NEGATIVE BINOMIAL ANALYSIS OF INTERSECTION-ACCIDENT FREQUENCIES

By Mark Poch[1] and Fred Mannering,[2] Associate Member, ASCE

(Reviewed by the Highway Division)

ABSTRACT: Traffic accidents at urban intersections result in a huge cost to society in terms of death, injury, lost productivity, and property damage. Unfortunately, the elements that effect the frequency of intersection accidents are not well understood and, as a result, it is difficult to predict the effectiveness of specific intersection improvements that are aimed at reducing accident frequency. Using seven-yr accident histories from 63 intersections in Bellevue, Washington (all of which were targeted for operational improvements), this paper estimates a negative binomial regression of the frequency of accidents at intersection approaches. The estimation results uncover important interactions between geometric and traffic-related elements and accident frequencies. The findings of this paper provide exploratory methodological and empirical evidence that could lead to an approach to estimate the accident reduction benefits of various proposed improvements on operationally deficient intersections.

INTRODUCTION

The effort that public agencies put into reducing traffic accidents is highly justifiable. Traffic accidents place a huge financial burden on society. As evidence of this, the National Safety Council (''Estimating'' 1994) provides estimates of the average costs of traffic accidents to the public. The calculable costs of motor-vehicle accidents are wage and productivity losses, medical expenses, administrative expenses, motor-vehicle damage, and employer costs. In 1993, the cost of all these items for each death and injury resulting from traffic accidents, as well as the property damage associated with accidents, was estimated to be $900,000 for a death, $32,800 for nonfatal disabling injuries, and $5,800 for property damage only. These figures appropriately represent the economic loss of society resulting from motor-vehicle accidents that have already occurred. However, because the figures do not include the value of a person's natural desire to live longer or to protect the quality of one's life, they actually underestimate the dollar value of future benefits derived from traffic safety measures that would prevent the accident from occurring. That is, these economic loss estimates do not include what people are willing to pay for improved safety. Studies that have addressed this willingness to pay arrive at a value of life of roughly $3,000,000 (Mannering and Winston 1995). It is clear that traffic accidents are a substantial economic burden on society.

The city of Bellevue, Washington, is one public agency that is very serious about reducing traffic accidents on city streets. Bellevue is situated just east of Seattle and is Washington's fourth largest city, with a population of just under 100,000. Using the figures provided by the National Safety Council (''Estimating'' 1994), it is estimated that the economic loss to society for traffic accidents in Bellevue in 1993 alone was approximately $21,000,000.

In 1990, the city of Bellevue developed the methodology for yearly intersection-accident studies. In 1993, the scope of the annual study was increased to include midblock street locations as well. As a result, many of the high-accident locations identified through intersection and midblock studies have had accident-reducing projects identified, funded, and implemented.

To further expand on the abilities of the city of Bellevue staff to both revise existing high-accident locations and also build safer new facilities, a need to further investigate accident-reducing techniques was identified. It was decided to further explore the geometric and traffic-related conditions on intersection approaches that contribute to increasing the frequency of intersection accidents.

This paper begins with a review of earlier research that has attempted to relate geometric and traffic conditions to accident frequencies. Based on this review, we present an appropriate modeling approach to study the relationship between roadway geometrics/traffic-related elements and accident frequencies at intersections. This is followed by a description of the available data and a presentation and discussion of model estimation results. Finally, an overall summary of model findings and their implications is given and some concluding remarks are provided.

PREVIOUS RESEARCH

Although many studies have addressed the relationship between traffic and geometric variables and accident frequency, a review of recent literature revealed that surprisingly few have studied the relationship between approach conditions (geometric and traffic-related) at intersections and accident frequency—with most focusing on accidents that occur along roadway segments rather than on intersection approaches. These studies have typically dealt with rural interstates and focused on accidents involving heavy trucks, instead of addressing urban and suburban streets and focusing on accidents involving all vehicles. Although such studies did not focus on the analysis of accident frequencies at intersections, they have provided methodological insights that are relevant to the intersection case. For example, Miaou and Lum (1993) evaluated the statistical properties of two conventional linear-regression models and two Poisson-regression models to determine their suitability for modeling vehicle accidents and highway geometric design relationships. They concluded that the two linear regression models lacked the distributional properties to adequately describe random, discrete, nonnegative, and sporadic vehicle accidents. Miaou and Lum also state that the Poisson regression models possessed the most desirable statistical properties in describing vehicle accident events, but

[1]Traffic Operations Engr., City of Bellevue Transp. Dept., Bellevue, WA 98009.
[2]Prof., Dept. of Civ. Engrg., Univ. of Washington, Seattle, WA 98195.
Note. Discussion open until September 1, 1996. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on February 27, 1995. This paper is part of the *Journal of Transportation Engineering*, Vol. 122, No. 2, March/April, 1996.

Paper No. 10216.

noted the limitation of the model in that the variance of the accident data is constrained to be equal to the mean. Overdispersion, which occurs when the variance of accident-frequency data is greater than its mean, can result in biased model coefficients and erroneous standard errors. The biased coefficients can result in an under- or overstatement of accident likelihood. The authors point out that the negative binomial or double Poisson distributions must be used to overcome the overdispersion concern.

Other authors have also dealt with similar methodological concerns. For example, Joshua and Garber (1990), Jovanis and Chang (1986), and Jones et al. (1991) have all empirically demonstrated the superiority of the Poisson regression, relative to standard linear regression, in analyzing accident frequencies; and Shankar et al. (1995) have addressed the overdispersion issue using a negative binomial regression.

Aside from previous studies' methodological contributions, findings with regard to the factors determining accident likelihoods in the nonintersection case can also contribute to our understanding of intersection-accident frequency. An important study in this regard was that of Wong and Nicholson (1992), who observed that modifications to roadway geometrics are an important concern because of the strong association between adverse geometric elements and high-accident locations. Other empirical relationships between vehicle accidents and highway geometrics have been studied through the use of statistical models to investigate accident involvement rate, accident probability, geometric design variables critical to safety, and the accident-reduction potential of geometric improvements (Hammerslag et al. 1982; Okamoto and Koshi 1989; Persaud 1991; Shankar et al. 1995).

A common criticism of many previous studies is that they do not usually include consideration of human factors. Treat (1980) and Sabey and Taylor (1980), show that human factors are involved in around 95% of all traffic accidents, either alone or combined with other factors. However, Massie et al. (1993) point out that the classical human-factors approach ignores the problem associated with classifying collisions and their related causes, be it human or otherwise, and fails to address the issue of helping drivers avoid collisions. By identifying geometric conditions that lend themselves to producing accidents, these conditions can be corrected. Further, it is clear that if motorists were cognizant of every geometric deficiency encountered and warned to be careful of these deficiencies, accident potential would be reduced. However, because this is an impossible task, correcting geometric deficiencies is an important step toward reducing accidents.

In summary, there have been many studies in recent literature analyzing the relationship between traffic/geometric variables and accident frequency. However, most of these studies have focused on roadway segments that tend to be in nonurban areas. Furthermore, for the most part, these studies did not analyze specific accident types, rather, accidents tended to be analyzed in whole or by the type of vehicle involved. Very little work has been done to study specific accident types on urban and suburban intersection approaches. Regarding the types of models used for accident-frequency studies, Poisson regression models have been shown to be more appropriate than conventional linear regression models. However, the inability of the Poisson model to handle overdispersed data is a lingering concern with regard to studying accident frequencies.

MODELING APPROACH

The objective of our study is to develop statistical models of the annual accident frequency on individual intersection approaches (i.e., a four-legged intersection-oriented north-south, east-west, would have four approaches: northbound, southbound, eastbound, and westbound). In so doing, we will actually estimate four different accident-frequency models that will predict: (1) Total accident frequency (all accident types); (2) rear-end accident frequency; (3) angle accident frequency; and (4) approach-turn accident frequency. In all cases the dependent variable (annual accident frequency) will be a nonnegative integer. On the basis of past research, the Poisson regression model is a natural first choice for modeling such data (i.e., discrete, nonnegative, and sporadic). However, as mentioned, a major limitation of the Poisson regression model is that the variance of the dependent variable (annual accident frequency) is constrained to be equal to its mean. There is a fairly large body of literature [see, e.g., Shankar et al. (1995)] that suggests that most accident data are likely to be overdispersed (i.e., the variance will likely be significantly greater than the mean). When the mean and variance of the data are not approximately equal, the variances of the estimated Poisson model coefficients tend to be underestimated and the coefficients themselves are biased. This limitation can be readily overcome by using the negative binomial distribution. The negative binomial distribution is well-suited to describe discrete, nonnegative events, and the Poisson requirement that the mean be equal to the variance is relaxed.

To derive the negative binomial model we first start with the Poisson model. For the Poisson model, the probability of having a specified number of accidents at intersection approach $i(n_i)$ per year (where n_i is a nonnegative integer) is

$$P(n_i) = \frac{\exp(-\lambda_i)\lambda_i^{n_i}}{n_i!} \tag{1}$$

where $P(n_i)$ = probability of an accident occurring on approach i, n_i times per year; and λ_i = Poisson parameter for approach i, which is equal to approach i's expected accident frequency per year [i.e., $E(n_i)$]. Poisson regressions are fitted to data by specifying the Poisson parameter λ_i to be a function of explanatory variables, which, in this case, could include signal control, sight distance, traffic volumes, and so on. This is done by specifying the Poisson parameters as

$$\ln \lambda_i = \boldsymbol{\beta}\mathbf{X}_i \tag{2}$$

where $\mathbf{X}_i$ = a vector of explanatory variables; and $\boldsymbol{\beta}$ = a vector of estimable coefficients.

The Poisson regression, as defined in (1) and (2), is estimable by standard maximum likelihood methods with the likelihood function

$$L(\boldsymbol{\beta}) = \Pi_i \frac{\exp[-\exp(\boldsymbol{\beta}\mathbf{X}_i)][\exp(\boldsymbol{\beta}\mathbf{X}_i)]^{n_i}}{n_i!} \tag{3}$$

As discussed, the Poisson regression is inappropriate if the mean and variance of the distribution is not approximately equal. If this equality does not hold, corrective measures must be taken to avoid model specification errors. The solution is to apply the negative binomial model. The negative binomial model is derived from the Poisson model by adding an independently distributed error term ε to (2), such that

$$\ln \lambda_i = \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i \tag{4}$$

where $\exp(\varepsilon_i)$ = a gamma-distributed error term with mean one and variance α. This gives a conditional probability

$$P(n_i|\varepsilon) = \frac{\exp[-\lambda_i \exp(\varepsilon_i)][\lambda_i \exp(\varepsilon_i)]^{n_i}}{n_i!} \tag{5}$$

Integrating ε out of this expression produces the unconditional distribution of n_i. The formulation of this distribution, which is used in maximum likelihood estimation, is

$$P(n_i) = \frac{\Gamma(\theta + n_i)}{[\Gamma(\theta)\cdot n_i!]}\cdot u_i^{\theta}(1 - u_i)^{n_i} \tag{6}$$

where $u_i = \theta(\theta + \lambda_i)$, and $\theta = 1/\alpha$. Eq. (6) can be easily used to write a likelihood function for coefficient estimation purposes [as in (3)]. The foregoing formulation results in a model (the negative binomial) that allows the mean to differ from the variance such that

$$\mathrm{var}(n_i) = E(n_i)[1 + \alpha E(n_i)] \tag{7}$$

where α (the variance of the gamma-distributed error term) = a measure of dispersion and is estimable by standard maximum likelihood techniques. The appropriateness of the negative binomial relative to the Poisson model is determined by the statistical significance of the estimated coefficient α. If α is not significantly different from zero (as measured by t-statistics), the negative binomial model simply reduces to a Poisson regression with $\mathrm{var}[n_i] = E[n_i]$. If α is significantly different from zero, the negative binomial is the correct choice and the Poisson model is inappropriate.

We will use the negative binomial to model the annual frequency of all accidents and three specific types of accidents (rear-end, angle, and approach-turn accidents). The estimation will be performed using standard maximum likelihood procedures [see Greene (1993)]. Careful attention will be given α to justify our choice of the negative binomial model. For an example of another study of accident frequencies using the negative binomial model, the reader is referred to Shankar et al. (1995).

DATA

In Bellevue, as in most cities, a large number of intersections are in residential areas and are characterized by low traffic volumes and few, if any, accidents. However, the city's interest lies primarily in problematic intersections with high accident potential. To address such intersections, we collected data from the list of intersections that were targeted for some sort of operational improvement between 1988 and 1992 (the selective nature of our sample must be kept in mind when interpreting the forthcoming empirical findings). An operational improvement was considered to be reconstruction (i.e., rebuilding and improving intersection), control change (i.e., adding stop signs, signal, etc.), phasing change (i.e., adding protective left-turn phasing, etc.), or channelization change (i.e., adding a left-turn pocket, etc.). Intersections with operational improvements were selected for study because they are intersections identified as being operationally deficient, and our chosen study period (1987–1993) will allow us to observe before-improvement and after-improvement accident frequencies at these intersection approaches. This will provide an important variation in operational conditions in our database. The resulting data gives a well-defined sample of city intersections to study, with all four different types of Bellevue street classifications (principal, minor, collector arterials, and local streets) well-represented. A total of 63 intersections were available for study.

Each intersection was divided into separate approaches and accident data were taken at each approach in one-yr intervals. In the event that the accident occurred in the intersection proper, accidents were assigned to the approach of the "at-fault" vehicle. The year in which the operational improvement at the intersection was made was not included because of the midyear change in conditions. Thus, a four-approach intersection with operational improvements made in 1989 would provide four observations (one for each approach) for the years 1987, 1988, 1990, 1991, 1992, and 1993. Such an intersection would provide $4 \times 6 = 24$ observations to the study. A total of 1,385 observations were provided by the 63 intersections studied. Taking repeat observations from each intersection (i.e., each intersection can generate as many as 24 observations) creates a possible correlation problem among observations. This issue will be empirically explored later in the paper.

For modeling purposes it was determined that possible explanatory variables [included in the vector $\mathbf{X}_i$ in (4)] had to be representative of an approach condition. An approach condition was determined to be a traffic or geometric entity that could be reasonably expected to influence the frequency of accidents. This included a wide range of variables such as approach volumes, number of approach lanes, speed limits, highway grades, signal-control characteristics, the presence of horizontal curves, sight-distance restrictions, and so on. In addition, we constructed indicator variables for the calendar year of the data and location of the intersection in the city (e.g., in the central business district, suburbs, etc.). Most variables relating to physical characteristics were readily available from intersection design plans. However, obtaining accurate annual traffic volume data was a more complex task. To generate annual traffic data by approach, morning peak period, afternoon peak period, and midday traffic volumes were gathered for available years. These counts were placed into standard expansion formulas [see Mannering and Kilareski (1990)] to obtain daily counts. If traffic data were missing in one of the years, the city of Bellevue's yearly factors were used to convert known traffic counts of a different vintage to traffic counts in missing years. The city of Bellevue has four different yearly expansion factors to account for different traffic growth rates in various parts of the city. While such disaggregation allows for some variance among intersections, some correlation of traffic volumes from one intersection to the next may exist (i.e., intersections in close geographical proximity are likely to have correlated traffic). Fortunately our intersections tend to be geographically dispersed because we are only studying a small subsample of the total number of intersections (i.e., only those targeted for operational improvements). Still, we will explore the empirical consequences of possibly correlated traffic data later in this paper.

For each observation (defined as an intersection approach), a total of 64 possible explanatory variables were collected. Recall that four separate models will be estimated with the following dependent variables: total number of annual accidents on the approach, total number of annual rear-end accidents on the approach, total number of annual angle accidents on the approach, and total number of annual approach-turn accidents on the approach. From 1987–93, a total 1,396 accidents were recorded with 26% being rear-end, 30% angle, 32% approach turn, and 12% all others. Because intersection approaches may have one of these accident types as the predominate type, separate models of rear-end, angle, and approach-turn accidents should provide valuable insights (that may or may not be uncovered in the model of total accident frequency) into the types of approach conditions that influence the frequency of these specific types of accidents.

MODEL FOR TOTAL INTERSECTION APPROACH ACCIDENTS

Total approach accidents include all reported accident types that occurred in the city of Bellevue. The accident types include angle, sideswipe/lane change, rear-end, head-on, fixed object/parked vehicle, approach turn, pedestrian/bicyclist, and others.

Negative binomial estimation results of annual accident frequency at intersection approaches are presented in Table 1. This table includes the estimated coefficients, associated t-statistics, and log-likelihood for the model. The table shows that all variables are of plausible sign (with a positive sign increasing accident frequency and a negative sign decreasing accident frequency) and highly significant (the lowest t-statistic is 1.64, $p = 0.05$). The variables included in this model (and subsequent models) are those that resulted in the highest t-statistics (after a systematic evaluation of all variables) and were selected from the 64 possible explanatory variables available.

Turning to the specific coefficient estimates presented in Table 1, we find that increasing left-turn volumes increase annual accident frequencies. The left-turn movement has always posed operational concerns at intersections. It is often the critical movement in signal timing, taking up valuable cycle time to serve few at the expense of many. Typically, the movement warrants exclusive single or dual lanes, which require large capital expenditures to construct, and when not provided, cause sight-distance, accident, and level-of-service concerns. The left-turn movement is inherently dangerous because, in most cases, it requires crossing the path of opposing traffic. To better grasp the relationship between left-turn volume (and other explanatory variables) and accident frequencies, elasticities are computed.

Elasticity is defined as

$$e_{x_{ik}}^{\lambda_i} = \frac{\partial \lambda_i}{\partial x_{ik}} \cdot \frac{x_{ik}}{\lambda_i} \tag{8}$$

where λ_i = mean accident frequency at intersection approach i (in a given year), as defined in (4); and x_{ik} = value of the explanatory variable k at intersection approach i. Differentiating (4) and applying (8) gives

$$e_{x_{ik}}^{\lambda_i} = \beta_k x_{ik} \tag{9}$$

where β_k = coefficient estimate of explanatory variable k (as presented in Table 1).

Average elasticities [(9) averaged over all intersection approach/year points] for all continuous explanatory variables are presented in Table 2. Elasticities for indicator variables (i.e., those variables that take values of zero or one) are not computed because their elasticity has no immediate interpretation.

With regard to left-turn volume, we find an elasticity of 2.28. Roughly stated, this means that a 1% increase in left-turn volumes will result in a 2.28% increase in total accidents. This is considered elastic (an absolute value greater than one) and underscores the importance of left-turn volumes in total accident frequency.

Right-turn volumes were also found to increase the likelihood of an accident. In general, high right-turn volumes are an indication of a higher-volume intersection, producing more conflicts and thus more accidents. Unlike the left turn, the right-turn movement is most often not provided an exclusive lane to complete the maneuver. Without an exclusive lane, right-turning vehicles are required to slow in a lane shared with speed-maintaining vehicles, and this speed differential tends to cause rear-end accidents. However, Table 2 shows that right-turn volumes have an elasticity of only 0.92 (meaning that a 1% increase in right-turn volumes only increases accident frequencies by 0.92%). Therefore, the role they play in total intersection approach accident frequency is much less than left-turn volumes.

The coefficient estimates show that the total volume on the opposing approach increases accident likelihoods. This is because the conflict for left-turning vehicles increase as fewer gaps are available for the maneuver. Higher opposing volume also reduces the percentage of cycle time at signals available for the approach. In addition, higher volumes of opposing vehicles cause greater numbers of conflicts as intersection-turning movements are made, and thus the strong tendency to increase the number of total approach accidents. The high elasticity (2.95 in Table 2) underscores the important explanatory power that opposing approach volume has on accident frequency.

Through, combined through-right, and right-turn lanes form the predominate classes of lane types at intersection approaches. With each additional opposing through lane of travel, the distance a left-turning vehicle is exposed to opposing traffic increases. In addition, more opposing through lanes indicates more opposing traffic, which again decreases gap availability and increases delay on the approach, which tends to increase total approach accidents. Thus, the positive coefficient in the model indicates that the greater the number of opposing lanes of this type, the higher the accident frequency.

The indicator variable for no control accounts for the difference between no control and control (i.e., a signal or stop sign present on the approach). Having no control on an approach was found to decrease accident frequencies. This finding is likely the outgrowth of the fact that no-control approaches usually signify low cross-street volumes, exceptionally good sight distances, and other sometimes unobserved

TABLE 1. Negative Binomial Estimation Results for Annual Accident Frequency (All Types) at Intersection Approaches

Variable (1)	Coefficient (t-statistic) (2)
Constant	−1.41 (−3.75)
Left-turn volume in thousands (average daily traffic)	0.251 (8.08)
Right-turn volume in thousands (average daily traffic)	0.0902 (3.27)
Total opposing approach volume in thousands (average daily traffic)	0.0523 (4.95)
Number of through, combined through-right, and right-turn lanes	0.153 (2.32)
No control on approach (1 if no control, 0 otherwise)	−0.753 (−4.40)
Signal control (1 if signal control, 0 otherwise)	−0.325 (−2.20)
Two-phase signal (1 if two-phase signal, 0 otherwise)	0.270 (2.76)
Eight-phase signal (1 if eight-phase signal, 0 otherwise)	0.610 (5.11)
Protected left turn (1 if protected left turn, 0 otherwise)	−0.468 (−4.48)
Local street approach (1 if local street approach, 0 otherwise)	−0.336 (−2.16)
All approaches local streets (1 if all approaches local streets, 0 otherwise)	−1.093 (−4.71)
Sight-distance restriction (1 if sight distance is restricted, 0 otherwise)	1.123 (4.35)
Combined through-left lane and two or more lanes on the approach (1 if there is a combined through-left lane and two or more lanes on the approach, 0 otherwise)	0.201 (1.64)
Left turns not aligned and not a single-lane approach, protected left, or stop control (1 if left turns are not aligned and the approach does not have a single lane, protected left, or stop control, 0 otherwise)	−0.899 (−6.19)
Approach speed limit (km/h)	0.0203 (2.78)
Opposing approach speed limit (km/h)	−0.0075 (−2.36)
α (Negative binomial dispersion parameter)	0.346 (5.96)
Number of observations	1,385
Log-likelihood at zero	−2,123.11
Log-likelihood at convergence	−1,698.26
ρ^2	0.200

TABLE 2. Annual Accident Frequency (All Types) at Intersection Approaches: Elasticity Estimates

Elasticity with respect to (1)	Elasticity (2)
Left-turn volume in thousands (average daily traffic)	2.28
Right-turn volume in thousands (average daily traffic)	0.92
Total opposing approach volume in thousands (average daily traffic)	2.95
Approach speed limit (km/h)	0.98
Opposing approach speed limit (km/h)	−0.34

factors that lead to a tendency toward low overall accident rates (which is often why no control is warranted). The negative value of the estimated coefficient for this variable confirms this.

The indicator variable for signal control compares signalized with unsignalized intersections. Signals are installed on the basis of warrants that analyze volumes on the major and cross street to determine whether volumes are at a level that exclusive right-of-way needs to be assigned to the cross street to reduce delay and increase safety. Signalization not only assigns right-of-way to cross traffic that otherwise would have to wait for gaps in major street traffic, but also can eliminate the effects of poor sight distance due to substandard horizontal or vertical alignment of fixed objects at previously unsignalized locations. The coefficient estimate in Table 1 shows that signal control results in a significant decrease in accident frequency.

The variable comparing two-phase signals with all others has a coefficient that shows that intersections that are signalized and run on two phases tend to increase the frequency of accidents. With two-phase signals, the left turns on each approach are allowed to proceed as "permissive," which requires these vehicles to yield to oncoming traffic. This conflict tends to increase accidents. In addition, two-phase signals tend to exist where signal warrants are just met, and often these locations have less than desirable geometrics such as the lack of exclusive turn pockets and properly aligned lanes. It is important to remember that a two-phase signal will have both the signal-control indicator variable (with its coefficient of -0.325) and the two-phase indicator variable (with its coefficient of 0.270) active. Thus the net effect of a two-phase signal is still a net reduction in accident frequency (with an effective coefficient of -0.055) relative to nonsignal-controlled intersections.

A similar indicator variable is that comparing eight-phase signals with all other types (signalized and nonsignalized). Although eight-phase signals operate left turns on each approach as "protected," which assigns these left turns exclusive right-of-way to complete the maneuver, we still see that, according to the model, eight-phase signals are associated with higher accident frequencies (i.e., a positive coefficient). This is true even when the signal-control indicator coefficient of -0.325 is taken into account because the net effect is still 0.285. This can be explained by looking at the conditions under which eight-phase signals are installed: high volume, high congestion intersections. In all likelihood this finding evolves from the correlation of the eight-phase indicator variable with important unobserved variables such as driver frustration at complex intersections, sensory overload, and so on. It does not necessarily imply that replacing an eight-phase signal with a two-phase signal will decrease accident likelihoods. The solution to this counterintuitive finding is to collect additional data in an attempt to find the variable(s) the eight-phase indicator variable is acting as a proxy for. However, if one were to omit this variable because of its counterintuitive explanation an omitted variables bias could result, causing a serious model misspecification.

The variable indicating a protected left turn has a coefficient that shows that if the approach has a protected left turn, then total accidents tend to decrease. Protected left-turn phasing allows the left-turn maneuver to proceed with no conflicts, thus vehicles do not have to yield to oncoming traffic and adjacent pedestrians as do vehicles turning left under the permissive left condition. This decreases the likelihood of collisions resulting from the left-turning vehicles failing to yield to oncoming traffic.

Approaches that were classified as a local street were found to have lower accident frequencies. By definition, local streets carry the lowest amount of volume of all the street classifications. This low amount of volume causes a lower opportunity for accidents on the approach, and thus tends to decrease the total number of approach accidents given by the model. The negative coefficient associated with the variable confirms this.

If all approaches at the intersection were classified as local streets, a significant reduction in accident frequency is indicated by the negative coefficient. This is again likely to be the outgrowth of very low traffic volumes. The low volume on each approach results in few conflicts for turning traffic and traffic that is often regulated by stop control.

The presence of a sight-distance restriction was found to significantly increase accident frequency. A sight-distance restriction could result at a stop-controlled or uncontrolled approach due to fixed objects in medians or at the corners, sight-restricting horizontal or vertical curvature of the roadway, or overgrown brush or other vegetation. In these cases, the restriction would be realized when the standard sight line based on the speed of traffic on the cross street is not provided from the stop point on the approach to cross traffic in both directions. A sight-distance restriction could also occur at a signalized approach due to horizontal or vertical curvature across the intersection, an object in a median area, or misaligned left turn lanes. Regarding horizontal or vertical curvature, or an object in the median, a restriction would be realized when the largest identifiable gap in oncoming traffic is not adequate to provide the left-turning vehicle time to identify the gap and complete the left-turn maneuver. Regarding misaligned left-turn lanes (which account for roughly 5% of the sight-distance restrictions), whenever misaligned left-turn lanes are present on the subject and opposing approach, we regard this as a sight-distance restriction because misaligned left-turn lanes result in a decrease in the gap time for conflicting movements. In general, the various sight-distance restrictions have numerous adverse consequences, including a reduction in the amount of time drivers have to identify and react to traffic-control devices and regulatory signs. Such restrictions translate into a significant increase in accident frequency as indicated by the positive coefficient shown in Table 1.

Intersection approaches with a combined through-left lane and two or more total lanes were found to have higher accident frequencies when compared to approaches that did not have these conditions. Through-left lanes cause accident potential because: (1) Left-turning vehicles that must stop and wait for a gap to complete the maneuver cause a high potential for rear-end accidents as through vehicles approach in the same lane at prevailing speed; and (2) stopped left-turning vehicles that face stopped left-turning vehicles on the opposing approach must overcome the sight restriction to the opposing through vehicles to successfully complete the maneuver. This concern is a significant cause of approach-turn accidents. The positive coefficient estimate for this variable shows that total accidents increase on approaches with this configuration because of the foregoing two reasons.

Intersection approaches that had left turns that were not aligned and did not have a single-lane approach, protected left, or stop control were found to have lower accident frequencies than intersection approaches that did not have these conditions. At first, thought, one might expect an increase in approach accidents for an intersection with these conditions. After all, when this indicator variable is 1, it is assured that both approaches are either signalized with permissive only-lefts or are uncontrolled, and contain two or more lanes. With this intersection approach condition, left-turning vehicles on one approach have restricted sight distance when there are stopped left-turning vehicles on the opposing approach. This situation is associated with the increased likelihood of approach-turn accidents. However, this intersection approach condition is typically associated with few left-turning vehicles on the approach, and the likelihood of simultaneous opposing left-turning vehicles is very low. Thus, even though the sight restricting simultaneous left-turn situation occurs under this approach condition, it does not happen enough to cause an increase in accidents. In addition, if the left-turning vehicle volume was high enough to begin causing more accidents, then it is likely that intersection modifications such as the installation of left-turn pockets or split phasing would have been made. The model shows a tendency toward lower intersection accidents under this approach condition because the frequency of accidents resulting from left-turn movements tend to be low and there are few other accident types causing concerns under these conditions.

Two variables capture speed-limit effects: the approach speed limit and the opposing approach speed limit. Both these speed limits are correlated with actual vehicle speeds. Simply stated, higher speeds increase the amount of distance a vehicle travels while stopping or perceiving and reacting to a stimulus. The ability to slow or stop a vehicle over a short distance is important to accident avoidance, especially when confronting typical intersection approach situations such as stopped vehicles, crossing vehicles, and changing signal indications.

The coefficient estimate for the approach speed limit shows that, as expected, the higher the approach speed limit, the higher the accident frequency. However, the opposing approach speed limit was found to have a negative coefficient, indicating that a higher approach speed results in fewer accidents. This finding is probably the outgrowth of the high correlation between the approach speed limit and the opposing approach speed limit. In the most common case, the two speed limits are equal, which results in an effective coefficient of 0.0206 (the approach speed-limit coefficient minus the opposing approach speed-limit coefficient) multiplied by the shared speed limit of the approach and the opposing approach. This produces a net positive correlation between speed and accident frequency. The estimation of two separate coefficients for approach and opposing approach speed limits allows us to fully capture the interesting interaction between the two.

TABLE 3. Negative Binomial Estimation Results for Annual Rear-End Accident Frequency at Intersection Approaches

Variable (1)	Coefficient (t-statistic) (2)
Constant	−4.35 (−5.44)
Total intersection volume in thousands (average daily traffic)	0.0308 (5.04)
Right-turn volume in thousands (average daily traffic)	0.0986 (2.35)
Opposing left-turn volume in thousands (average daily traffic)	−0.0986 (−3.20)
Two-phase signal (1 if two-phase signal, 0 otherwise)	0.778 (3.97)
Eight-phase signal (1 if eight-phase signal, 0 otherwise)	0.549 (3.89)
Permissive left turn (1 if permissive left turn, 0 otherwise)	−0.986 (−4.97)
Intersection in central business district (CBD) (1 if intersection is in the CBD, 0 otherwise)	−0.409 (−2.00)
All approaches local streets (1 if all approaches local streets, 0 otherwise)	−1.74 (−2.25)
Highest approach speed limit at the intersection (km/h)	−0.0239 (−1.60)
Greater than 5% uphill or downhill grade on approach (1 if greater than 5% uphill or downhill grade on approach, 0 otherwise)	0.454 (3.34)
Left-turn pocket, through, and combined through-right on approach (1 if there is a left-turn pocket, through, and combined through-right on approach, 0 otherwise)	0.519 (3.70)
Restricted sight distance for left-turning vehicles due to a fixed object (1 if restricted sight distance for left-turning vehicles due to a fixed object, 0 otherwise)	2.312 (4.07)
Approach speed limit (km/h)	0.064 (4.77)
α (Negative binomial dispersion parameter)	0.319 (2.26)
Number of observations	1,385
Log-likelihood at zero	−1,468.10
Log-likelihood at convergence	−726.43
ρ^2	0.505

TABLE 4. Negative Binomial Estimation Results for Annual Angle Accident Frequency at Intersection Approaches

Variable (1)	Coefficient (t-statistic) (2)
Constant	−5.00 (−5.90)
Left-turn volume in thousands (average daily traffic)	0.209 (3.46)
Total opposing approach volume in thousands (average daily traffic)	0.0345 (1.62)
Number of approach lanes	0.319 (2.20)
Number of opposing lanes	−0.279 (−1.40)
Number of through, combined through-right, and right-turn lanes	0.489 (2.53)
No control on approach (1 if no control, 0 otherwise)	−2.136 (−4.75)
Signal control (1 if signal control, 0 otherwise)	−1.151 (−4.00)
Two-phase signal (1 if two-phase signal, 0 otherwise)	0.830 (4.10)
Eight-phase signal (1 if eight-phase signal, 0 otherwise)	0.667 (2.809)
Protected left turn (1 if protected left turn, 0 otherwise)	−0.662 (−2.88)
Permissive left turn (1 if permissive left turn, 0 otherwise)	−0.503 (−2.09)
All approaches local streets (1 if all approaches local streets, 0 otherwise)	−0.763 (−2.07)
Number of intersection approach legs	1.123 (4.35)
Combined through-left, or left-turn drop lanes and two or more lanes on the approach (1 if there are combined through-left, or left-turn drop lanes and two or more lanes on the approach, 0 otherwise)	−0.443 (−2.58)
Restricted sight distance on stop-controlled approach (1 if the approach is stop-controlled and has restricted sight distance, 0 otherwise)	1.621 (5.60)
Horizontal curve on opposing approach (1 if horizontal curve on the opposing approach, 0 otherwise)	0.313 (1.34)
Greater than 5% uphill or downhill grade on approach (1 if greater than 5% uphill or downhill grade on approach, 0 otherwise)	−0.251 (−1.34)
Opposing approach speed limit (km/h)	−0.0206 (−2.74)
α (Negative binomial dispersion parameter)	0.696 (3.61)
Number of observations	1,385
Log-likelihood at zero	−1,523.53
Log-likelihood at convergence	−826.27
ρ^2	0.458

Table 1 also shows that the use of the negative binomial model is justified by the highly significant value of α (t-statistic = 5.96). Use of the Poisson regression would have produced considerable bias in coefficient estimates.

Finally, a measure of overall statistical fit, the likelihood ratio index ρ^2 (similar to R^2 in regression analysis) is computed as

$$\rho^2 = 1 - \mathcal{L}(\boldsymbol{\beta})/\mathcal{L}(0) \qquad (10)$$

where $\mathcal{L}(\boldsymbol{\beta})$ = log-likelihood at convergence; and $\mathcal{L}(0)$ = log-likelihood at zero (i.e., all elements of the vector $\boldsymbol{\beta} = 0$). The value of 0.200 is quite satisfactory considering the variance in the data and that ρ^2 values tend to be generally lower than typical R^2 values [see Ben-Akiva and Lerman (1985)].

TABLE 5. Negative Binomial Estimation Results for Annual Approach-Turn Accident Frequency at Intersection Approaches

Variable (1)	Coefficient (t-statistic) (2)
Constant	−3.66 (−5.32)
Left-turn volume in thousands (average daily traffic)	0.297 (4.605)
Total opposing approach volume in thousands (average daily traffic)	0.0623 (3.13)
Opposing left-turn volume in thousands (average daily traffic)	−0.114 (−1.31)
Eight-phase signal (1 if eight-phase signal, 0 otherwise)	0.522 (2.52)
Permissive left turn (1 if permissive left turn, 0 otherwise)	1.973 (10.24)
Protective/permissive left turn (1 if protective/permissive left turn, 0 otherwise)	1.72 (8.96)
Local street approach (1 if local street approach, 0 otherwise)	−0.336 (−2.16)
Highest approach speed limit at the intersection (km/h)	−0.0266 (−2.02)
Sight-distance restriction (1 if sight distance is restricted, 0 otherwise)	1.764 (7.05)
Combined through-left, or left turn drop lanes and two or more lanes on the approach (1 if there are combined through-left, or left-turn drop lanes and two or more lanes on the approach, 0 otherwise)	0.614 (3.65)
Left turns not aligned and not a single-lane approach, protected left, or stop control (1 if left turns are not aligned and the approach does not have a single lane, protected left, or stop control, 0 otherwise)	−1.23 (−4.97)
Left turn restriction (1 if left turns are not permitted, 0 otherwise)	−1.789 (−1.74)
Horizontal curve on opposing approach (1 if horizontal curve on the opposing approach, 0 otherwise)	0.322 (1.475)
Opposing approach speed limit (km/h)	0.0368 (2.82)
α (Negative binomial dispersion parameter)	0.505 (3.74)
Number of observations	1,385
Log-likelihood at zero	−1,581.79
Log-likelihood at convergence	−732.93
ρ^2	0.537

In addition to the total accident-frequency model presented in Table 1, models of the frequency of rear-end, angle, and approach-turn accidents were estimated. The negative binomial estimation results for these models are presented in Tables 3, 4, and 5. Table 6 summarizes the findings of all four negative binomial accident-frequency models presented in this paper.

Most of the results summarized in Table 6 are reasonable and self-explanatory, and follow the aforementioned explanations for the total accident model. However, at least one finding requires some additional explanation. This is the finding that intersections in the central business district (CBD) have a lower likelihood of rear-end accidents, which we attribute to the signal progression typically provided for through vehicles in the CBD area. Such progression decreases the amount of stops and starts vehicles have to make, decreasing rear-end accident potential. In future studies, the possibility of using progression should be explored on intersection approaches that are in close proximity to other signalized intersections. Unfortunately, our data did not provide extensive signal-progression information. Such information would be quite valuable if we were analyzing all intersections in the city but, because our sample of intersections is limited to intersections that have been targeted for operational improvements, and these intersections are likely to be geographically dispersed, a simple CBD indicator is all we were able to find to be significant.

TABLE 6. Summary of Findings of Negative Binomial Annual Accident Frequency Models (Tables 1, A1, A2, and A3)

Variable (1)	Finding (2)
Approach left-turn volume	Increase in total, angle, and approach-turn accidents
Approach right-turn volume	Increase in total and rear-end accidents
Total opposing approach volume	Increase in total, angle, and approach-turn accidents
Total number of opposing approach lanes	Increase in total accidents
Combined left-through lanes on approaches with two or more lanes and combined left-through, or left-turn drop lanes and two or more lanes on the approach	Increase in total accidents and approach-turn accidents and decrease in angle accidents
No control on approach	Decrease in total and angle accidents
Signal control	Decrease in total and angle accidents
Two-phase signal	Increase in total, rear-end, and angle accidents
Eight-phase signal	Increase in total, rear-end, and approach-turn accidents
Protected/permissive left turn	Increase in total and approach-turn accidents
Sight-distance restriction	Increase in total and approach-turn accidents
Restricted sight-distance on stop-controlled approach	Increase in angle accidents
Local street approach	Decrease in total and approach-turn accidents
All approaches local streets	Decrease in total, rear-end, and angle accidents
Through-left lane and two or more lanes on approach	Increase in total accidents
Horizontal curve on approach	Increase in total accidents
Approach speed limit	Increase in total and rear-end accidents
Intersection in the central business district (CBD)	Decrease in rear-end accidents
Greater than 5% uphill or downhill grade on approach	Increase in rear-end and angle accidents
Number of intersection approaches	Increase in angle accidents
Horizontal curve on opposing approach	Increase in angle and approach-turn accidents
Permissive left turn on approach	Increase in approach-turn accidents
Left-turn restriction on approach	Decrease in approach-turn accidents

SPECIFICATION ISSUES AND TESTS

The use of multiple observations from each intersection (caused by splitting the intersection into its approaches) and repeated observations from each intersection approach (caused by using data from different years) could give rise to a potentially serious model specification concern. That is, the gamma error term in the negative binomial model could be correlated from one observation to the next, which is a violation of the error-term independence assumption made to derive the model. The consequence of nonindependence of error terms is a loss in estimation efficiency (i.e., standard errors of estimated coefficients will become larger), and this could lead one to draw erroneous conclusions regarding coefficient estimates. To test for the extent of this problem in our models, we conduct a series of likelihood ratio tests. The basic idea of the tests is to segment the sample into subsets of data that are less likely to be afflicted correlation problems. If these smaller data subsets produce model estimation results that are not significantly different from those estimation results produced by the overall data sample, it can be concluded that any independence violations are not significantly affecting model results [see Ben-

Akiva and Lerman (1985), and Mannering and Winston (1991) for applications of this approach].

Formally, the test begins by segmenting the data into subsets $g = 1, \ldots, G$ (where G = total number of data subsets) so that

$$N = \sum_{g=1}^{G} N_g \quad (11)$$

where N = full sample size; and N_g = size of the gth subset. For testing purposes the likelihood ratio statistic is (Ben-Akiva and Lerman 1985)

$$-2\left[\mathcal{L}_N(\boldsymbol{\beta}) - \sum_{g=1}^{G} \mathcal{L}_{N_g}(\boldsymbol{\beta}^g)\right] \quad (12)$$

where $\mathcal{L}_N(\boldsymbol{\beta})$ = log-likelihood at convergence of the model estimated on all data N with a single coefficient vector $\boldsymbol{\beta}$; and $\mathcal{L}_{N_g}(\boldsymbol{\beta}^g)$ = log-likelihood at convergence of the model estimated on the gth subset of the data. This test statistic is χ^2 distributed with the degrees of freedom equal to

$$\sum_{g=1}^{G} K_g - K \quad (13)$$

where K_g = number of coefficients in the gth data subset model; and K = number of coefficients in the full-sample model.

The first test we conduct focuses on the possible correlation resulting from using separate approaches from the same intersection. To evaluate the effect this possible correlation is having on our estimation results, we split the data into four subsets: one each for northbound, southbound, eastbound, and westbound approaches. Such data segmentation ensures that, in each data subset, there is only one observation from each intersection. Using all accident types and 18 variables (see Table 1), the log-likelihoods at convergence for the northbound, southbound, eastbound, and westbound models are −445.43, −385.44, −420.98, and −422.81, respectively. Applying (12), with the full-sample log-likelihood being −1,698.26 (see Table 1) at convergence, gives a χ^2 value of 47.2. With 54 degrees of freedom [72 − 18, (13)], the corresponding p-value is 0.732. This means that we can only be about 27% confident that the coefficients estimated on the data subsets are significantly different from the full-sample coefficients. This is compelling evidence that possible correlation among intersection approaches is not significantly affecting our estimation results.

To test for possible correlation problems from one year to the next (which could result from the use of annual expansion factors to estimate traffic volumes, among other factors), we split the data into six subsets: one each for 1987, 1988, 1989, 1990, 1991, and 1992–93 (combined to obtain model convergence). Again, using all accident types and 18 variables (see Table 1), the log-likelihoods at convergence for the 1987, 1988, 1989, 1990, 1991, and 1992–93 models are −270.77, −218.60, −259.96, −219.76, −264.51, and −427.54, respectively. Applying (12), with the full-sample log-likelihood being −1,698.26 at convergence, gives a χ^2 value of 74.24. With 90 degrees of freedom [108 − 18, (13)] the corresponding p-value is 0.885. This means that we can only be about 12% confident that the coefficients estimated on the data subsets are significantly different from the full-sample coefficients. This shows that possible correlation from year to year is not significantly affecting our estimation results.

We also ran likelihood ratio tests on approach- and year-data subsets for the rear-end, angle, and approach-turn accident-frequency models shown in Tables 3, 4, and 5, respectively. The findings were the same and the possibility of coefficients varying on the data subsets could not be accepted at reasonable confidence levels (the highest was 22%). This indicates that possible correlations are not significantly affecting these models either.

Unfortunately, data limitations do not allow us to test for possible correlations among approaches and years simultaneously (i.e., at most we would have 63 observations in such data subsets, one for each intersection). However, the likelihood ratio tests that we were able to conduct provide strong evidence that such correlations are not playing a significant role in our estimations. This finding suggests that our models are reasonably well-specified (i.e., no major omitted variables that would contribute to correlations in error terms).

Another concern relates to the assumed log-linear relationship of accidents and explanatory variables [i.e., (4)]. From a practical point of view, this function form is necessary to estimate the likelihood function because the possibility of negative values can not be handled in the probability expression [(1)]. Still, the suitability of this form can be questioned. To informally test this, we regress the number of accidents on explanatory variables in a log-linear and a simple linear form using ordinary least-squares regression (omitting zero-accident observations because the log of zero is not defined). The estimation results showed an R^2 of 0.2186 for the log-linear model and 0.2129 for the linear model. Based on this informal test, there seems to be little difference between linear and log-linear forms.

SUMMARY OF FINDINGS

The results presented in this paper (Tables 1, 3, 4, and 5) demonstrate that negative binomial regression modeling of intersection approach accidents can be used to identify significant traffic and geometric elements that tend to increase or decrease accident frequency. The identification of these elements can be used to further understand the cause of intersection approach accidents, both in total and of specific type.

Understanding the elements that tend to increase or decrease intersection approach accidents can be beneficial to accident reduction in two ways. First, this knowledge can be applied during the project development stage of new intersection construction or intersection rebuilding through capital programs. Identifying and removing the elements from intersection design that increase accidents can be most easily accomplished in the plan review stages. Second, this knowledge can be used during periodic comprehensive high intersection-accident location reviews. As field reviews of individual high accident intersections are conducted, the reviewer can be looking for the presence of elements known to increase accidents at the intersection approach. Once these elements are identified, the feasibility of revising the approach can be determined and a project to correct the concern can be programmed and implemented.

CONCLUDING REMARKS

The high cost of traffic accidents to society clearly justifies the adoption of accident-reduction programs. Such programs, however, need some empirical basis to guide the allocation of funds to the most needy projects. This paper has attempted to provide some exploratory work in this regard.

The results obtained here by exploring the traffic and geometric variables that affect intersection accidents (using a negative binomial regression of accident frequency) provide an important first step toward developing a systematic, statistically defensible approach to identifying the impacts of possible intersection improvements. In terms of future work, an analysis similar to the one performed in this paper should be conducted on a large random sample of intersections. Such a sample would allow further exploration of the determinants of intersection-accident frequencies. Moreover, it would permit one to study different formulations of the accident frequency variable (e.g., a comparison of assigning accidents to an intersection versus assigning accidents to an intersection approach). Information from such work would provide a significant contribution to the accident-analysis field.

APPENDIX I. REFERENCES

Ben-Akiva, M., and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand.* MIT Press, Cambridge, Mass.

"Estimating the cost of accidents 1993." (1994). *Rep.*, National Safety Council Statistics Dept., Itasca, Ill.

Greene, W. H. (1983). *Economic analysis.* Macmillan Publishing Co., New York, N.Y.

Hammerslag, R., Roos, J. P., and Kwakernaak, M. (1982). "Analysis of accidents in traffic situations by means of multiproportional weighted Poisson model." *Transp. Res. Rec. 847*, Transp. Res. Board, Washington, D.C., 29–36.

Jones, B., Janssen, L., and Mannering, F. (1991). "Analysis of the frequency and duration of freeway accidents in Seattle." *Accident Anal. and Prevention*, 23(4), 239–255.

Joshua, S. C., and Garber, N. J. (1990). "Estimating truck accident rate and involvements using linear and Poisson regression models." *Transp. Plng. and Technol.*, 15(1), 41–58.

Jovanis, P., and Chang, H. L. (1986). "Modeling the relationship of accidents to miles traveled." *Transp. Res. Rec. 1068*, Transp. Res. Board, Washington, D.C., 42–51.

Mannering, F. L., and Kilareski, W. P. (1990). *Principles of highway engineering and traffic analysis.* John Wiley & Sons, Inc., New York, N.Y.

Mannering, F. L., and Winston, C. (1991). "Brand loyalty and the decline of American automobile firms." *Brookings Papers on Economic Activity: Microeconomics 1991*, Brookings Instn., Washington, D.C., 67–114.

Mannering, F. L., and Winston, C. (1995). "Automobile airbags in the 1990s: Market failure or market efficiency?" *The J. of Law and Economics*, 38(2), 265–279.

Massie, D. L., Campbell, K. L., and Blower, D. F. (1993). "Development of a collision typology for evaluation of collision avoidance strategies." *Accident Anal. and Prevention*, 25(3), 241–257.

Miaou, S.-P., and Lum, H. (1993). "Modeling vehicle, accidents and highway geometric design relationships." *Accident Anal. and Prevention*, 25(6), 689–709.

Okamoto, H., and Koshi, M. (1989). "A method to cope with random errors of observed accident rates in regression analysis." *Accident Anal. and Prevention*, 21(3), 317–332.

Persaud, B. N. (1991). "Estimating accident potential of Ontario road sections." *Transp. Res. Rec. 1327*, Transp. Res. Board, Washington, D.C., 47–53.

Sabey, B., and Taylor, H. (1980). "The known risks we run: the highway." *Supplementary Rep. SR 567*, Transport and Road Res. Lab., Crowthorne, U.K.

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

Treat, J. R. (1990). "A study of precrash factors involved in traffic accidents." *Res. Rev.*, 10(6)/11(1).

Wong, Y. D., and Nicholson, A. (1992). "Driver behavior at horizontal curves: risk compensation and the margin of safety." *Accident Anal. and Prevention*, 24(4), 425–436.

APPENDIX II. NOTATION

The following symbols are used in this paper:

$E[\,]$ = expected value;
e = elasticity;
G = total number of data subsets;
K = number of explanatory variables;
$L(\,)$ = likelihood function;
$\mathcal{L}(\,)$ = log-likelihood;
N = sample size;
n = number of accidents;
$P(\,)$ = probability of having a specified number of accidents;
p = probability value from a one-tailed t-test;
u = $\theta(\theta + \lambda_i)$;
$\mathbf{X}$ = vector of explanatory variables;
x = explanatory variable;
α = variance of gamma-distributed error term (negative binomial dispersion parameter);
$\boldsymbol{\beta}$ = vector of estimable coefficients;
$\Gamma(\,)$ = refers to the gamma distribution;
∂ = partial differential;
ε = error term;
θ = inverse of negative binomial dispersion parameter;
λ = Poisson parameter;
ρ = likelihood ratio index; and
χ = refers to the chi-squared distribution.

Subscripts

g = data subset indicator;
i = indicator for intersection approach; and
k = indicator for an explanatory variable.