Crash data analysis: Collective vs. individual crash level approach

Mohamed Abdel-Aty, Anurag Pande

Abstract

Introduction: Traffic safety literature has traditionally focused on identification of location profiles where *"more crashes are likely to occur"* over a period of time. The analysis involves estimation of crash frequency and/or rate (i.e., frequency normalized based on some measure of exposure) with geometric design features (e.g., number of lanes) and traffic characteristics (e.g., Average Annual Daily Traffic [AADT]) of the roadway location. In the recent past, a new category of traffic safety studies has emerged, which attempts to identify locations where a *"crash is more likely to occur."* The distinction between the two groups of studies is that the latter group of locations would change based on the varying traffic patterns over the course of the day or even within the hour. *Method:* Hence, instead of estimation of crash frequency over a period of time, the objective becomes real time estimation of crash likelihood. The estimation of real time crash likelihood has a traffic management component as well. It is a proactive extension to the traditional approach of incident detection, which involves analysis of traffic data recorded immediately after the incident. The units of analysis used in these studies are individual crashes rather than counts of crashes. *Results:* In this paper, crash data analysis based on the two approaches, collective and at individual crash level, is discussed along with the advantages and shortcomings of the two approaches.

1. Introduction

Traffic safety research includes an extensive array of research areas and the most prominent of them is crash data analysis. Songchitruksa and Tarko (2006) have recently pointed out some shortcomings in crash data based safety analysis and proposed other observable traffic characteristics, more frequent than crashes, as an alternative. Nonetheless, analysis of crash data remains the most widely adopted approach to assess safety of a transportation facility (e.g., freeways, arterials, intersections).

The conventional approach has been to establish relationships between crash frequency and the traffic characteristics, environmental conditions, and geometry of the roadway. This study is based on the premise that crashes are caused due to bad decisions made by the driver in an environment resulting from surrounding traffic conditions and the geometric design created by the engineer. The influence of geometric design on the likelihood of a driver making bad decision has been well documented in the traditional traffic safety literature. This direction of research is helpful in making decisions in such things as posting warning signs on roadway sections.

On the other hand, the attention given to the surrounding traffic conditions immediately preceding crash occurrence has almost been non-existent. The measures of traffic conditions used commonly in the literature are Average Annual Daily Traffic (AADT). AADT is a measure that is recorded by most agencies around the country/world and is available for all roadway sections and provides a measure of exposure for the particular roadway section. Crash frequency analysis based on AADT is an aggregate, cumulative, or collective way to look at the crash data where frequency of crashes is calculated by aggregating the crash data over specific time periods (months

or years) and locations (specific roadway sections; Golob, Recker, & Alvarez, 2004a).

In the recent past a tremendous growth has been observed in traffic management and information systems, especially on uninterrupted flow facilities. A critical part of these systems is the traffic surveillance apparatus that continuously records the traffic data. Due to recent advances in capabilities to collect and store the data through underground sensors, these data are available for many freeways. Availability of these data has inspired a new series of studies in traffic safety in which traffic conditions right before the crashes may be examined to detect patterns that commonly occur before crashes. It is no longer mandatory to combine the crash data in the form of crash frequency since traffic conditions preceding each crash are available as well.

These studies are also part of a new trend in the area of freeway traffic management. Before these studies, traffic management research had been focused on incident detection. The approach to incident detection was to analyze traffic data after several incidents to develop models that can then separate the real-time traffic conditions resulting from incidents from free-flow and/or recurring congestion. Advancements in mobile phone usage and video surveillance technology eventually forced the traffic management authorities to look for more proactive solutions. This has also contributed to some of the more recent studies related to freeway safety research. In this paper we summarize the traditional and the more recent approach to crash data analysis as well as their differences, advantages/disadvantages of each approach, and conclusive remarks on possible implications for the future of traffic safety research.

2. Collective approach to traffic safety

As discussed, the aggregate or collective approach to crash data analysis is characterized by crash frequency modeling. Crash frequency modeling enumerates the relationship between observed crash counts and existing geometric, roadway, and traffic conditions on a given stretch of a roadway. A sample format of the data for this type of analysis is shown in Table 1. It may be observed that crashes that occur in the same section of the roadway are combined in the form of their frequency, which in turn would be used as the dependent variable in the analysis. Also, note that the independent variables shown in the table are just examples and different studies have used varied sets of variables depending upon the scope of research. Focus of crash frequency models is generally two-fold on: (a) modeling methodology and (b) the parameters used as dependent and independent variables. In other words, the past research in this area has been diverse methodologically as well as empirically (Chang, 2005).

2.1. Crash frequency modeling: Methodological advances

The crash frequency data have been analyzed using a number of statistical methodologies. Initially multiple linear regression was used for model formulation. However, as pointed out by Joshua and Garber (1990), linear regression models do not describe the nature of the crash frequency data adequately. Poisson or Negative Binomial (NB) regression models, instead, are better suited for defining the random, discrete, and nonnegative nature of crash occurrence (Milton & Mannering, 1998). The log-linear model is the best known example of Poisson regression. It essentially is a generalized linear model (GLM) for Poisson-distributed data and specifies how the size of a cell frequency depends on the levels of categorical variables for that cell. The nature of this specification relates to the association and interaction structure among the categorical variables (Agresti, 2002).

It should be noted that the Poisson model formulation requires the mean and variance of the crash data to be equal. Therefore, the NB model, which has all the desirable statistical properties and also relaxes this constraint, is the most popular model formulation for crash frequency estimation. A detailed comparison between Poisson and Negative Binomial crash frequency models may be found in Miaou (1994). The findings suggested that since crash data tend to be overdispersed (i.e., variance ≫ mean), negative binomial modeling is the more appropriate technique of the two.

The findings from studies mentioned so far were based on the ability of the model formulation (such as Poisson or NB regression) to capture the underlying distribution of the crash frequency data. Recently some researchers have proposed 'distribution free' methodologies for the analysis of crash data. These methodologies include decision trees and artificial intelligence techniques such as the neural networks. No inherent assumptions about the distribution of the crash frequency data are needed to apply these techniques, which are essentially driven by observed data. For example, Chang and Chen (2005) and Abdel-Aty and Keller (2005) adopted

Table 1
Format of the crash data for crash frequency analysis (to identify the locations where "more crashes are likely to occur")

Section	Frequency/ Rate of crashes (y)	Traffic related factors *			Roadway geometry *		
		Speed limit	Truck percentage	V/C ratio	Curvature	Presence of ramps	Number of lanes
1	Y_1	…	…	…	…	…	…
2	Y_2	…	…	…	…	…	…
…	…	…	…	…	…	…	…
N	Y_n	…	…	…	…	…	…

* The factors shown here just for example and it is by no means a comprehensive list of factors associated with crashes.

Classification and Regression Tree (CART), the most commonly applied data mining technique, for crash frequency estimation. A comparison of the results from the CART, Artificial Neural Network (ANN) and NB regression models demonstrated that both CART and ANN are good alternatives to NB regression for estimation of freeway crash frequency (Chang, 2005; Chang & Chen, 2005). Since these data-driven techniques do not require any pre-defined underlying relationship between target (dependent) variable and predictors (independent variables), they are powerful data analysis tools. Based on this detailed review of the literature it may be concluded that while the researchers have employed a wide array of tools to model crash frequency/rate, more recent studies have explored the potential of 'data driven' techniques. In the next section we focus on the empirical explorations of the studies analyzing crash frequency.

2.2. Crash frequency modeling: Empirical explorations

From an empirical standpoint, crash frequencies are estimated as a function of a variety of factors including geometric characteristics (e.g., horizontal and vertical alignments, and shoulder width), traffic characteristics (e.g., AADT and percentage of trucks), and weather conditions (e.g., rain or snow; Shankar, Mannering & Barfield, 1995; Poch & Mannering 1996). In addition, Abdel-Aty and Radwan (2000) accounted for the demographic characteristics of the involved drivers (age and gender) by developing separate NB regression models for crashes involving drivers belonging to different age and gender groups.

Conventionally, most crash frequency models have used AADT to represent traffic characteristics. However, researchers are moving toward microscopic crash analysis, which includes analysis of hourly crash data (e.g., Ceder & Livneh, 1982a,b). These studies have been referred to as 'disaggregate' studies by Sullivan (1990). Measures such as hourly volumes are applied to cope with the uncertainty in the measurement of AADT values and incapability of this aggregate factor in capturing accurate traffic flow variations. These 'microscopic' traffic parameters not only include hourly volume but logical measures of congestion represented by v/c ratio (Frantzeskakis & Iordanis, 1987) and level of service (LOS; Persaud & Nguyen, 1998), along with distributional properties of variation in speed (Abdel-Aty, Pemmanaboina, & Hsia, 2006).

Based on the review of these studies it is revealed that since Ceder and Livneh (1982a,b) proposed the application of hourly traffic data, many studies have shown them to be preferable over AADT (e.g., Garber & Wu, 2001; Pasupathy, Ivan, & Ossenbruggen, 2000). The main difference between using the hourly traffic parameters and the AADT is that while the latter is readily available for most transportation facilities, the former had to be collected from the field for the purpose of any particular study. The availability of loop detector data for the freeways has not only helped overcome this limitation (Abdel-Aty et al., 2006), but also facilitated further disaggregating of the freeway crash data.

3. Individual crash level analysis of crash data and freeway traffic management

This approach to the analysis of crash data is characterized by each individual crash being the unit of analysis (Golob et al., 2004a). However, it is worth mentioning that the researchers have in fact analyzed the crash data in this form well before the advent of research directed toward proactive traffic management (e.g., Abdel-Aty, 2003; Duncan, Khattak, & Council, 1998; Abdel-Aty & Abdelwahab, 2000). The main focus of these studies was to associate the crash injury severity with driver and roadway characteristics. The basic premise was 'given a crash has occurred' estimate, for example, how severe would it be? Of course, these studies were not of use for proactive traffic management and the parameters used as independent variables in these studies were essentially static in nature. The focus of some of the more recent studies is to associate the crashes with real-time traffic characteristics and it is these studies that are relevant to proactive traffic management.

As discussed in the previous section, researchers have indeed disaggregated the crash data to assess the relationship between crashes and hourly traffic parameters (e.g., Ceder & Livneh, 1982a,b). The approach, however, is still aggregate in nature since crashes are still combined together albeit over a smaller interval. If the idea of disaggregating the crash data is taken to its extreme then each individual crash would become the unit of analysis. To assess the traffic conditions associated with crashes in this setting, traffic data right before the crashes is required. It is analogous to the collection of hourly traffic data to associate hourly crash counts with parameters such as hourly volume and V/C ratios.

Recently, collection of data representative of traffic conditions prevailing before individual crashes on instrumented freeways has become possible. These instrumented freeways are equipped with underground loop detectors that report traffic speed, volume, and lane-occupancy data at very short time intervals (from 1-minute to as short as 20-seconds). These data are collected for various operational purposes, such as for travel time estimation. It should be noted that the freeway crashes not only impact traffic safety, but also result in non-recurring congestion. Even the least severe of crashes impacts traffic operation in a considerable way and turn freeways into virtual parking lots. These facts signify that freeway crashes are not only critical from a traffic safety standpoint, but from an operational point of view. This is the reason why traffic management authorities have shown interest in traffic safety research based on the data disaggregated to individual crash level.

Research in freeway traffic management has focused on incident detection. The idea of incident detection involves analysis of patterns in the traffic surveillance data observed just after the incident. Since traffic data for the freeways are collected continuously it is possible to develop models using historical incident data and apply them in real-time to
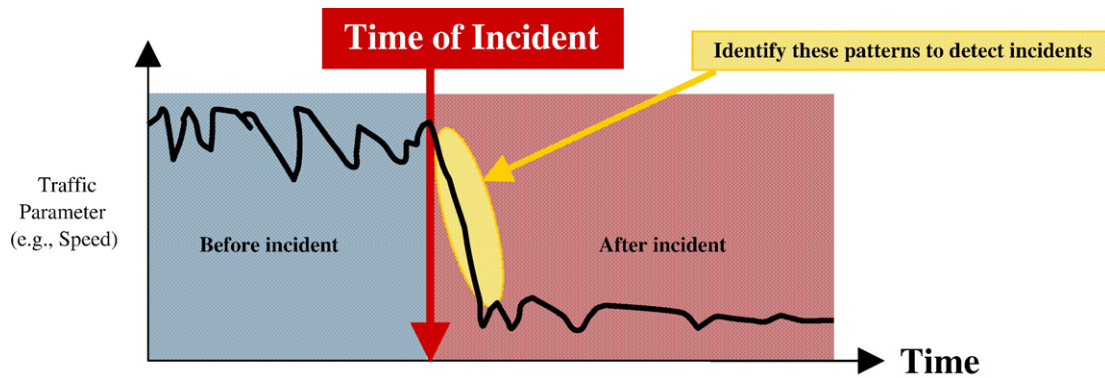
Fig. 1. Approach for incident detection.

examine the traffic data for any incident occurring on the freeway under surveillance. Fig. 1 shows, as an example, that the drop in speed following an incident may be used to detect an incident. This approach is reactive in nature and attempts to detect incidents in a timely fashion so that their impact can be minimized.

However, freeway data gathering is not the only field affected by the information technology revolution. The use of mobile phones has also increased manifolds along with the video surveillance of freeways. Due to these technological advancements the information on the incidents is immediately available to traffic management authorities. These advancements have rendered incident detection algorithms (e.g., Cheu & Ritchie, 1995; Abdulhai & Ritchie, 1999) somewhat irrelevant and traffic management authorities are becoming more interested in proactive strategies. These strategies would involve anticipating incidents along with strategies to avoid them altogether. In this regard, in their earlier studies the authors (Pande & Abdel-Aty, 2005) argued that crashes, in general, are more frequent and predictable than some other incidents such as a flat tire in the middle of rush hour. Therefore, to develop a proactive traffic management strategy, traffic data prior to individual historical crashes should be extracted and analyzed. Fig. 2 exemplifies what patterns would be of interest for such

analysis, assuming that the incident shown in Fig. 1 was a crash.

Based on the analysis of historical data, typical traffic patterns recorded prior to crashes may be identified. These patterns may then act as identifiers for real-time 'black-spots' on the freeway. This idea would require disaggregating the data to individual crash level. Table 2 depicts the format of the dataset that may be used for such analysis. It may be observed that each observation of this dataset, consisting of m observations, is an individual crash.

Hughes and Council (1999) overlaid the time of individual freeway crashes on time series of traffic data collected from nearby loop detector locations during the peak periods of the day. It was reported that macroscopic measures, such as AADT and even the hourly volume, correlate poorly with the real-time system performance as it relates to traffic safety.

Since their preliminary investigation, two streams of research work have been remarkable in this regard, one based on data from California freeways (Golob & Recker, 2001, Golob & Recker, 2004; Golob et al., 2004a; Golob, Recker, & Alvarez, 2004b) and the other based on data from the Interstate-4 corridor in Orlando, FL (Abdel-Aty & Pande, 2005; Pande & Abdel-Aty, 2005; Abdel-Aty, Uddin, Pande, Abdalla, & Hsia, 2004; Abdel-Aty, Uddin, & Pande, 2005; Abdel-Aty & Abdalla, 2004; Pande & Abdel-Aty, 2006a,b).
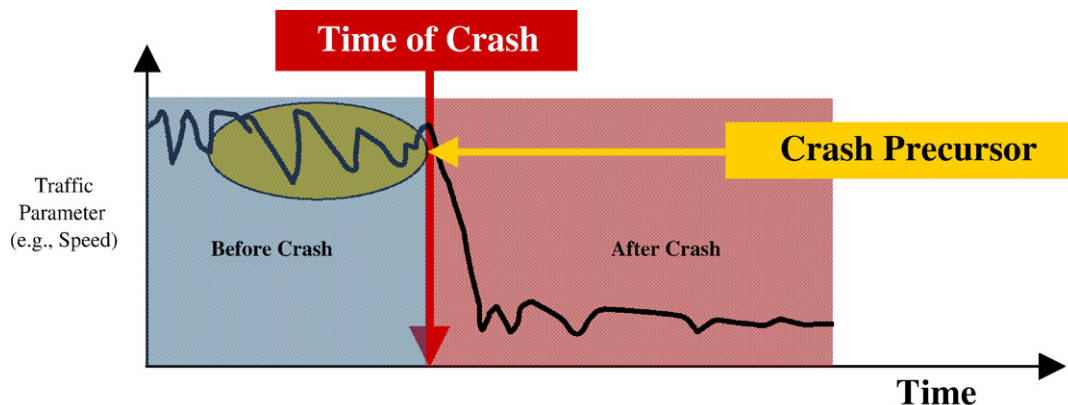

Fig. 2. Approach for proactive traffic management.

Table 2
Format of the crash data for analysis with data disaggregated to the extreme (to identify the locations where "crashes are likely to occur")

Crash	Traffic parameters from traffic surveillance system *			Roadway geometry *		
	Temporal variation of speed	Upstream occupancy	Differentialin speeds u/s and d/s	Curvature	Presence of ramps	Number of lanes
1	…	…	…	…	…	…
2	…	…	…	…	…	…
…	…	…	…	…	…	…
m	…	…	…	…	…	…

* The factors shown here just for example and it is by no means a comprehensive list of factors associated with crashes.

Both groups of studies used individual crashes as observations (thereby disaggregating the data to the extreme). The critical difference between the two is that while the research in California used only crash data, the authors compared the crash data with either stratified matched non-crash data (e.g., Abdel-Aty et al., 2004, 2005) or randomly selected non-crash data (e.g., Pande & Abdel-Aty, 2006a,b). Assembly of a large database by the researchers made it possible to separately analyze the crashes by first harmful event initiating the crash (i.e., by type such as rear-end, sideswipe).

Golob and Recker (2001) used non-linear (nonparametric) canonical correlation analysis (NLCCA). Research based on Interstate-4 in Orlando (FL) was conducted using a variety of statistical and data mining tools. These tools include Probabilistic Neural Network (PNN; Abdel-Aty & Pande, 2005), matched case-control Logistic Regression (Abdel-Aty et al., 2004), Generalized Estimation Equation (Abdel-Aty & Abdalla, 2004), along with classification tree/neural networks (Pande & Abdel-Aty, 2006a,b). The output of the models developed in these studies was a measure of crash risk for given traffic conditions and these models can be utilized for real-time crash risk assessment. It may be observed that even though this area of research is relatively new, significant progress is being made. This progress has been made possible by the successful application of data mining techniques for efficiently analyzing large databases.

4. Comparison of the two approaches to crash data analysis

According to Golob et al. (2004a), even though aggregate studies have been useful in identifying relationships between crash frequency/rates and traffic flow parameters, they can be susceptible to the problem of ecological fallacy. The ecological fallacy is a widely recognized error in the interpretation of statistical data, whereby inferences about the nature of individuals are based solely upon aggregate statistics collected for the group to which those individuals belong (Robinson, 1950). The studies analyzing data at individual crash level are, in theory, free from this fallacy.

While traditional studies attempt to estimate the crash frequency or rate, the research with individual crashes as the units of analysis have developed models to estimate some measure of the real-time likelihood of crash occurrence. It is worth mentioning that Lee, Saccomanno, and Hellinga (2002) and Lee, Hellinga, and Saccomanno (2003) did estimate the

real-time likelihood of freeway crash occurrence using a log-linear model based on the crash frequency analysis. To formulate the log-linear model, the continuous traffic parameters such as the density and coefficient of variation in speed were categorized resulting in loss of information due to categorization. Therefore, crash data disaggregated to individual crash level is more widely used for examining real-time traffic conditions on the freeway for their crash potential.

Based on a rudimentary analysis of traffic patterns observed prior to individual crashes, Hughes and Council (1999) observed that "traffic flow consistency" should be considered as a factor for crash causation. This observation is analogous to the conclusions from the aggregate studies, which indicate that the highway "design inconsistency" is perceived by the driver as an important factor associated with traffic safety from a human-factors standpoint.

Being data intensive, the individual crash level approach to analysis is suitable for the freeways where the traffic surveillance apparatus is already in place. It is also possibly the reason that even though the researchers did observe the advantages of disaggregating the data (Ceder & Livneh, 1982a,b), the idea of taking it to the extreme (to the individual crash level) did not get traction until recently. Enhancement of data collection, archiving, and analyzing capabilities and a significant push from the traffic management community led to research in this direction.

It should be noted that despite the advantages, there is a limit to what can be achieved by disaggregating the crash data to the level of individual crashes (Pande & Abdel-Aty, 2006a,b). These studies showed that the conditions prone to rear-end crashes may be more readily identified in real-time compared to the conditions for lane-change related crashes. In other words, measure of crash risk obtained from such analyses could be more reliable for crashes of a particular type than other crashes. For example, a rear-end crash would be more 'predictable' than a single-vehicle roll over. The reason for this difference in reliability could be the contribution of human error on the part of the involved drivers. Considering the fact that the rear-end crashes are the most frequent on freeways and some single-vehicle collisions might result from evasive driver actions under traffic conditions that are in fact prone to rear-end crashes, this concern is somewhat alleviated.

Also, note that the intervention measures for 'black-spots' identified based on the collective studies have been investigated extensively and agencies around the world have been using them for some time now (e.g., rumble strips,

median barrier). However, the remedies for real-time 'black-spots' are not entirely clear because it would require real-time intervention in the freeway operation with Intelligent Transportation System (ITS) strategies. Such strategies and assessment of their impacts on traffic operation and real-time crash risk remain formidable research problems (Abdel-Aty, Dilmore, & Hsia, 2006).

5. Conclusive remarks

A comprehensive review of the literature shows that research toward proactive traffic management based on analysis of crash data disaggregated to individual crash level has gained swift momentum. One cautionary note about these studies is that even though a reliable link between real-time traffic conditions and freeway crashes has been established, it is not entirely clear as to how that information might be used. There could be two potential ways to affect the real-time freeway traffic conditions: (a) one with strategies such as VSL and ramp metering with the goal of reducing the measure of the real-time risk; and (b) the other with the specific warnings to the motorists on the freeway if the estimated crash risk goes beyond a specified threshold.

The implementation of these strategies is not trivial, whereas the output of crash frequency models is readily understood in terms of how it could be implemented. For example, if it is found that a freeway section with a horizontal curve is experiencing high frequency of crashes then some kind of warning message sign or smoothening of the curve may be adopted. However, no such measures are available for real-time application. Even warning messages delivered to the drivers through Variable Message Signs might not have a desired impact.

Another caveat is that traffic data requirements for crash sample disaggregated to individual crash level are prohibitively extensive for roadways other than instrumented freeways. Hence, the aggregate crash frequency approach is still the most applicable for intersections and arterials where the traffic data are not collected continuously. One trend that is visible in both approaches dealing with crash data analysis is the advent of data mining techniques, such as the neural networks, classification tree, and unsupervised clustering. These techniques have been successfully employed in other data intensive industries such as insurance and banking. As the researchers explore the potential of data mining in crash data analysis, a better insight into the crash patterns can be expected.

Finally, it is worth mentioning that the studies with individual crash level approach to data analysis have been referred to as "disaggregate studies" in the literature (e.g., Golob et al., 2004a). However, since the traffic parameters are being aggregated over time and space (albeit a shorter span of time), these studies are not truly "disaggregate." Using crashes as units of analysis helps in avoiding ecological fallacies and has led to significant advancements in traffic safety research. However, strictly "disaggregate studies" would have to involve microscopic approaches (i.e., data obtained from individual vehicles).

References

Abdel-Aty, M. (2003). Analysis of driver injury sev erity levels at multiple locations using ordered probit models. *Journal of Safety Research*, *34*(5), 597−603.

Abdel-Aty, M., & Abdelwahab, H. (2000). Exploring the relationship between alcohol and the driver characteristics in motor vehicle accidents. *Accident Analysis and Prevention*, *32*(4), 473−482.

Abdel-Aty, M., & Abdalla, F. (2004). Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes using generalized estimating equations for correlated data. Presented at the 83rd Annual Meeting of the Transportation Research Board (TRB), Washington D.C.

Abdel-Aty, M., Dilmore, J., & Hsia, L. (2006). Applying Variable Speed Limits and the Potential for Crash Migration. *Transportation Research Record*, *1953*, 21−30.

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. (2005). Identifying crash propensity using specific traffic speed conditions. *Journal of Safety Research*, *36*(1), 97−108.

Abdel-Aty, M., Pemmanaboina, R., & Hsia, L. (2006). Assessing crash occurrence on urban freeways by applying a system of interrelated equations. *Transportation Research Record*, *1953*, 1−9.

Abdel-Aty, M., & Radwan, E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis and Prevention*, *32*(5), 633−642.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., & Hsia, L. (2004). Predicting freeway crashes based on loop detector data using matched case-control logistic regression. *Transportation Research Record*, *1897*, 88−95.

Abdel-Aty, M., Uddin, N., & Pande, A. (2005). Split models for predicting multi-vehicle crashes during high-speed and low-speed operating conditions on freeways. *Transportation Research Record*, *1908*, 51−58.

Abdulhai, B., & Ritchie, S. (1999). Enhancing the universality and transferability of freeway incident detection using a Bayesian-based neural network. *Transportation Research Part C: Emerging Technologies*, *7*(5), 261−280.

Agresti, A. (2002). *Categorical data analysis.* New York, NY: John Wiley and Sons, Inc.

Ceder, A., & Livneh, M. (1982). Relationship between road accidents and hourly traffic flow-I. *Accident Analysis and Prevention*, *14*(1), 19−34.

Ceder, A., & Livneh, M. (1982). Relationship between road accidents and hourly traffic flow II. *Accident Analysis and Prevention*, *14*(1), 34−44.

Chang, L. -Y. (2005). Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science*, *43*(8), 541−557.

Chang, L. -Y., & Chen, W. -C. (2005). Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, *36*(4), 365−375.

Cheu, R., & Ritchie, S. (1995). Automated detection of lane-blocking freeway incident using artificial neural networks. *Transportation Research Part C: Emerging Technologies*, *3*(6), 371−388.

Duncan, C., Khattak, A., & Council, F. (1998). Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. *Transportation Research Record*, *1635*, 63−71.

Frantzeskakis, J., & Iordanis, D. (1987). Volume to capacity ratio and traffic accidents on interurban four-Lane highways in Greece. *Transportation Research Record*, *1112*, 29−38.

Garber, N., & Wu, L. (2001). Stochastic models relating crash probabilities with geometric and corresponding traffic characteristics data. Research Report No. UVACTS-5-15-74, Center for Transportation Studies at University of Virginia, Charlottesville, VA.

Golob, T., & Recker, W. (2001). Relationships among urban freeway accidents, traffic flow, weather and lighting Conditions. California PATH Working Paper UCBITS-PWP-2001-19, Institute of Transportation Studies, University of California, Berkeley, CA.

Golob, T., & Recker, W. (2004). A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part. A*, *38*(1), 53−80.

Golob, T., Recker, W., & Alvarez, V. (2004). Freeway safety as a function of traffic flow. *Accident Analysis and Prevention*, *36*(6), 933−946.

Golob, T., Recker, W., & Alvarez, V. (2004). Tool to evaluate the safety effects of changes in freeway traffic flow. *Journal of Transportation Engineering*, *130*(2), 222−230.

Hughes, R., & Council, F. (1999). On establishing relationship(s) between freeway safety and peak period operations: Performance measurement and methodological considerations. Presented at the 78th annual meeting of Transportation Research Board, Washington, D.C.

Joshua, S., & Garber, N. (1990). Estimating truck accident rate and involvement using linear and Poisson regression models. *Transportation Planning and Technology*, *15*, 41−58.

Lee, C., Hellinga, B., & Saccomanno, F. (2003). Real-time crash prediction model for the application to crash prevention in freeway traffic. *Transportation Research Record*, *1840*, 67−78.

Lee, C., Saccomanno, F., & Hellinga, B. (2002). Analysis of crash precursors on instrumented freeways. *Transportation Research Record*, *1784*, 1−8.

Miaou, S. P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention*, *26*(4), 471−482.

Milton, J., & Mannering, F. (1998). The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation*, *25*(4), 395−413.

Pande, A., & Abdel-Aty, M. (2005). Freeway safety strategy for advanced proactive traffic management. *Journal of Intelligent Transportation Systems*, *9*(3), 145−158.

Pande, A., & Abdel-Aty, M. (2006). A comprehensive analysis of the relationship between real-time traffic surveillance data and rear-end crashes on freeways. *Transportation Research Record*, *1953*, 41−49.

Pande, A., & Abdel-Aty, M. (2006). Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis and Prevention*, *38*(5), 936−948.

Pasupathy, R., Ivan, J., & Ossenbruggen, P. (2000). Single and multi-vehicle prediction models for two-lane roadways. Final Report, Project UCNR9-8, United States Department of Transportation, Washington, D.C.

Persaud, B., & Nguyen, T. (1998). Disaggregate safety performance models for signalized intersections on Ontario provincial roads. *Transportation Research Record*, *1635*, 113−120.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection accident frequencies. *Journal of Transportation Engineering*, *122*(2), 105−113.

Robinson, W. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*(3), 351−357.

Shankar, V., Mannering, F., & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis and Prevention*, *27*(3), 371−389.

Songchitruksa, P., & Tarko, A. (2006). The extreme value theory approach to safety estimation. *Accident Analysis and Prevention*, *38*(4), 811−822.

Sullivan, E. C. (1990). Estimating accident benefits of reduced freeway congestion. *Journal of Transportation Engineering*, *116*(2), 167−180.