Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 4798400, 18 pages
https://doi.org/10.1155/2023/4798400

Research Article

Crash Frequency Minimization with Severity Mitigation in Road Geometric Design Using Chance Constraint Programming Optimization

Wei Li,[1] Minjiao Zhang ,[2] and Anuj Sharma[3]

[1]*Office of Transportation Data, Georgia Department of Transportation, Atlanta, GA 30308, USA*
[2]*Department of Economics, Finance and Quantitative Analysis, Kennesaw State University, Kennesaw, GA 30144, USA*
[3]*Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA 50010, USA*

Correspondence should be addressed to Minjiao Zhang; mzhang16@kennesaw.edu

Received 13 February 2023; Revised 3 June 2023; Accepted 7 June 2023; Published 4 July 2023

Academic Editor: Yajie Zou

Crash frequency and crash severity are two major aspects of transportation safety. In this paper, we propose a decision-making scheme combining statistical analysis and optimization modeling to be used in transportation safety study. We conduct a safety analysis, a travel speed analysis, and an optimization analysis to develop a two-stage decision scheme to minimize crash frequency while mitigating crash severity, using data collected in urban environments in Lincoln, Nebraska. In the safety analysis and the travel speed analysis, we study the impact of lane width and other related road geometric design parameters on annual crash frequency and vehicle travel speed using count models and linear regression models, respectively. In the optimization analysis, the proposed two-stage stochastic programming model determines the lane width and other road geometric design parameters in the first stage, and then the posted road speed limit in the second stage for each scenario. To mitigate crash severity, we use a chance constraint to restrict a certain percentile of the vehicle travel speed to comply with the posted road speed limit. This two-stage decision scheme is shown to be effective for the data collected in Lincoln, Nebraska, when restricting the vehicle travel speed of up to two positive standard deviations from the mean travel speed to be under the posted speed limit. The application of a stochastic programming model that utilizes regression analysis results serves as an innovative decision scheme that effectively connects statistical analysis and optimization studies in road geometric design for transportation safety. Its objective is to minimize crash frequency while simultaneously mitigating crash severity. This methodology has extensive potential for application in various environments to assist in the reduction of both crash frequency and crash severity.

1. Introduction

Transportation safety is typically defined in terms of crashes, in which at least two aspects should be considered: crash frequency and crash severity (TRB Special Report 254 [1]). It is a traditional way to mitigate crash frequency as well as crash severity by improving road geometric design, in which lane width plays an important role in affecting the performance and cost of a roadway. Complete Streets, a transportation policy and design approach, requires streets to be planned, designed, and maintained to be safe, convenient, and comfortable for all street users, regardless of their transportation mode. With the recent trend in designing complete streets, the use of reduced lane widths instead of the 12 ft standard lanes has increased significantly, posing an urgent need of guidelines to quantify the trade-offs between the safety and efficiency of operations and the economics of the right for way savings. The document "A Policy on the Geometric Design of Highways and Streets" in 2018 recommends 10 to 12 ft wide lanes for urban and suburban arterials and urban collectors. Specifically, 10 ft wide lanes generally are for roadways with little or no truck traffic; 11 ft wide lanes are for urban arterial streets; and 12 ft wide lanes are for higher speed, free-flowing principal arterials. According to our five-state (Wyoming, Missouri, California, Kansas, and Iowa) survey regarding the lane width policy in urban settings, the right of way limitations were the key reason for the implementation of narrow lane widths in these states' roadway design. In this paper, the first part of our study is to conduct a statistical analysis for the impact of road width as well as other critical road geometric design parameters on road crash frequency, using the data collected in Lincoln, Nebraska.

Shinar [2] indicated that there is ample, but not unequivocal, evidence to show that, on a given road, the crash involvement rates of individual vehicles rise with their speed of travel. While the true travel speed is not necessarily the same or even close to the posted speed limit, many existing studies showed a strong relationship between the average travel speed with the posted speed limit and the road geometric design parameters. The appropriate posted speed limit needs to be chosen based on the road geometric design parameters to improve transportation safety. Incorrectly posted speed limits on streets and highways might lead to driver's noncompliance and a speed differential, which may lead to accident occurrences (Najjar et al. [3]). The second part of our study is to investigate the impact of the road geometric design parameters on the vehicle travel speed. In the statistical analysis of road geometric design parameters for both the crash frequency and vehicle travel speed, we differentiate the collected data according to their corresponding posted speed limits to facilitate the analysis and determination of the optimal posted speed limit.

Crash severity apparently heavily depends on the true precrash speed. The more severe injuries and fatalities are especially concerned for the vulnerable road users, such as motorcyclists, bicyclists, and pedestrians. Although most of the drivers are aware of the negative effects and potential dangers resulting from speeding, it is an unfortunate fact that drivers exceed posted speed limits on any kinds of roads. According to the National Safety Council, in 2019, the total number of fatal motor vehicle crashes attributable to speeding was 8,544 with 9,478 traffic fatalities. Concerning the importance of the aftermath of speeding, the National Transportation Safety Board (NTSB) listed "implement a comprehensive strategy to eliminate speeding-related crashes" as one of the 2021-2022 most wanted transportation safety improvements (NTSB [4]). In practice, the true precrash speed is nearly impossible to obtain. So to take the crash severity into consideration to improve transportation safety, we would like to restrict the possibility that the drivers' average travel speed obeys the posted speed limit to be above a certain reliability level, e.g., 80%. In the optimization analysis in this paper, we formulate the drivers' compliance with posted speed limits using a chance constraint, where the uncertainty comes from the daily and 5-minute travel volumes.

In this paper, we collect data at mid-block segments between signalized intersections in an urban environment in Lincoln, Nebraska, and conduct a combination of statistical and optimization analyses to study the impact of the lane width as well as other critical road geometric design parameters on the annual crash frequency. This novel decision scheme is developed through the utilization of a stochastic programming model that incorporates regression analysis results, aiming to minimize crash frequency while also mitigating crash severity. With its broad applicability, this methodology has the potential to be implemented in diverse environments to help reduce both crash frequency and crash severity.

The remainder of our paper is outlined as follows. We review relevant literature in Section 2. In Section 3, the statistical safety analysis studies the impact of road geometric design parameters on annual crash frequencies. In Section 4, the operational analysis evaluates the impact of road geometric design parameters on vehicle travel speed. In Section 5, we propose a two-stage stochastic programming model to determine the optimal road geometric design parameters to minimize the expected crash frequency while keeping the vehicle travel speed compliant to the posted speed limits to mitigate the crash severity. At the end, Section 6 concludes this study by summarizing the key findings, highlighting the ongoing work, and providing clear directions for future research.

2. Literature Review

Many studies have investigated the impact of narrow lane widths on the crash frequency of bicyclists, heavy vehicles, and passenger vehicles, however, without consistent findings in the advantages and disadvantages. For example, Noland [5] found that the narrow lane widths decrease the crash frequency of bicyclists, heavy vehicles, and passenger vehicles. But Potts et al. [6] concluded that the narrow lane widths, in contrast, increase the crash frequency of bicyclists, heavy vehicles, and passenger vehicles. To quantify the impact, Zegeer et al. [7] conducted a statistical testing along with an accidental prediction model and determined that lane widening of the 10 ft lanes can reduce related accidents by 12%, 23%, 32%, and 40% when it is 1 ft, 2 ft, 3 ft, and 4 ft wider, respectively. Hauer [8] presented six crash frequency models for urban four-lane undivided mid-block segments to find that lane width has some association with on-the--road property-damage-only crashes but no significant impact on off-the-road accident frequencies. Strathman et al. [9] found that the average lane width was positively related to crash frequency on urban freeway segments and negatively related to crash frequency on rural nonfreeway segments based on the Oregon state highway system. Harwood [10] indicated that the preferred lane width for urban arterial mid-block segments under most circumstances was 11 ft or 12 ft, while suggesting that narrower lane widths may bring traffic operational or safety benefits in some situations. Potts et al. [6] analyzed multiple roadway segments in Minnesota and Michigan but did not find a general indication that the use of lane widths narrower than 12 ft on urban and suburban arterials increased crash frequency. In another paper, Potts et al. [11] reported a possible indication that accident frequencies may be higher on four-lane undivided arterials with 9- to 10 ft lanes than on four-lane undivided arterials with 11 to 12 ft lanes in Minnesota. Zegeer et al. [12] analyzed bus and motor vehicle accident characteristics and provided recommendations for reducing bus-related highway crashes, such as keeping wide lane widths to minimize the chances of bus sideswipe collisions, and providing a lane width of at least 11 feet, but preferably 12 feet whenever possible. They found that the narrower the lanes, the larger the potential of sideswipe accidents. Sando and Moses [13] also indicated that narrow lane widths, especially lane widths of 10 ft or narrower, were overrepresented in the occurrences of bus sideswipe crashes. They recommended that 12 ft wide lanes be provided if possible for roadways located on transit routes. A recent study by Dai et al. [14] also concurred with the conclusion that a minimum 11 ft lane width is preferable for bus travel lanes, but also brought the reader's attention to the fact that the roadway design is on a case-by-case basis and it is not always feasible to have wide lanes for buses. Our study is conducted based on the data collected in Lincoln, Nebraska, where the Nebraska Administrative Code (Title 428) requires 11 ft wide lanes for local and collector roads on municipal streets and 10 ft wide concessions for local and collector roads on rural roads.

Speed harmony is the situation where operating speeds are consistent with the intended function of the roadways, which is therefore favorable with respect to safety and mobility. Donnell et al. [15] collected field data on two-lane and multi-lane urban and rural roads to explore speed harmony. They concluded that it is useful to include methods of operating speed prediction during the road design process. A lot of studies have been carried out to investigate the relationship between operating speed and road geometric design parameters and to propose a prediction model for free-flow speed in terms of road geometry. For example, a model to predict the 85th percentile speed on horizontal curves for high-speed highways was presented in Lamm and Choueiri [16]. Ali et al. [17] studied 35 four-lane urban streets in Fairfax County, Virginia, and showed that the posted speed limit, median width, and segment length all had significant effects on free-flow speed on urban streets. They proposed linear regression estimation of the mean and 85th percentile aggregated free-flow speeds. In Poe and Mason [18], vehicle travel speed data were collected in 27 urban collectors in Pennsylvania and the mixed modeling approach were used. They discovered that the lane width is significantly related to the operating speed at the midpoint of curve, while within a horizontal curve, increased lane width brought lower operating speed for low-speed urban streets. Furthermore, Gitelman et al. [19] indicated that the shoulder width and recovery zone width positively impact the travel speed. Gitelman et al. [20] adopted negative binomial statistical models to demonstrate a positive relationship between mean speeds and the number of accidents.

Speed not only affects the transportation safety in terms of the crash frequency, but also the severity, which is straightforward from the kinetic energy that is released during a collision. Using national crash databases, Campbell et al. [21] found that vehicle travel speed has a direct correlation with the severity of the crash. Kloeden et al. [22] concluded that there is a significant correlation between exceeding the speed limit, even by a small margin, and an increase in the likelihood of crashes and injury severity. Elvik et al. [23] proposed power functions to depict the relationship between speed and road safety, which is characterized by fatal injuries, fatal and serious injuries, all injuries, fatal accidents, fatal and serious accidents, and all injury accidents. Their analysis of road safety problems in Sweden indicated that even a small variation in traffic speed (by a factor of 1.35) can lead to a large variation in the number of fatalities (by a factor of 3.92) as well as all injuries. Similarly, Nilsson [24] suggested that even small reductions in vehicle speed can result in significant improvements in crash outcomes. Therefore, promoting speed compliance has been shown to be a critical way to reduce the severity of crashes on roads. However, to the best of our knowledge, there is limited quantitative study on how road geometric design and posted speed limits can be optimized to promote speed compliance. To bridge this research gap, in this paper, we propose a stochastic programming model that incorporates statistical analysis results on road geometric design for traffic safety and travel speed. The model provides an innovative decision scheme for determining the optimal road geometric design and posted speed limit that promote speed compliance to minimize the annual crash frequency while mitigating crash severity.

3. Safety Data Analysis

In this section, we conduct a safety data analysis using the count model to evaluate the impact of lane widths and other road geometric design parameters on annual crash frequency at mid-block segments. The ount-data modeling technique was shown by Lord and Mannering [25] to be an appropriate approach for crash frequency data analysis. Poisson regression or its derivatives, such as the negative model and zero-inflated model, are usually used to model the count data. But Lord and Mannering [25] indicated that, when the mean is much lower than the variance, a Poisson model would result in biased parameter estimates. The negative binomial model (Washington et al. [26]) is often used for the overdispersed crash data. In the existing studies, Poisson or negative binomial regression models are the two most frequently used regression models for evaluating the impact of lane widths on crash frequency at mid-block segments. However, using these traditional count models, unobserved heterogeneity between seemingly homogenous conditions may bring inconsistent estimates of the parameters. In this paper, we apply random parameters to account for unobserved heterogeneity in the regression models for the effects of lane widths on annual crash frequency.

3.1. Data Collection for Safety Analysis. The geometry and crash data were collected and processed in four steps: (1) data collection site selection, (2) geometry data collection based on field measurements and Google Earth measurements, (3) reduction of ten years of crash data, and (4) combining geometry and crash data. All the roads were identified as urban collectors, urban minor arterials, urban principal arterials-other nonconnecting link, and urban principal arterials-other connecting link, based on the National Functional Classification on the map of Lincoln offered by the Nebraska Department of Roads (NDOR, now Nebraska Department of Transportation). The data collection sites consist of all urban mid-block segments that are of 9 to 12 feet width with a speed limit of 45 MPH or lower and located within the city boundary of Lincoln, Nebraska.

For the collection of geometry data in the safety analysis, we considered the following parameters, as presented in Table 1. Considering the excessive amount of time needed for the field data collection, we collected data using both field measurements and Google Earth measurements. With Google Earth, the lane and shoulder widths were measured by the ruler function, and median types, shoulder types, speed limits, and so on, were observed by the street view function. The field data were used to validate the accuracy of the Google Earth data. All the crash data in Lincoln, Nebraska, from 2003 to 2012 was provided by NDOR. To improve the readability of the data, we converted the original crash data file from .txt to .xlsx format.

The combination of geometry data collected from Google Earth and crash data obtained from NDOR was crucial, with a key step being the matching of data collection sites to their corresponding historical crashes. This matching was achieved by comparing street names in both the crash and geometry information datasets. Subsequently, Microsoft Access was utilized to segregate accidents into segment approaches by comparing the vehicle driving direction in the crash dataset to the direction of segment observation in the geometry info dataset. At last, we computed the crash frequency for each data collection site.

3.2. Statistical Results for Safety Analysis. Based on the National Functional Classifications, the roadway types were classified as 14-urban principal arterial other connecting link; 15-urban principal arterial other nonconnecting link; 16-urban minor arterial; or 17-urban collector. The range of lane widths for the mid-block segments was from 9 to 12 feet. Since the sample sizes of segments that were less than 9 feet or more than 12 feet were too small to make the estimation model, those observations were not included in the analysis dataset. The effects of 9 ft, 10 ft, 11 ft, and 12 ft lanes on crash frequency were analyzed. In addition, to analyze the impact of road geometric design parameters on the crash frequency, this research did not count heavy vehicle or alcohol-related crashes, crashes that were not caused by road surface conditions, and crashes whose first leading event was motor vehicles in transit.

In the preliminary data processing, we found a high correlation between the posted speed limit and all other variables. Therefore, we develop the crash frequency models specifically for different posted speed limits. Together with the area type used to determine if the observed segments were in a central business district, we separated the observed mid-block segments into five groups: 25CBD (posted speed limit of 25 MPH and located within a central business district), 30NCBD (posted speed limit of 30 MPH and located outside of a central business district), 35NCBD (posted speed limit of 35 MPH and located outside of a central business district), 40 MPH (posted speed limit of 40 MPH), and 45 MPH (posted speed limit of 45 MPH). In addition, we observed that the number of lanes for each mid-block segment and the average daily traffic on the segment were highly correlated. So, a new variable, directional average daily traffic per lane was used to represent the average daily travel volume per lane for each direction instead of the average daily traffic.

Table 2 summarizes the number of mid-block segments of each lane width in each of the five groups. Considering the inadequate sample size of group 30NCBD, only the other four groups were included in the following statistical regression analysis.

For crash data in 25CBD, we used the Poisson regression model, which is a typical regression model for count data (Lord and Mannering [25]). However, we found that, in 35NCBD, 40 MPH, and 45 MPH, the variance of the dependent variable is higher than the mean, which indicates overdispersion. Therefore, we used the negative binomial regression models (Washington et al. [26]) for 35NCBD, 40 MPH, and 45 MPH to avoid biased parameter estimates from such overdispersed data (Lord and Mannering [25]). We tested the effect of 9, 10, and 11 ft lane widths on the annual crash frequency compared to the 12 ft wide lanes. The variables listed in Table 3 were found to be significantly related to the annual crash frequency. The descriptive statistics of these significant variables, as well as the variable of annual crash frequency are also included in Table 3.

The regression model results, including constants and coefficient estimates of the significant dependent variables, are shown in Tables 4–7. The variable notation, which will be referred to in Section 5.2 for the optimization model, are also included. In the regression analysis results for models 35NCBD, 40 MPH, and 45 MPH (Tables 5–7), the statistical significance of the dispersion parameter indicates that it is significantly different from zero. This suggests that the negative binomial model is appropriate for the data of 35NCBD, 40 MPH, and 45 MPH.

4. Travel Speed Analysis

In this section, we conduct a travel speed analysis to evaluate the impact of lane widths and other road geometric design parameters on vehicle travel speed.

4.1. Data Collection for Travel Speed Analysis. To collect data for the travel speed analysis, a total of 14 directional mid-block segment observations were randomly selected from the safety data collection sites in Lincoln, Nebraska, as listed in Table 8. We collected vehicle travel speed data at these 14 sites using a Wavetronix HD Sensor during a two-hour nonpeak period (1–3 pm) and a two-hour peak period (3:30–5:30 pm).

4.2. Statistical Results for Travel Speed Analysis. We conducted a linear regression analysis to study the impact of the road geometric design parameters on the vehicle travel speed. The linear regression models test the effect of 9, 10, and 11 ft lane widths on the vehicle travel speed, in

Table 1: Collected parameters.

Segment parameter	Description
Through lane width	Through lane width (ft)
Average daily traffic	Average daily traffic on the street
Shoulder presence indicator	0, if there is no shoulders on the street; 1, if street has shoulder
Shoulder width	Shoulder width (ft)
Shoulder type indicator	1, if street has paved shoulder; 2, if street has gravel shoulder; 3, if street has turf shoulder; 4, if street has composite shoulder
Median presence indicator	0, if there is no median on the street; 1, if there is median on the street
Median type	0, if street has painted or shared median; 1, if street has curbed median
Median width	Median width (ft)
On-street parking presence indicator	0, no on-street parking; 1, has on-street parking
Road speed limit	Road speed limit (MPH)
Number of lanes	Number of through lanes in one direction on the street

Table 2: Sample size for each data group.

	25CBD	30NCBD	35NCBD	40 MPH	45 MPH
9 ft	5	0	42	0	0
10 ft	23	0	88	32	2
11 ft	2	0	72	19	37
12 ft	7	0	27	32	54

comparison to 12 ft wide lanes. The independent variables include through lane width, shoulder indicator, shoulder type, shoulder width, median indicator, median type, median width, number of through lanes, segment length, and five-minute real-time volume (heavy vehicles excluded) in the tested segments. Categorized by the four speed limits (25CBD, 35NCBD, 40 MPH, and 45 MPH) and two data collection time periods (1–3 pm and 3:30–5:30 pm), eight linear regression models were studied in total.

Tables 9 and 10 present the linear regression results for the travel speed analysis. The last columns contain the notation for the significant variables, which are consistent with those in Tables 4–7, for the optimization model in Section 5.2. The standard errors of the linear regression reported in the tables will be needed in the optimization model in Section 5.2, too.

When comparing the results between the two time segments in the 25CBD model, the traffic volume has a significant impact on lowering the travel speed due to possible congestion during peak hours. While 9 ft wide lanes lead to a significant reduction in travel speed in both time segments compared to 12 ft wide lanes, 10 ft wide lanes increase travel speed during peak hours. This could be due to the fact that vehicles can maneuver more easily through slightly narrower lanes to avoid peak-hour congestion. Results of the 35NCBD models indicate that narrow lanes, compared to 12 ft wide lanes, generally lead to higher travel speeds in the noncentral business district with a posted speed limit of 35 miles per hour, in both nonpeak and peak hours. Moreover, shoulders increase travel speed and additional lanes decrease travel speed during nonpeak hours. In the 40 MPH model, 10 ft lanes lead to a decrease in travel speed, while 11 ft lanes increase travel speed compared to 12 ft lanes in both time segments. This could be due to the fact that fairly wider lanes provide extraspace for drivers to move at a fast speed. In the 45 MPH model, for both time segments, a 10 ft lane decreases the travel speed, while an 11 ft lane increases the travel speed compared to a 12 ft wide lane. However, a higher traffic volume increases travel speed for the 1–3 pm time segment but decreases travel speed for the 3:30–5:30 pm time segment. This could be due to the fact that during nonpeak hours, the road can accommodate more vehicles without causing congestion, which leads to a more efficient use of the available road capacity. Overall, the results indicate that 9, 10 and 11 ft wide lanes do not have a consistent effect in lower speed limit models 25CBD and 35NCBD. However, 10 ft wide lanes consistently decrease travel speed and 11 ft wide lanes consistently increase travel speed compared to 12 ft wide lanes at higher speed limit models of 40 MPH and 45 MPH, which may be because narrow lanes require drivers to be more cautious when driving in narrow lanes at high speed limit zones.

Table 3: Descriptive statistics of crash frequency-related variables.

Model	Variable	Descriptive value (std. dev.) (min) (max)
25CBD	Average annual crash frequency for each direction	0.38 (0.72) (0) (4)
	Average of directional average daily traffic per lane	3009.45 (1520.898) (1100) (7200)
	Percentage of each direction on M st from 11th st to Centennial Mall st	13
	Percentage of each direction on N st from Centennial st to 9th st	20
	Percentage of 10 ft lane width for each direction	77
35NCBD	Average annual crash frequency for each direction	1.29 (2.08) (0) (23)
	Average of directional average daily traffic per lane	5989.56 (2630.39) (1250) (14846.69)
	Percentage of each direction on 27th st between Nebraska highway and Cornhuskers highway	11
	Percentage of each direction on 40th st between Van Dorn st and Pioneers Blvd	1
	Percentage of 1 lane for each direction	61
	Percentage of the one way road	69
	Percentage of 9 ft lane width for each direction	18
	Percentage of 10 ft lane width for each direction	38
	Percentage of 11 ft lane width for each direction	31
40 MPH	Average annual crash frequency for each direction	1.15 (1.48) (0) (9)
	Average of directional average daily traffic per lane	5858.79 (2721.52) (2300) (19480.37)
	Average segment lengths	0.41 (0.27) (0.11) (1)
	Percentage of each direction on Cornhusker highway between N 29th st and N 33rd st	2
	Percentage of each direction with average daily travel in vehicles per lane less than 10000	94
	Percentage of 2 lanes for each direction	72
	Percentage of 10 ft lane width for each direction	39
	Percentage of 11 ft lane width for each direction	23
45 MPH	Average annual crash frequency for each direction	0.8 (1.21) (0) (10)
	Average of directional average daily traffic per lane	5291.65 (2223.07) (862.5) (11650)
	Average segment lengths	0.52 (0.34) (0.09) (2)
	Percentage of each direction on 27th st between Old Dairy rd and Kmart Dr	2
	Percentage of each direction on Nebraska highway from S 33rd st to S 27th st	1
	Percentage of 11 ft lane width for each direction	41

Table 4: 25CBD model estimation results.

Significant variable	Coefficient estimate	t-statistic	Variable notation
Constant	−1.17	−3.82	
Indicator of 10 ft lane width for each direction	−0.73	−3.51	w_{10}
Indicator of each direction on M st from 11th st to Centennial Mall st	1.19	3.84	q_1
Indicator of each direction on N st from Centennial st to 9th St	1.4	6.33	q_2
Directional average daily traffic per lane*	−0.36D-05 (0.00012)	−0.05 (4.26)	D^i
Number of observations		300	
Log-likelihood with constant only		−249.29	
Log-likelihood at convergence		−218.29	
McFadden pseudo R-squared		0.12	
Chi squared		62.01	
Info. criterion: AIC		1.50	
Finite sample: AIC		1.50	
Info. criterion: BIC		1.57	

*Random parameter, the standard deviation of parameter distribution is shown in parentheses.

5. Optimization Analysis

Section 3 provides us with the regression results of the crash frequency specifically for different posted speed limits, and Section 4 studies the impact of lane widths and other significant road geometric design parameters on the average vehicle travel speed. In this section, we propose a methodology to integrate the statistical analysis results with a stochastic programming model to determine the optimal choices of the posted speed limits, lane widths, and other significant road geometric design parameters to minimize the crash frequency while mitigating the crash severity by promoting speed compliance.

From the statistical analysis in Sections 3.2 and 4.2, we see there is significant uncertainty involved in two dependent variables: the directional average daily traffic per lane for the crash frequency regression model and the 5-minute travel volume for the travel speed regression model. Instead of using the average values of these uncertain variables, which might cause a large error to depict the real situation, or considering for the worst case to use a robust optimization model, which might lead to overconservative decisions, in this paper, we develop a stochastic programming model to optimize for the expected crash frequency, taking all possible scenarios of the random variables into consideration. The scenarios will be generated using historical data of the random variables.

In the construction or the long-term use of a road, some variables considered in this paper have to be determined at the beginning, such as the lane width or other road geometric design parameters. But some variables can be determined or adjusted later, after the uncertainty has been revealed. For example, we can adjust the speed limit of a road after an accurate forecast of the 5-minute travel volume of a day is obtained or change the posted speed limit in different periods of a day or different days in a week. Instead of making all the decisions at the beginning and keeping them unchanged, we can wait until the uncertainty is revealed to make recourse decisions to achieve optimality for the rest of the planning horizon. Therefore, we apply the dynamic decision-making scheme to develop a two-stage stochastic programming model for our problem, where the decisions of road geometric design parameters are made in the first stage and the posted speed limit is determined in the second stage.

Depending on the values of the random variables and the posted speed limit, vehicle travel speeds vary. The more compliant a driver is while driving on a road, the less severe a potential crash could be. To mitigate the severity of the potential crashes across all the scenarios, we apply a chance constraint (Miller and Wagner [27] and Prékopa [28]) to guarantee the probability of a certain percentage of the vehicles (e.g., the 75 percentile vehicle travel speed) not exceeding the posted speed limit to be beyond a certain probability (e.g., 85% reliability). Below, we will discuss the chance-constrained two-stage stochastic programming model for our paper.

5.1. Optimization Model. We define W to be the set of all lane widths under consideration (9 ft, 10 ft, and 11 ft) in addition to the 12 ft lane width and Q the set of variables of other road geometric design parameters. In the first stage, we determine the lane width as well as other road geometric design parameters. Let binary variable $w_k = 1$ if lane width $k \in W$ is selected and $w_k = 0$ otherwise. Notice that $\sum_{k \in W} w_k = 0$ implies the selection of the 12 ft lane width. We also have variable q_i for $i \in Q$ represent the value of road geometric design parameter $i \in Q$.

The random vector in our problem consists of the directional average daily traffic per lane and the 5-minute travel volume, denoted as (Δ, δ). Each realization of the random vector (Δ, δ) is referred to as a scenario, with a certain probability of being realized. For example, (8000, 60) could be a scenario of (Δ, δ) with probability 0.1, representing a 10% chance to observe the directional average daily traffic per lane as 8000 and the 5 minute travel volume as 60. We assume that the random vector (Δ, δ) has finitely many scenarios or can be sufficiently described by a finite collection of scenarios, denoted as $\{(D^i, d^i)\}_{i \in S}$, where (D^i, d^i) is the realization of (Δ, δ) in scenario $i \in S$. After observing a realization of (Δ, δ), in the second stage, we determine a scenario-specific posted speed limit to minimize the annual crash frequency while controlling the compliance of the travel speed with the posted speed limit. Define set L to be the set of all speed limits under consideration. Let binary variable $u_l^i = 1$ if posted speed limit $l \in L$ is adopted in scenario $i \in S$ and $u_l^i = 0$ otherwise. Moreover, we use a chance constraint to ensure the λ-th percentile vehicle travel speed not to exceed the posted speed limit with a certain probability, which is referred to as the required reliability level and denoted as τ.

Table 5: 35NCBD model estimation results.

Significant variable	Coefficient estimate	t-statistic	Variable notation
Constant	−1.25	−15.35	
Indicator of 9 ft lane width for each direction	−0.36	−4.05	w_9
Indicator of 10 ft lane width for each direction*	−0.28 (0.52)	−4.17 (14.26)	w_{10}
Indicator of 11 ft lane width for each direction	−0.30	−4.18	w_{11}
Indicator of 1 lane for each direction	−0.62	−11.94	q_3
Indicator of one way road	0.39	3.37	q_4
Indicator of each direction on 27th st between Nebraska highway and Cornhusker highway	0.93	15.26	q_5
Indicator of each direction on 40th st between Van Dorn st and Pioneers Blvd	1.6	10.61	q_6
Directional average daily traffic per lane*	0.0003 (0.89D-04)	23.635 (26.518)	D^i
Dispersion parameter	7.14	6.70	
Number of observations		2290	
Log-likelihood with constant only		−5544.58	
Log-likelihood at convergence		−3001.87	
McFadden pseudo R-squared		0.46	
Chi squared		5085.41	
Info. criterion: AIC		2.63	
Finite sample: AIC		2.63	
Info. criterion: BIC		2.66	

*Random parameter, the standard deviation of parameter distribution is shown in parentheses.

Table 6: 40 MPH model estimation results.

Significant variable	Coefficient estimate	t-statistic	Variable notation
Constant	−3.39	−12.07	
Indicator of 10 ft lane width for each direction*	0.06 (0.39)	0.56 (6.71)	w_{10}
Indicator of 11 ft lane width for each direction*	−0.33 (0.71)	−2.27 (6.14)	w_{11}
Directional average daily traffic per lane	0.0002	12.88	D^i
Indicator of 2 lanes for each direction	−0.38	−3.28	q_7
Segment length	1.17	7.08	q_8
Indicator of each direction on Cornhusker highway between N 29th st and N 33rd st	1.01	0.09	q_9
Indicator of each direction with average daily travel in vehicles per lane less than 10000	1.79	10.145	q_{10}
Dispersion parameter	5.03	1.08	
Number of observations		830	
Log-likelihood with constant only		−1649.98	
Log-likelihood at convergence		−1113.51	
McFadden pseudo R-squared		0.33	
Chi squared		1072.95	
Info. criterion: AIC		2.71	
Finite sample: AIC		2.71	
Info. criterion: BIC		2.77	

*Random parameter, the standard deviation of parameter distribution is shown in parentheses.

Table 7: 45 MPH model estimation results.

Significant variable	Coefficient estimate	t-statistic	Variable notation
Constant	−1.86	−17.91	
Indicator of 11 ft lane width for each direction*	−0.06 (0.52)	−0.68 (7.49)	w_{11}
Directional average daily traffic per lane	0.0001	11.73	D^i
Segment length	0.65	5.35	q_8
Indicator of each direction on 27th st between Old Dairy rd and Kmart Dr	1.58	7.51	q_{11}
Indicator of each direction on Nebraska highway from S 33rd st to S 27th st	0.94	2.71	q_{12}
Dispersion parameter	7.21	2.76	
Number of observations		910	
Log-likelihood with constant only		−1331.99	
Log-likelihood at convergence		−1004.13	
McFadden pseudo R-squared		0.25	
Chi squared		655.72	
Info. criterion: AIC		2.22	
Finite sample: AIC		2.22	
Info. criterion: BIC		2.27	

*Random parameter, the standard deviation of parameter distribution is shown in parentheses.

Table 8: Travel speed data collection sites.

Main street	Direction		From	To	Group	Lane width (ft)
	From	To				
12th	S	N	O st	P st	25CBD	9
M st	W	E	12th st	13th st	25CBD	10
12th	S	N	N st	O st	25CBD	11
12th	S	N	M st	N st	25CBD	12
Van Dorn st	W	E	S 40th st	S 48th st	35NCBD	9
16th st	N	S	A st	south st	35NCBD	10
70th	N	S	S Wedgewood Dr	Teton Dr	35NCBD	11
27th st	S	N	Capitol Pkwy	Randolph st	35NCBD	12
W O st	E	W	N 70th st	N68th st	40 MPH	10
Pine Lake rd	E	W	S 27th st	Ridge rd/Helen Witt Dr	40 MPH	11
Superior st	W	E	N 14th st	N 20th st	40 MPH	12
27th st	S	N	Hwy6	K mart Dr	45 MPH	10
Pine Lake rd	E	W	Beaver Creek ln	S 40th st	45 MPH	11
27th st	N	S	Superior st	Old Dairy rd	45 MPH	12

Table 9: Results of linear regression models for travel speed during 1–3 pm

Model summary for 25CBD			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	1456	0.177	5.440
	Coefficient estimate		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	23.651	41.539	
Indicator of 9 ft lane width	−6.186	−14.445	w_9
Indicator of 10 ft lane width	−4.236	−8.654	w_{10}
Model summary for 35NCBD			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	3099	0.055	5.200
	Coefficient estimate		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	36.972	31.009	
Indicator of 9 ft lane width	3.213	3.795	w_9
Indicator of shoulder appearance	2.503	7.390	q_{13}
Number of through lanes	−0.832	−2.165	q_{14}
Number of vehicles in every five minutes	−0.024	−3.998	d^i
Model summary for 40 MPH			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	5005	0.379	4.992
	Coefficient estimate		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	42.885	113.809	
Indicator of 10 ft lane width	−6.279	−24.451	w_{10}
Indicator of 11 ft lane width	1.209	5.339	w_{11}
Number of vehicles in every five minutes	−0.029	−5.129	d^i
Model summary for 45 MPH			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	4415	0.448	5.326
	Coefficient estimate		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	34.885	64.940	
Indicator of 10 ft lane width	−3.587	−18.952	w_{10}
Indicator of 11 ft lane width	9.901	28.769	w_{11}
Number of vehicles every five minutes	0.029	5.934	d^i

At last, we use $\mathscr{F}_l^i(\mathbf{w},\mathbf{q})$ to denote the regression function for the annual crash frequency and $\mathscr{V}_l^{i\lambda}(\mathbf{w},\mathbf{q})$ to denote the vehicle travel speed at λ-th percentile with posted speed limit $l \in L$ in scenario $i \in S$, where $\mathbf{w} = (\{w_k\}_{k\in W})$ and $\mathbf{q} = (\{q_i\}_{i\in Q})$. Below, we present the generic two-stage stochastic programming model with a chance constraint for our problem.

$$\min \quad \mathbb{E}_{(\Delta,\delta)} \sum_{l\in L} u_l^i \mathscr{F}_l^i(\mathbf{w},\mathbf{q}), \tag{1}$$

$$\text{s.t.} \sum_{k\in W} w_k \leq 1, \tag{2}$$

$$\sum_{l\in L} u_l^i = 1 \quad i \in S, \tag{3}$$

$$\mathbb{P}_{i\in S}\left[\left\{\mathscr{V}_l^{i\lambda}(\mathbf{w},\mathbf{q}) \leq l u_l^i + M\left(1-u_l^i\right)\right\}_{l\in L}\right] \geq \tau, \tag{4}$$

$$q_i \in \Theta \quad i \in Q, \tag{5}$$

$$w_k \in \mathbb{B} \quad k \in W, \tag{6}$$

$$u_l^i \in \mathbb{B} \quad l \in L, i \in S, \tag{7}$$

where set Θ includes additional necessary restrictions for the road geometric design parameter variables $\{q_i\}_{i\in Q}$. In this model, objective function (1) minimizes the expected annual crash frequency over all scenarios of (Δ,δ). Constraint (2) is to select exactly one lane width. Note that 12 ft lane width is selected when $\sum_{k\in W} w_k = 0$. Constraints (3) select one posted

Table 10: Results of linear regression models for travel speed during 3:30–5:30 pm

Model summary for 25CBD			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	2250	0.226	5.363
	Coefficient estimation		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	23.717	64.830	
Indicator of 9 ft lane width	−7.934	−19.961	w_9
Indicator of 10 ft lane width	1.323	4.115	w_{10}
Indicator of 11 ft lane width	−3.497	−9.501	w_{11}
Number of vehicles in every five minutes	−0.026	−2.924	d^i
Model summary for 35NCBD			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	4864	0.119	6.143
	Coefficient estimation		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	36.193	64.337	
Indicator of 9 ft lane width	1.962	3.349	w_9
Indicator of 10 ft lane width	3.947	13.738	w_{10}
Indicator of 11 ft lane width	7.636	19.598	w_{11}
Number of vehicles in every five minutes	−0.067	−9.623	d^i
Model summary for 40 MPH			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	7754	0.205	4.734
	Coefficient estimation		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	40.056	487.711	
Indicator of 10 ft lane width	−5.211	−39.248	w_{10}
Indicator of 11 ft lane width	1.147	7.239	w_{11}
Model summary for 45 MPH			
Dependent variable	Number of observations	Adjusted R square	Standard error
Vehicles' travel speed	6082	0.588	4.651
	Coefficient estimation		
Independent variable	Coefficient	t-statistic	Variable notation
Constant	43.394	77.442	
Indicator of 10 ft lane width	−6.839	−48.637	w_{10}
Indicator of 11 ft lane width	6.064	23.430	w_{11}
Number of vehicles every five minutes	−0.031	−6.433	d^i

speed limit in each scenario $i \in S$. Chance constraint (4) ensures that, across all scenarios $i \in S$, the probability of the λ-th percentile travel speed not exceeding the selected posted speed limit is beyond the reliability level τ, where M is a big enough constant. Constraints (5) are additional necessary restrictions for the road geometric design parameters, such as binary restrictions for variables $\{q_i\}_{i \in Q \setminus \{8,14\}}$ and integer restriction for variable q_{14}. At the end, constraints (6) and (7) are the binary restrictions.

5.2. Case Study. In this section, we apply the model proposed in Section 5.1 to the traffic environment in Lincoln, Nebraska, using the data collected and regression models developed in Section 4. Assuming a new road of 0.264 miles (the average segment length of the observed sites) will be constructed in our data collection area, we would like to identify the optimal road geometric design to minimize the expected annual crash frequency while keeping the vehicle travel speed compliant to the posted speed limit to mitigate the crash severity.

As discussed in Sections 3 and 4, the posted speed limits applicable in our study in the urban area of Lincoln, Nebraska, are 25, 35, 40, and 45 MPH. Recall that in the regression analysis for the crash frequency in Section 3.2, the significant dependent variables with their coefficient estimates and notation are presented in Tables 4–7. In the regression analysis for the average vehicle travel speed in Section 4.2, the significant dependent variables with their

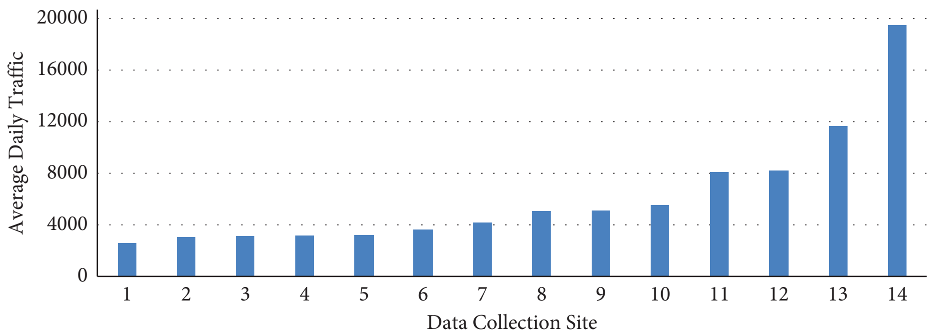


FIGURE 1: Average daily traffic at 14 data collection sites.

coefficient estimates and notation are presented in Tables 9 and 10.

The randomness of the two variables, directional average daily traffic per lane and 5-minute travel volume, introduces uncertainty to this case. To implement the stochastic programming model presented in Section 5.1, a finite collection of scenarios is needed to represent the possible cases of these random variables. We cluster historical data values into appropriate ranges that best depict the possible cases of the random variables. This approach enables us to capture the range of possible traffic volumes while also balancing the computational complexity of the model.

The directional average daily traffic per lane at the 14 data collection sites is shown in Figure 1. We sort and group them into three cases: low directional average daily traffic per lane with an average of 3868.4 at probability 0.714, medium directional average daily traffic per lane with an average of 9308.33 at probability 0.214, and high directional average daily traffic per lane with an average of 19480.37 at probability 0.072.

The 5-minute travel volumes of the 14 data collection sites for 1–3 pm and 3:30–5:30 pm time segments are displayed in histograms in Figure 2. For the 1–3 pm segment, it is observed that about 98% of the 5-minute travel volumes fall between 0 and 120. Based on this, we divide the data into three groups according to the 5-minute travel volume: low 5-minute travel volume (less than 40) at 1–3 pm with an average of 22.13 at probability 0.4344, medium 5-minute travel volume (between 40 and 80) at 1–3 pm with an average of 58.93 at probability 0.3031, and high 5-minute travel volume (greater than 80) at 1–3 pm with an average of 100.69 at probability 0.2625. These three groups provide three cases for the 5-minute travel volume in the 1–3 pm segment. Similarly, for the 3:30–5:30 pm segment, it is found that about 93.5% of the 5-minute travel volumes are between 0 and 150. We also divide this data into three groups: low 5-minute travel volume (less than 50) at 3:30–5:30 pm with an average of 24.95 at probability 0.3388, medium 5-minute travel volume (between 50 and 100) at 3:30–5:30 pm with an average of 76.43 at probability 0.3420, and high 5-minute travel volume (greater than 100) at 3:30–5:30 pm with an average of 130.87 at probability 0.3192. These three groups provide three cases for the 5-minute travel volume in the 3:30–5:30 pm segment. By combining the three cases of the directional average daily traffic per lane and the three cases of the 5-minute travel volume for each time segment, we generate nine scenarios for the 1–3 pm segment and nine scenarios for the 3:30–5:30 pm segment.

Below, we summarize the parameters and variables, and present the models for our case study.

Parameters are as follows:

(i) $W = \{9, 10, 11\}$: the set of possible lane widths in addition to 12 ft lane width
(ii) $L = \{25, 35, 40, 45\}$: the set of posted speed limits under consideration
(iii) $S = \{1, \ldots, 9\}$: the set of scenarios
(iv) p^i: probability of scenario $i \in S$
(v) τ: required reliability level.
(vi) ρ: the number of standard errors from the average vehicle speed corresponding to the λ-th percentile speed, i.e., $\rho = \text{NORM.S.INV}(\lambda)$
(vii) D^i: the directional average daily traffic per lane in scenario $i \in S$
(viii) d^i: the average 5-minute travel volume in scenario $i \in S$

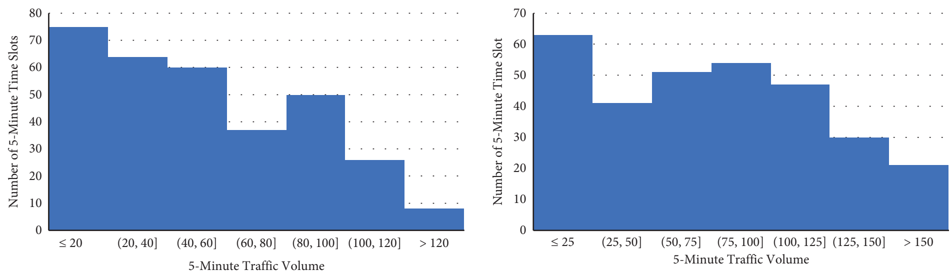
First-stage decision variables are as follows:

(i) $w_k = 1$ if the lane width $k \in W$ is adopted, $w_k = 0$ otherwise
(ii) q_i: the value of the i-th road geometric design parameter variable as defined in Tables 4–7 and 9, 10 with $i \in Q$

Second-stage decision variables are as follows:

(i) $u_l^i = 1$ if road speed limit $l \in L$ is adopted in scenario $i \in S$, $u_l^i = 0$ otherwise
(ii) n_l^i: the expected annual crash frequency with lane width $l \in L$ in scenario $i \in S$ (auxiliary variables for ease of model presentation)
(iii) $z_l^i = 1$ if the λ-th percentile vehicle travel speed is under the speed limit $l \in L$ in scenario $i \in S$, $z_l^i = 0$ otherwise

The model for the 1–3 pm time segment



(a) (b)

Figure 2: 5-Minute travel volume: (a) 1–3 pm and (b) 3:30–5:30 pm.

$$\min \sum_{i\in S} p^i \sum_{l\in L} u_l^i n_l^i,$$
$$\text{s.t. } (2), (3), (6), (7), \tag{8}$$

$$n_{25}^i = e^{-1.17-0.73w_{10}+1.19q_1+1.4q_2+0.0000036D^i} \quad i \in S, \tag{9}$$

$$n_{35}^i = e^{-1.25-0.36w_9-0.28w_{10}-0.3w_{11}-0.62q_3+0.39q_4+0.93q_5+1.6q_6+0.0003D^i} \quad i \in S, \tag{10}$$

$$n_{40}^i = e^{-3.39+0.06w_{10}-0.33w_{11}+0.0002D^i-0.38q_7+1.17q_8+1.01q_9+1.79q_{10}} \quad i \in S, \tag{11}$$

$$n_{45}^i = e^{-1.86-0.06w_{11}+0.0001D^i+0.65q_8+1.58q_{11}+0.94q_{12}} \quad i \in S, \tag{12}$$

$$23.651 - 6.186w_9 - 4.236w_{10} + 5.440\rho \le 25z_{25}^i + M\left(1 - z_{25}^i\right) \quad i \in S, \tag{13}$$

$$36.972 + 3.213w_9 + 2.503q_{13} - 0.832q_{14} - 0.024d^i + 5.200\rho \le 35z_{35}^i + M\left(1 - z_{35}^i\right) \quad i \in S, \tag{14}$$

$$42.885 - 6.279w_{10} + 1.209w_{11} - 0.029d^i + 4.992\rho \le 40z_{40}^i + M\left(1 - z_{40}^i\right) \quad i \in S, \tag{15}$$

$$34.885 - 3.587w_{10} + 9.901w_{11} + 0.029d^i + 5.326\rho \le 45z_{45}^i + M\left(1 - z_{45}^i\right) \quad i \in S, \tag{16}$$

$$z_l^i \le u_l^i \quad l \in L, i \in S, \tag{17}$$

$$\sum_{i\in S} p^i \sum_{l\in L} z_l^i \ge \tau \tag{18}$$

$$z_l^i \in \mathbb{B} \quad l \in L, i \in S, \tag{19}$$

$$q_3 + q_7 \le 1, \tag{20}$$

$$q_{14} = 3 - 2q_3 - q_7, \tag{21}$$

$$q_8 = 0.264, \tag{22}$$

$$q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_9, q_{10}, q_{11}, q_{12}, q_{13} \in \mathbb{B}, \tag{23}$$

where constraints (9)–(12) use auxiliary variables to simplify the objective function and constraints (13)–(19) are the deterministic equivalent of constraint (4) for the 1–3 pm time segment. Constraints (20)–(23) are the additional necessary restrictions for the road geometric design variables $\{q_i\}_{i\in Q}$.

Table 11: Computational results of models for 1–3 pm and 3:30–5:30 pm time segments.

Time segment	Reliability level τ	ρ, number of standard deviations from mean speed (percentile) ← High speed and high severity −3 (0.1%)	−2 (2.2%)	−1 (15.8%)	0 (50%)	Low speed and low severity ⟶ 1 (84.1%)	2 (97.7%)	3 (99.9%)
1–3 pm	0.5	0.08670	0.08670	0.08670	0.09259	0.09259	0.19312	Infeasible
	0.6	0.08670	0.08670	0.08670	0.09259	0.09259	0.21410	Infeasible
	0.7	0.08670	0.08670	0.08670	0.09259	0.10004	0.23528	Infeasible
	0.8	0.08670	0.08670	0.08670	0.09259	0.10794	0.26316	Infeasible
	0.9	0.08670	0.08670	0.08670	0.09259	0.10847	0.29785	Infeasible
	1	0.08670	0.08670	0.08670	0.09259	0.11592	0.38713	Infeasible
3:30–5:30 pm	0.5	0.08670	0.08670	0.08670	0.09259	0.09259	0.18508	Infeasible
	0.6	0.08670	0.08670	0.08670	0.09259	0.09259	0.22457	Infeasible
	0.7	0.08670	0.08670	0.08670	0.09259	0.09259	Infeasible	Infeasible
	0.8	0.08670	0.08670	0.08670	0.09259	0.09860	Infeasible	Infeasible
	0.9	0.08670	0.08670	0.08670	0.09259	0.10608	Infeasible	Infeasible
	1	0.08670	0.08670	0.08670	0.09259	0.18976	Infeasible	Infeasible

The model for the 3:30–5:30 pm time segment

$$\min (8),$$

$$\text{s.t.}\ (2), (3), (6), (7), (9)-(12), (17)-(23), \tag{24}$$

$$23.717 - 7.934w_9 + 1.323w_{10} - 3.497w_{11} - 0.026d^i + 5.363\rho \le 25z^i_{25} + M\left(1 - z^i_{25}\right) \quad i \in S,$$

$$36.193 + 1.962w_9 + 3.947w_{10} + 7.636w_{11} - 0.067d^i + 6.143\rho \le 35z^i_{35} + M\left(1 - z^i_{35}\right) \quad i \in S, \tag{25}$$

$$40.056 - 5.211w_{10} + 1.147w_{11} + 4.734\rho \le 40z^i_{40} + M\left(1 - z^i_{40}\right) \quad i \in S, \tag{26}$$

$$43.394 - 6.839w_{10} + 6.064w_{11} - 0.031d^i + 4.651\rho \le 45z^i_{45} + M\left(1 - z^i_{45}\right) \quad i \in S, \tag{27}$$

where constraints (24)–(27) together with (17)–(19) are the deterministic equivalent of constraint (4) for the 3:30–5:30 pm time segment.

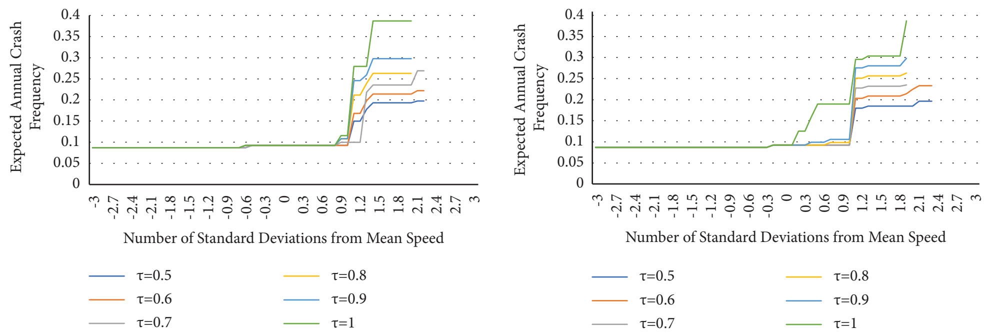
To have a comprehensive computational study of possible situations, we test for a variety of reliability levels ($\tau = 0.5, 0.6, 0.7, 0.8, 0.9$ and 1) and up to three standard deviations from the estimated mean vehicle travel speed (equivalently, 0.1 to 99.9 percentiles).

Table 11 presents the optimal objective function values of the two-stage stochastic programming model, i.e., the minimum expected annual crash frequencies, with the corresponding percentile of vehicle travel speed under the posted speed limit for time segments 1–3 pm and 3:30–5:30 pm with different reliability levels. "Infeasible" cases are the models for which no feasible solution exists because of the overrestrictive requirement for traffic speed limit control.

The solution of each model is composed of the optimal values of the studied road geometric design parameters as well as the selected posted speed limit for each scenario. For example, in time segment 1–3 pm with 80% reliability level ($\tau = 0.8$), if we would have 84.1% of the vehicles ($\rho = 1$) compliant to the posted speed limit, the minimum crash frequency is 0.10794 with the optimal solution as: 10 ft road with 2 lanes each direction, 40 MPH posted speed limit for scenarios 2 and 3, 25 MPH posted speed limit for all other scenarios, and value 0 for all other parameters.

Due to the length of the complete computational results, in Table 11 we report only the computational results with the speed control at 0.1%, 2.2%, 15.8%, 50%, 84.1%, 97.7%, and 99.9%, i.e., $\rho = -3, -2, -1, 0, 1, 2$, and 3, respectively. The complete computational results can be found in Appendix A. Also, Figure 3 depicts the changes in crash frequency as the road geometric design shifts to promote lower average vehicle travel speed.

As presented, with the increase of the reliability level (i.e., with a higher chance that the speed compliance can be met across all scenarios), the expected annual crash frequency goes up. This is because the chance constraint brings to the model an additional restriction, which leads the priority of the road geometric design to shift from crash frequency toward crash severity. The higher the reliability level is, the more restrictive the model is. In other words, the chance constraint trades off the crash frequency for the lower severity of crashes in the road design. Similarly, at the same reliability level, the higher the speed compliance percentile is (i.e., the less severe the crashes may be), the more expected annual crashes there are. This is also because the road design of the lowest crash frequency is sacrificed to comply with the



(a) (b)

Figure 3: Expected annual crash frequency with different reliability and speed control levels: (a) 1–3 pm and (b) 3:30–5:30 pm.

Table 12: Expected annual crash frequency for 1–3 pm with different τ and ρ.

ρ	Reliability level τ					
	0.5	0.6	0.7	0.8	0.9	1
−3.0 to −1.0	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.9	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.8	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.7	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.6	0.0867	0.09259	0.09259	0.09259	0.09259	0.09259
−0.5	0.0867	0.09259	0.09259	0.09259	0.09259	0.09259
−0.4	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
−0.3	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
−0.2	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
−0.1	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.1	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.2	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.3	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.4	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.5	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.6	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.7	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.8	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.9	0.09259	0.09259	0.10004	0.10794	0.10847	0.11592
1.0	0.09259	0.09259	0.10004	0.10794	0.10847	0.11592
1.1	0.15001	0.16805	0.10004	0.21135	0.24603	0.27936
1.2	0.15001	0.16805	0.10004	0.21135	0.24603	0.27936
1.3	0.17774	0.19794	0.21814	0.23896	0.25916	0.27936
1.4	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
1.5	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
1.6	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
1.7	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
1.8	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
1.9	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
2.0	0.19312	0.2141	0.23528	0.26316	0.29785	0.38713
2.1	0.19771	0.2217	0.26892	Infeasible	Infeasible	Infeasible
2.2	0.19771	0.2217	0.26892	Infeasible	Infeasible	Infeasible
2.3	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible
2.4	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible
2.5 to 3.0	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible

speed control to have a higher percentile of the traffic flow to be under the posted speed limit to mitigate the crash severity. According to the computational results in Appendix A, when no more than 24.2% of the vehicle speeds (−0.7 standard deviation from the mean speed) at 1–3 pm and no more than 38.2% of the vehicle speeds (−0.3 standard deviation from the mean speed) at 3:30–5:30 pm are required to be lower than the posted speed limit, the objective function values remain the same, at 0.0867, which is the lowest expected annual crash frequency. This is because at

Table 13: Expected annual crash frequency for 3:30–5:30 pm with different τ and ρ.

ρ	Reliability level τ					
	0.5	0.6	0.7	0.8	0.9	1
−3.0 to −1.0	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.9	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.8	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.7	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.6	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.5	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.4	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.3	0.0867	0.0867	0.0867	0.0867	0.0867	0.0867
−0.2	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
−0.1	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.1	0.09259	0.09259	0.09259	0.09259	0.09259	0.09259
0.2	0.09259	0.09259	0.09259	0.09259	0.09259	0.12546
0.3	0.09259	0.09259	0.09259	0.09259	0.09259	0.12546
0.4	0.09259	0.09259	0.09259	0.09259	0.09923	0.15833
0.5	0.09259	0.09259	0.09259	0.09259	0.09923	0.18976
0.6	0.09259	0.09259	0.09259	0.09259	0.09923	0.18976
0.7	0.09259	0.09259	0.09259	0.0986	0.10608	0.18976
0.8	0.09259	0.09259	0.09259	0.0986	0.10608	0.18976
0.9	0.09259	0.09259	0.09259	0.0986	0.10608	0.18976
1.0	0.09259	0.09259	0.09259	0.0986	0.10608	0.18976
1.1	0.18019	0.20293	0.22782	0.25078	0.27567	0.29587
1.2	0.18019	0.20293	0.22782	0.25078	0.27567	0.29587
1.3	0.18508	0.20864	0.2322	0.25648	0.28005	0.30361
1.4	0.18508	0.20864	0.2322	0.25648	0.28005	0.30361
1.5	0.18508	0.20864	0.2322	0.25648	0.28005	0.30361
1.6	0.18508	0.20864	0.2322	0.25648	0.28005	0.30361
1.7	0.18508	0.20864	0.2322	0.25648	0.28005	0.30361
1.8	0.18508	0.20864	0.2322	0.25648	0.28005	0.30361
1.9	0.18508	0.2141	0.23528	0.26316	0.29785	0.38713
2.0	0.18508	0.22457	Infeasible	Infeasible	Infeasible	Infeasible
2.1	0.19664	0.23295	Infeasible	Infeasible	Infeasible	Infeasible
2.2	0.19664	0.23295	Infeasible	Infeasible	Infeasible	Infeasible
2.3	0.19664	0.23295	Infeasible	Infeasible	Infeasible	Infeasible
2.4	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible
2.5 to 3.0	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible	Infeasible

least 24.2% of the traffic at 1–3 pm and 38.2% of the traffic at 3:30–5:30 pm are naturally lower than the selected posted speed limit. When we aim to have more than 98.6% of the vehicle speed (2.2 standard deviation from the mean speed) at 1–3 pm or more than 98.9% of the vehicle speeds (2.3 standard deviation from the mean speed) at 3:30–5:30 pm to be lower than the speed limit, the models are infeasible. This is because the speed control requirement is so restrictive that there is no of combination of road geometric design parameters and posted speed limit to satisfy this requirement. Therefore, adapting road geometric design parameters for speed compliance in order to mitigate crash severity is only applicable on a certain range of the percentiles of the vehicle travel speed, i.e., between 24.2–98.6% for 1–3 pm and 38.2–98.9% for 3:30–5:30 pm.

6. Conclusions and Future Research

In this study, we first examined the safety and operational effects of the road geometric design parameters. The safety analysis highlighted the impact of lane widths as well as other road geometric design parameters on annual crash frequency in the urban environments of Lincoln, Nebraska, by using Poisson and negative binomial regressions. The operational analysis studied the effect of the lane width on vehicle's travel speed using linear regressions. Integrating the regression results, we proposed a two-stage stochastic programming model to determine the optimal lane widths, posted speed limits and other-related road geometric design parameters for the possible scenarios while improving vehicle travel speed compliance with the posted speed limit to mitigate crash severity. The proposed model has been tested on the data collected at 14 data collection sites in Lincoln, Nebraska. Depending on the reliability and percentile of speed compliance, the model provided us with different sets of road geometric design parameters to balance the goals of minimizing the annual crash frequency and mitigating the crash severity. The computational study validated the effectiveness of our decision scheme. We are also collaborating with the government to implement this methodology and experiment with our findings in practice, further proving its effectiveness.

The integration of regression analysis results into a stochastic programming model creates an innovative decision-making framework for road geometric design to improve transportation safety. This methodology has the potential for widespread use, as it can be applied in diverse environments to help reduce both crash frequency and severity. For example, with the increasing emergence of connected and autonomous vehicles, incorporating platoon strategies into road geometric design is a crucial aspect of transportation planning (Yang et al. [29], Zhong et al. [30]). Taking into consideration the characteristics of platoon strategies in the safety and travel speed analyses, such as platoon percentage, platoon size, intraplatoon distance, platoon speed, formation, and coordination time, the proposed stochastic programming model can be applied to determine the road geometric design that minimizes crash frequency while mitigating crash severity, accounting for the existence of platoon strategies. Furthermore, we also found many interesting future research directions to explore. For example, the safety analysis can be done for the crashes at the intersections, and a similar stochastic programming model can help determine the optimal design to improve safety at the intersections. Also, it would be interesting to incorporate more traffic violation events into consideration when analyzing the traffic safety, such as the vehicle lane violations.

Appendix

A. Complete Computational Results

In this appendix, we include the complete computational results in Tables 12 and 13 for the case study in Section 5.2.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Part of the safety analysis and travel speed analysis in this study was based on the project titled "Safety and Operational Analysis of Lane Widths in Mid-Block Segments and Intersection Approaches in the Urban Environment in Nebraska," which was funded by the Nebraska Department of Transportation.

References

[1] National Academy Press, *Managing Speed: Review of Current Practices for Setting and Enforcing Speed Limits*, TRB Special Report 254, National Academy Press, Washington, DC, USA, 1998.

[2] D. Shinar, "Speed and crashes: a controversial topic and an elusive relationship," Appendix B of TRB Special Report 254, pp. 221–276, National Academy Press, Washington, DC, USA, 1998.

[3] Y. Najjar, R. Stokes, and E. Russell, "Setting speed limits on Kansas two-lane highways: neuronet approach," *Transportation Research Record*, vol. 1708, no. 1, pp. 20–27, 2000.

[4] Ntsb, "2021-2022 NTSB most most wanted list of transportation safety improvements," 2021, https://www.ntsb.gov/Advocacy/mwl/Pages/default.aspx.

[5] R. Noland, "Traffic fatalities and injuries: the effect of changes in infrastructure and other trends," *Accident Analysis & Prevention*, vol. 35, no. 4, pp. 599–611, 2003.

[6] I. B. Potts, D. W. Harwood, and K. R. Richard, "Relationship of lane width to safety on urban and suburban arterials," *Transportation Research Record*, vol. 2023, no. 1, pp. 63–82, 2007.

[7] C. V. Zegeer, D. W. Reinfurt, J. Hummer, L. Herf, and W. Hunter, *Safety Effects of Cross-Section Design for Two-Lane Roads*, Transportation Research Board, Washington, DC, USA, 1987.

[8] E. Hauer, *Safety Models for Urban Four-Lane Undivided Road Segments*, Transportation Research Board, Washington, DC, USA, 2004.

[9] J. G. Strathman, K. J. Dueker, J. Zhang, and T. Williams, *Analysis of Design Attributes and Crashes on the oregon Highway System*, Oregon Department of Transportation, Center for Urban Studies, Portland State University, Portland, OR, USA, 2001.

[10] D. W. Harwood, *Effective Utilization of Street Width on Urban Arterials*, Transportation Research Board, Washington, DC, USA, 1990.

[11] I. B. Potts, D. W. Harwood, D. J. Torbic et al., "Lane widths, channelized right turns, and right-turn deceleration lanes in urban and suburban areas," Final Report of NCHRP Project 3-72, Midwest Research Institute, Kansas City, MO, USA, 2006.

[12] C. V. Zegeer, H. F. Huang, J. C. Stutts, E. Rodgman, and J. E. Hummer, *Commercial Bus Accident Characteristics and Roadway Treatments*, Transportation Research Board, Washington, DC, USA, 1994.

[13] T. Sando and R. Moses, "Integrating transit into traditional neighborhood design policies– the influence of lane width on bus safety," Final Report, Florida Department of Transportation Transit Office, Tallahassee, FL, USA, 2010.

[14] B. Dai, J. G. Hudson, B. Ettelman, and E. S. Park, "Influence of lane width on bus crashes," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2675, no. 2, pp. 87–97, 2021.

[15] E. T. Donnell, S. C. Himes, K. M. Mahoney, and R. J. Porter, "Understanding speed concepts: key definitions and case study examples," *Transportation Research Record*, vol. 2120, no. 1, pp. 3–11, 2009.

[16] R. Lamm and E. M. Choueiri, "Recommendations for evaluating horizontal alignment design consistency based on investigations in the state of New York," *Transportation Research Record*, vol. 1122, pp. 68–78, 1987.

[17] A. Ali, A. Flannery, and M. Venigalla, *Prediction Models for Free Flow Speed on Urban Streets*, Transportation Research Board, Washington, DC, USA, 2007.

[18] C. M. Poe and J. M. Mason, "Analyzing influence of geometric design on operating speeds along low-speed urban streets: mixed model approach," *Transportation Research Record*, vol. 1737, no. 1, pp. 18–25, 2000.

[19] V. Gitelman, F. Pesahov, R. Carmel, and S. Bekhor, "The identification of infrastructure characteristics influencing travel speeds on single-carriageway roads to promote

self-explaining roads," *Transportation Research Procedia*, vol. 14, pp. 4160–4169, 2016.

[20] V. Gitelman, E. Doveh, and S. Bekhor, "The relationship between free-flow travel speeds, infrastructure characteristics and accidents, on single-carriageway roads," *Transportation Research Procedia*, vol. 25, pp. 2026–2043, 2017.

[21] B. N. Campbell, J. D. Smith, and W. G. Najm, "Examination of crash contributing factors using national crash databases," Department of Transportation, Washington, DC, USA, DOT HS 809 664, DOT-VNTSC- NHTSA-02-07, 2003.

[22] C. N. Kloeden, A. J. McLean, and G. Glonek, *Reanalysis of Traveling Speed and the Risk of Crash Involvement in Adelaide South australia*, Department of Transport and Regional Services, Australian Transport Safety Bureau, Canberra, Australia, 2002.

[23] R. Elvik, P. Christensen, and A. Helene Amundsen, "Speed and road accidents: an evaluation of the power model," TØI report 740/2004, Institute of Transport Economics TOI, Oslo, Norway, 2004.

[24] G. Nilsson, "Traffic safety dimensions and the power model to describe the effect of speed on safety," *Bulletin*, Lund Institute of Technology, Lund, Sweden, 2004.

[25] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.

[26] S. Washington, M. Karlaftis, and F. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2nd edition, 2011.

[27] B. L. Miller and H. M. Wagner, "Chance constrained programming with joint constraints," *Operations Research*, vol. 13, no. 6, pp. 930–945, 1965.

[28] A. Prékopa, "Contributions to the theory of stochastic programming," *Mathematical Programming*, vol. 4, no. 1, pp. 202–221, 1973.

[29] X. Yang, Y. Zou, and L. Chen, "Operation analysis of freeway mixed traffic flow based on catch-up coordination platoon," *Accident Analysis & Prevention*, vol. 175, Article ID 106780, 2022.

[30] Z. Zhong, E. E. Lee, M. Nejad, and J. Lee, "Influence of CAV clustering strategies on mixed traffic flow characteristics: an analysis of vehicle trajectory data," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102611, 2020.