A novel approach for analyzing severe crash patterns on multilane highways

Anurag Pande, Mohamed Abdel-Aty

A B S T R A C T

This study presents a novel approach for analysis of patterns in severe crashes that occur on mid-block segments of multilane highways with partially limited access. A within stratum matched crash vs. non-crash classification approach is adopted towards that end. Under this approach crashes serve as units of analysis and it does not require aggregation of crash data over arterial segments of arbitrary lengths. Also, the proposed approach does not use information on non-severe crashes and hence is not affected by under-reporting of the minor crashes. Random samples of time, day of week, and location (i.e., milepost) combinations were collected for multilane arterials in the state of Florida and matched with severe crashes from the corresponding corridor to form matched strata consisting of severe crash and non-crash cases. For these cases, geometric design/roadside and traffic characteristics were derived based on the corresponding milepost locations. Four groups of crashes, severe rear-end, lane-change related, pedestrian, and single-vehicle/off-road crashes, on multilane arterials segments were compared separately to the non-crash cases. Severe lane-change related crashes may primarily be attributed to exposure while single-vehicle crashes and pedestrian crashes have no significant relationship with the ADT (Average Daily Traffic). For severe rear-end crashes speed limit, ADT, *K*-factor, time of day/day of week, median type, pavement condition, and presence of horizontal curvature were significant factors. The proposed approach uses general roadway characteristics as independent variables rather than event-specific information (i.e., crash characteristics such as driver/vehicle details); it has the potential to fit within a safety evaluation framework for arterial segments.

1. Introduction

This study presents a novel approach to assess patterns in severe crashes on multilane arterial segments. Multilane arterials for this study are defined as highways that have (i) at least two lanes in each direction and (ii) are *not* limited access facilities (i.e., no expressways/freeways). These arterials consist of signalized/unsignalized intersections joined by mid-block segments. Assessment of safety on multilane arterials (or any other roadway for that matter) is traditionally based on two broad criteria, namely, crash counts or crash rate (i.e., counts normalized by vehicle-mile travelled), and crash injury severity. Crash counts are traditionally estimated using negative binomial regression models (e.g., Abdel-Aty, 2003; Knuiman et al., 1993). Severity based analysis, on the other hand, relies on classification of crash outcomes in terms of levels of injury severity (e.g., Yau, 2004; Abdel-Aty, 2003). Injury severity outcomes may be formulated as binary (severe vs. non-severe; e.g., Yau, 2004) ordi-

nal (Abdel-Aty, 2003) or multinomial target variable (Shankar and Mannering, 1996).

The objective of this study is to identify traffic and highway design parameters significantly associated with severe crashes on *segments* of multilane arterials. Towards that end an alternative to these traditional approaches is proposed. In crash frequency analysis the dependent variable, i.e., frequency of crashes, is calculated by aggregating the crash data over specific time periods (months or years) and locations (Golob et al., 2004; Abdel-Aty and Pande, 2007). In terms of locations, signalized or unsignalized intersections are well defined entities within the roadway infrastructure. The individual intersections act as logical units for aggregating the crash data in the form of crash frequencies (e.g., Wang and Abdel-Aty, 2006). Crash frequency analysis for roadway segments, on the other hand, requires aggregation of crash data over segment(s) of certain length(s). For example, Caliendo et al. (2007) divided each direction of a four-lane divided arterial into segments with constant horizontal curvature and longitudinal slope. Donnell and Mason (2006) analyzed the crash frequencies for ½-mile segments. The selection of the length(s) of segments used to aggregate the crash data is arguably arbitrary. The results obtained from crash frequency analysis are likely to be sensitive to the lengths over which data are analyzed. The comparisons of non-crash data with

crash data proposed in this study allow for using crashes themselves as the unit of analysis without having to select arbitrary segment length over which to aggregate crash data.

Studies analyzing severity outcomes of crashes (e.g., Yau, 2004; Abdel-Aty, 2003) have used crashes as units of analysis to assess “given a crash has occurred how severe would it be?” However, comparative analysis between severe and non-severe crashes is affected by under-representation of the least severe crashes in the documented crash data (Abdel-Aty and Keller, 2005). Furthermore, as Milton et al. (2008) have pointed out; the insights provided by these models have limited application in safety improvement programs since these models require *event-specific* explanatory variables to produce useable estimates of injury severity outcomes. The approach proposed herein has the advantages of the methodology used by Milton et al. (2008) as it uses *non-event* (i.e., *crash*)-specific factors affecting severe crashes on roadway sections. Under-representation of non-severe crashes in the database is a non-issue for the proposed approach as it relies on comparisons only between *severe* crashes and *non-crash* cases. For binary classification of *severe* crashes vs. *non-crash* cases the target variable is 1 for severe crashes and 0 for the non-crash cases. The procedure to extracting crash and non-crash data for the corridors can also be easily implemented as discussed in the next section.

2. Data extraction and exploration

The analysis presented herein is based on 6857 crashes (reported during the year 2006) from 151 multilane arterial corridors of length between 5 and 15 miles in the state of Florida. These corridors consist of signalized intersections as well as access points without signal control (i.e., unsignalized intersections). The analysis focuses on segment crashes that are not affected by the intersecting traffic streams and may be attributed only to the segments of corresponding roadways. To compare the crashes with, a sample of non-crash cases is generated by randomly selecting milepost location, time of day, and day of week combination on these corridors. These randomly selected time and locations on the arterials (when no crash was observed) are then used to form matched strata of severe crashes and non-crash cases for each of the 151 arterials. Detailed procedure for data collection and preparation is described in this section.

2.1. Crash data collection

As mentioned earlier, the severe crashes attributable to arterial segments are the focus of this investigation. These segment crashes are defined as the crashes that are not related with the traffic on the intersecting streets. In other words, vehicles involved in the crash were neither coming from nor going to the intersecting roads/driveways. First of all crashes with first harmful event characterized as “Collision with Motor Vehicle in Transport (Left-turn)” and “Collision with Motor Vehicle in Transport (Right-turn)” were eliminated from the sample. The next step was to identify which of the *remaining* crashes may be attributable to arterial segments and *not* to (signalized or unsignalized) the intersections. A detailed review of crash reports revealed that the variable “Site location” from the crash reports by itself was a weak indicator for the same. It was observed that it is possible for a crash to be not attributable to a signalized intersection even if it may have occurred very close to one. In fact, “traffic control” in combination with the “site location” did a superior job in attributing crashes to one of the three roadway elements (i.e., segments, signalized intersections, and unsignalized intersections) associated with the event of crash. Also, crashes with “Collision with Motor Vehicle in Transport (Angle)” as the identified first harmful event were excluded from the sample if the contribut-

ing cause for the crash was noted as “Improper turn” or “Failed to yield Right of Way”. These crashes are caused by vehicles making right/left turn turns and/or by vehicles that fail to yield right of way to through vehicles. Crashes remaining in the database are not attributable to signalized/unsignalized intersections and may be solely attributed to the segments of the multilane highways.

The segment crash data consisted of 6857 events with 10.69% of them resulting in fatal or incapacitating injury. The remaining 89.31% of the crashes were non-severe crashes. For the 6208 crashes for which type of crash (i.e., first harmful event) information was available, the breakdown of severe vs. non-severe crashes by type and their overall share in the crash data is depicted in Fig. 1. It is worth repeating that one may expect the share of non-severe crashes within each of the crash types to be even higher than shown in Fig. 1 due to the well documented problem of under-reporting of these crashes (Abdel-Aty and Keller, 2005).

In Fig. 1 the crash data are divided into five collision types namely, rear-end, single-vehicle/off-road, lane-change related, pedestrian, and head-on. This categorization is obtained by logically combining categories of “first harmful event” in the crash database. For example, crashes with first harmful events “Motor vehicle ran into Ditch/Culvert” and “Ran off-road into water” were part of the crash type Single-vehicle off-road. The five types of crashes are arranged from left to right in Fig. 1 by descending share in the overall crash data. Note that head-on collisions are rare on these multilane highways and makeup only 2% of the data even though 27% of them are severe. Collisions involving pedestrians have the highest percentage of severe crashes followed by head-on and single-vehicle/off-road crashes, respectively. Lane-change related and rear-end collisions have the least percentage of severe crashes, respectively. Lane-change related crashes consist of crashes with first harmful event as “Collision with Motor Vehicle in Transport (Sideswipe)” and “Collision with Motor Vehicle in Transport (Angle)” where the contributing cause is neither “Improper turn” nor “Failed to yield Right of Way”. Hence, we are considering only the angle crashes attributable to the arterial segments, which by definition are not affected by traffic streams (either from or turning on to) intersecting roadways. The authors postulated that these crashes would never be right angle crashes. Therefore, the crashes for which the first harmful event has been noted as “Collision with Motor Vehicle in Transport (Angle)” (by the law enforcement personnel on crash site) are essentially lane-change related crashes. This postulation was verified by manually reviewing 70 randomly selected crash reports for such crashes.

2.2. Extraction of non-crash cases

A sample of non-crash cases has been used in the analysis which acts as control within strata defined by the corridors. These non-crash cases were drawn randomly from each corridor. The year 2006 may be divided into 35,040 15-min periods ($4(15\text{-min periods per hour}) \times 24\text{ h} \times 365\text{ days} = 35,040\text{ 15-min periods}$), which would be the number of options available to choose the “time of non-crash”. Similarly, pool of possible milepost locations for each corridor consisted of mileposts starting at beginning milepost and culminating at the ending milepost with an increment 0.001 mile. Once the time and locations for non-crash cases were available; Roadway Characteristics Inventory database was then used to derive roadway characteristics for these non-crash cases based on the milepost. The increment level of 0.001 mile was chosen since the crash location milepost in the crash database also reported with three decimal places.

With the selected increments of time and location for non-crash cases consider a corridor with beginning milepost 0.0 and ending milepost 6.0, there would be 210,240,000 ($35,040 \times (6/0.0001) = 210,240,000$) options to select (day, time, and

Fig. 1. Severity of crashes by collision type and their share in the overall crash data.

location of) non-crash cases. 0.5% of these non-crash cases were drawn randomly from the available options for each corridor. These selected non-crash cases for each of the corridors are matched with the severe crashes from the same corridors to create 151 strata for within stratum matched sampling framework. Details of the sampling and modeling procedure are described later in the paper. It is worth mentioning that this process of creating random time/day/location sets for non-crash events may be replicated easily for any highway. Based on the set of locations selected; the relevant highway design data may be derived from the corresponding roadway characteristics inventory maintained by most agencies.

2.3. Traffic/geometric information for crash and non-crash cases

The next step was to extract geometric design features such as the curvature, median type, sidewalk, etc. for crash and non-crash cases. These relevant variables were extracted from the Roadway Characteristics Inventory (RCI) database (RCI Features and Characteristics Handbook, 2001). The extraction of traffic/geometric information was based on the milepost locations and the roadway IDs for the arterial corridors. For crashes, it was the actual milepost location of the crash from the FDOT crash database and for non-crash cases it was assigned using the procedure described in the last section.

RCI database provides information on Florida's state maintained road network indexed by data segments. RCI features are listed in the handbook (RCI Features and Characteristics Handbook, 2001) published by FDOT (Florida Department of Transportation) and Table 1 details the relevant variables extracted from this database. Note that most of the variables tabulate are not in exactly the same form as the original database. The original categories of the variables in the RCI database were combined to have a meaningful sample size within crash and non-crash data. Table 1 also provides the percentages of crash and non-crash cases for all categories of the variables listed. It may be compared to the overall percentage of crash and non-crash cases in the database (found in the header row of Table 1) to get a descriptive estimate of the variables or categories associated with crash occurrence on multilane highways.

ADT, percentage of trucks (*T*-factor), and *K*-factor (design hour volume as a percentage of ADT) were measured on a continuous scale in the original database. For the analysis these variables were divided into categories since we do not expect their relationships with severe crash occurrence to be monotonous in nature. As we shall observe later, the results actually substantiate the reasoning behind the categorization. The categorization of the continuous variable ADT is such that the four resulting categories have same number of observations. *T*-factor and *K*-factor were divided into three categories such that the resulting categories have same num-

ber of observations. This method of categorization makes it more likely that resulting coefficients for all categories (in the models estimated) of a variable are equally reliable.

Time of crash (and non-crash cases), along with day of week, was combined into one variable representing day of week and time of day. The four categories of this variable include weekday morning peak hour, weekday afternoon peak hour, Friday/Saturday night, and other off-peak periods. Note that the weekend night times was separated from the other off-peak periods because of the increased likelihood of alcohol impaired driving. It is worth mentioning that this time of day information could be derived using increments larger than 15 min (e.g., 60 min) when creating the pool of all possible values for "time of non-crash".

The RCI database also provides pavement condition information in the form of Present Serviceability Rating (PSR) based on the AASHTO Road Test (RCI Features and Characteristics Handbook, 2001). The final variable used for pavement condition had the following three categories: very poor/poor pavements ($PSR < 3.00$), fair pavements ($3.0 \leq PSR < 3.9$), and good/very good pavements ($PSR \geq 4.0$).

Three binary variables representing the presence of horizontal curvature, roadside parking, and crash attenuators were also used in the analysis along with type of median and presence/width of sidewalk. Median types were consolidated into nine categories shown in Table 1. The presence and width of the sidewalk was represented by the variable "sidewalk" and its categories are also listed in Table 1.

One of the variables considered but not included in the analysis was Sun glare. It was available for crash cases from the event reports but was missing for the non-crash cases. Presence/possibility of Sun glare could be 'derived' from the location and time of day for the non-crash cases. It was, however, observed, that the total number of crashes for which "Glare" was noted as a vision obstruction were only 19. To further investigate this factor the crash reports with "Other (Explain)" as the vision obstruction were read one by one. It was found that among 99 such crashes only two were affected by Glare. Sample size of 21 out of the total 6857 crashes was not sufficient to examine Sun glare as a factor. The variables shown in Table 1 are not event-specific characteristics (such as driver characteristics, seat-belt use, etc.) which, as Milton et al. (2008) argued, allows for a more general, non-event-specific interpretation of factors.

3. Modeling methodology

For modeling crash vs. non-crash cases, within stratum matched case-control sampling is implemented. The purpose of the matched sampling based analysis is to explore the effects of variables of interest while controlling for the confounding variables through

Table 1
Variables used in the analysis.

Variable description	Categories	Percentage of crash cases (10.52) (%) ^a	Percentage of non-crash cases (89.48) (%) ^a
Posted speed limit	Speed limit < 40 MPH	10.44	89.56
	40 ≤ speed limit < 50 MPH	15.13	84.87
	50 ≤ speed limit < 60 MPH	7.45	92.55
	Speed limit ≥ 60 MPH	3.91	96.09
ADT (annual daily traffic)	ADT < 14,900	2.49	97.51
	14,900 ≤ ADT < 26,500	6.27	93.73
	26,500 ≤ ADT < 40,000	12.57	87.43
	ADT ≥ 40,000	20.77	79.23
Average K-factor	K-factor < 9.35	13.53	86.47
	9.35 ≤ K-factor < 10.14	12.67	87.33
	K-factor ≥ 10.14	5.86	94.14
Average truck factor	T-factor < 4.84	13.33	86.67
	4.84 ≤ T-factor < 8.75	12.26	87.74
	T-factor ≥ 8.75	5.97	94.03
Combination of day of week and time of day	Afternoon peak weekday	20.82	79.18
	Friday or Saturday Night	9.74	90.26
	Morning peak weekday	13.54	86.46
	Off-peak	9.17	90.83
Pavement condition (PSR)	PSR < 3.0 very poor/poor	5.82	94.18
	3.00 ≤ PSR < 3.90 fair	10.37	89.63
	PSR ≥ 4.00 good/very good	11.18	88.82
Median type	Two-way left-turn lane (TWLTL)	11.60	88.40
	Grass/lawn	18.87	81.13
	Guard rail	17.85	82.15
	Barrier other than guard rail	6.72	93.28
	Canal or ditch	9.78	90.22
	Curb < 6 in.	17.19	82.81
	Curb ≥ 6 in.	14.28	85.72
	Paved not for travel	8.62	91.38
	No median	2.47	97.53
Sidewalk	No sidewalk,	7.34	92.66
	Sidewalk ≤ 6 ft	13.81	86.19
	Sidewalk > 6 ft	14.38	85.62
Presence of traffic crash attenuators	Yes	14.56	85.44
	No (binary)	10.48	89.52
Presence of on-street parking	Yes	14.50	85.50
	No (binary)	8.11	91.89
Presence of horizontal curvature	Yes	4.54	95.46
	No (binary)	10.73	89.27

^a Overall percentage of crash and non-crash cases in the database.

the design of the study (Abdel-Aty et al., 2004). It is used in this study to compare sample of crashes with non-crash cases within the data stratified by the corridors. Crash based case-control and cohort method has also been used to estimate Accident Modification Factors (AMF) by Jovanis and Gross (2007). Harb et al. (2008) used the $m:n$ matched sampling to compare work-zone with non-work-zone crashes.

Under the proposed matched study design, crash and non-crash cases from each of the 151 arterials form an individual stratum. Each stratum consists of severe crashes and non-crash cases from the corresponding corridor. The sampling is referred as $m:n$ matching and each corridor (i.e., stratum) can have different number of crash (m) and non-crash cases (n). Within stratum differences between characteristics of crash and non-crash cases may then be utilized for estimation of statistical model(s) for the binary target variable.

In the present analysis there would be 151 strata (equal to the number of corridors) with m crashes and n non-crash cases in each stratum j . Lets stipulate $p_j(x_{ij})$ to be the probability that i th observation in the j th stratum is a crash with $x_{ij} = (x_{1ij}, x_{2ij}, \dots, x_{kij})$ being the vector of k variables $x_1, x_2, \dots, x_k$; $i = 0, 1, 2, \dots, m+n-1$; and $j = 1, 2, \dots, N$. The probability $p_j(x_{ij})$ of an observation being a crash may

be modeled as follows:

$$\log it(p_j(x_{ij})) = \alpha_j + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} \quad (1)$$

The intercept term α_j is different for different strata. It summarizes the effect of variables used to form strata on the probability of crash. In order to take account of the stratification in the analysis of the observed data, one constructs a conditional likelihood. The likelihood function $L(\beta)$ is independent of the intercept terms $\alpha_1, \alpha_2, \dots, \alpha_N$. Hence, the effects of matching variables cannot be estimated using Eq. (1) even as it is not required to estimate the relative odds of crash occurrence. Also, note that the sampling strategy *cannot* be used to estimate crash probabilities and only relative odds of crash occurrence may be estimated. However, the values of the β parameters that maximize the likelihood function are in fact the estimates of β coefficients in Eq. (1). Further details on the derivation of maximum likelihood function may be found in Collett (2003).

The first step in the analysis was to estimate simple (with only one covariate) models. For estimating simple models two comparisons were considered: (1) crash vs. non-crash and (2) severe crash vs. non-crash cases. The severe crashes are defined as the crashes with

incapacitating injury and/or fatal injury. The other injury severity levels including possible and non-incapacitating injuries are considered to be non-severe crashes. It is worth mentioning that the first of these two comparisons is problematic since the sample consisting of all crashes would be biased with least severe crashes under-represented due to the reporting bias mentioned earlier (Abdel-Aty and Keller, 2005). Therefore, the preliminary analysis presented herein as well as the multivariate analysis was limited to the comparison of severe crashes vs. non-crash cases. Logistic regression based on within stratum matched sampling of crash and non-crash cases is used as the statistical tool. The simple (one covariate) models are estimated using the SAS procedure TPHREG (SAS/STAT® 9.1 User's Guide, 2004). Note that the categorized variables based on the ranges of *T*-factor, *K*-factor, and ADT along with speed limits, width of sidewalk, and pavement surface conditions are used as nominal variables (as opposed to ordinal variables). The nominal scale ensures that one is able to capture the non-monotonous nature of the relationship between these variables and crash occurrence.

4. Preliminary analysis: simple models

Table 2 shows the coefficients of the simple models for each of the variables along with corresponding *p*-values. Significantly positive (*p*-value in bold font), significantly negative (*p*-value in bold-italic font), and statistically insignificant coefficients (*p*-value in regular font) are distinguished in the table. It may be observed that sections with “40 ≤ speed limit < 50 MPH” are more likely to have severe crashes compared to sections with speed limit greater than 60 MPH (i.e., the base case). These results provide justification for using this variable on a nominal scale and not on a continuous or ordinal scale. Similarly, for severe crashes segments with ADT from 26,500 through 40,000 are not significantly different from segments with ADT greater than 40,000. Hence, it seems that the severe crash occurrences may not be directly attributable to exposure. It should be noted that while *K*-factor is not a factor when comparing severe crashes with non-crash cases it might be related to certain types of severe crashes. Based on the coefficients for the two levels of the nominal variable representing the truck factor (compared to the base case “*T*-factor ≥ 8.75”) the segments with *T*-factor < 4.84% are more likely to have severe crash. It may be related with the fact that multilane arterials with low truck factor are expected to have higher pedestrian traffic, which increases the likelihood of pedestrian related crashes (which in turn have disproportionately high severe to non-severe crash ratio). This inference, however, further highlights the need for segregating crash data by type. Also, of note is the fact that Friday/Saturday nights have the most significant coefficient compared to all other categories of the variable representing day of week/time of day.

Corridors without a sidewalk are less likely to have severe crashes. The results also show that the improved ride quality (represented by levels of pavement condition) improves safety with pavements classified as very good/good having a significantly negative coefficient compared to poor/very poor pavements. Also, for multilane arterials with attenuators, at 90% confidence interval, attenuators' presence has no significant effect on the occurrence of severe crashes.

The median types are divided into nine types with the no median being the base case. The coefficient for grassed median is almost as significant as the curbs. The results are reflective of the fact that rural sections (with grass medians) are more likely to have severe crashes. The next section outlines some of the issues with this preliminary analysis and then expands on it to estimate logistic regression models with multiple covariates for severe crash data segregated by type. The extended analysis is useful for drawing more precise inferences.

5. Analysis with multiple covariates

The sample used for preliminary analysis consisted of all different types of severe crashes. The different group of severe crashes may in fact have different traffic/geometric design variables associated with them. The analysis in this section is carried out for severe crashes segregated by crash type. Among the severe crashes, single-vehicle/off-road crashes were in a plurality with more than 35% of the data followed by rear-end, pedestrian related, and lane-change related crashes. The head-on collisions were less than 4% of the severe crashes. Head-on crashes were not used in the analysis since the absolute number of severe crashes for this type was too low to provide any meaningful sample size for analysis. Hence, the analysis was limited to four different models with severe crashes of four types individually compared with non-crash cases. These crash types are rear-end, lane-change related, pedestrian related, and single-vehicle/off-road crashes. Potential independent variables used in the analysis were shown in Table 1. Backward variable selection procedure was used for identifying the most significant variables from the initial set of potential variables.

Table 3 shows the results of backward variable selection procedure with within stratum matched case-control logistic regression. For backward selection procedure parameters for the complete model (with all potential variables) are first estimated. Then the results of the Wald test for individual parameters are examined and the least significant effect, which does not meet the *p*-value criterion for keeping a variable in the model, is removed. The removed effect is permanently excluded from the model and the process is repeated until no other effect may be removed based on the *p*-value threshold (*p*-value > 0.15) (SAS/STAT® 9.1 User's Guide, 2004). The backward selection model is preferred because it starts with the complete set of variables included in the model. For more details on the backward variable selection and its advantages one may consult Vittinghoff (2005).

The results in Table 3 are tabulated in eight columns with two columns each (representing coefficient and corresponding *p*-values) for the four types of crashes. Note that the rows corresponding to some of the variables show “X” since these variables (or all the categories of nominal variables) were *not* found to be significantly associated with the occurrence of corresponding severe crash type. The likelihood ratio test statistic as well as corresponding *p*-values as measures of goodness-of-fit have also been provided (at the bottom of the table) for all four models. The test statistics as well the corresponding *p*-values indicate that all models are statistically significant. Based on the *p*-values for likelihood ratio test, the model explaining severe lane-change related crashes seems to have the least ‘explanatory power’. The results tabulated in Table 3 are discussed in the following two subsections: first the significant factors associated with each of the four crash types are discussed individually and then, a discussion on differences in significance of factors among the different crash types is provided.

5.1. Significant variables for each crash type

For severe rear-end crashes speed limit, ADT, *K*-factor, time of day/day of week, median type, pavement condition, and presence of horizontal curvature were significant. Severe rear-end crashes are more likely to occur on sections with “40 ≤ speed limit < 50 MPH”. On multilane arterial sections with speed limit less than 40 MPH speeds are likely too low to have severe rear-end crashes; while on sections with speed limit greater than or equal to 50 MPH rear-end crashes are less likely to occur. Compared to the base case (“ADT ≥ 40,000”) the sections with lower ADT are more likely to observe severe rear-end crashes. However, the magnitudes of coefficients (even though they are negative for all three categories with lower ADTs) show that sections with “ADT < 14,900” (category with

Table 2
Results from the preliminary analysis (one covariate).

Objective Variables	Severe Crash vs. non-crash comparison		
	Parameter	Standard error	p-value
<i>Traffic related parameters</i>			
Speed limit<40 MPH	0.154	0.253	0.5425
40≤Speed limit<50 MPH	0.629	0.217	0.0038
50≤Speed limit<60 MPH	0.340	0.216	0.116
<i>Speed limit≥60 MPH</i>	.	.	.
ADT<14900	-1.041	0.228	<.0001
14900≤ADT<26500	-0.651	0.154	<.0001
26500≤ADT<40000	-0.187	0.124	0.1336
<i>ADT≥40000</i>	.	.	.
K-Factor<9.35	0.466	0.287	0.1053
if 9.35≤ K-Factor < 10.14	0.200	0.286	0.4851
<i>K-Factor ≥ 10.14</i>	.	.	.
T-Factor<4.84	0.383	0.176	0.0295
4.84≤ T-Factor < 8.75	0.186	0.160	0.2435
<i>T factor≥8.75</i>	.	.	.
Afternoon Peak Weekday	0.497	0.121	<.0001
Friday or Saturday Night	0.700	0.122	<.0001
Morning Peak Weekday	0.049	0.147	0.7386
<i>Off-peak period</i>	.	.	.
<i>Highway design/Pavement related Parameters</i>			
Two-way left-turn lane	1.536	0.358	<.0001
Lawn	1.847	0.345	<.0001
Guard rail	2.466	0.816	0.0025
Barrier other than guard rail	2.185	0.492	<.0001
Canal or Ditch	2.769	0.759	0.0003
Curb < 6 in	1.937	0.348	<.0001
Curb ≥ 6 in	1.900	0.358	<.0001
Paved not for travel	1.478	0.408	0.0003
<i>No Median</i>	.	.	.
No Side Walk	-0.293	0.154	0.0566
Side Walk ≤ 6ft.	0.008	0.143	0.9573
<i>Side Walk > 6ft.</i>	.	.	.
Good/very good pavement	-0.908	0.285	0.0014
Fair pavement	-0.169	0.117	0.1502
<i>Poor/very poor pavement</i>	.	.	.
Horizontal curvature	-0.142	0.306	0.6423
Attenuators	0.448	0.348	0.1977
Presence of on-street Parking	0.324	0.111	0.0035

lowest ADT) are in fact *more* likely to experience a severe rear-end crash compared to sections with “14,900 < ADT ≤ 26,000”. It indicates that occurrence of severe rear-end crash occurrences are not directly related with exposure. The results once again highlight the importance of measuring these variables on a nominal scale. The sections with “K-factor < 9.35” are more likely to have severe rear-end crashes. With a lower K-factor relatively less traffic is served during the design peak hour and hence the vehicles are more likely to interact during off-peak hours, possibly at higher speeds, leading to increased likelihood of severe rear-end crashes. In terms of severe rear-end crashes Friday and Saturday nights are statistically not different than other off-peak periods. Severe rear-end crashes are less likely to occur on the fair pavements compared to poor pavements. Note that the coefficient for good/very good pavement is also negative compared to the base case (poor/very poor pavements) even though it is not statistically significant. Presence of horizontal curvature was negatively associated with likelihood of severe rear-end crashes. Median types related with likelihood of severe rear-end crash occurrence are discussed in the next section since it requires some context from their relationship (or lack thereof) with occurrence of severe crashes of other types.

ADT, K-factor, T-factor, and pavement condition are significantly related with *severe lane-change related crashes*. As such a monotonic trend may be observed in the coefficients for the three classes of the ADT variable. It indicates that the severe lane-change related crashes on arterial segments may be explained in terms of exposure. Segments with K-factor greater than 10.14% are more likely to have severe lane-change related crashes. Severe lane-change related crashes are also more likely to occur on sections with “4.84 ≤ T-factor < 8.75”. The sections with even lower truck factor also have a negative coefficient compared to sections with “T-factor ≥ 8.75” but it is not statistically significant. The results indicate that arterial sections with higher percentage of trucks are more likely to have severe lane-change related crashes. It is an expected result since the lane-change related collisions involving large trucks are likely to be more severe. Hence, lane-change related warnings on sections with high truck traffic may be an effective countermeasure for such crashes. Fair pavements reduce the likelihood of severe lane-change related crashes and the coefficient for good/very good pavement is also negative with p-value only slightly higher than 0.10. It shows that improving pavement condition can lead to reduction in severe lane-change related crashes.

Table 3

The parameters significantly affecting severe crashes of different types.

Type of crashes	<i>Rear-end</i>		<i>Lane-change</i>		<i>Pedestrian</i>		<i>Single vehicle</i>	
Variable	Parameter	p-value	Parameter	p-value	Parameter	p-value	Parameter	p-value
Speed limit<40 MPH	0.526	0.4367	X	X	X	X	X	X
40<= Speed limit<50 MPH	0.990	0.0837	X	X	X	X	X	X
50<= Speed limit<60 MPH	0.327	0.5299	X	X	X	X	X	X
<i>Speed limit>=60 MPH</i>	.	.	X	X	X	X	X	X
ADT<14900	-0.949	0.0975	-4.898	0.0004	X	X	X	X
14900<=ADT<26500	-1.305	0.0002	-1.245	0.0321	X	X	X	X
26500<=ADT<40000	-0.533	0.0305	-0.833	0.0621	X	X	X	X
<i>ADT>=40000</i>	.	.	.	.	X	X	X	X
K-Factor<9.35	1.884	0.0018	-3.914	0.0496	X	X	X	X
9.35<= K-Factor < 10.14	0.213	0.7201	-3.636	0.0188	X	X	X	X
<i>K-Factor>= 10.14</i>	.	.	.	.	X	X	X	X
T-Factor<4.84	X	X	-0.426	0.5061	0.783	0.0524	X	X
4.84<= T-Factor < 8.75	X	X	-1.160	0.0484	0.294	0.4501	X	X
<i>T factor>=8.75</i>	X	X	.	.	.	.	X	X
Afternoon Peak Weekday	1.060	<.0001	X	X	-0.281	0.4179	0.344	0.1327
Friday or Saturday Night	0.196	0.5372	X	X	0.582	0.0225	1.061	<.0001
Morning Peak Weekday	0.588	0.0141	X	X	-1.397	0.0170	-0.101	0.7168
<i>Off-peak period</i>	.	.	X	X	.	.	.	.
TWLT	1.147	0.3164	X	X	X	X	2.250	0.0360
Lawn	3.096	0.0051	X	X	X	X	3.721	0.0003
Guard rail	3.189	0.0552	X	X	X	X	4.298	0.0013
Barrier other than guard rail	1.556	0.3128	X	X	X	X	4.565	<.0001
Canal or Ditch	4.339	0.0067	X	X	X	X	6.040	0.0034
Curb<6 in	2.837	0.0104	X	X	X	X	3.016	0.0039
Curb >=6 in	2.715	0.0153	X	X	X	X	3.040	0.0038
Paved not for travel	1.907	0.1191	X	X	X	X	2.992	0.0065
<i>No Median</i>	.	.	X	X	X	X	.	.
No Side Walk	X	X	X	X	-0.947	0.0059	X	X
Side Walk <= 6ft.	X	X	X	X	-0.190	0.4758	X	X
<i>Side Walk > 6ft.</i>	X	X	X	X	.	.	X	X
Good or very good	-0.653	0.2955	-2.045	0.1215	X	X	X	X
Fair	-0.575	0.0168	-0.910	0.0548	X	X	X	X
<i>Poor or very poor pavement</i>	.	.	.	.	X	X	X	X
Presence of horizontal curvature	-2.613	0.0551	X	X	0.751	0.0880	X	X
Attenuators	X	X	X	X	1.219	0.0471	X	X
Parking	X	X	X	X	0.521	0.0385	0.673	0.0027
<i>Likelihood ratio test for goodness-of-fit</i>	119.086	<.0001	34.592	0.0005	57.354	<.0001	88.849	<.0001
Base case categories	Binary variable	Statistically significant at 90% C.I.			X=Variable not included based on Backward Selection			

For pedestrian related severe crashes on arterial segments T-factor, time of day/day of week, along with presence of sidewalk, attenuators, and roadside parking were significant factors. Segments with very low truck traffic are more likely to have pedestrian related crashes, since the sections with high pedestrian traffic are expected to have very little to no large trucks. This is why the relationship between crash likelihood and T-factor is not monotonous. The corridors with " $4.84 \leq T\text{-factor} < 8.75$ " and " $T\text{-factor} \geq 8.75$ " are not statistically different from each other, while the corridors with " $T\text{-factor} < 4.84$ " are significantly more likely to have pedestrian related severe crashes. Sidewalks greater than or less than 6 ft are not statistically different in terms of occurrence of severe pedestrian crashes. It indicates that widening the sidewalk may not lead to a reduction in risk of severe pedestrian related crashes. However, the roadways with no sidewalk are in fact less likely to have these crashes likely due to low pedestrian traffic. Presence of roadside parking is significantly related to increased likelihood of pedestrian related crashes. With roadside parking one expects significant number of mid-block road crossings/pedestrian traffic and hence increased likelihood of pedestrian related severe crashes. As expected, pedestrian related crashes are also likely to occur during Friday/Saturday nights. Significantly positive coefficients for presence of attenuators and horizontal curve are explained in the

next section since the relevant discussion requires the context of insignificance of these parameters in the other three models.

For severe single-vehicle/off-road crashes time of day/day of week, median type, and presence of parking are significant. Friday/Saturday nights have significantly higher likelihood of severe single-vehicle crashes compared to other periods of the day. All eight median types are significantly more crash prone compared to no median. Canal or ditch as median increase the likelihood of severe single-vehicle crashes and have the largest coefficient. "Barrier other than guard rail" has the most significant coefficient in terms of the (smallest) p-value. It indicates that the presence of median barriers other than guard rail may increase the likelihood of severe crashes. The finding appears to be consistent with Elvik (1995) who noted that median barriers (other than guard rail) lead to a 30% increase in crash rate without a corresponding reduction in severity given a crash has occurred. It is also worth mentioning that presence of attenuators is not a significant factor for severe single-vehicle/off-road crashes. A significant proportion of severe single-vehicle crashes involve hitting roadside signs, and roadside objects. In the area where roadside parking is present one is more likely to find and hit such objects and hence presence of roadside parking is positively related with the severe crashes involving single vehicle.

Fig. 2. Comparison between crash types based on chi-sq. statistic corresponding to coefficients for speed limit, ADT, and presence of horizontal curve.

5.2. Differences among crash types

It is interesting to note that none of the factors are significantly associated with *all four* groups of crashes. Even the categories of the variable day of week/time of day, which is significantly related to three types of crashes, have widely varying coefficients. In this section we discuss the differences in coefficients of the same variables for different crash types. Figs. 2–5 compare different crash types by plotting a combination of the sign (positive above the x-axis and negative below) and strength of the coefficients (in terms of the chi-sq. test statistic value corresponding to the coefficients) for each variable.

In Fig. 2 it may be observed that speed limit on the arterial segments is not a significant factor for any group of severe crashes except for the rear-end crashes. ADT is *not* a significant factor for pedestrian related and single-vehicle/off-road crashes. Since ADT has an effect on inter-vehicle interactions it is reasonable that this variable only affects the severe crash types involving more than one moving vehicles. The coefficients for three categories of the ADT provide an interesting contrast between rear-end and lane-change related crashes. The contrast is clearly visible in Fig. 2. With increasingly negative coefficients for the three categories with lower ADT; it is apparent that severe lane-change related crashes on arterial segments are actually better explained by exposure compared to the severe rear-end crashes.

Most drivers drive slower on the curved sections which leads to the presence of horizontal curve either not being significant (for

severe lane-change related and single-vehicle crashes) or even negatively related with likelihood of severe rear-end crashes. However, slower speeds do not reduce the severity of pedestrian crashes. Therefore, presence of horizontal curvature is positively related with likelihood of severe pedestrian related crashes.

The contrast between coefficients of *T*-factor for pedestrian and lane-change related severe crashes is interesting (Fig. 3). While corridor sections with lowest percentage of trucks (*T*-factor < 4.84%) are *more* likely to have severe pedestrian related crashes; the sections highest percentage of trucks (*T*-factor $\geq 8.75\%$) are more likely to have severe lane-change related crashes. The former is likely related with higher pedestrian exposure on arterials with low truck traffic; while the latter could be the basis for warning motorists about being cautious while changing lanes on sections with high *T*-factor.

Note that with poor/very poor pavements as the base case both remaining categories have a negative coefficient indicating that improving pavement condition may actually reduce the likelihood of both severe rear-end and lane-change related crashes (Fig. 4). One may suspect that improved ride quality would increase the travel speed thereby increasing the likelihood of severe crashes. That concern, however, is somewhat alleviated by the fact that that pavement condition is *not* significant for severe single-vehicle/off-road crashes. Presence of crash attenuators is a significant factor associated with severe pedestrian related crashes. It indicates that attenuators are installed at high crash risk locations but since these attenuators can only reduce the severity of impact for the vehi-

Fig. 3. Comparison between crash types based on chi-sq. statistic corresponding to coefficients for *K*-factor, *T*-factor and presence/width of sidewalk.

Fig. 4. Comparison between crash types based on chi-sq. statistic corresponding to coefficients for day of week/time of day, pavement condition, presence of attenuators, and roadside parking.

Fig. 5. Comparison between crash types based on chi-sq. statistic corresponding to coefficients for median type.

cles; they are unlikely to reduce the severity of pedestrian related crashes. It explains their significantly positive association with likelihood of severe pedestrian crashes while no significant association with severe crashes of other three types. The results show that in addition to the crash attenuators some countermeasures for pedestrian related crashes also need to be considered.

Median type is not a significant factor for severe lane-change related or pedestrian crashes, which is why Fig. 5 only shows bars corresponding to rear-end and single-vehicle/off-road crashes. Among the severe rear-end crashes barrier other than guard rail are not significantly different from the roadway sections without a median. However, for severe single-vehicle crashes barrier other than guard rail is the most significant category. Sections with paved median not for travel and TWLTL are not significantly associated with severe rear-end crashes but are more likely to have severe single-vehicle/off-road crashes. Sections with lawn/grass median and with canal and ditch are significantly associated with both of these groups of severe crashes. Their association with severe rear-end crashes is likely explained by the fact that these medians are generally found on sections with high travel speeds where drivers are likely to be caught unaware of the traffic ahead. On the other

hand these sections are *also* prone to severe single-vehicle/off-road type collisions either because of drivers trying to avoid a rear-end collision and/or losing control of the vehicles due to excessive speeds.

6. Concluding remarks

This study provides a new approach for identifying significant factors related with severe crashes on segment (or mid-blocks) of multilane arterials with partially limited access. The fundamental difference between this approach and crash frequency analysis is that crash counts do not need to be aggregated over roadway segments of arbitrarily selected length value that may influence the results. The approach also improves on another traditional approach, where severity of crash (given a crash has occurred) is estimated, since our approach does not use information on non-severe crashes. Non-severe crashes are usually under-reported and hence under-represented in the data sources (Abdel-Aty and Keller, 2005).

The methodology to derive non-crash control set may be easily implemented for any highway corridor to conduct a similar analy-

sis. To get more specific insights into severe crashes on a particular corridor one may consider carrying out the analysis for a single corridor as well. It is worth mentioning that this approach is limited in that it is not suitable for comparing intersections' crash patterns. Since individual intersections provide logical units for aggregating the crash data a frequency approach is still best suited for analysis of intersection crashes. Comparisons between selected non-crash cases with the signalized (or unsignalized) intersection related crashes using the proposed approach, for example, would yield information that would mostly reflect the characteristics belonging to *locations of the signalized intersection* and not much else. However, with segment crashes the comparisons yield important geometry/traffic related parameters that significantly relate with crash occurrence on the segments.

The analysis yielded some interesting relationships between severe crash occurrence and presence of crash attenuators, times of day/day of week, and horizontal curvature. The relationship between exposure (represented by ADT) and severe lane-change related crashes on arterial sections was found to be more apparent compared to relationship between ADT and severe rear-end crashes. The information not used explicitly for the analysis is the driver related factors and a within stratum matched sampling technique was used to implicitly control for these factors. One way to account for these factors is to use induced exposure to derive the driver related factors for crash and non-crash locations and then include them as independent variables. Once one accounts for those factors explicitly the crash vs. non-crash classification approach may also be suitable for a data mining type analysis.

Acknowledgement

The authors wish to thank the Florida Department of Transportation for funding this research.

References

- Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research* 34 (5), 597–603.
- Abdel-Aty, M., Keller, J., 2005. Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis and Prevention* 37 (3), 417–425.
- Abdel-Aty, M., Pande, A., 2007. Crash data analysis: collective vs. individual crash level approach. *Journal of Safety Research*.
- Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., Hsia, L., 2004. Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record* 1897 (1897), 88–95.
- Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. *Accident Analysis and Prevention* 39 (4), 657–670.
- Collett, D., 2003. *Modelling Binary Data*. Chapman & Hall/CRC.
- Donnell, E., Mason, J., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. *Accident Analysis and Prevention* 38 (3), 590–599.
- Elvik, R., 1995. The safety value of guardrails and crash cushions: a meta-analysis of evidence from evaluation studies. *Accident Analysis and Prevention* 27 (4), 523.
- Golob, T., Recker, W., Alvarez, V., 2004. Freeway safety as a function of traffic flow. *Accident Analysis and Prevention* 36 (6), 933–946.
- Harb, R., Radwan, E., Yan, X., Pande, A., Abdel-Aty, M., 2008. Freeway work-zone crash analysis and risk identification using multiple and conditional logistic regression. *Journal of Transportation Engineering* 134, 203.
- Jovanis, P., Gross, F., 2007. Estimation of safety effectiveness of changes in shoulder width with case control and cohort methods. *Transportation Research Record: Journal of the Transportation Research Board* 2019 (1), 237–245.
- Knuiman, M., Council, F., Reinfurt, D., 1993. Association of median width and highway accident rates. *Transportation Research Record*, 70–170.
- Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. *Accident Analysis and Prevention* 40 (1), 260–266.
2001. *RCI Features and Characteristics Handbook*. Florida Department of Transportation, Tallahassee, FL.
2004. *SAS/STAT® 9.1 User's Guide*. SAS Institute Inc., Cary, NC.
- Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research* 27 (3), 183–194.
- Vittinghoff, E., 2005. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. Springer.
- Wang, X., Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis and Prevention* 38 (6), 1137–1150.
- Yau, K., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. *Accident Analysis and Prevention* 36 (3), 333–340.