Assessment of the Impact of Lane Width on Arterial Crashes

Ziaur Rahman[1], Arezoo Memarian[1], Sunil Madanu[1]
Gazi Iqbal[2], Hadis Anahideh[2],
Stephen P. Mattingly[1], Jay M. Rosenberger[2]

1. Introduction

Safety represents a major concern in roadway design and management. For the last 20 years, research has considered the safety efficacy of lane width, but it has provided little statistical insight. Although AASHTO guidelines recommend using standard 12-foot lane widths, previous research has shown both positive and negative safety outcomes due to lane width variation, which makes definitive conclusions difficult. Zegeer et al. (1) use variables such as type of development, terrain, section length, average daily traffic, speed, horizontal curvature, vertical grade, side slope ratio, lane width, shoulder width, and parking to analyze the safety effects of cross-section design. Their results show that one foot of lane widening (from 9 feet) reduces accidents by 12%, and it almost doubles for every extra foot of widening. Wang et al. (2) build a Poisson Regression model by using functional class, number of lanes, road surface width, divided/undivided highway, median width and type, intersection type, access control, and area type as independent variables. Harwood et al. (3) and Harkey et al. (4) develop lane width (12 feet as reference) crash modification factors for two-lane rural highways, which indicate that widening lanes reduces single-vehicle run-off-road crashes, multiple-vehicle head-on, opposite-direction sideswipe, and same-direction sideswipe collisions. On the contrary, some studies like NCHRP report 330 (5) states that "narrower lane widths (less than 11 feet) can be used effectively in urban arterial street improvement projects where the additional space can be used to relieve traffic congestion or address specific accident patterns." Potts et al. (2007) find no safety risk for lane widths narrower than 12 feet on urban and suburban arterials (6). Although the impact of narrow lanes on vehicular crash rates appears inconclusive, narrower lane widths provide opportunities for other potential safety and operational benefits, which include reducing pedestrian crossing distance, auxiliary lanes, bicycle lanes, and buffer areas. Furthermore, cities and counties resurface their streets based on assumptions that narrower lane widths increase accidents. This paper uses several different statistical approaches to explore the relationship between lane width and crash rates, and it is organized into five sections. The second section discusses the given data set and modeling approaches. The third section describes the data set and data processing required for analysis. The fourth section explains the different modeling approaches with their results. Finally, the fifth section draws conclusions from the analysis.

2. Case Study Description

This paper assesses the impact of lane width on the safety of arterial roads. The study uses ten years of crash data from four cities in Nebraska provided by the Transportation Research Board (TRB). This data set contains midblock segment details, such as speed limit (*SL*), presence of a median (*M*), presence of a shoulder (*S*), lane width (*LW*), presence of on-street parking (*OSP*), indication of a one-way segment (*OW*), indication of a segment in a central business district (*CBD*), segment length (*SGLT*), the number of through lanes (*NTL*), and annual average daily traffic per lane (*AADTpLn*). The data set also includes the segments' yearly crash frequencies from 2003 to 2012 for different categories such as gender, age, severity, type, movement, and leading cause. This paper develops three different statistical modeling approaches for four response variables. The first approach estimates the relationship between the likelihood of a crash with the aforementioned midblock segment variables, while the second approach analyses the frequency of crashes. The third approach aggregates the data into five-year observations and then determines the relationships between injury and non-injury related crashes with the same independent variables.

3. Data Description and Processing

The TRB data set contains 19,600 observations, which is reduced to 18,227 observations due to missing data. AADTpLn, SL, NTL, and SGLT are considered continuous or multinomial variables, whereas S, M, OSP, and OW are examples of binary variables. LW is considered a 4-level categorical variable (*LW9*, *LW10*, *LW11*) because there is no reason to presume a linear or log linear relationship between lane width and crash frequency; the researchers selected 12 feet lanes as the reference condition because it is the standard for most highway classes.

[1] Department of Civil Engineering, The University of Texas at Arlington, Arlington TX
[2] Department of Industrial Engineering, The University of Texas at Arlington, Arlington TX

According to the NHTSA American National Standards (8), an accident is classified by the most serious injury. The data set contains crash frequency in six different severity categories, namely non-reportable, property damage, visible injury, possible injury, disabling injury, and fatal injury. A continuous dependent variable referred to as *non-injury* is the total of the first two variables while *injury* is the sum of the latter four variables.

Four data sets are derived from the 18,227 observations. Using the 18,277 observations, 3,645 observations (20%) are randomly selected for model validation and are referred to as the *Annual Test Data*. The *Annual Training Data* uses the remaining 14,582 observations for model development. The selection of these data sets appears unbiased based on the distributions of age and gender within the crash data. In addition, five years of crash data are aggregated into single observations and are referred to as the *5-Year Data*, which includes 3,734 observations. Similarly, the 5-Year Data is split into *5-Year Testing Data* (746 observations) and *5-Year Training Data* (2,988 observations). The crash frequency distributions of the 5-Year Testing and Training Data for LW10 (16% to 22%) and LW11 (36% to 29%) are a little different; however, the chi-squared tests conducted on the injury and non-injury crashes with 95% confidence do not reject the hypotheses that the subsets are similar.

4. Modeling and Analysis

This section describes five statistical models that use Classification and Regression Trees (CART). CART, developed by Breiman et al. in the early 1980's (9), is a decision tree tool to construct a predictive model; it starts from a root node and splits the data set into two branches based on least squares and cross validation. The splitting process continues until reaching terminal nodes. Each observation falls into exactly one terminal node based upon the tree logic (10). In addition, CART yields variable importance scores that rank the importance of each dependent predictor variable.

This research develops a pure CART model and four treed regression models. The *pure CART model* predicts whether a crash occurs in the *Annual Training Data* and is constructed using Salford Systems software. The model has no restriction on the required number of observations in the terminal nodes, and some of the resulting terminal nodes have as few as two observations, which suggests that this model overfits the data. Nonetheless, six variables (SGLT, AADTpLn, SL, NTL, S, and NFRC) have greater importance than any lane width variable. The study removes NFRC from further analysis because its definition appears subjective.

The four treed regression models use CART to build an initial tree based on the training data. Each training data set is split within each terminal node of these CART models, and stepwise logistic and linear regression models using all of the predictor variables are estimated for the appropriate response variables. The CART models combined with the stepwise regression models are considered *treed regression models*. The study uses the *Annual Training Data* to build a logistic treed regression model to predict the likelihood of a crash and another treed regression model to predict the number of crashes. To control for the five most important variables, these models include them as the only predictor variables and require at least 500 observations per terminal node. Two other treed regression models predict the number of injury and non-injury crashes from the 5-Year Data. The CART model for the logistic treed regression model has 22 terminal nodes, an R^2 value of 0.13, a root node of AADTpLn (at the value 4383), and a variable importance score ranking of AADT, SGLT, SL, NTL, and S, which is similar to the pure CART model. Similarly, the CART model for the treed regression CART model has 21 terminal nodes with an R^2 value of 0.16, a root node split on AADTpLn (at 7482), and the same ranking in the variable importance scores. The non-injury and injury CART models (Figure 1) are constructed using R-studio software. Although these CART models consider all predictor variables, only SL, NTL, AADTpLn and SGLT appear in the trees, and the root node splits on AADTpLn, which is consistent with the previous CART models. The non-injury and injury CART models have training R^2 values of 0.18 and 0.16. Stepwise regression is conducted within each terminal node of each tree. Within each terminal node, the logistic treed regression model uses the statistical software SAS; the PROC LOGISTIC function includes a stepwise selection where a significance level of 0.30 is required for variable entry into the model, and 0.35 is required to retain a variable. Within each terminal node, the treed regression model fits a step-wise linear regression model where all variables are statistically significant at the 5% level to predict the annual number of crashes. The 5-year models also build stepwise linear regression models for each terminal node. A significance level of 0.30 is required for variable entry, and a level of 0.35 is required for variable retention. The R^2 values for the treed regression models are larger than those of the CART models, suggesting that the stepwise regression in the terminal nodes improves these models' predictive power. The logistic treed regression model R^2

and Tjur R^2 are both 0.18, while the CART R^2 is 0.13. For the treed regression model, the R^2 value is 0.27 compared to only 0.16 for the CART model. R^2 values for the non-injury and injury models are 0.36 and 0.30, which are 0.18 and 0.14 higher than those of the CART models. Moreover, testing data is used to validate the treed regression models; all have training and testing R^2 values within 0.02 of each other, suggesting that the treed regression models do not appear overfit.

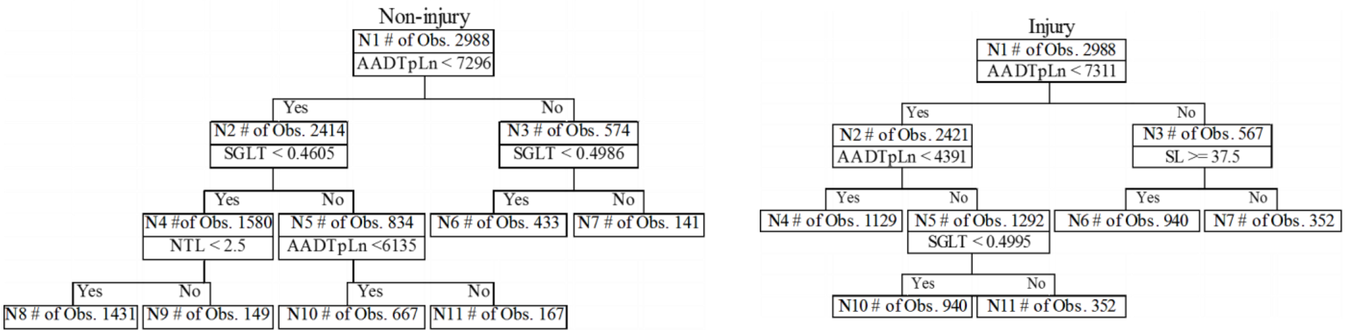

Figure 1. CART models for injury and non-injury treed regression models.

Node characteristics for the logistic treed regression model, the treed regression model, and the injury and non-injury models are shown in Table 1, Table 2, and Table 3, respectively. Although these models have other significant variables, because of the study focus, only estimates for lane widths are shown in the tables. All four treed regression models show that in most nodes, either the R^2 is quite small or the lane width coefficients or odds ratios are either insignificant or small. Only Node 3 in the logistic treed regression model has all four R^2 values greater than 0.10. In the treed regression model, Nodes 18 and 21 are the only nodes to have R^2 values greater than 0.16 and include lane width coefficients greater than 1 crash per year. Node 3 in the injury model and Nodes 3 and 4 in the non-injury model exclude the lane width variables. Although Nodes 1 and 2 in the injury model and Node 2 in the non-injury model include coefficients on lane width variables, their 95% confidence intervals include zero, suggesting that impacts may be insignificant. Consequently, in most cases, factors other than lane width account for crash likelihood, frequency, and severity; however, narrower lane widths can increase crashes for high volume segments. For instance, Node 3 in the treed logistic regression model with the highest R^2 (0.16) shows that 9-foot lane widths triple the probability of a crash, while 10-foot lane widths and 11-foot lane widths increase the probability by 73% and 53%, respectively, when compared to 12-foot lane width sections. Node 21 in the treed regression model is defined by higher AADT per lane, speed limits, and longer segments, and in it, 10-foot lane widths and 11-foot lane widths increase the number of crashes by 3.74 and 1.81, respectively. Since 9-foot lane widths only appear in 5% of the observations, they do not register a significant effect. Node 18 similarly has higher AADT per lane but only 30-35 mph speed limits and shorter segments. For these sections, both 10-foot lane widths and 9-foot lane widths but the 10-foot lane width's increase (2.44) appears greater than 9-foot lane width's increase (1.48). In the injury and non-injury models, all high volume nodes (Nodes 4 and 5 in the injury model and Nodes 5 and 6 in the non-injury model) show positive significant coefficients for at least one lane width variable. On segments with higher volume and high speed limits (Node 4 in injury model), 11-foot lane widths increase the number of injury crashes by 0.705 per 5 years. On those with 30-35 mph speed limits (Node 5), 10-foot lane widths increase injury crashes by 3.703. On short segments with higher volume (Node 5 in non-injury model), 10-foot lane widths increase the number non-injury crashes by 5.76 per 5 years. On long segments (Node 6), 10-foot and 11-foot lane widths increase non-injury crashes by 14.94 and 5.69, respectively.

Conclusions

The purpose of this study is to identify the impact of lane width on safety. Analysis from the four treed regression models indicates that lane width has limited impact on safety in most cases. Nonetheless, all four models also show that narrower lanes increase the likelihood and the frequency of both injury and non-injury crashes on high-volume segments; these results may be used to identify potentially risky volumes. These models are built using data that omits potential confounding effects such as weather conditions and visibility. Consequently, more robust data may improve the quality of the models and possibly further isolate the effects of lane width on safety.

Table 1 Treed regression characteristics and lane width odds ratios from logistic regression analysis

Node	# of Obs.	Range	SL	NTL	AADTpln	SGLT	Dummy Variable (=1)				Odds Ratio			*P*-value			Training		Testing	
							LW9	LW10	LW11	Crash	LW9	LW10	LW11	LW9	LW10	LW11	R^2	Tjur R^2	R^2	Tjur R^2
1	767	Min	25	1	111	0.03	32	156	227	195	3.03	-	1.53	0.009	-	0.002	0.11	0.08	0.10	0.09
		Max	35	4	3,250	0.15														
2	632	Min	20	1	150	0.15	30	196	165	225	-	-	1.64	-	-	0.021	0.07	0.05	0.04	0.05
		Max	35	3	3,255	0.36														
3	601	Min	20	1	3,258	0.03	28	140	283	294	3.01	1.73	1.53	0.028	0.074	0.026	0.16	0.12	0.17	0.15
		Max	35	6	4,378	0.36														
4	605	Min	40	1	100	0.03	0	31	165	110		0.27	1.42		0.029	0.138	0.09	0.05	0.09	0.07
		Max	45	3	3,616	0.37														
5	571	Min	40	1	3,625	0.07	0	8	120	160		-	1.52		-	0.065	0.11	0.08	0.00	0.04
		Max	55	4	4,375	0.37														
6	647	Min	25	1	500	0.37	17	124	298	269	3.33	-	2.81	0.026	-	<.0001	0.13	0.10	0.03	0.07
		Max	40	1	4,350	2.01														
7	823	Min	25	2	453	0.37	0	267	254	489		1.27	-		0.131	-	0.10	0.07	0.03	0.05
		Max	40	5	4,375	1.01														
8	794	Min	45	1	110	0.38	0	36	417	288		-	0.71		-	0.031	0.06	0.04	0.00	0.02
		Max	60	4	4,375	3.88														
9	793	Min	20	1	4,411	0.02	0	78	221	287		3.28	1.44		<.0001	-	0.04	0.03	0.05	0.05
		Max	40	2	8,050	0.14														
10	776	Min	25	1	4,425	0.15	14	86	253	385	2.14	2.78	-	0.284	<.0001	-	0.06	0.05	-0.02	0.02
		Max	40	2	8,077	0.24														
11	544	Min	30	1	4,400	0.24	34	110	159	346	2.88	-	-	0.081	-	-	0.06	0.04	0.03	0.04
		Max	40	2	8,075	0.26														
12	803	Min	30	1	4,388	0.27	46	142	299	438	3.99	2.43	1.77	0.001	0.001	0.132	0.10	0.07	0.06	0.07
		Max	40	2	7,957	0.38														
13	542	Min	45	1	4,470	0.09	0	35	229	212		-	-		-	-	0.02	0.01	0.00	0.01
		Max	55	2	8,075	0.25														
14	635	Min	45	1	4,400	0.25	0	26	268	179		-	-		-	-	0.03	0.02	0.03	0.02
		Max	55	2	8,075	0.38														
15	605	Min	30	1	8,088	0.02	0	53	230	355		3.88	1.43		0.000	0.056	0.06	0.04	0.05	0.05
		Max	45	2	16,650	0.24														
16	512	Min	35	1	8,088	0.24	0	63	159	373		5.91	1.72		0.001	0.045	0.16	0.11	0.07	0.11
		Max	45	2	16,650	0.38														
17	510	Min	30	3	4,416	0.07	0	92	146	402		-	-		-	-	0.07	0.04	0.06	0.06
		Max	45	5	11,017	0.35														
18	1,079	Min	25	1	4,388	0.39	95	231	353	674	1.65	-	-	0.055	-	-	0.06	0.05	0.05	0.06
		Max	40	3	6,425	2.01														
19	568	Min	45	1	4,418	0.39	0	21	258	292		-	0.68		-	0.030	0.08	0.06	0.04	0.05
		Max	55	3	6,429	3.88														
20	570	Min	25	1	6,450	0.39	30	49	180	454	0.51	-	0.51	0.151	-	0.004	0.06	0.04	-0.03	0.01
		Max	40	3	7,475	1.00														
21	519	Min	35	1	7,488	0.39	50	64	199	457	-	7.26	1.36	-	0.056	0.280	0.06	0.02	-0.01	0.01
		Max	40	3	19,480	1.01														
22	686	Min	45	1	6,450	0.39	0	15	206	439		-	1.57		-	0.016	0.09	0.06	0.05	0.06
		Max	55	3	14,900	1.20														

Note: Grey areas show non-occurrence of lane width in the dataset

Table 2 Treed regression characteristics and lane width coefficients from linear regression analysis

Node	# of Obs.	Range	AADTpln	SL	SGLT	NTL	Dummy Variable (=1)			Coefficient			R^2	R^2
							LW9	LW10	LW11	LW9	LW10	LW11	Training	Testing
1	1465	Min	111	20	0.03	1	73	329	357	0.32	-	-	0.05	0.06
		Max	3,600	40	0.25	4								
2	532	Min	3,608	20	0.03	1	0	101	226		-	-	0.19	0.19
		Max	4,411	40	0.25	6								
3	903	Min	624	25	0.25	1	51	166	310	0.36	-	-	0.14	0.07
		Max	4,411	40	0.46	3								
4	743	Min	100	45	0.05	1	0	17	211		-	-	0.03	0.04
		Max	4,400	55	0.44	4								
5	703	Min	500	25	0.47	1	17	141	334	-	-	0.37	0.09	0.01
		Max	4,400	45	2.01	1								
6	624	Min	453	25	0.46	2	0	216	187		-	-	0.27	0.36
		Max	4,403	40	1.01	5								
7	571	Min	110	45	0.46	2	0	0	309			-	0.04	0.09
		Max	4,375	60	3.88	4								
8	820	Min	4,413	20	0.02	1	14	138	330	-	0.43	0.48	0.22	0.19
		Max	7,477	35	0.25	5								
9	600	Min	4,450	30	0.25	1	47	234	107	-	-	-0.69	0.06	0.11
		Max	7,463	35	0.49	3								
10	623	Min	4,416	40	0.07	1	0	80	173		-	-0.30	0.09	0.12
		Max	5,418	55	0.35	4								
11	502	Min	5,425	40	0.06	2	0	14	129		1.31	-	0.07	0.06
		Max	7,449	40	0.20	3								
12	523	Min	5,425	40	0.20	1	0	13	259		-	-	0.01	-0.01
		Max	7,475	40	0.35	3								
13	706	Min	5,425	45	0.09	1	0	56	304		-	-	0.05	-0.02
		Max	7,469	55	0.35	4								
14	600	Min	4,425	40	0.35	1	0	26	221		-	-	0.07	0.07
		Max	7,475	45	0.49	3								
15	552	Min	4,450	25	0.49	1	109	100	234	-	-	-	0.09	0.06
		Max	7,464	40	2.01	1								
16	683	Min	4,438	30	0.49	2	0	142	194		-	-0.68	0.08	0.08
		Max	7,438	40	1.01	3								
17	767	Min	4,418	45	0.49	1	0	27	308		-	-	0.05	0.00
		Max	7,475	55	3.88	3								
18	692	Min	7,488	30	0.02	1	31	131	237	1.48	2.44	-	0.17	0.11
		Max	14,847	35	0.49	3								
19	820	Min	7,488	40	0.06	1	0	27	294	-	0.73	-0.27	0.03	0.03
		Max	8,675	45	0.50	3								
20	571	Min	8,688	40	0.08	1	0	1	161			0.92	0.07	0.22
		Max	16,650	45	0.50	3								
21	582	Min	7,500	35	0.50	1	34	64	204	-	3.74	1.81	0.29	0.26
		Max	19,480	50	1.01	3								

Note: Grey areas show non-occurrence of lane width in the dataset; all lane widths are significant at 5 percent level; - represents lane width is insignificant at 5 percent level.

Table 3 Injury and non-injury treed regression characteristics and lane width linear regression coefficients

		Injury					Non-Injury					
Node		1	2	3	4	5	1	2	3	4	5	6
#of Obs.	Training	1129	940	352	369	198	1431	149	667	167	433	141
	Testing	269	252	83	83	59	374	36	141	52	114	29
Avg. crash /node		0.686	1.130	2.051	2.176	4.040	2.502	6.188	4.112	7.090	6.989	11.369
SL	Min	20	20	25	40	30	20	25	25	30	30	35
	Max	60	55	55	50	35	55	45	60	55	45	50
NTL	Min	0	1	1	1	1	0	3	0	1	1	1
	Max	6	5	3	3	3	2	6	5	3	3	3
AADTLn	Min	48	4400	4400	7340	7313	48	48	486	6146	7298	7299
	Max	4381	7298	7310	17783	13700	7295	7277	6125	7283	16500	17783
SGLT	Min	0.028	0.028	0.500	0.063	0.025	0.028	0.065	0.461	0.461	0.025	0.499
	Max	3.878	0.499	2.007	1.007	1.013	0.460	0.377	3.878	1.204	0.499	1.013
Count	LW9	19	24	19	0	13	31	1	17	13	6	7
	LW10	192	117	51	9	31	186	41	123	10	27	13
	LW11	418	324	124	116	79	475	54	271	62	147	52
Coefficient	LW9											
	LW10		0.244			3.703	1.112	1.377			5.67	14.94
	LW11	0.119			0.705							5.69
LW10 CI	Min		-0.043			2.23	0.69	-0.18			2.71	9.485
	Max		0.531			5.17	1.53	2.93			8.62	20.405
LW11 CI	Min	-0.03			0.106							2.097
	Max	0.26			1.305							9.287
R^2 Training		0.083	0.080	0.091	0.081	0.339	0.155	0.437	0.161	0.249	0.137	0.421
R^2 Testing		0.106	0.073	0.100	0.091	0.188	0.129	0.721	0.164	0.186	0.132	0.240
Global R^2 Training		0.3					0.36					
Global R^2 Testing		0.33					0.37					
CART R^2 Training		0.16					0.18					
CART R^2 Testing		0.17					0.2					

Reference

1. Zegeer, C.V., D.W. Reinfurt, J. Hummer, L. Herf, and W. Hunter. *Safety Effects of Cross-Section Design for Two-Lane Roads*. Publication FHWA-RD-87-008. FHWA, U.S. Department of Transportation, 1987.
2. Wang, J., Hughes, E. W., and Stewart, R., Safety Effects of Cross-section Design on Rural Multilane Highways. *Transportation Research Circular*, (E-C003): 18:1-14. 1998.
3. Hardwood, D., Rabbani, E., Ricard, K., McGee, H., and Gittings, G. "System-wide Impact of Safety and Traffic Operations Design Decisions for 3R Projects," NCHRP Report 486, Transportation Research Board, Washington, D.C. (2003).
4. Harkey, D.L., S. Raghavan, B. Jongdea, and F.M. Council . Crash Reduction Factors for Traffic Engineering and ITS Improvement. Highway Safety Research Centre, University of North Carolina, Raleigh, NC. (2007).
5. Harwood, D. W., Effective Utilization of Street Width on Urban Arterials. *Transportation Research Board*, NCHRP-330, Washington, D.C., 1998.
6. Potts, B. I., Harwood, W. D., and Richard, R. K., Relationship of Lane Width to Safety for Urban and Suburban Arterials. In *TRB 2007 Annual Meeting* CD-ROM. Transportation Research Board, Washington, D.C., 2007.
7. American Association of State Highway and Transportation Officials, A Policy on Geometric Design of Highways and Streets, Washington, DC 2001.
8. American National Standards, Manual on Classification of Motor Vehicle Traffic Accidents, 7th Edition, ANSI D16.1-2007.
9. Breiman, L., Friedman, J., Olshen, R., and Stone, C., Classification and Regression Trees. Wadsworth, Belmont, CA, 1984.
10. Wikipedia, Decision tree learning. URL: http://en.wikipedia.org/wiki/Decision_tree_learning. Accessed November, 2014.