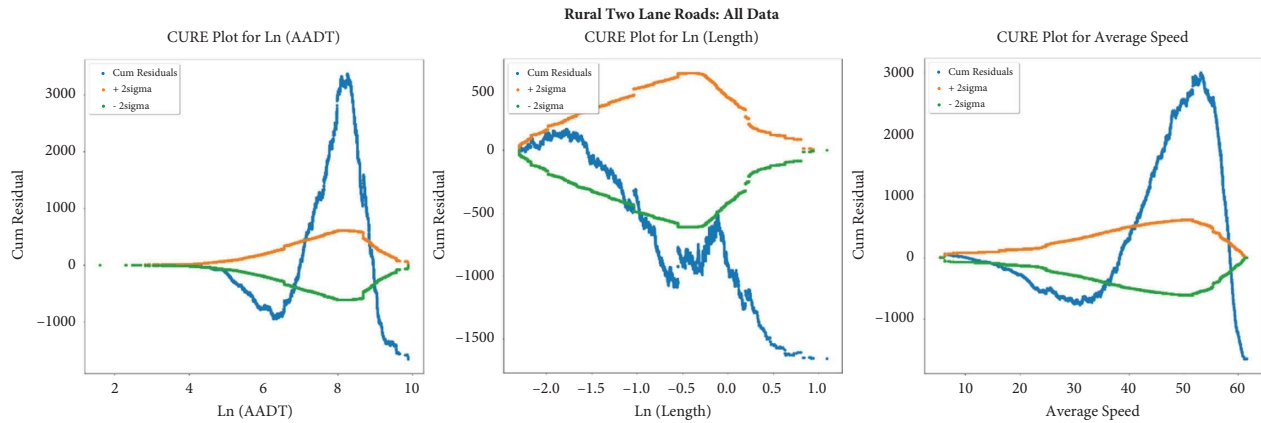
Figure 3: CURE plots for the model based on average speed.

Table 3: Model specification for low-speed roads.

	Coefficient	Std. error	95% CI
Intercept (ε)	−4.95	0.107	(−5.16, −4.74)
Ln (AADT)	0.93	0.019	(0.89, 0.97)
Ln (L)	0.92	0.033	(0.85, 0.98)
Model form	$\mu = e^{-4.95} \cdot \text{AADT}^{0.93} \cdot L^{0.92}$		
MAPE	59.79%		
RMSE	1.64		
MAD	0.86		

*Note: p value <0.0001 for all the variables.

in equation (11). The goal was to illustrate how utilizing separate models using speed and volume enhances the overall accuracy of crash prediction for rural two-lane roadways. To achieve this, all of the predictions made by the low-speed, medium-speed, and high-speed road models, which are based on speed and AADT, were aggregated. Subsequently, error metrics were computed to assess prediction accuracy. The performance of the combined model was also compared to that of the conventional model (Table 7), which incorporates only AADT and length variables. Table 7 demonstrates that when speed is utilized as a categorizer, and the model is then subdivided based on AADT within the medium-speed group, there is a notable reduction in the prediction error of up to 11.3%.

To further evaluate the performance of our models across different crash ranges, Figure 7 displays the confusion matrices for both the single average speed-based model (left) and the combined models (right). These matrices depict the accuracy of predictions for each range, with the diagonal line showing the percentage of correct predictions. Although both models perform similarly in terms of accurately predicting crashes, the combined models exhibit fewer predictions that deviate significantly from the actual values. For instance, the combined models predict only 0.14% and 0.62% of locations with zero and 1–3 crashes, respectively, to have more than 10 crashes, as opposed to 0.27% and 1.5% predicted by the single model. Moreover, for locations with more than 10 crashes, the combined models mistakenly predict only 6.7% to have 1–3 crashes, whereas the single model erroneously predicts 8.1% to have 1–3 crashes. These findings demonstrate the

advantage of the combined models over the single model in practical applications that aim to identify high-risk segments and inform improvement decisions.

Overall, the findings of this study indicated that the performance of the crash prediction model for rural two-lane roadways can be improved. This improvement was accomplished by using the actual dataset to estimate speed metrics and by taking speed and AADT into consideration as the categorizers.

5. Discussion and Summary

The objective of this study was to examine how speed contributes to the crashes of rural two-lane highways. This was achieved by integrating measured speed data into the crash prediction model. We examined the impact of four distinct speed metrics on crashes. The findings revealed that all four speed metrics exhibited statistical significance in their respective models. Subsequently, we opted to conduct a more comprehensive examination of average speed in conjunction with AADT and segment length, as average speed more accurately depicts the prevailing operating conditions encountered by drivers on these roadways.

Upon conducting a more thorough investigation, it was discovered that there exists a negative correlation between the average speed and frequency of crashes on rural two-lane roadways.

This negative correlation aligns with prior research findings that crashes tend to occur less when average speed is higher [8, 26, 30, 60, 61]. One possible justification for this observed relationship is that rural two-lane highways with higher speeds are typically the primary routes in the area, often with improved geometric characteristics [30].

In addition, it was revealed that the importance of speed crash prediction seems to increase with speed. This observation prompted us to categorize the entire dataset based on speed into three subsets: below 30 mph, between 30 mph and 50 mph, and above 50 mph. The analysis showed that speed was not significant for roads in the low-speed category but was significant for roads in both medium- and high-speed categories. While the effect of speed on crash prediction was shown to be statistically significant within the medium-

TABLE 4: Comparison of models for medium-speed roads.

|	Traditional model			Speed-based model		
	Coefficient	Std. error	95% CI	Coefficient	Std. error	95% CI
Intercept (ε)	−4.58	0.057	(−4.69, −4.47)	−4.32	0.067	(−4.46, −4.19)
Ln (AADT)	0.88	0.008	(0.87, 0.90)	0.91	0.009	(0.89, 0.92)
Ln (L)	1.06	0.013	(1.03, 1.08)	1.07	0.013	(1.05, 1.09)
Average speed	−			−0.01	0.001	(−0.01, −0.007)
Model form	$\mu = e^{-4.58}.\text{AADT}^{0.88}.L^{1.06}$			$\mu = e^{-4.32}.\text{AADT}^{0.91}.L^{1.07}.e^{-0.01V_a}$		
MAPE	61.04%			60.89%		
RMSE	2.85			2.84		
MAD	1.52			1.52		

*Note: p value <0.0001 for all the variables.
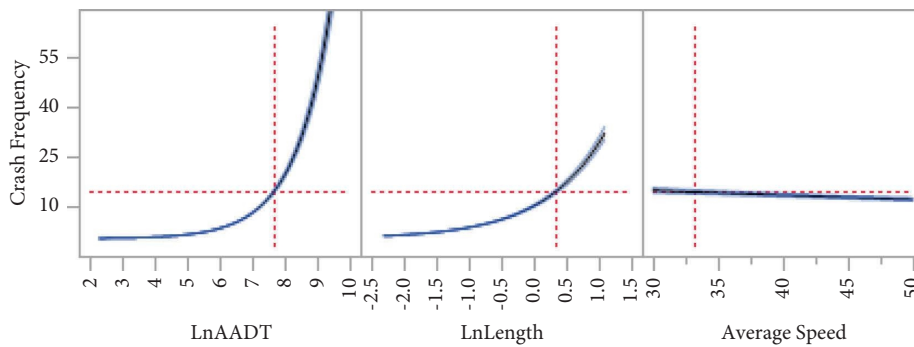
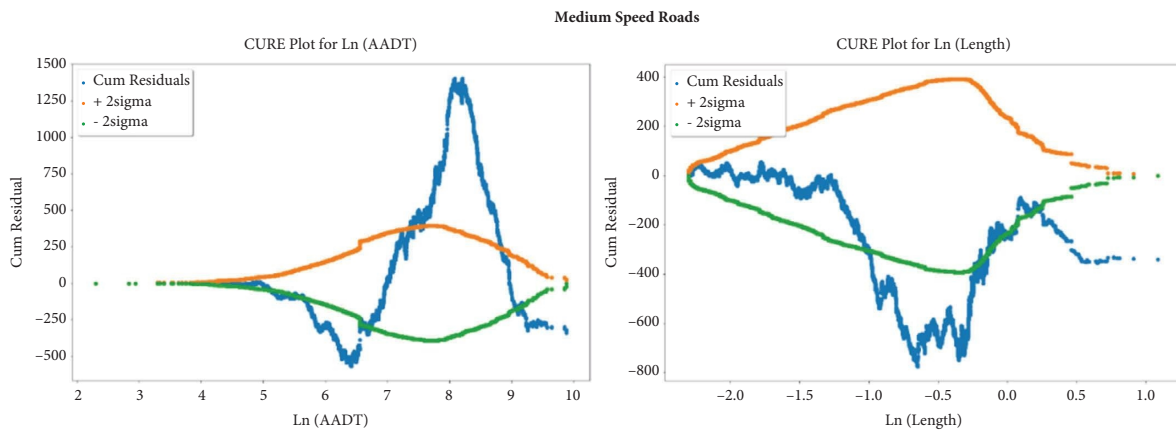FIGURE 4: Marginal model plots for a medium-speed model.

FIGURE 5: CURE plots for the independent variables in the medium-speed model.

TABLE 5: Models based on AADT categorizer and performance comparison.

Models tested for medium-speed roads	No of segments for training	No of segments for testing	MAPE (%)	RMSE	MAD
Low-volume submodel ($\mu = e^{-4.99}.\text{AADT}^{0.95}.L^{1.07}$)	18,342	6,114	60.81	2.29	1.31
High-volume submodel ($\mu = e^{-3.31}.\text{AADT}^{0.72}.L^{0.99}$)	1,964	654	80.26	5.00	3.18
Two submodels combined	20,307	6,768	63.03	2.73	1.50
Single model	20,307	6,768	61.04	2.85	1.52

speed group, its overall influence was not particularly pronounced. In contrast, speed had a greater impact on crashes occurring on high-speed roads. According to the study dataset, high-speed roads exhibited better geometric characteristics (such as wider lanes and shoulders) than low- and medium-speed roads. This observation implies that speed can serve as an indicator of the geometric condition of rural two-lane highways.

Medium Speed Roads with Low Volume

(a)

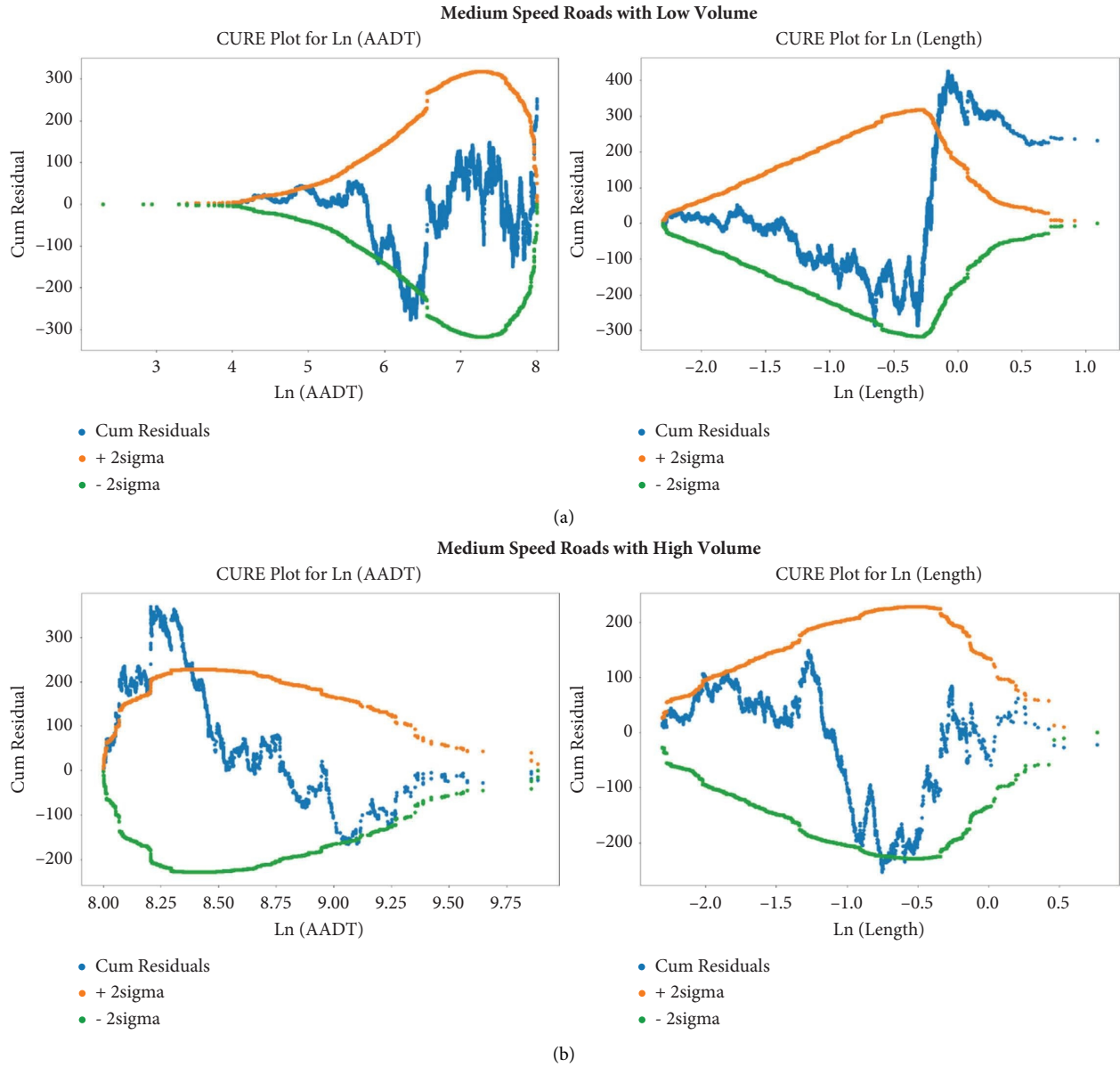**Medium Speed Roads with High Volume**

(b)

FIGURE 6: CURE plots of the models with an AADT categorizer. (a) Low-volume roads. (b) High-volume roads.

TABLE 6: Comparison of models for high-speed roads.

	Traditional model			Speed-based model		
	Coefficient	Std. error	95% CI	Coefficient	Std. error	95% CI
Intercept (ε)	−2.96	0.136	(−3.23, −2.69)	1.12	0.255	(0.62, 1.62)
Ln (AADT)	0.62	0.017	(0.59, 0.65)	0.73	0.018	(0.69, 0.77)
Ln (L)	0.96	0.019	(0.92, 0.99)	0.98	0.019	(0.95, 1.03)
Average speed	—			−0.09	0.005	(−0.10, −0.08)
Model form	$\mu = e^{-2.96}.\text{AADT}^{0.62}.L^{0.96}$			$\mu = e^{1.12}.\text{AADT}^{0.73}.L^{0.98}.e^{-0.09V_{\text{avg}}}$		
MAPE	75.23%			71.06%		
RMSE	7.23			7.16		
MAD	2.51			2.39		

*Note: p value <0.0001 for all the variables.

TABLE 7: Performance comparisons.

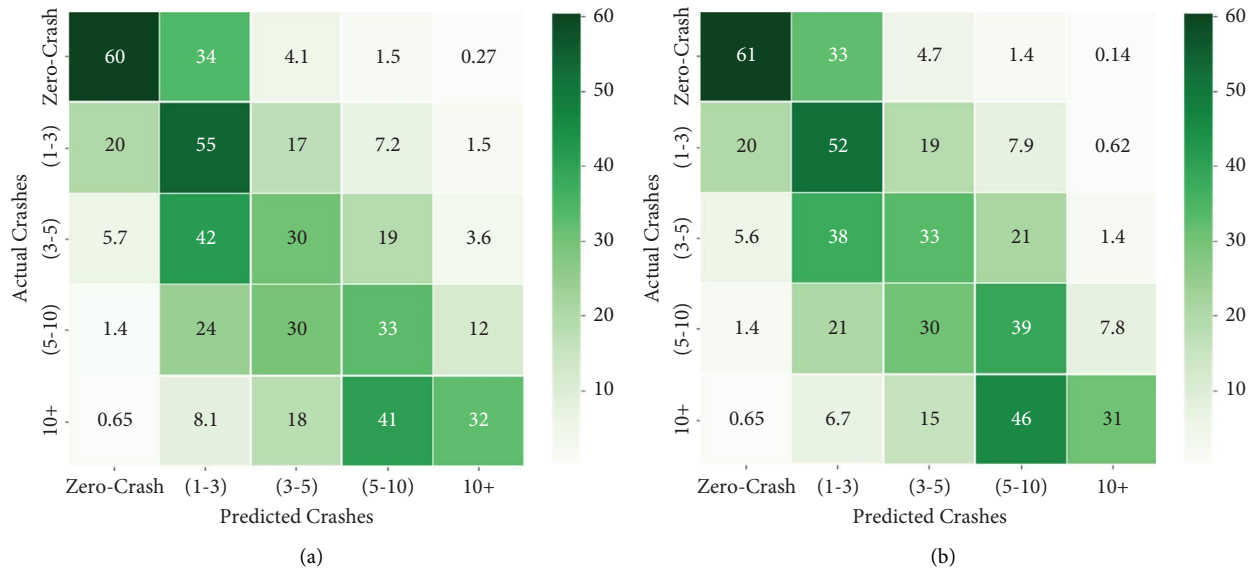|	MAPE (%)	RMSE	MAD
Traditional model	68.48	3.71	1.61
Single average speed-based model	66.86	3.66	1.59
Combined models with speed and AADT as categorizers	64.49	3.29	1.50

FIGURE 7: Confusion matrix to compare performances between single and combined models: (a) single average speed-based model; (b) combined models.

Furthermore, our study has revealed that incorporating an additional categorizer based on AADT in conjunction with speed and developing submodels under each speed group leads to improved predictions compared to a single model. While developing models for predicting crashes of rural two-lane highways, it is important to consider both speed and AADT as categorizers, provided that the available data are sufficient for separate models.

Overall, the findings of this study suggest that the effect of speed in predicting crash frequency can differ based on the speed ranges of rural two-lane highway sections. Such an analysis of speed on rural two-lane highways can provide valuable insights into the geometric and operational features of the roadway. This information can be effectively utilized to assess the safety performance of these highways under different circumstances. Consequently, appropriate countermeasures can be implemented to improve safety on these roads. Moreover, the developed submodels can be a valuable tool for transportation planners and policymakers to locate high-risk segments and allocate budgets to improve safety on those roads.

Currently, the sample size for roadways with higher operating speeds is relatively limited. We will continue to collect data on these roads to further test the performance of the model. In addition, better speed data coverage on rural low-volume roads is necessary to have a reliable estimate of the 85th percentile speed for assessing safety

from a design consistency perspective. To further understand the role of speed specific to rural two-lane highways, it would be interesting to incorporate additional geometric variables and possibly crash severity into the model. Furthermore, in light of the concerns raised regarding the ZINB model, it is important to explore alternative statistical approaches that can handle the issue of excess zeros, such as the random parameters negative binomial, random parameters negative binomial-generalized exponential, random parameters negative binomial-Lindley, and extreme value models [47–49]. In addition, considering more advanced techniques, such as machine models, could potentially enhance the overall performance and predictive capability of the model.

Data Availability

The crash data used in this study are available at Kentucky State Police (https://crashinformationky.org/). Road attributes are available at Highway Information System (https://transportation.ky.gov/Planning/Pages/HIS-Extracts.aspx). Lastly, HERE speed data are proprietary and would not be made available due to the restriction of data use agreement.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

References

[1] Assessing and managing the ecological impacts of paved roads, *Ch. 2 History and Status of the US Road System*, National Research Council, Washington, DC, USA, 2005.

[2] K. R. Agent, J. G. Pigman, and N. Stamatiadis, "Countermeasures for fatal crashes on two-lane rural roads," KTC-01-11/SPR211-00-2F, p. 55, Kentucky Transportation Cabinet, Frankfort, KY, USA, 2001.

[3] F. Rahman, "Incorporating speed into crash modeling for rural two-lane highways," Theses and Dissertations-Civil Engineering, Uknowledge, Frankfort, KY, USA, 2022.

[4] D. Solomon, *Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle*, U.S. Department of Commerce, Washington, DC, USA, 1964.

[5] Aashto, *Highway Safety Manual Part C*, AASHTO, Washington, DC, USA, 1st edition, 2010.

[6] P. J. Cooper, "The relationship between speeding behaviour (as measured by violation convictions) and crash involvement," *Journal of Safety Research*, vol. 28, no. 2, pp. 83–95, 1997.

[7] B. Fildes, G. Rumbold, and A. Leening, *Speed Behavior and Drivers' Attitude to Speeding (Report 16)*, Monash University Accident Research Center, Victoria, Australia, 1991.

[8] N. J. Garber and R. Gadiraju, "Factors affecting speed variance and its influence on accidents," *Transportation Research Record*, vol. 1213, pp. 64–71, 1989.

[9] M. G. Karlaftis, S. P. Latoski, N. J. Richards, and K. C. Sinha, "ITS impacts on safety and traffic management: an investigation of secondary crash causes," *ITS Journal- Intelligent Transportation Systems Journal*, vol. 5, no. 1, pp. 39–52, 1999.

[10] C. Kloeden, A. McLean, V. Moore, and G. Ponte, "Traveling speed and the risk of crash involvement," in *NHMRC Road Accident Research Unit*, p. 69, The University of Adelaide, Adelaide, Australia, 1997.

[11] K. M. Kockelman and J. Ma, "Freeway speeds and speed variations preceding crashes, within and across lanes," *Journal of the Transportation Research Forum*, p. 43, 2007.

[12] F. Hossain and J. Medina, "Effects of operating speed and traffic flow on severe and fatal crashes using usRAP," in *Proceedings of the 99th Annual Meeting of the Transportation Research Board*, p. 18, Washington, D.C, USA, January, 2020.

[13] J. Hutton, D. Cook, J. Grotheer, and M. Conn, "Evaluation of the relationship between driving speed and crashes on urban and suburban arterials," in *Proceedings of the 99th Annual Meeting of the Transportation Research Board*, Washington, D.C, USA, January, 2020.

[14] C. N. Kloeden, J. McLean, and G. F. V. Glonek, *Reanalysis of Travelling Speed and the Risk of Crash Involvement in Adelaide South Australia*, Australian Transport Safety Bureau, Canberra, Australia, 2002.

[15] J. Kononov, C. Lyon, and B. K. Allery, "Relation of flow, speed, and density of urban freeways to functional form of a safety performance function," *Transportation Research Record*, vol. 2236, no. 1, pp. 11–19, 2011.

[16] Y. J. Kweon and K. M. Kockelman, "Safety effects of speed limit changes: use of panel models, including speed, use, and design variables," *Transportation Research Record*, vol. 1908, no. 1, pp. 148–158, 2005.

[17] M. C. Taylor, D. Lynam, and A. Baruya, *The Effects of Drivers' Speed on the Frequency of Road Accidents*, Transport Research Laboratory Crowthorne, Crowthorne, UK, 2000.

[18] I. B. Anderson and R. A. Krammes, "Speed reduction as a surrogate for accident experience at horizontal curves on rural two-lane highways," *Transportation Research Record*, vol. 1701, no. 1, pp. 86–94, 2000.

[19] S. Cafiso, A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud, "Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables," *Accident Analysis & Prevention*, vol. 42, no. 4, pp. 1072–1079, 2010.

[20] J. De Oña and L. Garach, "Accidents prediction model based on speed reduction on Spanish two-lane rural highways," *Procedia-Social and Behavioral Sciences*, vol. 53, pp. 1010–1018, 2012.

[21] D. Llopis-Castelló, F. Bella, F. J. Camacho-Torregrosa, and A. García, "New consistency model based on inertial operating speed profiles for road safety evaluation," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 4, p. 10, 2018.

[22] J. C. W. Ng, *Quantifying the Relationship between Geometric Design Consistency and Road Safety*, Department of Civil Engineering, University of British Columbia, Vancouver, Canada, 2002.

[23] S. Boonsiripant, "Speed profile variation as a surrogate measure of road safety based on GPS-equipped vehicle data," Civil Engineering, Georgia Institute of Technology, Atlanta, GA, USA, 2009.

[24] S. Das, S. Geedipally, R. Avelar et al., *Rural Speed Safety Project for USDOT Safety Data Initiative*, Texas A & M Transportation Institute, Bryan, TX, USA, 2020.

[25] S. Das, S. R. Geedipally, and K. Fitzpatrick, "Inclusion of speed and weather measures in safety performance functions for rural roadways," *IATSS Research*, vol. 45, no. 1, pp. 60–69, 2021.

[26] N. Dutta and M. D. Fontaine, "Improving freeway segment crash prediction models by including disaggregate speed data from different sources," *Accident Analysis & Prevention*, vol. 132, Article ID 105253, 2019.

[27] D. Ederer, M. Rodgers, M. Hunter, and K. Watkins, "Probe-speed based safety performance metrics in Georgia: a case study," in *Proceedings of the 99th Annual Meeting of the Transportation Research Board*, p. 16, Washington, D.C, USA, January, 2020.

[28] M. Banihashemi, M. Dimaiuta, A. Zineddin, B. D. Spear, S. Omar, and Z. Hans, "Using linked SHRP2 RID and NPMRDS data to study speed- safety relationships on urban interstates and major arterials," in *Proceedings of the 98th Annual Meeting of the Transportation Research Board*, p. 21, Washington, D.C, USA, January, 2019.

[29] E. S. Park, K. Fitzpatrick, S. Das, and R. Avelar, "Exploration of the relationship among roadway characteristics, operating speed, and crashes for city streets using path analysis," *Accident Analysis & Prevention*, vol. 150, Article ID 105896, 2021.

[30] X. Pei, S. Wong, and N.-N. Sze, "The roles of exposure and speed in road safety analysis," *Accident Analysis & Prevention*, vol. 48, pp. 464–471, 2012.

[31] J. Stipancic, E. Racine, A. Labbe, N. Saunier, and L. Miranda-Moreno, "Relating traffic flow to crashes using massive GPS data: smartphones and usage-based insurance data agree," in

Proceedings of the 99th Annual Meeting of Transportation Research Board, Washington, D.C, USA, January, 2020.

[32] X. Wang, Q. Zhou, M. Quddus, T. Fan, and S. Fang, "Speed, speed variation and crash relationships for urban arterials," *Accident Analysis & Prevention*, vol. 113, pp. 236–243, 2018.

[33] R. Yu, M. Quddus, X. Wang, and K. Yang, "Impact of data aggregation approaches on the relationships between operating speed and traffic safety," *Accident Analysis & Prevention*, vol. 120, pp. 304–310, 2018.

[34] J. Yuan, M. Abdel-Aty, J. Fu, Y. Wu, L. Yue, and N. Eluru, "Developing safety performance functions for freeways at different aggregation levels using multi-state microscopic traffic detector data," *Accident Analysis & Prevention*, vol. 151, Article ID 105984, 2021.

[35] E. R. Green, *Segmentation Strategies for Road Safety Analysis*, Civil Engineering, University of Kentucky, Frankfort, KY, USA, 2018.

[36] E. Hauer and J. Bamfo, "Two tools for finding what function links the dependent variable to the explanatory variables," in *Proceedings of the ICTCT 1997 Conference*, p. 18, Lund, Sweden, November 1997.

[37] Here, "The state of U.S. EV charging," 2023, https://www.here.com/.

[38] X. Zhang and M. Chen, "Methodology for conflating large-scale roadway networks," *Transportation Research Record*, vol. 2677, Article ID 3611981221115085, 2022.

[39] S. Li, K. Zhu, B. Van Gelder, J. Nagle, and C. Tuttle, "Reconsideration of sample size requirements for field traffic data collection with global positioning system devices," *Transportation Research Record*, vol. 1804, no. 1, pp. 17–22, 2002.

[40] J. L. Crunkleton and P. J. Tarnoff, "Statistical patterns of traffic data and sample size estimation," *Center for advanced transportation technology*, vol. 2015, 2015.

[41] J. Q. Li, W. Liu, X. Yang, P. Lu, and K. C. Wang, "Statistical safety performance models considering pavement and roadway characteristics," *Journal of Advanced Transportation*, vol. 2022, Article ID 5871601, 12 pages, 2022.

[42] M. Sun, X. Sun, M. A. Rahman, M. Akter, and S. Das, "Modeling two-way stop-controlled intersection crashes with zero-inflated models on Louisiana rural two-lane highways," *IATSS Research*, vol. 45, no. 3, pp. 303–309, 2021.

[43] G. Gamaleldin, H. Al-Deek, A. Sandt, J. McCombs, A. El-Urfali, and N. Uddin, "Developing context-specific safety performance functions for Florida intersections to more accurately predict intersection crashes," *Journal of Transportation Safety & Security*, vol. 14, no. 4, pp. 607–629, 2022.

[44] P. C. Anastasopoulos, "Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis," *Analytic methods in accident research*, vol. 11, pp. 17–32, 2016.

[45] A. Abdulhafedh, "Incorporating zero-inflated Poisson (ZIP) regression model in crash frequency analysis," *International Journal of Novel Research in Interdisciplinary Studies*, vol. 10, 2023.

[46] T. Pew, R. L. Warr, G. G. Schultz, and M. Heaton, "Justification for considering zero-inflated models in crash frequency analysis," *Transportation Research Interdisciplinary Perspectives*, vol. 8, Article ID 100249, 2020.

[47] D. Lord and S. R. Geedipally, "The negative binomial–Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1738–1742, 2011.

[48] D. Lord, S. Washington, and J. N. Ivan, "Further notes on the application of zero-inflated models in highway safety," *Accident Analysis & Prevention*, vol. 39, no. 1, pp. 53–57, 2007.

[49] R. Rusli, M. M. Haque, A. P. Afghari, and M. King, "Applying a random parameters Negative Binomial Lindley model to examine multi-vehicle crashes along rural mountainous highways in Malaysia," *Accident Analysis & Prevention*, vol. 119, pp. 80–90, 2018.

[50] G. E. Box, A. Luceño, and M. del, "Carmen paniaguaquinones," *Statistical Control by Monitoring and Adjustment*, John Wiley & Sons, Hoboken, NJ, USA, 2011.

[51] Y. Ob, T. Bello, and O. Gureje, "Zero inflated Poisson and zero inflated negative binomial models with application to number of falls in the elderly," *Biostatistics and Biometrics Open Access Journal*, vol. 1, no. 4, 2017.

[52] J. Xu, K. M. Kockelman, and Y. Wang, "Modeling crash and fatality counts along mainlanes and frontage roads across Texas: the roles of design, the built environment, and weather," in *Proceedings of the 93rd Annual Meeting of the Transportation Research*, p. 24, Washington, D.C, USA, January 2014.

[53] X. Yan, B. Wang, M. An, and C. Zhang, "Distinguishing between rural and urban road segment traffic safety based on zero-inflated negative binomial regression models," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 789140, 11 pages, 2012.

[54] M. Ridout, J. Hinde, and C. G. Demétrio, "A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives," *Biometrics*, vol. 57, no. 1, pp. 219–223, 2001.

[55] I. Anderson, K. Bauer, D. Harwood, and K. Fitzpatrick, "Relationship to safety of geometric design consistency measures for rural two-lane highways," *Transportation Research Record*, vol. 1658, no. 1, pp. 43–51, 1999.

[56] C. Bornheimer, *Finding a New Safety Performance Function for Two-Way, Two-Lane Highways in Rural Areas*, Civil Engineering, University of Kansas, Lawrence, KS, USA, 2011.

[57] A. Montella and L. L. Imbriani, "Safety performance functions incorporating design consistency variables," *Accident Analysis & Prevention*, vol. 74, pp. 133–144, 2015.

[58] B. Hariharan, *Functional Form and Heterogeneity Effects on Safety Performance Function Estimation*, Department of Civil and Environmental Engineering, Pennsylvania State University, Pennsylvania, PA, USA, 2015.

[59] R. D. Cook and S. Weisberg, "Graphics for assessing the adequacy of regression models," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 490–499, 1997.

[60] A. Baruya, "Speed-accident relationships on European roads," in *Proceedings of the 9th International Conference on Road Safety in Europe*, pp. 1–19, Cologne, Germany, September 1998.

[61] J. Stipancic, E. B. Racine, A. Labbe, N. Saunier, and L. Miranda-Moreno, "Relating traffic flow to crashes using Massive GPS Data: smartphones and usage-based insurance data agree," in *Proceedings of the 99th Annual Meeting of the Transportation Research Board*, Washington, D.C, USA, January 2020.