Final Report VTRC 08-R1

Virginia Transportation Research Council

research report

Development of a Safety Evaluation Procedure for Identifying High-Risk Signalized Intersections in the Virginia Department of Transportation's Northern Virginia District

http://www.virginiadot.org/vtrc/main/online_reports/pdf/08-r1.pdf

YOUNG-JUN KWEON, Ph.D.
Associate Research Scientist

Virginia Transportation Research Council, 530 Edgemont Road,
Charlottesville, VA 22903-2454, www.vtrc.net, (434) 293-1900

Standard Title Page - Report on Federally Funded Project

1. Report No. FHWA/VTRC 08-R1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Development of a Safety Evaluation Procedure for Identifying High-Risk Signalized Intersections in the Virginia Department of Transportation's Northern Virginia District		5. Report Date September 2007 6. Performing Organization Code
7. Author(s) Young-Jun Kweon, Ph.D.		8. Performing Organization Report No. VTRC 08-R1
9. Performing Organization and Address Virginia Transportation Research Council 530 Edgemont Road Charlottesville, VA 22903		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 80576
12. Sponsoring Agencies' Name and Address Virginia Department of Transportation 1401 E. Broad Street Richmond, VA 23219	Federal Highway Administration 400 North 8th Street, Room 750 Richmond, VA 23219-4825	13. Type of Report and Period Covered Final 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This research was undertaken to develop an evaluation procedure to identify high-risk four-legged signalized intersections in VDOT's Northern Virginia district by traffic movements and times of day. By using the developed procedure, traffic engineers are expected to be able to identify signalized intersections where the traffic crash occurrences under different traffic conditions for different times of day are more frequent than would normally be expected.

Using generalized linear models such as negative binomial models, one safety performance function was estimated for each of nine crash population reference groups formed by three traffic crash patterns (crash patterns 1, 4, and 6) and four times of day (A.M. peak, mid day, P.M. peak, and evening off peak). Crash pattern 1 is a same-direction crash (rear-end, sideswipe or angle crash) that occurs after exiting the intersection; crash pattern 4 is a right-angle crash between two adjacent straight-through vehicle movements in the intersection; and crash pattern 6 is an angle or head-on or opposite sideswipe crash between a straight-through vehicle movement and an opposing left-turn vehicle movement in the intersection.

The procedure developed in this study is based on the empirical Bayes (EB) method. Additional data do not need to be collected in order to use the EB procedure because all the data required for applying the EB procedure should be obtainable from VDOT's crash database and from Synchro input data that are already available to traffic engineers for traffic signal phase plans. Thus, the EB procedure is cost-effective and readily applicable. For easy application of the EB procedure, an EB spreadsheet was developed using Microsoft Excel, and a users' guide was prepared. These are available from the author upon request.

17 Key Words Crash analysis, empirical Bayes, high risk, identification, evaluation, intersection		18. Distribution Statement No restrictions. This document is available to the public through NTIS, Springfield, VA 22161.	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 78	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

FINAL REPORT

DEVELOPMENT OF A SAFETY EVALUATION PROCEDURE FOR IDENTIFYING HIGH-RISK SIGNALIZED INTERSECTIONS IN THE VIRGINIA DEPARTMENT OF TRANSPORTATION'S NORTHERN VIRGINIA DISTRICT

Young-Jun Kweon, Ph.D.
Associate Research Scientist

Virginia Transportation Research Council
(A partnership of the Virginia Department of Transportation
and the University of Virginia since 1948)

In Cooperation with the U.S. Department of Transportation
Federal Highway Administration

Charlottesville, Virginia

September 2007
VTRC 08-R1

DISCLAIMER

The contents of this report reflect the views of the author, who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Virginia Department of Transportation, the Commonwealth Transportation Board, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

EXECUTIVE SUMMARY

Introduction

A traffic signal is commonly installed to address traffic operations at intersections, and it plays an important role in achieving traffic safety at intersections. The Virginia Department of Transportation (VDOT) has a process to identify and ameliorate safety deficiencies at signalized intersections. By improving the physical design and signal phases of intersections, VDOT contributes to a safer environment for drivers and non-motorized users at these intersections. However, VDOT's funds for safety improvements are limited, and not all identified safety deficiencies can be mitigated in one fiscal year. In order to maximize the impact of safety improvements using limited resources, VDOT needs to identify which intersections require more attention for safety improvements based on traffic crash risk measures such as traffic crash frequency and rate.

At present, traffic engineers in Virginia have no easy method to determine quickly whether a particular intersection is associated with an unusually high crash risk. For example, VDOT's Northern Virginia (NOVA) District staff cannot easily determine which intersections with a permissive or protected left-turn signal phase are operated under a high risk of traffic crashes. Thus, they do not know which intersections should be studied in greater detail to identify potential crash countermeasures. To meet this need, a procedure is needed to determine whether a particular signalized intersection suffers from an unusually large number of crashes. Such a procedure would need to be flexible enough to accommodate various real intersection conditions such as continuous volume range.

Purpose and Scope

This project was undertaken to develop a procedure to identify high-risk signalized intersections in Virginia whereby traffic engineers could identify an intersection where traffic crash occurrences were more frequent than would normally be expected taking into account different traffic movements and times of day. The scope of this project was limited to traffic safety evaluations of four-legged signalized intersections in VDOT's NOVA District.

Methods

The method used to achieve the research objectives is presented in Figure ES-1. The research was divided into three stages. In Stage 1, data collection and preparation and preliminary data analysis were performed. In Stage 2, traffic crash prediction models including mean and variance models were developed. In Stage 3, the empirical Bayes (EB) method was applied using the models developed in Stage 2.

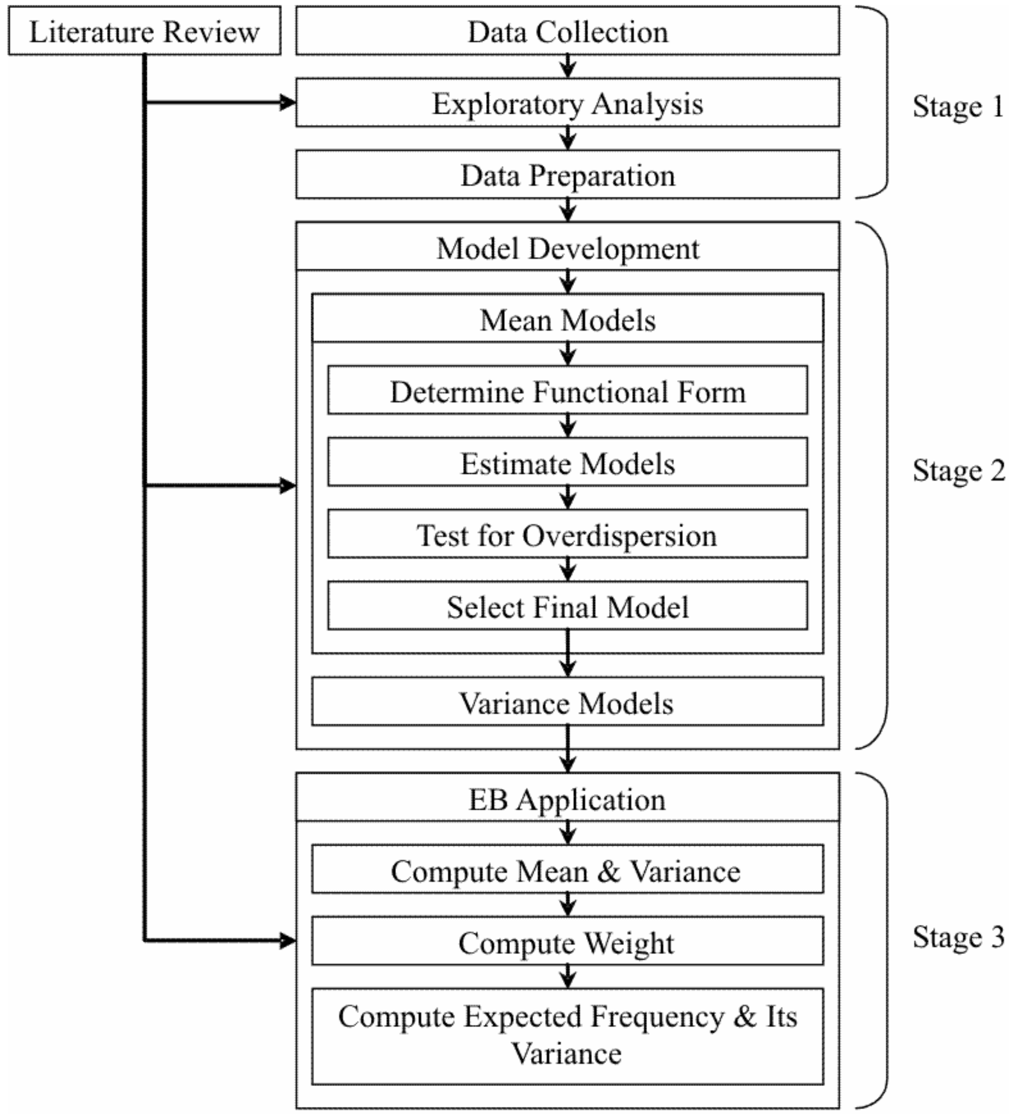


Figure ES-1. Method Used to Develop the Procedure

Although it might be tempting to select the intersections with the highest crash rate from previous years for safety improvement, the high crash rate at those intersections might be attributable to chance. To ensure that intersections are identified where the high crash rate is attributable to contributing geometric and/or human factors (e.g., poor visibility, speeding, poor channelization, etc.) and not to chance and/or bias (e.g., regression-to-mean bias or sample selection bias), an expected crash frequency of the intersection should be computed using the three stages proposed here. When the observed number of crashes exceeds the confidence limit of the expected number of crashes, it is likely that the high crash risk is not attributable to chance.

In Stage 1, an exploratory analysis of potentially useful variables was performed after traffic crash, traffic flow, and signal phase data were collected. The main goal of the analysis was to classify traffic crashes properly by defining traffic crash patterns for model development; descriptive statistics, histograms, cross-tabulations, and regression models were used for the analysis. A crash occurrence mechanism (i.e., relationship between crash occurrence and a given condition) is unlikely to be the same across different patterns of traffic crashes. For example, single-vehicle crashes are different from multi-vehicle crashes in a crash occurrence mechanism. In general, as traffic volume increases, single-vehicle crashes decrease and multi-vehicle crashes increase. Therefore, these two crash patterns should be analyzed separately.

In addition to crash patterns, time of day was believed to affect the crash occurrence mechanism because traffic patterns and travel purposes vary across four times of day (i.e., A.M. peak, mid day, P.M. peak, and evening off peak). For example, commuting trips occur during A.M. and P.M. peak periods whereas non-commuting trips occur during mid day and evening off peak periods. Therefore, for the purpose of model development, time of day was also used to form a crash population reference group, which is formed by a combination of different crash patterns and different times of day. At the end of Stage 1, a separate dataset was prepared for each crash population reference group.

In Stage 2, using the datasets prepared in Stage 1, mean and variance models were developed for each crash population reference group. For the mean model, a proper relationship between crash frequency and traffic flows was determined for each group. Between two types of count response models, Poisson assuming equidispersion and negative binomial (NB) assuming overdispersion, an appropriate model was selected through tests for overdispersion. After a final mean model was estimated, a final variance model was developed.

In Stage 3, the final mean and variance models from Stage 2 served as inputs to the EB method to produce the EB estimates of the expected crash frequency and its variance.

Results

EB Procedure

A 10-step EB procedure was developed in this study.

Step 1. Select a crash pattern for safety evaluation from crash patterns 1, 4, and 6. These crash patterns are explained in Table 3 in the full report.

Step 2. Select a time of day for the safety evaluation: A.M. peak, mid day, P.M. peak, or evening off peak. For crash pattern 4, the period from the beginning of A.M. peak until the end of evening off peak should be used.

Step 3. Determine a crash population reference group from groups 1 through 9. A crash population reference group is automatically determined when a crash pattern and a time of day are selected. The nine crash population reference groups are explained in Table 4 in the full report.

Step 4. Collect the data required for the selected crash population reference group. The data required for each of the nine crash population reference groups are listed in Table 5 in the full report.

Step 5. Select the correct models. The final mean and variance models for each of the nine crash population reference groups are listed in Table 6 in the full report.

Step 6. Calculate the mean and variance using the selected models. The calculation should be done for each of the four pairs of conflict vehicle movements. For example, for crash pattern 6 (a collision between straight-through traffic and opposing left-turning traffic), there are four pairs of such conflict movements at a four-legged intersection: (1) northbound straight-through and southbound left-turn, (2) southbound straight-through and northbound left-turn, (3) eastbound straight-through and westbound left-turn, and (4) westbound straight-through and eastbound left-turn.

Entering required inputs into the selected mean and variance models will produce estimates of the mean and variance of the expected crash frequency for each of the four pairs of conflict movements. The input values should be within the valid ranges presented in Table 7 in the full report. Because all models were developed using the data within the ranges shown in Table 7, the results from the models will be valid only when the input values fall within the specified ranges. Although results can still be obtained using inputs outside the ranges, the validity of the results will be questionable.

Step 7. Calculate the EB weight. An EB weight is calculated for each of the four pairs of conflict movements using the following equation:

$$\hat{\omega}_i = \frac{1}{1 + \dfrac{\hat{V}(\kappa_i)}{\hat{E}(\kappa_i)}}$$

where i indexes the four pairs of conflict movements (i = 1, 2, 3, and 4; see step 6 for an example) and $\hat{E}(\kappa_i)$ and $\hat{V}(\kappa_i)$ are estimates of the mean and the variance, respectively, from step 6.

Step 8. Calculate the expected crash frequency. An expected crash frequency is calculated for each of the four pairs of conflict movements using the following equation:

$$\hat{\kappa}_i = \hat{E}(\kappa_i \,|K_i) = \hat{\omega}_i \cdot \hat{E}(\kappa_i) + (1 - \hat{\omega}_i) \cdot K_i$$

where K is the number of recorded crashes of the specified crash pattern in the past 4 years during the specified time period.

Step 9. Calculate the variance of the expected crash frequency. A variance of the expected crash frequency is calculated for each of the four pairs of conflict movements using the following equation:

$$\hat{\sigma}_i^2 = \hat{V}(\kappa_i \mid K_i) = \begin{cases} \hat{\kappa}_i, & \text{if Poisson} \\ (1-\hat{\omega}_i)\cdot\hat{\kappa}_i, & \text{if NB} \end{cases}$$

Step 10. Calculate the expected crash frequency and its variance for an intersection. An expected crash frequency and its variance for an entire intersection are calculated by summing the expected crash frequencies and their variances over the four pairs of conflict movements:

$$\hat{\kappa} = \sum_{i=1}^{4} \hat{\kappa}_i \text{ and } \hat{\sigma}^2 = \sum_{i=1}^{4} \hat{\sigma}_i^2$$

Note that independent variances are assumed for the summation.

Use of the EB Procedure by Traffic Engineers

Traffic engineers can apply the EB procedure to identify high-risk intersections (and high-risk conflicting traffics within such intersections). Safety evaluations can be conducted at an intersection level using the expected crash frequency and its variance from step 10. An evaluation can also be conducted at a conflict movement level within an intersection using the expected crash frequency and its variance from steps 8 and 9. To allow a better understanding of the EB procedure, an example of such an evaluation is provided in the full report.

Conclusions

- *The EB procedure developed in this study can be used by traffic engineers to evaluate the safety of a four-legged signalized intersection.* Traffic engineers can follow the procedure using field data and will obtain the expected crash frequency and its variance for different crash patterns and different times of day. By using fundamental statistical methods such as a confidence interval or a hypothesis test, traffic engineers can determine whether the intersection of interest is associated with an abnormally high crash risk.

- *Additional data do not need to be collected in order to apply the EB procedure.* Because all the data required for applying the EB procedure should be obtainable from VDOT's crash database and from Synchro input data that are already available to traffic engineers for traffic signal phase plans, the EB procedure is cost-effective and readily applicable.

- *The EB procedure is valid for use with only four-legged signalized intersections in VDOT's NOVA District within the valid input ranges.* The data used to develop the estimated mean and variance models in Table 6 in the full report were collected from four-legged signalized intersections in the district. If the intersection geometry, traffic patterns, and driver behavior were similar to those in VDOT's NOVA District, the EB procedure might be usable for other areas. However, the results for such areas might not be valid; a proper validation process using local data would be necessary to confirm the results.

- *The EB procedure may not be very useful for some of the nine crash population reference groups.* Traffic crashes for some crash population reference groups, such as reference group 9 (i.e., crash pattern 6 during the evening off peak period), were rare during the 4-year data period. The expected crash frequency for such reference groups would be less than 1 crash per 4 years over the entire range of input values. Thus, even 1 crash in 4 years is likely to lead to a conclusion that an intersection is associated with an abnormally high crash risk (e.g., reference groups 6 and 9, corresponding to crash pattern 6 in mid day and off peak periods, respectively, as shown in the EB case study in the full report and Appendix D).

- *An EB Spreadsheet, which aids in the application of the EB procedure, and a users' guide were developed.* For easier application of the EB procedure, an EB spreadsheet was developed using Microsoft Excel, and a users' guide was prepared. They are available from the author upon request.

Recommendations

1. *VDOT's Information Technology Division (IT Division), VTRC, and VDOT's NOVA District should facilitate the application of the developed EB procedure for the NOVA District.* Although the EB procedure is not difficult for traffic engineers to follow, it can be cumbersome and time-consuming for them to apply to the many intersections that would need to be evaluated for traffic safety. Thus, the IT Division, VTRC, and the NOVA District should collaborate to automate the application of the EB procedure to assess the safety of four-legged signalized intersections in the NOVA District.

   - *The IT Division should integrate data for calibration of the EB procedure and automate the application of the EB procedure.* The IT Division should extract the necessary data (i.e., traffic volumes, left-turn signal types, times of data, traffic crash characteristics and counts, and vehicle information) from Synchro files, time-based coordinate event sheets, and VDOT's crash database and integrate them into datasets in a format suitable for calibrating the EB procedure. After the calibration is done by VTRC, the IT Division should automate the application of the calibrated EB procedure for the NOVA District.

   - *VTRC should calibrate the EB procedure using the datasets that the IT Division integrates.* Using the datasets that the IT Division integrates, VTRC should calibrate all the model parameters embedded in the EB procedure. In addition, it should develop new models if necessary to enhance the reliability and accuracy of the EB procedure.

   - *The NOVA District should provide assistance to the IT Division and VTRC.* In the process of data integration and/or procedure calibration, practical insights and local information will more than likely be needed from the NOVA District.

2. *VTRC and VDOT's NOVA District should update the EB procedure when traffic characteristics of the four-legged signalized intersections change.* The EB procedure is based on the data believed to represent the prevailing traffic characteristics of the four-legged signalized intersections in the NOVA District during the years from 2001 through 2004. Intersection geometry, traffic patterns, and driver behaviors continue to change over time; as a consequence, the traffic characteristics influencing crash occurrence change. Thus, when the traffic characteristics of these intersections become significantly different from those used in this study, the EB procedure should be updated using newly collected data representing contemporary prevailing traffic characteristics. There are no established criteria for determining when the results should be updated. Engineers' judgment will certainly play a major role in such a determination.

Costs and Benefits Assessment

This study provided an explicit procedure whereby traffic engineers in VDOT's NOVA District can quickly evaluate the safety of four-legged signalized intersections. Such an intersection carrying a crash risk higher than normally expected can be identified by following the EB procedure with input data.

Using the EB procedure, traffic engineers can identify not only which intersections carry a high risk but also what traffic movements at the intersection and which time of day carry a high crash risk for the intersection. Thus, they can focus only on the identified movements and time of day to improve the safety of the intersection. In addition, when a site visit to the identified high-risk intersection is needed, the most appropriate time for the visit (e.g., A.M. peak or P.M. peak) can be identified using the results from the EB procedure.

Moreover, the EB procedure does not require additional data collection efforts as long as Synchro input data are available, which is common for signalized intersections in Virginia. Use of the EB procedure is likely to save a considerable amount of time and cost involved with field data collection whenever VDOT's NOVA District conducts a safety evaluation of its four-legged signalized intersections.

If the entire procedure from data preparation to application of the EB procedure were automated, traffic engineers could instantly assess the safety of the four-legged signalized intersections at any time just by choosing intersections of interest without manually entering the input. Moreover, calibrating the models and updating the results should be much more efficient and much less time-consuming. When the automated process is established, development and application of the EB procedure can be readily achieved by other VDOT districts as long as the appropriate data are provided.

FINAL REPORT

DEVELOPMENT OF A SAFETY EVALUATION PROCEDURE FOR IDENTIFYING HIGH-RISK SIGNALIZED INTERSECTIONS IN THE VIRGINIA DEPARTMENT OF TRANSPORTATION'S NORTHERN VIRGINIA DISTRICT

Young-Jun Kweon, Ph.D.
Associate Research Scientist

INTRODUCTION

Installing a traffic signal is a common measure used to address traffic operations at intersections. The signal serves two main purposes: (1) it increases efficiency by maximizing throughput of traffic volumes exiting an intersection, and (2) it enhances safety by spatially and/or temporally separating conflicting vehicular movements at the intersection. Traffic signals play an important role in making intersections safe.

Each year, however, a significant portion of traffic crashes occurs at intersections with traffic signals. According to the National Highway Traffic Safety Administration (NHTSA) (2005), in 2003, intersection and intersection-related traffic crashes made up about 41% of total crashes and about 46% of fatal and injury crashes in the United States. Of these, half occurred at signalized intersections. Further, according to the Virginia Department of Motor Vehicles (2006), 153,849 crashes occurred in Virginia in 2005 (875 fatal crashes and 55,041 injury crashes), and of these, 19.1% occurred at signalized intersections.

The Virginia Department of Transportation (VDOT) has a process to identify and ameliorate safety deficiencies at signalized intersections. By improving the physical design and signal phasing of intersections, VDOT contributes to a safer environment for drivers and non-motorized users at signalized intersections. However, VDOT's funds for safety improvements are limited, and not all identified safety deficiencies can be mitigated in one fiscal year. In order to maximize the impact of safety improvements using limited budgets and resources, VDOT needs a method to identify which intersections require more attention for safety improvements.

In 1991, VDOT's Traffic Engineering Division developed tables of expected traffic crash numbers for VDOT's nine districts (VDOT, 1991). Expected crashes (per 3 years) were estimated for three volume categories (less than 10,000; 10,000 to 20,000; and more than 20,000 total entering vehicles per day) and four types of intersections (three- or four-legged signalized or unsignalized intersections). Although these tables could be used to help traffic engineers identify intersections with unusually high crash risk, they are likely to be inappropriate for intersections in areas that have seen a large growth in traffic volume and/or considerable changes in traffic patterns (e.g., lane use and traffic concentration during peak hours) since 1991. For example, among about 1,200 signalized intersections maintained by VDOT's Northern Virginia (NOVA) District, the traffic characteristics of many intersections have changed significantly, as they have in most urban areas in other districts.

Thus, traffic engineers in the NOVA District have no easy method for quickly determining whether a particular signalized intersection carries an unusually high crash risk. Such a procedure is needed, and the procedure needs to be flexible enough to accommodate various real intersection conditions, such as continuous traffic volume.

PURPOSE AND SCOPE

The purpose of this project was to develop a procedure to identify high-risk signalized intersections in Virginia whereby traffic engineers could identify an intersection where traffic crash occurrences were more frequent than would normally be expected taking into account different traffic movements and times of day.

The scope of this project was limited to traffic safety evaluations of four-legged signalized intersections in VDOT's NOVA District.

METHODS

Overview

The method used to achieve the research objectives is presented in Figure 1. As shown, the method was composed of three stages. Stage 1 included data collection, initial data analysis, and data preparation for model development. Stage 2 included the development of mean and variance regression models of traffic crashes using SAS 9.1. Stage 3 included application of the empirical Bayes (EB) procedure using the final mean and variance models developed in Stage 2.

In Stage 1, an exploratory analysis was performed after the data were collected. The main goal of the analysis was to classify traffic crashes properly by defining traffic crash patterns for model development. Because a crash occurrence mechanism is unlikely to be the same for different patterns of traffic crashes, traffic crashes should be analyzed separately for different crash patterns. For example, single-vehicle crashes are different from multi-vehicle crashes in a crash occurrence mechanism. Therefore, these two crash types should be investigated separately.

Time of day was believed to affect the crash occurrence mechanism. Therefore, in conjunction with traffic patterns, it was used to form a crash population reference group for the purpose of model development. A separate dataset was prepared for each crash population reference group for model development.

In Stage 2, once the dataset was prepared for each crash population reference group in Stage 1, mean and variance models were developed for each group. For the mean model, a proper relationship (i.e., functional form) between traffic crash frequency and traffic flow variables was determined. Between two types of count response models, Poisson assuming

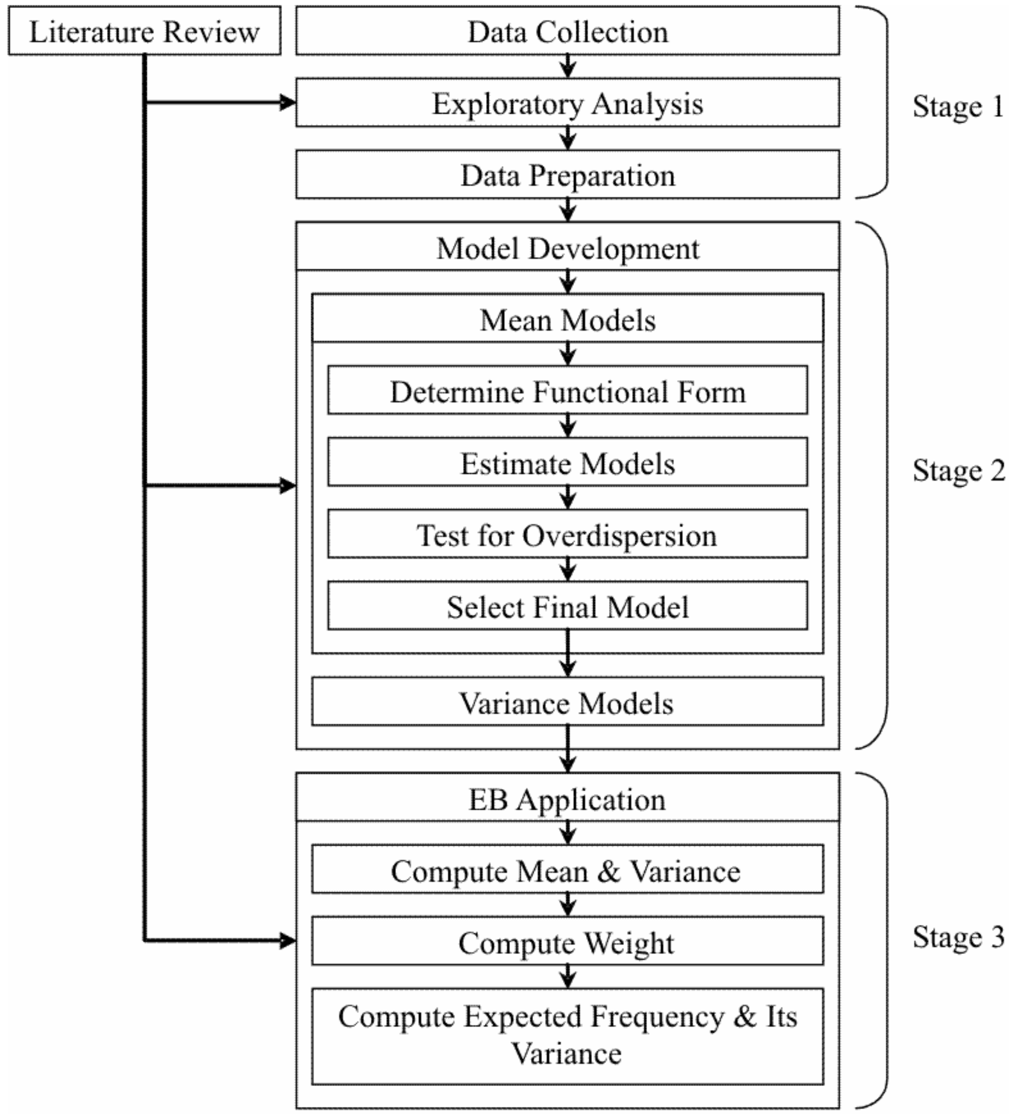


Figure 1. Research Method

equidispersion and negative binomial (NB) assuming overdispersion, the most appropriate model was selected through tests for overdispersion. After a final mean model was developed, a final variance model was developed.

In Stage 3, with the final mean and variance models at hand from Stage 2, the EB method was applied to produce the EB estimates of the expected crash frequency and its variance.

Literature Review

The author surveyed the traffic crash modeling literature pertaining to intersection crash analysis and the application of the EB method. Hakkert and Mahalel (1978) related the number of traffic crashes at intersections and the new measure of traffic volumes using a linear relationship between the number of conflict points and traffic crash counts suggested by

Schaechterle et al. (1970) and Matson et al. (1955). Their unit of analysis was an intersection. Interestingly, they used the index of traffic flows as the measure of traffic volumes, which was the sum of the products of the traffic volumes at each of the 24 vehicular conflict points at a four-legged intersection. The data included traffic flows at 16 daytime hours on a weekday and traffic crashes at 202 urban intersections and 20 inter-urban intersections from 1971 through 1972 in Israel.

Ceder and Livneh (1982a,b) examined the relationship between crash density and rate and hourly traffic flow in a power functional form, which was found to be proper by their previous study in 1978 using average daily traffic for single-vehicle and multi-vehicle crashes under free-flow and congested flow conditions. They developed models using time-sequence and cross-sectional data collected for 8 years from eight four-lane roadway segments in Israel.

Hauer et al. (1988) focused on intersections, and they analyzed traffic crashes by traffic flow movements. They argued that only traffic flows that are involved in a crash should be included in a crash analysis. They categorized two-vehicle traffic crashes at four-legged intersections into 15 crash patterns based on the movements of two vehicles preceding a collision. Manually collected hourly traffic volumes were entered into the exponential mean function of the NB models in two forms: log-transformed flow without and with the restriction of its coefficient being equal to 1. As Lau and May (1989) mentioned, the practical use of this study is questionable mainly because of the data requirements. This type of study requires accurate hourly traffic volumes by turning movements, and most agencies might not have such data or the resources to collect them.

Mountain and Fawaz (1996) estimated traffic crash frequencies at intersections using readily available input data including total entering traffic volumes for major and minor roads, control type (e.g., priority control, traffic signal, and roundabout), number of approaches, and speed limits. A total of 622 intersections with 111 traffic signal intersections in England were included in the study. Among several alternative functional forms in the NB model, the cross-product model with separate exponents for the major and minor entering flows produced the best fit to the data.

Persaud and Nguyen (1998) studied crashes at signalized intersections by injury severity, the number of vehicles in a crash, and peak time periods using data from Ontario, Canada. They performed two levels of analysis: aggregate level and disaggregate level. For the disaggregate level, they employed the classification of crash types similar to that of Hauer et al. (1988) yet classified crashes into 25 types by turning movements. Since they did not collect enough data at the disaggregate level by peak time periods, they developed models using daily data such as average annual daily traffic.

Data Collection

Intersection Selection

A total of 49 signalized intersections in VDOT's NOVA District were initially selected for data collection for model development based on the availability of traffic volume data, validity of crash data, and judgment of traffic engineers in the district. Of the 49 intersections, 3 are in Loudoun County, 9 are in Prince William County, and 37 are in Fairfax County. All 49 intersections are equipped with actuated signals, and many of them are coordinated in their signal phase plans with neighboring signalized intersections. A list of the 49 intersections is provided in Appendix A.

Traffic Crash Data

Traffic crash data for the 49 intersections from 2001 through 2004 were obtained from VDOT's Oracle crash database. Three tables, CRASHDOCUMENT, CRASHINTERSECTION, and CRASHVEHICLE, in the database, which was updated most recently at the time of data collection, were used to extract the necessary data. By the definition of *intersection crash* in Virginia (crashes occurring within 150 ft of an intersection), all records of crashes within 150 ft of the selected intersections were extracted.

Traffic Volume and Signal Phase Data

Synchro files of the 49 intersections were obtained with the help of traffic engineers in the NOVA District. Four Synchro files by time of day (A.M. peak, mid day, P.M. peak, and off peak) were acquired for each intersection. The files for about half of the intersections were provided by the traffic engineers, and those for the other half were downloaded from the NOVA GIS Applications through the VDOT intranet. Hourly traffic volume data by turning movement and left-turn signal phase data were extracted from the Synchro files. According to the traffic engineers in the district, the traffic volume data were collected from 2001 through 2002. An example of the raw Synchro data report is shown in Appendix B.

Signal Plan Data

Time-based coordination event sheets for the 49 intersections were obtained from a traffic engineer at the Northern Virginia Smart Traffic Center using the MIST client/server system. The sheet for each intersection contained information regarding the start and end times of signal plans for different times of day (A.M. peak, mid day, P.M. peak, off peak, and free operation) on each of 7 days in a week. These sheets are typically used to operate traffic signal plans and coordinate the signal phases of neighboring intersections. An example of an event sheet is shown in Appendix C.

Exploratory Analysis

Using the combined raw data and the individual raw datasets, a preliminary data analysis was performed for exploratory purposes. By computing basic statistics and producing cross-tabulation, insights about how to analyze the data were gained (e.g., what variables can be used for model development, if necessary, after variables were transformed by recoding and/or combining).

One of the main goals of this analysis was to determine classification criteria for traffic crash patterns. The crash patterns delineated by the criteria were helpful in defining crash population reference groups for model development and EB application. It was necessary to develop one safety performance function (SPF) for each of the crash reference population groups in order to apply the EB method properly. For clarification, a crash pattern is different from a collision type that is available in VDOT's crash database. Examples of collision type are head-on, rear-end, and angle. A *crash pattern* is defined later in this report.

Factors potentially useful for classification criteria were examined, such as the number of vehicles in a crash (e.g., single-, two-, and multi-vehicle crash), injury severity (e.g., fatal, injurious, and total crash), and crash counts versus victim counts (e.g., the number of injury crashes versus the number of traffic injuries).

If crash patterns are defined by the number of vehicles in a crash, crashes belonging to the same reference group should involve the same number of vehicles. The assumption behind this grouping definition is that a crash occurrence mechanism is differentiated by the number of vehicles in a crash; thus, the relationship between crash occurrence and traffic flow should be identified separately by the number of vehicles. Single-vehicle crashes are thought to be different from multi-vehicle crashes (Ceder and Livneh,1982a,b; Kockelman and Kweon, 2002; Persaud and Nguyen, 1998). In general, single-vehicle crashes appear to occur more often as traffic volume increases up to a particular level, and they appear to occur less often as traffic volume increases above that level. However, multi-vehicle crashes appear to continue to increase as traffic volume increases.

For this study, the number of vehicles (single- versus multiple-vehicle crashes) and turning movements (left-turn, right-turn, and straight-through) were selected to form the classification criteria for traffic crash patterns. Crash reference population groups were formed by times of day (e.g., A.M. peak and P.M. peak) in conjunction with the crash patterns defined by the crash classification criteria (i.e., a combination of the number of vehicles and the turning movements). The information used to define the reference population group is described here.

Traffic Crash Patterns

A total of 17 crash patterns were identified through logical consideration and the literature review (Hauer et al., 1988; Persaud and Nguyen, 1998); 16 patterns excluding the single-vehicle crash type, shown in Figure 2, were used for this study.

These 16 crash patterns were proposed to reflect the two conflicting traffic movements contributory to crash occurrence. It is reasonable to relate crash occurrence to the traffic flows that are involved in a crash, which is suggested by Hauer et al. (1988). If there were more than two vehicles in a crash, the first two contributing vehicles were used to classify the crash in accordance with one of the 16 patterns for this study. Because the crash occurrence mechanism is believed to differ for these crash patterns (Hauer et al., 1988), the crash patterns are expected to offer insights into the nature of the different crash occurrence mechanisms.

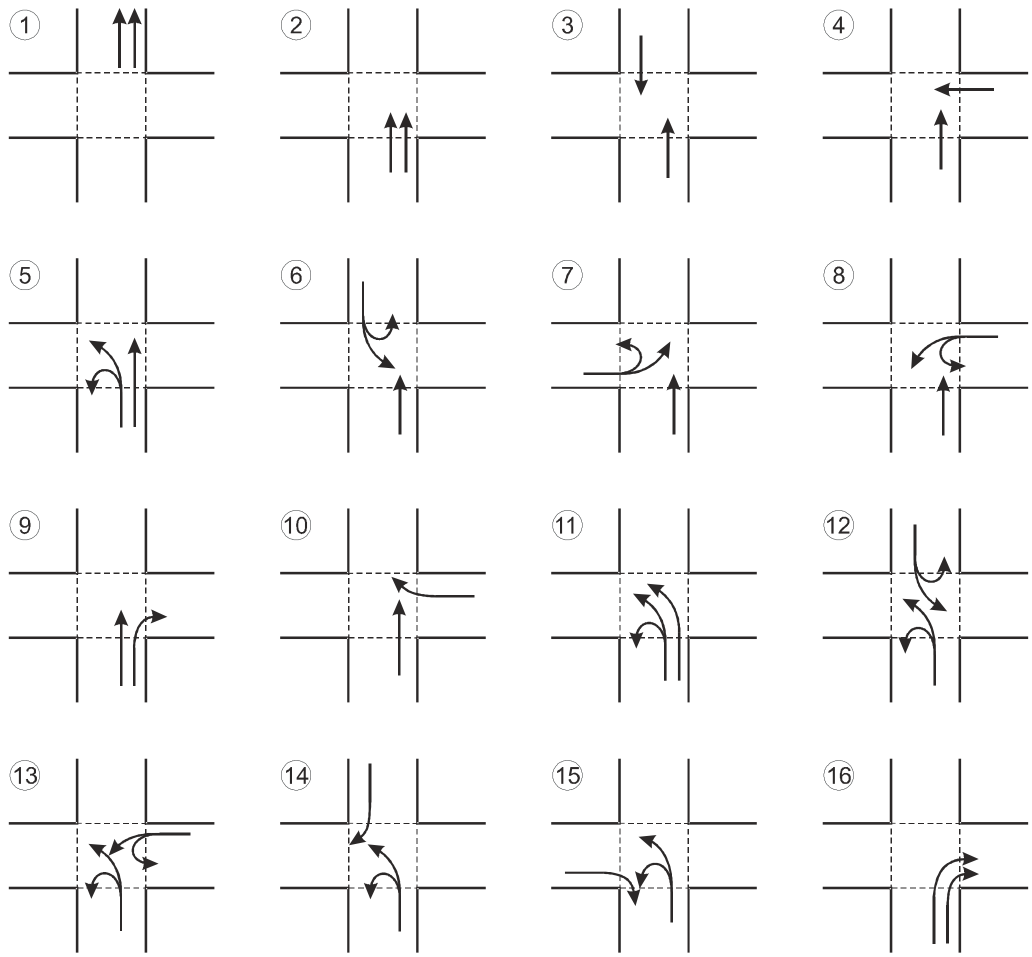


Figure 2. Traffic Crash Patterns of Multiple Vehicles by Vehicle Movements. The dotted lines indicate the box of an intersection physically dividing the inside from the outside of the intersection.

Crash pattern 1 represents a crash of more than one vehicle with the first two contributory vehicles involved in the crash having passed through an intersection and proceeding in the same direction. The crash occurs within 150 feet of the intersection. The first two vehicles may start from different approaches of an intersection. Crash pattern 2 represents a crash between the first two vehicles moving straight from the same approach and having a collision inside the intersection box. All other crash patterns are understandable from the figure. Except for crash pattern 1, all crash patterns represent a crash that occurred inside the intersection box.

The 16 crash patterns can be categorized with three fields in the tables CRASHINTERSECTION and CRASHVEHICLE in VDOT's crash database: (1) OFFSET (distance from the intersection), (2) VEHICLEMANEUVER (vehicle's movement at the time of a collision such as "making a right turn" and "slowing or stopping"), and (3) VEHICLPLACEMENT (vehicle's traveling direction prior to a collision: east, west, south, and north). The definitions of the fields are provided in Table 1.

Using the three fields, crash pattern 1 is defined by a positive OFFSET value, VEHICLPLACEMENT codes of the first two vehicles in the collision as the same, and the

Table 1. Definitions of VEHICLEMANEUVER and VEHICLPLACEMENT

Code	Description
VEHICLEMANEUVER	
01	Going straight ahead
02	Making right turn
03	Making left turn
04	Making U-turn
05	Slowing or stopping
06	Starting in traffic lane
07	Starting from parked position
08	Stopped in traffic lane
09	Ran off road–right
10	Ran off road–left
11	Parked
12	Backing
13	Passing
14	Changing lanes
15	Other
16	Not stated
VEHICLPLACEMENT	
N	North
S	South
E	East
W	West

VEHICLEMANEUVER codes of 01, 05, 06, 08, 13, and 14 for the first two vehicles. Crash pattern 4 is defined by a zero OFFSET, the VEHICLPLACEMENT codes of the first two vehicles perpendicular to each other (e.g., north and east), and the 01 VEHICLEMANEUVER code of the two vehicles. Crash pattern 6 is defined by a zero OFFSET, the opposite VEHICLPLACEMENT codes of the first two vehicles (e.g., north and south), and the 01 VEHICLEMANEUVER code for one vehicle and 03 or 04 VEHICLEMANEUVER code for the other vehicle.

Time of Day

The crash occurrence mechanism for the same crash pattern might be different by time of day; thus, time of day in addition to the crash pattern was used to define a crash population reference group. Four times of day were found in the data: A.M. peak, mid day (between the end of A.M. peak and the start of P.M. peak), P.M. peak, and off peak (between the end of P.M. peak and the start of night-time non-coordinating signal operation, called “free operation”). The start and end times of these four times of day differ by intersections. The typical signal time schedule of the intersections used in this study was 6 A.M.–9 A.M. (A.M. peak), 9 A.M.–3 P.M. (mid day), 3 P.M.–7 P.M. (P.M. peak), 7 P.M.–10 P.M. (evening off peak), and 10 P.M.–6 A.M. (free operation).

A separate dataset for each crash reference population group defined by the crash pattern and the time of day was prepared for model development and EB application, and the details of such data preparation are described here.

Data Preparation

As described previously, a total of 49 intersections were initially selected for the study from about 1,200 signalized intersections in the NOVA District. Of those, 48 are four-legged and 1 is three-legged. The intersection with three legs was excluded from the study. For the 48 four-legged intersections, the directional orientation matching process between traffic volume data (from Synchro files) and crash data revealed that the police crash reports for 2 intersections showed inconsistencies indicating the north in the crash diagrams; i.e., crash reports referencing the same intersection indicated the north differently. Thus, these 2 intersections were removed from the study. Of the 46 intersections, 1 intersection was not identified in VDOT's crash database by its crossing street names; thus, it was removed. Of the 45 intersections, Synchro files for 2 intersections could not be obtained; thus, these 2 were removed from the study.

The types of left-turn signal phase and traffic flows were extracted from Synchro files by traffic turning movement and time of day. Because there are 12 different turning movements at a typical four-legged intersection (4 movements for each left-turn, right-turn, and straight-through flow) and 4 different times of day in a typical weekday (A.M. peak, mid day, P.M. peak, and off peak), 48 turning traffic flows (12 movements $\times$ 4 times of day) and 16 left-turn phases (4 approaches $\times$ 4 times of day) were recorded for each intersection.

Vehicle maneuver (e.g., moving straight and turning left) and directional information (e.g., heading east/west/north/south) for each vehicle in a crash were extracted from the crash vehicle table, and traffic crash information (e.g., time/date/year and number of vehicles) for each crash was extracted from the crash document table, both of which are in VDOT's crash database. These two extracted datasets were combined, and crashes occurring outside 150 feet of an intersection were removed.

Time of day (e.g., start and end time of each time of day) and type of operation (e.g., free operation) for each intersection were extracted from time based coordination event sheets. Nine intersections (intersection sequential numbers 23, 25, 32, 33, 35, 36, 40, 42, and 47) were operated under free operation (i.e., signal operation under non-coordinated actuated signal time plan having a setup of only minimum and maximum green times); thus, their start and end times for the four times of day could not be identified from the event sheets. These nine intersections could not be analyzed by different times of day and thus were removed from the analysis. However, some of the nine intersections were already removed because of other reasons addressed previously. In the end, 32 to 35 intersections were used for model development depending on crash population reference groups, and the data from them were prepared for the analysis.

All of the distinct datasets were merged into a single dataset. The merged dataset was aggregated by intersection, time of day, and the classification criteria determined in the exploratory analysis. In the end, one dataset for each of crash reference population groups was prepared for model development; e.g., theoretically, 64 crash reference population groups exist for multi-vehicle crashes.

Regarding left-turn signal phases, about 14% of intersection approaches were permissive, about 21% were protected, about 12% were split, and about 53% were permissive plus protected. About 20% of traffic crashes occurred in the A.M. peak, about 31% occurred in mid day, about 33% occurred in the P.M. peak, and about 17% occurred in the evening off peak.

Model Development

Three aggregate levels of analysis (high, intermediate, and low) were initially proposed depending on how the data would be aggregated for model development. The high aggregate level aggregates the data for an entire intersection; thus, the analysis unit is an intersection as a whole. The intermediate level aggregates the data for an approach of an intersection; thus, the analysis unit is an approach. The low level, which is the most disaggregate level in the study, aggregates the data for a pair of conflicting vehicle movements of the first two vehicles contributing to crash occurrence (e.g., a straight-through vehicle colliding with an opposing left-turning vehicle); thus, the analysis unit is a pair of vehicle movements. Among the three levels, the high and low levels of analysis were performed and the low level of analysis turned out to be the only level useful for the project. Thus, only the model development procedure using the low aggregate level of data is described here.

Step MD1: Quick Diagnostic Check for Overdispersion

The number of crashes is very often analyzed by count data regression models because it is a non-negative integer (i.e., count response) in its characteristics. The Poisson regression model is the standard model for count response data (e.g., number of crashes per intersection; number of patients' visits to a physician's office per day) and requires the equidispersion assumption, meaning that the conditional mean and variance of the count response are equal. In other words, the mean and variance of errors from the Poisson model are assumed to be equal.

It is very important to understand how the count response (e.g., crash frequency per year for an intersection) that is a dependent variable in count data models is dispersed (i.e., under-, equi-, and overdispersion) because the appropriate functional form of the model depends on the level of dispersion of the count response. For example, the Poisson model is appropriate for equidispersed data, a binomial model is appropriate for underdispersed data, and an NB model is appropriate for overdispersed data. If an inappropriate model were chosen for data, the base assumption of the model would be violated, probably resulting in bias in estimates of the model's parameters. Thus, the use of those estimates would likely lead to erroneous conclusions.

As a quick diagnostic check of the level of dispersion, a comparison between the unconditional mean and variance of the count response data was suggested by Cameron and Trivedi (1998). If the variance of the count data exceeds twice the unconditional mean, the data are likely to be overdispersed. The factor of two for the mean is used to account for potential inclusion of explanatory variables in count response models so that the likely conditional mean and variance of the dependent variable (i.e., count response) in the presence of the explanatory variables can be roughly compared.

If the mean exceeds the variance, the data are certainly underdispersed. If the variance is between the mean and twice the mean, the data can be one of under-, equi-, or overdispersion, depending on how much unconditional overdispersion can be taken out by explanatory variables in count response models.

Although this diagnostic check is a useful tool to identify potential overdispersion in the data quickly, it is somewhat preliminary in nature; thus, more rigorous approaches to test the presence of overdispersion should also be conducted to verify the level of dispersion. For this study, this diagnostic check was used to gain insights for developing mean models in step MD2.

Step MD2: Mean Model Development

Typical linear regression models are usually not appropriate for count response data unless the mean of the count response is relatively high (e.g., the analysis of a value of 10 or greater). This is because the normality and homoskedasticity (i.e., constant variance) assumptions required for the linear regression models are likely to be violated with most count data (especially traffic crash data) and because the prediction of the linear models can be negative. Count responses are non-negative and typically skewed (to the right) in distribution, which implies nonnormal distribution, and their variances increase as their means increase, implying nonconstant variance. This nonconstant variance, which is a violation of the homoscedasticity assumption, does not affect unbiasedness, which is a statistical property of a parameter whose estimate is not biased from a true value, of the parameter estimates. However, it results in biased estimates of their variances, which, in turn, affect the validity of significance tests (e.g., *t* tests) for those unbiased parameter estimates. Its consequence is greater than that in linear models.

The standard model for count data is the Poisson regression model, assuming equidispersion. However, crash count data are often overdispersed; thus, the Poisson model is likely to be improper. On such occasions, there are several alternative models designed to handle overdispersion of the data, including overdispersed Poisson, NB, zero-inflated Poisson, and zero-inflated NB models (see Kweon and Kockelman [2005] for the application of those models in traffic safety and Slymen et al. [2006] for their application in the health sciences). Among them, the NB model with the mean dispersion function (i.e., NegBin Type II according to Cameron and Trivedi [1998]) is the most frequently employed in crash analysis and is the most suitable for the application of the EB method (Hauer, 1997). For this study, the Poisson and NB regression models were used (more information regarding count data models may be found in the report by Cameron and Trivedi [1998]).

The Poisson regression model consisting of the Poisson probability mass function (pmf) and an exponential mean function can be expressed as follows:

$$\Pr(Y_i = y_i \mid \mathbf{x}_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \text{ and } \lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i)\text{, where } y_i = 0, 1, 2, \ldots$$

where i = index for a subject (e.g., intersection); Y_i = count response variable (e.g., number of crashes in 4 years); $\mathbf{x}_i$ = vector of explanatory variables (e.g., traffic volume and types of left-turn signal phase); $\boldsymbol{\beta}$ = vector of parameters to be estimated; and λ_i = mean level of the count response (e.g., average number of crashes). The equidispersion condition of the Poisson model can be mathematically expressed as:

$$E(Y_i \mid \mathbf{x}_i) = V(Y_i \mid \mathbf{x}_i).$$

The NB regression model with the mean dispersion function can be viewed as an expansion of the Poisson model because it adds a random disturbance to the exponential mean function of the Poisson model as follows:

$$\mu_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i) = \exp(\boldsymbol{\beta}'\mathbf{x}_i) \times u_i = \lambda_i u_i$$

where ε_i is a random disturbance, and $\varepsilon_i = \ln(u_i)$. The added random term, $u_i = \exp(\varepsilon_i)$, is often assumed to follow a gamma distribution with a single parameter for mathematical tractability so that a closed form expression can be readily derived. Then, the NB model can be expressed as follows:

$$\Pr(y_i \mid \mathbf{x}_i, \varepsilon_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \text{ and } \mu_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i) = \lambda_i \cdot u_i$$

where $u_i \sim Gamma(b,b)$, and integrating ε_i out leads to the following model equation:

$$\Pr(Y_i = y_i \mid \mathbf{x}_i) = \frac{\Gamma(y_i + k^{-1})}{\Gamma(y_i + 1)\Gamma(k^{-1})} \left(\frac{k^{-1}}{k^{-1} + \lambda_i} \right)^{k^{-1}} \left(\frac{\lambda_i}{k^{-1} + \lambda_i} \right)^{y_i}$$

where $k = \frac{1}{b}$ is called the dispersion parameter (of the NB model), and $\Gamma(\cdot)$ is the gamma function.

The overdispersion condition can be mathematically shown as

$$V(Y_i \mid \mathbf{x}_i) = \lambda_i (1 + \alpha \lambda_i) > E(Y_i \mid \mathbf{x}_i) = \lambda_i.$$

More information about this and other count data models may be found in the report by Cameron and Trivedi (1998).

Step MD2.1: Determination of Functional Form of Flow Variables

Two visual approaches were applied to determine appropriate functional forms of traffic flow variables: (1) plots of average aggregated crash counts by traffic flow groups (i.e., a set of flow ranges) and (2) plots of cumulative sum of residuals. In addition, literature reviews with a specific focus on functional forms of traffic flow variables in SPFs were conducted to identify the most frequently adopted functional forms. The reviewed studies included Ardekani et al. (2002), Belanger (1994), Bauer and Harwood (2000), Forbes and Belluz (2003), Harwood et al.

(2000), Lyon et al. (2005), McGee et al. (2003), Mountain and Fawaz (1996), Poch and Mannering (1996), Regional Municipality of Durham (in Ontario, Canada) (2001), Regional Municipality of Halton (in Ontario) (2001), Sayed and Rodriguez (1999), Vogt (1999), Vogt and Bared. (1999), and Washington et al. (2005). The form employed most often among those studies is a log-transformed form of traffic flow variables inside the exponential mean function of a count response model as follows:

$$\mu = \exp(\beta_0 + \beta_1 \cdot \log f_1 + \beta_2 \cdot \log f_2) = \alpha \cdot f_1^{\beta_1} \cdot f_2^{\beta_2}$$

where μ is an expected crash count, f_1 and f_2 are traffic flow variables (e.g., average annual daily traffic or hourly traffic volumes for major and minor roads), and $\alpha = \exp(\beta_0)$.

Ardekani et al. (2002) described the importance of logical considerations in determining the functional form of traffic flow variables and provided different forms representing different shapes of the relationship between traffic flow and expected crash count. One of the desired properties for the flow-crash function is that the curve drawn by the function should go through the origin, thereby guaranteeing that zero crashes are expected when there is no traffic. The log-transformed form discussed here satisfies this property.

Using Plots of Average Aggregated Crash Counts by Traffic Flow Groups

The response values (i.e., the number of crashes) for this study have low mean values and contain many zeros, which, in some cases, is due to the nature of the crash data. Thus, typical plots between crash counts and traffic volumes do not usually provide good insight about their relationship (i.e., functional form of the relation), as shown in Figure 3.

In such cases, one way to obtain a visual clue about the relationships is to group the crash data by traffic flows (e.g., 0-500, 500-1,000, etc., in vehicles per day), calculate an average crash count for each group, and plot the average crash count against the traffic flow groups as shown in Figure 4. This was successfully done by Hauer et al. (1988) when they examined traffic crash and volume data similar to the data collected for this study. Figure 4 presents three typical functional forms used in this study.

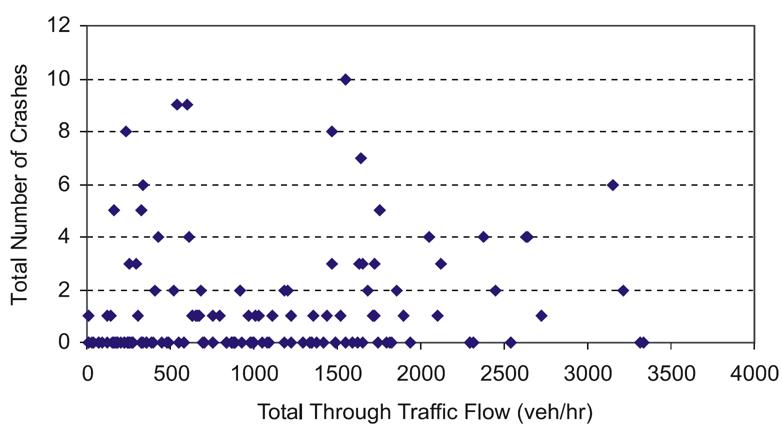
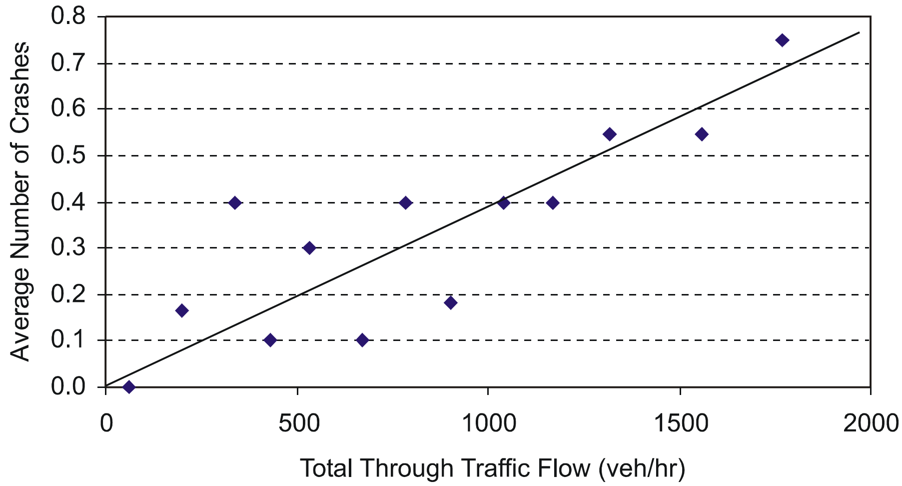
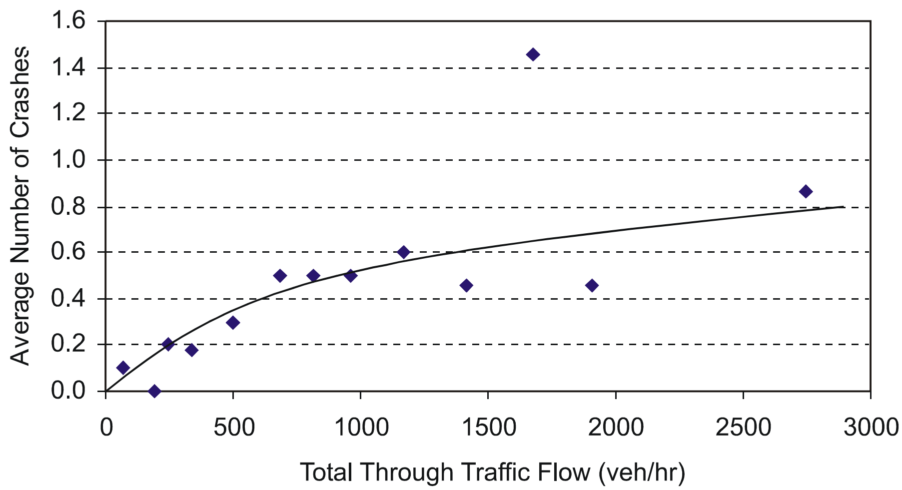


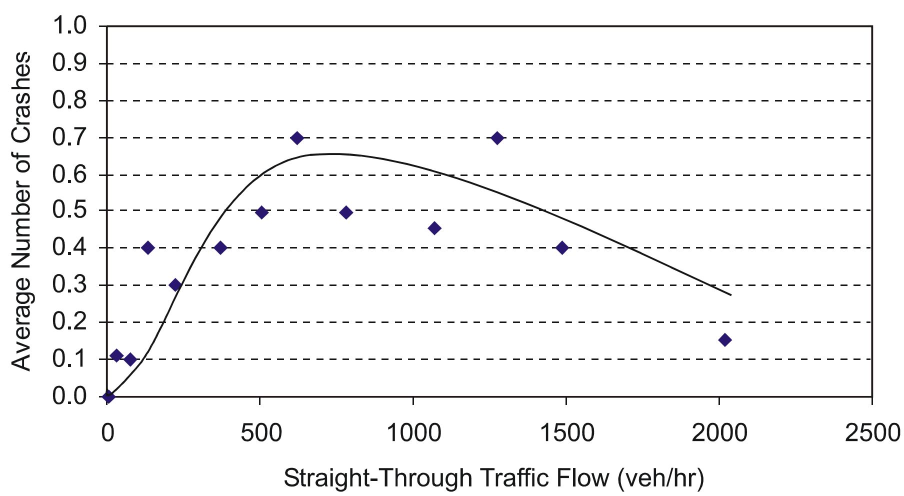
Figure 3. Raw Crash Count Versus Traffic Flow



(a) $\mu = \exp(\beta_0 + 1 \cdot \log Flow_{through}) = \alpha \cdot Flow_{through}$



(b) $\mu = \exp(\beta_0 + \beta \cdot \log Flow_{through}) = \alpha \cdot Flow_{through}^{\beta}$



(c) $\mu = \exp(\beta_0 + \beta_1 \log Flow + \beta_2 Flow) = \alpha \cdot Flow^{\beta_1} \cdot \exp(\beta_2 \cdot Flow)$

Figure 4. Average Crash Count Versus Traffic Flow

Using Plots of Cumulative Sum of Residuals

Lin et al. (2002) proposed graphical and numerical model assessment methods using cumulative sums of residuals over certain explanatory variables in regression models. By visually and numerically comparing residual patterns generated from simulation, appropriate functional forms of explanatory variables can be inferred. Figure 5 shows two examples of plots of cumulative sum of residuals from 10,000 simulation runs using the data for this study.

Step MD2.2: Overdispersion Tests and Model Type Selection

After a final model is developed for each of the Poisson and NB models, one of these two types of models should be selected. The deciding factor is whether there an overdispersion portion remains after the explanatory variables in the models eliminate overdispersion from the count responses. Of the several ways to examine overdispersion in the data, four tests were applied in this study: (1) dispersion parameter using deviance and Pearson's chi-square statistics, (2) regression-based equidispersion test, (3) confidence interval of the NB dispersion parameter, and (4) Lagrange multiplier (or score) test for overdispersion. The first is rather subjective in determining overdispersion because no definite criteria exist in indicating overdispersion; the other three are objective.

For an illustration of the four tests and the quick diagnostic check stated in step MD1, Table 2 provides test results for crash pattern 1 during mid day hours. As is obvious, the NB model was selected over the Poisson model by all five test statistics for this case.

Dispersion Parameters Based on Deviance and Pearson's Chi-Square Statistics

The deviance or Pearson's chi-square statistics divided by its degrees of freedom can be used to estimate a dispersion parameter (ϕ) that is often used to indicate roughly the level of dispersion of the count response. For the Poisson and NB distributions, the dispersion parameter, ϕ, using deviance and Pearson's chi-square statistics can be obtained as follows (McCullagh and Nelder, 1991):

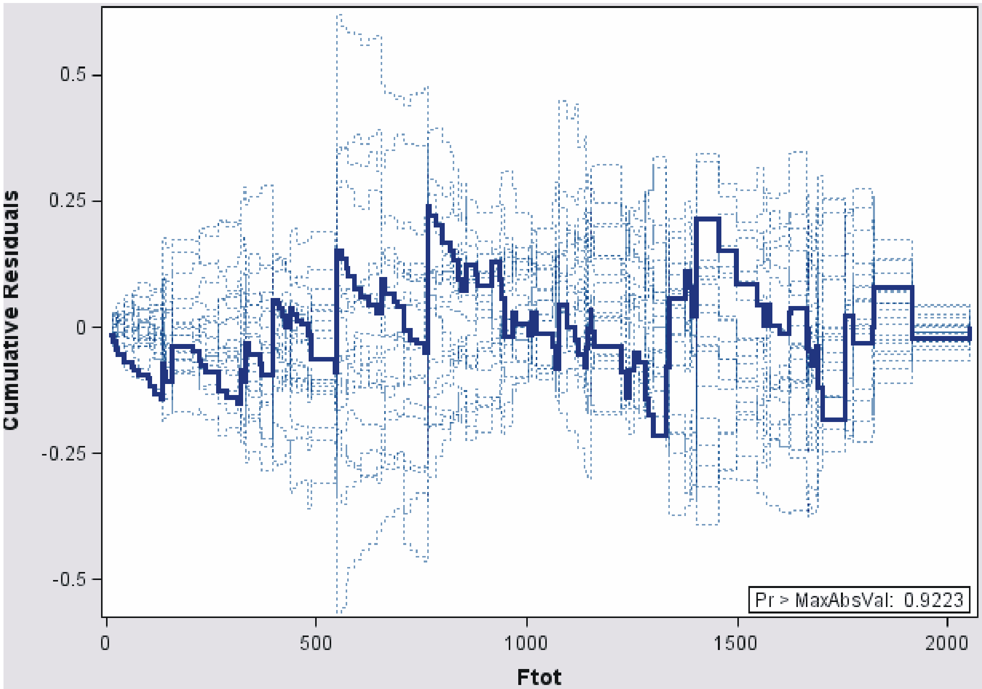
Using deviance statistics,

$$\phi = \frac{D}{(n - p)}$$

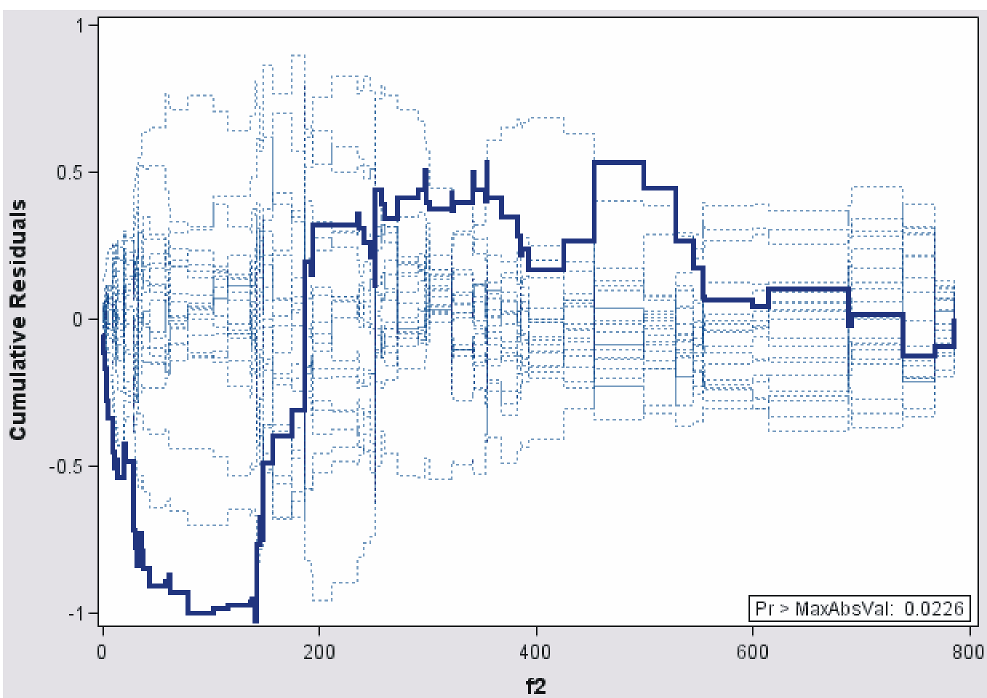
where D = deviance, n = number of observations, and p = number of parameters (so $n - p$ is the degree of freedom). D is written as follows:

Poisson distribution: $D = 2\sum \left[y \log\left(\frac{y}{\mu}\right) - \left(y - \mu\right) \right]$ and

NB distribution: $D = 2\sum \left[y \log\left(\frac{y}{\mu}\right) - \left(y + \frac{1}{k}\right) \log\left(\frac{y + 1/k}{\mu + 1/k}\right) \right]$;



(a) $\mu = \exp(\beta_0 + \beta \cdot f_{tot})$



(b) $\mu = \exp(\beta_0 + \beta \cdot f_2)$

Figure 5. Cumulative Sum of Residuals. Larger *p*-values suggest more appropriate form of variable of interest in model, and observed residuals (dark line) do not show typical shape for appropriate form of variable. According to these criteria, *plot (a)* presents example of appropriate functional form (high p-value and atypical shape of dark line), whereas *plot (b)* presents example of inappropriate functional form (very small *p*-value and typical wave shape of dark line). For *plot (b),* log-transformed variable appears more appropriate. Light dotted lines represent simulated residuals.

Table 2. Example of Overdispersion Tests

Overdispersion Test	Statistics	Test Result	Selected Model
Dispersion parameters based on deviance and Pearson's chi-square statistics	ϕ = Pearson's chi-square/degree of freedom =3.665 >> 1.0	Overdispersion	NB model
Regression-based equidispersion test	p-values of coefficients = 0.0003 and 0.0010 < 0.05	Overdispersion	NB model
Confidence interval of NB dispersion parameter	95% confidence interval of k = [1.3529, 3.4148] not containing zero	Overdispersion	NB model
Lagrange multiplier test	p-value of test = 0.00001<0.05	Overdispersion	NB model
Quick diagnostic check	Variance = 4.38 > 2 $\times$mean = 2.41	Overdispersion	NB model

Using Pearson's chi-square statistics,

$$\phi = \frac{\chi^2}{(n-p)}$$

where $\chi^2 = \sum \frac{(y-\mu)^2}{V(\mu)}$ with $V(\mu)$ is a variance model. $V(\mu)$ is written as follows:

Poisson distribution: $V(\mu) = \mu$ and
NB distribution: $V(\mu) = \mu + k\mu^2$.

If the estimate of the dispersion parameter (ϕ) is substantially larger than 1, overdispersion might be present. However, it should be noted that such deviation from 1 might result not only from overdispersion in the data but also from other problems such as incorrect model specification, missing explanatory variables, and outliers in the data (Heinzl and Mittlbock, 2003). Therefore, the dispersion parameter should be used with caution, and it is recommended that other tests for overdispersion be used also when overdispersion is detected by the dispersion parameter.

Regression-Based Equidispersion Test

One common way of testing equidispersion is a nested chi-square (χ^2) test comparing the Poisson model against an alternative model that has the Poisson model as a nested case, such as the NB model. However, this test requires a particular parametric assumption for the density of the response values (e.g., NB distribution). Cameron and Trivedi (1990) proposed a regression-based test for equidispersion that does not require a parametric assumption such as to this, but requires only the specification of the relationship between the mean and the variance of the response variable.

The following regression models are estimated to perform the test:

$$\frac{\left((y-\hat{y})^2 - y\right)}{\sqrt{2}\hat{y}} = \alpha \text{ or } \beta \times \frac{\hat{y}}{\sqrt{2}\hat{y}} \text{ or } \gamma \times \frac{\hat{y}^2}{\sqrt{2}\hat{y}}$$

where y = observed count value (e.g., number of crashes in 4 years); $\hat{y}$ = estimated count value by the Poisson model; and α, β, and γ are regression parameters to be estimated and used for the equidispersion test. If α, β, and γ are *not* statistically different from zero (i.e., statistically *in*significant), there is no overdispersion remaining from the estimated Poisson model, which means that the Poisson model is appropriate. Otherwise, overdispersion still exists; thus, the current Poisson model is not appropriate. This implies that more variables with the Poisson model are needed to explain the remaining overdispersion or that other models handling overdispersion such as the NB model are needed. Since more variables could not be obtained in most real circumstances and the most frequently used model with overdispersed data is the NB model, the statistical significance of the coefficients (α, β, and γ) implies that the NB model is more appropriate than the Poisson model.

Confidence Interval of NB Dispersion Parameter

One of the parameters in the NB model estimated using the maximum likelihood technique is the NB dispersion parameter, *k*, which appears in the variance function, $V(\mu) = \mu + k\mu^2$. If the parameter equals zero, the model reduces to the Poisson model. The 95% Wald confidence interval of the dispersion parameter was produced from the model estimation. If the confidence interval includes zero, the Poisson model is more appropriate than the NB model. If not, the NB model is appropriate.

Lagrange Multiplier (or Score) Test for Overdispersion

Cameron and Trivedi (1998) proposed a score (or Lagrange multiplier) statistic to test overdispersion in the Poisson model against the NB model (NegBin II Type) with the mean dispersion function (i.e., variance function), $V(\mu) = \mu + k\mu^2$, where μ is the conditional mean and *k* is the dispersion parameter of the NB model. This test is designed to detect a specific type of overdispersion: the NegBin II Type.

$$LM = \frac{s^2}{V}$$

where *s* = score statistics of the test for overdispersion in the Poison model against the NegBin II Type model with the restriction of $k = 0$. This test is similar to the regression-based equidispersion test. This test is detailed in a report by Cameron and Trivedi (1998).

Step MD2.3: Goodness of Fit

Once the final model type (i.e., Poisson versus NB model) is determined in step MD2.2 and the final model estimates are obtained, a goodness of fit model is examined using several fit measures and a graphical comparison. As for the fit measures, four pseudo R-squared measures were used. As for the visual comparison, mean observed and predicted probabilities for each count response were calculated and drawn in one graph so that the closeness of the predictions made by different models to observed responses could be visually compared.

Pseudo R-squared Measures

The four fit measures are (1) likelihood-ratio index, R_L^2; (2) correlation-based R-square, R_C^2; (3) deviance-based R-square, R_D^2; and (4) dispersion parameter-based R-square, R_k^2. The fourth is applicable for the NB model, not the Poisson model. Pseudo R-squares in non-linear models are analogous to the R-square in linear models. However, the interpretation of the R-square, the percentage of total variation of data explained by the model fitted to the data, does not carry over to the pseudo R-squares because of the non-linear nature of the models for which the pseudo R-squares are computed.

The four pseudo R-squared measures are written as follow:

$$R_L^2 = 1 - \frac{\ln L_1}{\ln L_0}$$

$$R_C^2 = r^2$$

$$R_D^2 = 1 - \frac{KL_1}{KL_0}$$

$$R_k^2 = 1 - \frac{k_1}{k_0}$$

where L_1 is the likelihood value of the estimated model; L_0 is the likelihood value of the intercept-only model; r^2 is a square of the correlation coefficient between the count response and its predicted mean, $[corr(y, \hat{\mu})]^2$; KL_1 is the Kullback-Leibler (KL) divergence of the estimated model, $2[\ln L_S - \ln L_1]$ with L_S being the likelihood value of the saturated model; KL_0 is the KL divergence of the intercept-only model, $2[\ln L_S - \ln L_0]$; k_1 is a dispersion parameter of the estimated model with an NB error structure; and k_0 is a dispersion parameter of the intercept-only model with an NB error structure.

The deviance-based R-square was proposed by Cameron and Windmeijer (1997), and the dispersion parameter-based R-square was proposed by Miaou (1996). However, these measures might be biased if the sample size is small. For example, Miaou (1996) tested the validity of the dispersion parameter-based R-square with large simulation data and recommended its use for a large sample size. Heinzl and Mittlbock (2003) found that the deviance-based R-square for the Poisson model could be significantly biased in the case of a small sample size.

Each pseudo R-square measure has a different interpretation, yet none of their interpretations is as intuitive as the R-square for linear models. The likelihood-ratio index (R_L^2) means the ratio of improvement in log-likelihood values between the model with only intercept and the model of interest. The correlation-based R-square (R_C^2) is basically meant to show the degree of correlation between the predicted counts by the model and the observed counts. The deviance-based R-square (R_D^2) is somewhat similar to, but more complicated than, the likelihood-ratio index. It shows the ratio of improvement in KL information values between the model with only intercept and the model of interest. The KL information for the model of interest, for example, implies improvement in log-likelihood values between the saturated model and the model of interest. The dispersion parameter-based R-square (R_k^2) shows the percentage of total dispersion of data explained by the model of interest.

As indicated, these pseudo R-square measures do not have the good statistical interpretation that the R-square has for linear models. Moreover, there seems to be no consensus with regard to using one measure over others; thus, more than one measure is typically presented in a study. Although several measures are computed and presented, they do not seem to carry much importance partly because there is no good interpretation of them and partly because of the absence of thresholds generally accepted for judging how well models are fitted to data. This study used the four measures presented because they can still offer some sense of a relative goodness of fit among different models and the use of the pseudo R-square measures appears to be common in many studies using non-linear models.

All pseudo R-squared measures should have two properties; (1) they should range from 0 to 1 and (2) they should not decrease as explanatory variables are added to a model (Demaris, 2004). The four measures have these two properties. However, other characteristics should also be noted. The likelihood-ratio index cannot approach 1; thus, it may underestimate the fitness of the model. The correlation-based R-square may violate the "nondecreasing" property.

Graphical Comparison of Mean Observed and Predicted Probabilities

The mean observed and predicted probabilities can be calculated respectively as follows:

$$\overline{\mathrm{Pr}}_{observed}(m) = \frac{1}{N}\sum_{i=1}^{N} I(y_i = m)$$
$$\overline{\mathrm{Pr}}_{pred}(m) = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathrm{Pr}}(y_i = m \mid \mathbf{x}_i)$$

where m is a unique value in the count responses (e.g., crash count); $I(\cdot)$ is an indicator function equaling 1 if the condition inside the parenthesis is true and 0 otherwise; i is an index for an entity (e.g., intersection) with $i = 1,\ldots, N$; y_i is the value of the count response of the entity i; and $\mathbf{x}_i$ is a set of explanatory variables. The mean observed probability is just a proportion of count responses equaling a certain value from the set of unique count response values. In the predicted probability, the individual conditional probability, $\hat{\mathrm{Pr}}(y_i = m \mid \mathbf{x}_i)$, is calculated using the estimated count model. For the Poisson model, the individual conditional probability is calculated by

$$\hat{\Pr}(y_i = m \mid \mathbf{x}_i) = \frac{\exp(-\hat{\lambda}_i)\hat{\lambda}_i^m}{m!} \text{ with } \hat{\lambda}_i = \exp(\hat{\boldsymbol{\beta}}'\mathbf{x}_i)$$

where $\hat{\boldsymbol{\beta}}$ is a set of estimated parameters of the Poisson model.

Figure 6 illustrates the difference between the mean observed and predicted probabilities. The comparison of the four mean probabilities indicates that the mean probability predicted by the NB model fits very well with the observed probability and is much better fitted than those of the Poisson model or the univariate Poisson. The visual comparison in Figure 6 clearly shows how well models fit data. In the end, the visual comparison turned out to be more useful than the pseudo R-squares in assessing goodness of fit models. See Demaris (2004) and Long (1997) for more illustrations of the graphical comparison using general count data.

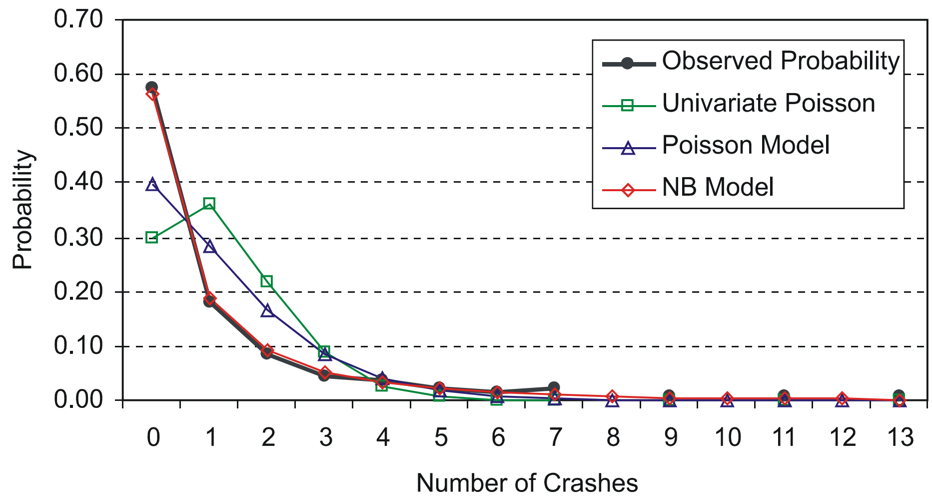


Figure 6. Mean Observed and Predicted Probabilities

Step MD3: Variance Model Development

To apply the EB method, the variance model should also be estimated along with the mean model in step MD2. A typical linear regression model can be used to estimate the variance model. Hauer (1997) reported that the following specification for variance models works well for traffic crash data:

$$V(\kappa_i) = \sigma_{\kappa_i}^2 = \beta \cdot E(\kappa_i)^2 = \frac{E(\kappa_i)^2}{b}$$

where κ_i is the true expected crash frequency, V is the variance, and $\beta = 1/b$ is the coefficient parameter of $E(\kappa)^2$ to be estimated. Note that an intercept is excluded from the model; thus, the model is forced to go through the origin. When the Poisson model was selected as the final model from step MD2.2, there was no need to develop the variance model because in theory,

$V(\kappa_i)$ equals zero; thus, the conditional variance of the count response equals the conditional mean.

Empirical Bayes (EB) Application

The EB method can be implemented in four steps once the estimates of the final mean and variance models are obtained from the model development procedure.

1. Calculate the mean and variance of the entity of interest.
2. Calculate the weight for the entity of interest.
3. Calculate the expected crash frequency of the entity of interest.
4. Calculate the variance of the expected crash frequency of the entity of interest.

Step EB1: Calculate Mean and Variance of Entity of Interest

The values of the explanatory variables of the entity of interest (e.g., a pair of two conflicting vehicle movements) enter into the estimated mean and variance models, producing the estimated mean and variance of the expected crash frequency of the entity. The following equations are used for the calculation:

$$\hat{E}(\kappa_j) = \hat{\mu}_j = \exp(\hat{\boldsymbol{\beta}}'\mathbf{x}_j)$$

$$\hat{V}(\kappa_j) = \hat{\sigma}^2_{\kappa_i} = \frac{\hat{E}(\kappa_j)^2}{\hat{b}}$$

where j is an index of an entity of interest; $\hat{\boldsymbol{\beta}}$ and $\hat{b}$ are obtained from the final mean model and the final variance linear model, respectively; and $\mathbf{x}_j$ is a set of values of the explanatory variables in the final mean model.

Step EB2: Calculate Weight for Entity of Interest

The weight for the EB method is computed using the estimates of the mean and the variance as follows:

$$\hat{\omega}_j = \frac{1}{1 + \hat{V}(\kappa_j) \big/ \hat{E}(\kappa_j)}.$$

If the Poisson model was selected, the variance should be zero; thus, the weight equals 1.

Step EB3: Calculate Expected Crash Frequency of Entity of Interest

The expected crash frequency is computed using the estimates of the mean and the weight and the actual crash count as follows:

$$\hat{\kappa}_j = \hat{E}(\kappa_j \mid K_j) = \hat{\omega}_j \cdot \hat{E}(\kappa_j) + (1 - \hat{\omega}_j) \cdot K_j$$

where K_j is the number of actual traffic crashes of the entity of interest, *j*.

If the Poisson model was selected, the weight equals 1; thus, the actual crash count does not affect the calculation of the expected crash frequency and the estimated crash frequency equals the estimate of the mean crash count, $\hat{E}(\kappa_j)$.

Step EB4: Calculate Variance of Expected Crash Frequency of Entity of Interest

The variance of the expected crash frequency is computed using the estimates of the expected crash frequency and the weight as follows:

$$\hat{\sigma}_j^2 = \hat{V}(\kappa_j \mid K_j) = (1 - \hat{\omega}_j) \cdot \hat{\kappa}_j .$$

If the Poisson model was selected, the variance equals zero in theory.

Using the estimated expected crash frequency from step EB3 and the estimated variance from step EB4, the probability that the entity of interest carries an abnormally high risk can be calculated based on the normal distribution assumption.

RESULTS AND DISCUSSION

This section provides the EB procedure developed in this study and describes how it was developed. The subsection on the EB procedure summarizes the finalized procedure for identifying high-risk four-legged intersections in VDOT's NOVA District, and the subsection of the development of the procedure provides the results of the steps described in the Methods section. This subsection will be useful for researchers and/or engineers when they develop or update the EB procedure.

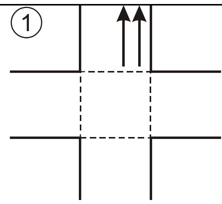
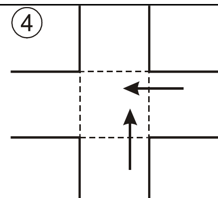
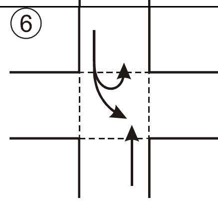
EB Procedure

This section describes the 10-step EB procedure developed to identify high-risk four-legged signalized intersections (and high-risk conflicting traffics within such intersections), the use of the procedure by traffic engineers, and a case study applying the procedure.

Final EB Procedure

Step 1. Select a crash pattern for safety evaluation from crash patterns 1, 4, and 6. These crash patterns are explained in Table 3.

Table 3. Descriptions of Crash Patterns

Crash Pattern	Configuration	Description
1		• Collision of multiple vehicles with the first two contributing vehicles having exited from an intersection and moving forward after the intersection. • Collision occurs outside the intersection box after the intersection. • Collision type could be a rear-end, same-direction sideswipe, or angle crash.
4		• Collision of multiple vehicles with the first two contributing vehicles entering straight-through into an intersection from two approaches perpendicular to each other. • Collision occurs inside the intersection box. • Collision type would be a right-angle crash.
6		• Collision of multiple vehicles with one of the first two contributing vehicles moving straight-through from an approach and the other one left-turning from an opposing approach. • Collision occurs inside the intersection box. • Collision type could be a head-on, angle, or opposite-direction sideswipe crash for a left-turn movement, and a rear-end, angle, right-angle, same-direction sideswipe crash for a U-turn movement.

Step 2. Select a time of day for the safety evaluation: A.M. peak, mid day, P.M. peak, or evening off peak. For crash pattern 4, the period from the beginning of A.M. peak until the end of evening off peak should be used.

Step 3. Determine a crash population reference group from groups 1 through 9. A crash population reference group is automatically determined when a crash pattern and a time of day are selected. The nine crash population reference groups are explained in Table 4.

Step 4. Collect the data required for the selected crash population reference group. The data required for each of the nine crash population reference groups are listed in Table 5.

Step 5. Select the correct models. The final mean and variance models for each of the nine crash population reference groups are listed in Table 6.

Table 4. Crash Population Reference Group

Crash Population Reference Group	Crash Pattern	Time of Day
1	1	A.M. peak
2	1	Mid day
3	1	P.M. peak
4	1	Evening off peak
5	4	From A.M. peak until evening off peak
6	6	A.M. peak
7	6	Mid day
8	6	P.M. peak
9	6	Evening off peak

Table 5. Required Input Data

Crash Population Reference Group	Crash Pattern	Time Period	Basic Inputs	Additional Input
1	1	A.M. peak	Straight-through volume Left-turn volume Right-turn volume Number of hours of A.M. peak	Number of crashes of specified crash pattern in past 4 years during specified time period
2		Mid day	Straight-through volume Left-turn volume Right-turn volume	
3		P.M. peak	Straight-through volume Left-turn volume Right-turn volume Number of hours of P.M. peak	
4		Evening off peak	Straight-through volume Left-turn volume Right-turn volume Number of hours of evening off peak	
5	4	A.M. peak through evening off peak	Straight-through volume	
6	6	A.M. peak	Straight-through volume Left-turn volume Left-turn signal type (permissive-plus-protected or others)	
7		Mid day	Straight-through volume Left-turn volume Left-turn signal type (permissive-plus-protected or split or permissive)	
8		P.M. peak	Straight-through volume	
9		Evening off peak	Left-turn volume	

Entering required inputs into the selected mean and variance models will produce estimates of the mean and variance of the expected crash frequency for each of the four pairs of conflict movements. The input values should be within the valid ranges presented in Table 7. Because all models were developed using the data within the ranges shown in Table 7, the results from the models will be valid only when the input values fall within the specified ranges. Although results can still be obtained using inputs outside the ranges, the validity of the results will be questionable.

Step 7. Calculate the EB weight. An EB weight is calculated for each of the four pairs of conflict movements using the following equation:

$$\hat{\omega}_i = \frac{1}{1 + \dfrac{\hat{V}(\kappa_i)}{\hat{E}(\kappa_i)}}$$

where *i* indexes the four pairs of conflict movements (*I* = 1, 2, 3, and 4; see step 6 for an example) and $\hat{E}(\kappa_i)$ and $\hat{V}(\kappa_i)$ are estimates of the mean and the variance, respectively, from step 6.

Table 6. Mean and Variance Models

Crash Population Reference Group	Model	
	Type	Estimate
1	Mean/Poisson	$\hat{E}(\kappa) = \exp(-9.1012 + 0.6566 \log Flow_{total} + 1.088 AMPeakHours)$
	Variance	$\hat{V}(\kappa) = 0$
2	Mean/NB	$\hat{E}(\kappa) = \exp(-1.8844 + 0.3113 \log Flow_{total})$
	Variance	$\hat{V}(\kappa) = \hat{\mu}^2 / 0.6661$
3	Mean/NB	$\hat{E}(\kappa) = \exp(-9.2153 + 0.08012 \log Flow_{total} + 0.8731 PMPeakHours)$
	Variance	$\hat{V}(\kappa) = \hat{\mu}^2 / 1.6529$
4	Mean/Poisson	$\hat{E}(\kappa) = \exp(-8.8123 + 0.7337 \log Flow_{total} + 0.8731 OffPeakHours)$
	Variance	$\hat{V}(\kappa) = 0$
5	Mean/NB	$\hat{E}(\kappa) = \exp(-3.4740 + 0.5780 \log Flow_{minor})$
	Variance	$\hat{V}(\kappa) = \hat{\mu}^2 / 2.4022$
6	Mean/Poisson	$\hat{E}(\kappa) = \exp\begin{pmatrix} -6.8155 + 0.7736 \log Flow_{through} - 0.0013 Flow_{through} \\ +0.3212 \log Flow_{left-turn} + 0.7209 PmPt_{left-turn} \end{pmatrix}$
	Variance	$\hat{V}(\kappa) = 0$
7	Mean/NB	$\hat{E}(\kappa) = \exp\begin{pmatrix} -14.9690 + 1.6388 \log Flow_{through} - 0.0026 Flow_{through} \\ +1.3497 \log Flow_{left-turn} - 0.0096 Flow_{left-turn} \\ +1.5125 Split_{left-turn} + 2.7704 Perm_{left-turn} + 1.2094 PmPt_{left-turn} \end{pmatrix}$
	Variance	$\hat{V}(\kappa) = \hat{\mu}^2 / 0.6739$
8	Mean/NB	$\hat{E}(\kappa) = \exp(-2.1953 + 0.3309 \log Flow_{through})$
	Variance	$\hat{V}(\kappa) = \hat{\mu}^2 / 0.5561$
9	Mean/NB	$\hat{E}(\kappa) = \exp(-7.1126 + 1.6355 \log Flow_{left-turn} - 0.0102 Flow_{left-turn})$
	Variance	$\hat{V}(\kappa) = \hat{\mu}^2 / 0.3685$

$Flow_{total}$ is a sum of straight-through volume, left-turn volume, and right-turn volume. $\hat{E}(\kappa) = \hat{\mu}$.

Step 8. Calculate the expected crash frequency. An expected crash frequency is calculated for each of the four pairs of conflict movements using the following equation:

$$\hat{\kappa}_i = \hat{E}(\kappa_i | K_i) = \hat{\omega}_i \cdot \hat{E}(\kappa_i) + (1 - \hat{\omega}_i) \cdot K_i$$

where K is the number of recorded crashes of the specified crash pattern (from step 1) in the past 4 years during the specified time period (from step 2).

Step 9. Calculate the variance of the expected crash frequency. A variance of the expected crash frequency is calculated for each of the four pairs of conflict movements using the following equation:

$$\hat{\sigma}_i^2 = \hat{V}(\kappa_i | K_i) = \begin{cases} \hat{\kappa}_i, & if\ Poisson \\ (1 - \hat{\omega}_i) \cdot \hat{\kappa}_i, & if\ NB \end{cases}$$

Table 7. Valid Range of Input Values

Crash Population Reference Group	Input Variable	Minimum	Maximum
1	$Flow_{total}$	12	3,562
	AMPeakHours	2.5	4
2	$Flow_{total}$	8	3,338
3	$Flow_{total}$	3	3,268
	PMPeakHours	3.5	5.1
4	$Flow_{total}$	9	2,053
	OffPeakHours	1.5	3
5	$Flow_{minor}$	1	785
6	$Flow_{through}$	1	2,406
	$Flow_{left\text{-}turn}$	4	584
	$PmPt_{left\text{-}turn}$	0	1
7	$Flow_{through}$	1	1,860
	$Flow_{left\text{-}turn}$	1	568
	$Split_{left\text{-}turn}$	0	1
	$Perm_{left\text{-}turn}$	0	1
	$PmPt_{left\text{-}turn}$	0	1
8	$Flow_{through}$	1	2,958
9	$Flow_{left\text{-}turn}$	1	756

$Flow_{total}$ is a sum of straight-through volume, left-turn volume, and right-turn volume.

Step 10. Calculate the expected crash frequency and its variance for an intersection. An expected crash frequency and its variance for an entire intersection are calculated by summing the expected crash frequencies and their variances over the four pairs of conflict movements:

$$\hat{\kappa} = \sum_{i=1}^{4} \hat{\kappa}_i \text{ and } \hat{\sigma}^2 = \sum_{i=1}^{4} \hat{\sigma}_i^2$$

Note that independent variances are assumed for the summation.

Use of the EB Procedure by Traffic Engineers

Traffic engineers can apply the EB procedure to identify high-risk four-legged signalized intersections (and high-risk conflicting traffics within such intersections). Safety evaluations can be conducted at an intersection level using the expected crash frequency and its variance from step 10. An evaluation can also be conducted at a conflict movement level within an intersection using the expected crash frequency and its variance from steps 8 and 9.

Example Application of the EB Procedure

An example of an application of the EB procedure is presented here.

Description of Example

At a four-legged signalized intersection, there were two crashes involved with a straight-through vehicle and an opposing left-turn vehicle in the past 4 years. One crash occurred at 5:30 P.M.: the first vehicle was going straight from the east approach and the second was turning left from the west approach. The other crash occurred at 8:00 P.M: the first vehicle was turning left from the north approach and the second was going straight from the south approach. The typical P.M. peak hours of this intersection are from 5:00 P.M. through 8:30 P.M. Straight-through traffic volumes during a typical P.M. peak hour are 700 vehicles per hour heading north, 900 heading south, 1,500 heading east, and 750 heading west. A traffic engineer wants to know if this intersection is associated with an abnormally high crash risk.

Application of the EB Procedure

Step 1. Select a crash pattern for safety evaluation from crash patterns 1, 4, and 6. The two crashes in this example are classified as crash pattern 6 because straight-through and opposing left-turning traffics contributed to their occurrence (see Table 3).

Step 2. Select a time of day for the safety evaluation: A.M. peak, mid day, P.M. peak, or evening off peak. The two crashes occurred during the P.M. peak hours.

Step 3. Determine a crash population reference group from groups 1 through 9. According to Table 4, crash pattern 6 and P.M. peak corresponds to crash population reference group 8.

Step 4. Collect the data required for the selected crash population reference group. According to Table 5, the required data for crash population reference group 8 are straight-through traffic volumes during a typical P.M. peak hour. Because there are four approaches in this intersection, there are four straight-through volumes: 700, 900, 1,500, and 750 heading north, south, east, and west, respectively.

Step 5. Select the correct models. In accordance with Table 6, for crash population reference group 8, the following two models are selected for a mean model and a variance model, respectively:

$$\hat{E}(\kappa) = \exp\left(-2.1953 + 0.3309 \log Flow_{through}\right) \text{ with NB and } \hat{V}(\kappa) = \hat{\mu}^2 / 0.5561$$

The mean model is a negative binomial (NB) type.

Step 6. Calculate the mean and variance using the selected models. According to Table 7, the straight-through traffic volume should be between 1 and 2,958. Because the four traffic volumes given in the example are within the valid range, the following results are valid for use in the steps that follow.

$$\hat{E}(\kappa_{north}) = \exp\left(-2.1953 + 0.3309 \log 700\right) = 0.2854, \; \hat{V}(\kappa_{north}) = 0.2854^2 / 0.5561 = 0.1465$$

$\hat{E}(\kappa_{south}) = \exp(-2.1953 + 0.3309 \log 900) = 0.2959$, $\hat{V}(\kappa_{south}) = 0.2959^2 / 0.5561 = 0.1575$
$\hat{E}(\kappa_{east}) = \exp(-2.1953 + 0.3309 \log 1500) = 0.3184$, $\hat{V}(\kappa_{east}) = 0.3184^2 / 0.5561 = 0.1823$
$\hat{E}(\kappa_{west}) = \exp(-2.1953 + 0.3309 \log 750) = 0.2882$, $\hat{V}(\kappa_{east}) = 0.2882^2 / 0.5561 = 0.1494$

Step 7. Calculate the EB weight.

$$\hat{\omega}_{north} = 1 / \left[1 + \hat{V}(\kappa_{north}) / \hat{E}(\kappa_{north})\right] = 1 / [1 + 0.1465 / 0.2854] = 0.6608$$
$$\hat{\omega}_{south} = 1 / [1 + 0.1575 / 0.2959] = 0.6527$$
$$\hat{\omega}_{east} = 1 / [1 + 0.1823 / 0.3184] = 0.6359$$
$$\hat{\omega}_{west} = 1 / [1 + 0.1494 / 0.2882] = 0.6586$$

Step 8. Calculate the expected crash frequency.

$$\hat{\kappa}_{north} = \hat{\omega}_{north} \cdot \hat{E}(\kappa_{north}) + (1 - \hat{\omega}_{north}) \cdot K_{north} = 0.6608 \times 0.2854 + 0.3392 \times 1 = 0.5278$$
$$\hat{\kappa}_{south} = 0.6527 \times 0.2959 + 0.3473 \times 0 = 0.1931$$
$$\hat{\kappa}_{east} = 0.6359 \times 0.3184 + 0.3641 \times 0 = 0.2025$$
$$\hat{\kappa}_{west} = 0.6586 \times 0.2882 + 0.3414 \times 1 = 0.5312$$

There were two traffic crashes. One crash occurred between the first vehicle going straight from the east approach (i.e., heading west) and the second vehicle turning left from the west approach (i.e., heading east). Thus, $K_{west} = 1$. The other crash occurred between the first vehicle turning left from the north approach (i.e., heading south) and the second vehicle going straight from the south approach (i.e., heading north). Thus, $K_{north} = 1$.

Step 9. Calculate the variance of the expected crash frequency.

$$\hat{\sigma}^2_{north} = (1 - \hat{\omega}_{north}) \cdot \hat{\kappa}_{north} = (1 - 0.6608) \times 0.5278 = 0.1790$$
$$\hat{\sigma}^2_{south} = (1 - 0.6527) \times 0.1931 = 0.0671$$
$$\hat{\sigma}^2_{east} = (1 - 0.6359) \times 0.2025 = 0.0731$$
$$\hat{\sigma}^2_{west} = (1 - 0.6586) \times 0.5312 = 0.1814$$

Step 10. Calculate the expected crash frequency and its variance for an intersection.

$$\hat{\kappa} = \hat{\kappa}_{north} + \hat{\kappa}_{south} + \hat{\kappa}_{east} + \hat{\kappa}_{west} = 0.5278 + 0.1931 + 0.2025 + 0.5312 = 1.4546$$
$$\hat{\sigma}^2 = \hat{\sigma}^2_{north} + \hat{\sigma}^2_{south} + \hat{\sigma}^2_{east} + \hat{\sigma}^2_{west} = 0.1790 + 0.0671 + 0.0731 + 0.1814 = 0.5006$$

For the variance resulting from the summation of the four variances, an independence assumption is required. This assumption was verified with the data used for this study.

Interpretation of Results

Once expected crash frequencies (means; $\hat{\kappa}$'s) and their variances ($\hat{\sigma}^2$'s) are obtained, the traffic safety associated with each of the four pairs of conflict movements and an entire intersection for crash population reference group 8 can be evaluated (i.e., crash pattern 6 during P.M. peak hours). For the entire intersection, the expected crash frequency during P.M. peak hours in crash pattern 6 in 4 years is 1.4546, and the observed crash frequency is 2. The traffic engineer might be tempted to conclude that the intersection is associated with a higher crash risk than would normally be expected because what was observed was greater than what was expected. However, in order to draw such a conclusion, a variance should be taken into account.

One way of accounting for a variance is a confidence interval. A confidence interval can be computed using the following equation:

$$C.I. = \hat{\kappa} \pm z_{\alpha/2} \times \hat{\sigma}$$

Using the results of the example, the confidence interval for the entire intersection is constructed at a 90% confidence level (i.e., 10% significance level; $\alpha = 0.1$) as follows:

$$C.I. = 1.4546 \pm z_{0.05} \times \sqrt{0.5006} = 1.4546 \mp 1.645 \times 0.7075 = [0.2908, 2.6184]$$

This interval implies that the expected crash frequency can be located between 0.2908 and 2.6184 at the 90% confidence level. Because the upper limit of the interval, 2.6184, is greater than the observed crash frequency, 2, the traffic engineer can conclude that the intersection is not associated with an abnormally high risk of crash pattern 6 during P.M. peak hours.

The other way of accounting for a variance is a *z*-test. The null and alternative hypotheses can be set as follows:

Ho: No difference (implying that the expected crash frequency, 1.4546, is not statistically different from the observed frequency, 2, for the example)

Ha: Difference

A *z*-score can be calculated using the following equation:

$$z\ score = \frac{\hat{\kappa} - c}{\hat{\sigma}}.$$

Basically, the *z*-test will lead to the same conclusion as that drawn using the confidence interval.

EB Case Study

The EB procedure was applied to the data used in developing the procedure to illustrate the use of the study results. Appendix D presents the outcomes of the EB procedure for crash pattern 6 using the data of the 35 intersections used for model development. By entering straight-through and left-turn traffic volumes and the type of left-turn signal phase of those intersections, the estimated expected crash frequency, $\hat{\kappa} = E(\kappa \mid K)$, corresponding to the input condition and the predicted probability that the condition is unsafe was calculated.

Since occurrences of crash pattern 6 were rare during the mid day and evening off peak periods, all pairs of conflict movements of crash pattern 6 (i.e., straight-through and opposing left-turn traffic movements) with at least 1 crash in 4 years were marked as being "unsafe." This means that when at least one crash between the two traffic movements occurred during the mid day period (or during the evening off peak period) in 4 years, that condition is believed to be associated with abnormally high risk even without the application of the EB procedure. This finding was foreseeable from Figures 15 and 19 discussed later in this report. Most of the points in these figures were below 1 in terms of the average number of crashes, and this implies that the expected number of crashes is likely to be less than 1 in 4 years over all ranges of traffic flows. Therefore, even 1 crash in 4 years between the two traffic flows is higher than what is expected, regardless of traffic volume and type of left-turn signal.

However, most pairs of traffic conflict movements of crash pattern 6 associated with 1 crash in 4 years were determined to be safe (at the 95% confidence level) for the P.M. peak period whereas two pairs of the movements with 1 crash were determined to be safe (at the 95% confidence level) for the A.M. peak period. The remaining pairs with 1 crash during the A.M. peak were determined to be unsafe at the 95% confidence level.

Development of EB Procedure

Exploratory Analysis

After combining the 16 crash patterns and the four times of day described in the Methods section, a total of 64 crash population reference groups were identified for multi-vehicle crashes. The original intention was to develop an SPF for each of the 64 groups, and the intention was to apply the EB method for each group using the estimated SPF.

However, many of the 64 groups had too few crash counts throughout the 4 years of the study period (2001–2004); thus, they were likely to be excluded from model development because it would to be difficult to estimate their SPFs. Although the function might be estimated, it is not likely to be useful for the EB application because crashes of those groups are rare so that even 1 crash occurrence in a group is considered to constitute an abnormally high risk. In such cases, a daily aggregation approach might be used. Aggregating the data over 1 day (i.e., total traffic crash count over the entire time period covering the four times of day and hourly traffic volume averaged over the hours of the time period) might be used to estimate an SPF.

Table 8 shows the traffic crash counts from the 33 intersections for 4 years from 2001 through 2004 by crash patterns and times of day. A total of 643 traffic crashes occurred at the 33 intersections in the 4 years. Of the 643 crashes including 36 single-vehicle crashes, 35% (223 crashes) match crash pattern 1, 25% (163 crashes) match crash pattern 6, and 10% (66 crashes)

Table 8. Crash Counts for Four Years by Reference Population Group

Crash Pattern	Time of Day				Total
	A.M. peak	Mid day	P.M. peak	Off peak	
0	4	8	16	8	36
1	47	71	66	39	223
2	0	0	0	0	0
3	0	1	1	0	2
4	14	23	19	10	66
5	3	1	5	0	9
6	37	38	54	34	163
7	0	5	7	0	12
8	9	9	7	4	29
9	1	4	6	0	11
10	6	10	8	6	30
11	1	6	4	3	14
12	1	3	2	0	6
13	1	1	1	1	4
14	1	2	3	1	7
15	1	1	3	1	6
16	5	13	6	1	25
Total	131	196	208	108	643

Crash Pattern 0 represents a single-vehicle crash.

match crash pattern 4. These three crash patterns make up 70% of all multi-vehicle crashes at the intersections, and they were the main focus of the analysis.

Only crash patterns 1 and 6 appeared to be eligible for a separate analysis by the four times of day, and the other crash patterns seemed to have too few crashes to capture a meaningful relationship (i.e., SPF) between crash occurrence and traffic flow (and other variables) by a separate time of day. However, for crash pattern 4, it appears that a daily aggregate analysis probably would identify a useful relationship. In the end, crash patterns 1 and 6 were analyzed by the four times of day, and crash pattern 4 was analyzed over the entire period of the four times of day. Thus, for the nine crash reference population groups (i.e., crash patterns 1 and 6 by four times of day and crash pattern 4 over the period of the four times of day), the final estimated SPFs were estimated, and the expected crash frequencies were computed based on the EB method using the SPFs. The SPFs of the nine reference groups were estimated for each group. The three crash patterns that were included in the analysis, and they can be defined as follow:

- *Crash pattern 1:* Same-direction crash (angle, rear-end, or sideswipe) that occurs after exiting the intersection.

- *Crash pattern 4:* Right-angle crash between two adjacent straight-through vehicle movements in the intersection.

- *Crash pattern 6:* Angle, head-on, or opposite sideswipe crash between a straight-through vehicle movement and an opposing left-turn vehicle movement in the intersection.

Model Development

The procedure for developing final mean and variance models was described in steps MD1 through MD3. The model development results of the same procedure applied to the nine crash reference population groups are summarized here.

Crash Pattern 1 in A.M. Peak

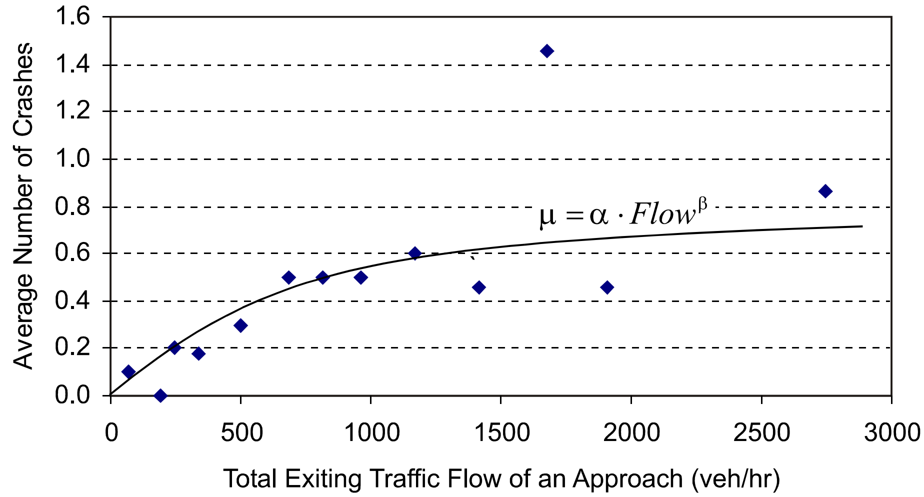
A total of 35 intersections were included for this reference population group. Because there are four approaches per intersection, 140 (35×4) observations (or approaches) can be used in the model development in theory. However, one approach had to be removed because a data value was missing. Therefore, a total of 139 observations were used to develop an SPF for this reference population group. The quick diagnostic check for overdispersion (variance = 0.846 versus $2 \times$ mean = 0.987) described in step MD1 turned out to be indecisive in indicating the presence of overdispersion.

In order to propose an appropriate functional form for the relationship between traffic crash and the traffic flow exiting from an intersection and the ratio of straight-through traffic, scatter plots of average aggregated crash counts, described in step MD2.1(a), were created and are shown in Figure 7. Plots of the cumulative sum of residuals, described in step MD2.1(b), confirmed that all the functional forms were appropriate for use in the mean model (i.e., SPF). However, the ratio variable in the model turned out to be statistically insignificant; thus, it was removed from the final specification.

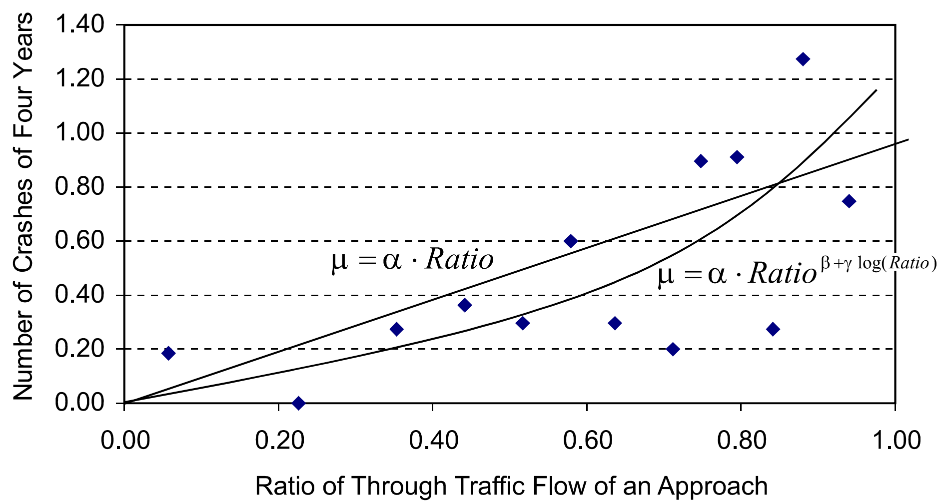
The final models with the Poisson and NB error structures were estimated separately. Using the results of these two models, four different overdispersion tests, described in step MD2.2, were performed: (1) dispersion parameters using deviance and Pearson's Chi-square statistics, (2) regression-based equidispersion test using the final Poisson model, (3) confidence interval of the NB dispersion parameter of the NB model, and (4) Lagrange multiplier (or score) test using the final NB model.

The dispersion parameters estimated by deviance and Pearson's Chi-square statistics suggested no or very weak overdispersion; the regression-based equidispersion test using the final Poisson model suggested the presence of overdispersion. The confidence interval of the NB parameter included zero implied no overdispersion, and the Lagrange multiplier test using the final NB model confirmed the absence of overdispersion. Thus, the Poisson model was selected for this reference group: crash pattern 1 in the A.M. peak.

To evaluate the goodness of fit, three pseudo R-squared measures, described in step MD2.3 (a), were computed, but deviance-based R-square (R_D^2) was excluded because of a problem in estimating the fully saturated model. The likelihood-ratio index (R_L^2) read 0.151, and the correlation-based R-square (R_C^2) read 0.168, which seemed fine as compared to those in other population reference groups. As discussed in the Methods section, the pseudo R-squared measures do not have a good statistical interpretation, such as the proportion of an explained



(a) Total traffic flow of an approach exiting from an intersection



(b) Ratio of straight-through traffic flow in total traffic flow of an approach exiting from an intersection

Figure 7. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 1 in A.M. Peak. Plot (a) suggests the functional form of $\mu = \alpha \cdot Flow^{\beta}$ for total traffic flow exiting from an intersection. The ratio of the straight-through traffic flow in the total exiting traffic flow was also considered for the SPF, and two forms are suggested by plot (b): $\mu = \alpha \cdot Ratio^{\beta}$ or $\mu = \alpha \cdot Ratio^{\beta+\gamma \log(Ratio)}$. (See Ardekani et al. [2002] for an illustration of these functional forms.)

variance for a linear model. Thus, they should not be compared with typical values found in studies using linear models but rather should be used for relative comparisons among non-linear models.

For visual comparison, mean observed probabilities, univariate Poisson probabilities, Poisson model probabilities, and NB model probabilities were computed using the final Poisson and NB models described in step MD2.3(b); they are drawn in Figure 8. It seems that the NB

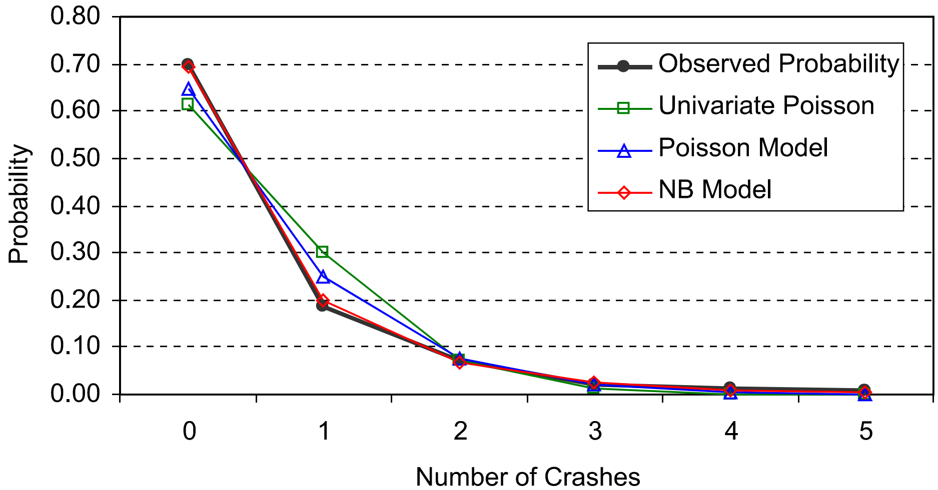


Figure 8. Mean Observed and Predicted Probabilities for Crash Pattern 1 in A.M. Peak

model fits the data slightly better than does the Poisson model. However, the probabilities by the Poisson model are not significantly different from the observed probabilities (underestimated for the zero-crash case and overestimated for the one-crash case), and the tests for overdispersion suggested that the Poisson model is appropriate.
The estimated final Poisson mean model (or SPF) is written as follow:

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\left(-9.1012 + 0.6566 \log Flow_{total,i} + 1.088 AMPeakHours_i\right)$$

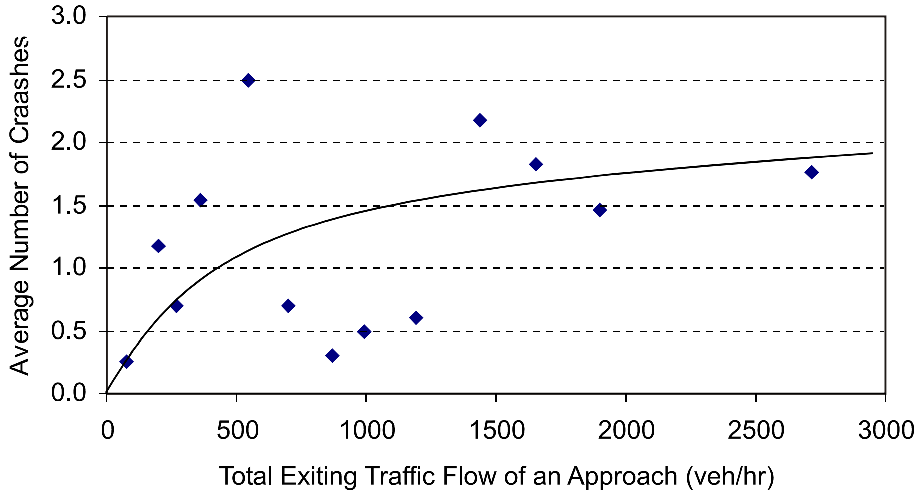
where i is an index of an approach; $\hat{\mu}_i$ is the estimated mean crash count in 4 years; $Flow_{total,i}$ is the total traffic flow on an approach i exiting from an intersection (vehicles per hour [veh/hr]); and $AMPeakHours_i$ is the number of hours of the A.M. peak period. $Flow_{total,i}$ (⮥↑⮤) is summation of left-turn (⮥), right-turn (⮤), and straight-through (↑) traffic flows coming from three different approaches and moving toward an approach i. A variance model was not needed because the Poisson model was selected over the NB model.

Crash Pattern 1 at Mid Day

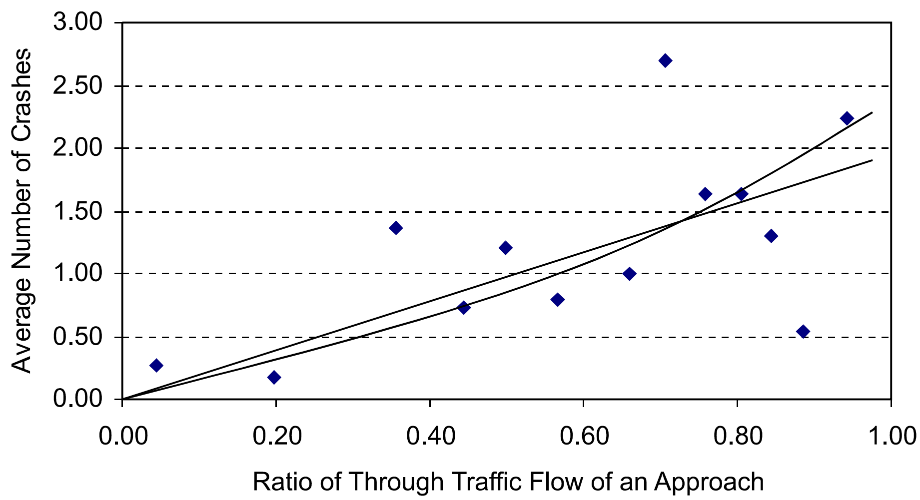
A total of 35 intersections were included, and 140 observations (or 140 approaches) were used to estimate the SPF for this population reference group. The quick diagnostic check indicated probable presence of overdispersion (variance = 4.38 > 2 × mean = 2.414).

Scatter plots of average aggregated crash counts were created against the total traffic flow and the ratio of straight-through traffic and are presented in Figure 9.

The final Poisson and NB models were estimated, and the four tests for overdispersion were performed using the parameter estimates of these models. All four tests confirmed that the crash count in this population reference group is overdispersed. Therefore, the NB model was chosen as the final model for the reference group of crash pattern 1 for mid day.



(a) Total traffic flow of an approach exiting from an intersection



(b) Ratio of straight-through traffic flow in total traffic flow of an approach exiting from an intersection

Figure 9. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 1 in Mid Day. Plots (a) and (b) suggest the functional forms of $\mu = \alpha \cdot Flow^{\beta}$ for traffic flow and $\mu = \alpha \cdot Ratio^{\beta}$ or $\mu = \alpha \cdot Ratio^{\beta + \gamma \log(Ratio)}$ for the ratio, which were supported by the plots of the cumulative sum of residuals.

Three pseudo R-squared measures (excluding R_D^2 because of an estimation problem of the fully saturated model) were computed, and they were lower than those for crash pattern 1 in the A.M. peak: $R_L^2 = 0.031$, $R_C^2 = 0.026$, and $R_k^2 = 0.067$. Figure 10 allows a visual evaluation of the goodness of fit. The NB model fitted the crash counts very nicely, and the two Poisson probabilities deviate greatly from observed probabilities at the crash counts, 0, 1, and 2.
The final NB mean model (or SPF) and the final variance model were estimated as follows:

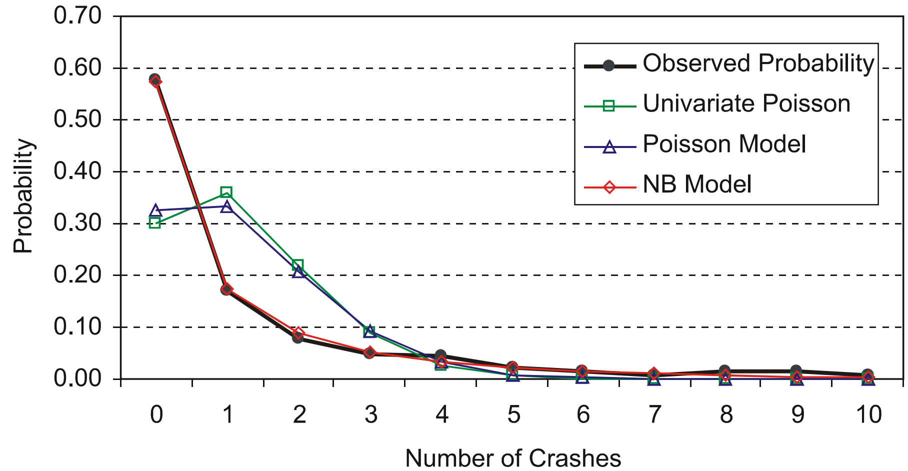


Figure 10. Mean Observed and Predicted Probabilities for Crash Pattern 1 in Mid Day

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp(-1.8844 + 0.3113 \log Flow_{total,i}) \text{ with } \hat{k} = 2.3839 \text{ for the mean model}$$

and

$$\hat{\sigma}^2_{\kappa_i} = \hat{V}(\kappa_i) = \frac{\hat{\mu}_i^2}{0.6661} \text{ for the variance model}$$

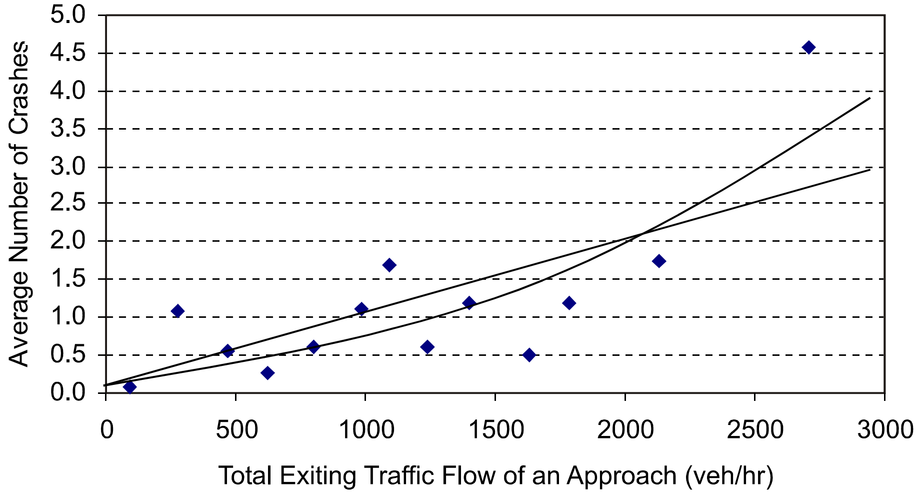
where $Flow_{total,i}$ is the total traffic flow on an approach i exiting from an intersection (veh/hr) and $\hat{k}$ is an estimate of the NB dispersion parameter.

Crash Pattern 1 in P.M. Peak

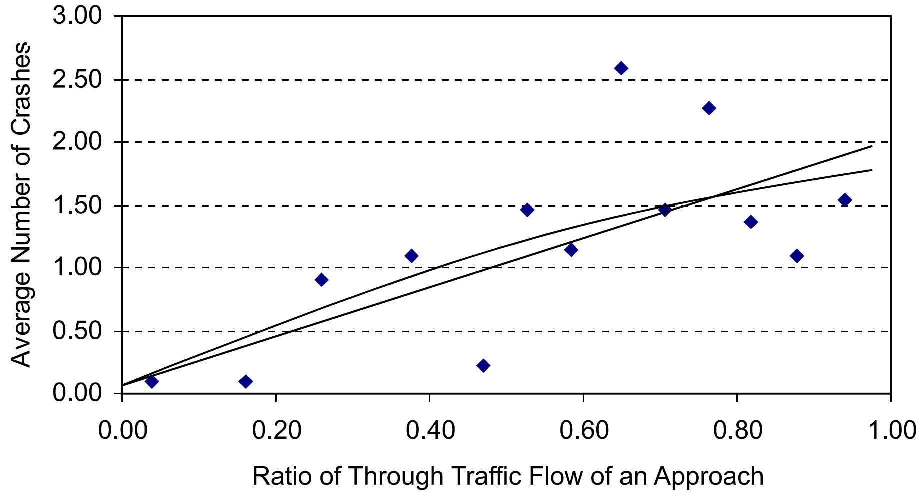
A total of 35 intersections were included in this population reference group, and 139 observations (or approaches) (excluding 1 approach with missing information) were used for developing the final SPF (or mean model). A quick diagnostic check for overdispersion indicated possible overdispersion in the crash counts in this reference group.

Figure 11 shows scatter plots for identifying probable functional forms between crash counts and the total traffic flow and the ratio of straight-through traffic flow. All of the forms were found to be acceptable by the plots of cumulative sum of residuals. However, the estimations of the models with the above variable forms revealed that $\mu = \alpha \cdot Flow^{\beta}$ is best suited for this population reference group according to the statistical significance of coefficient parameters corresponding to the forms and the goodness of fit models.

The final Poisson and NB models were estimated, and the four overdispersion tests were performed using the final models. All four tests suggested a remaining overdispersion, which implies that the NB model is preferred over the Poisson model for the reference group of crash pattern 1 in the P.M. peak.



(a) Total traffic flow of an approach exiting from an intersec tion



(b) Ratio of straight-through traffic flow in total traffic flow of an approach exiting from an intersection

Figure 11. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 1 in P.M. Peak. Plot (a) suggests $\mu = \alpha \cdot Flow^{\beta}$ (with β being close to 1) or $\mu = \alpha \cdot Flow^{\beta+\gamma \log(Flow)}$, and plot (b) suggests $\mu = \alpha \cdot Ratio^{\beta}$ (with β being close to 1) or $\mu = \alpha \cdot Ratio^{\beta}$.

Three pseudo R-squared measures were computed:

1. $R_L^2 = 0.198$
2. $R_C^2 = 0.168$
3. $R_k^2 = 0.341$.

All three measures are relatively high compared to those for other population reference groups, and the dispersion parameter-based R-square is especially high. Figure 12 allows a

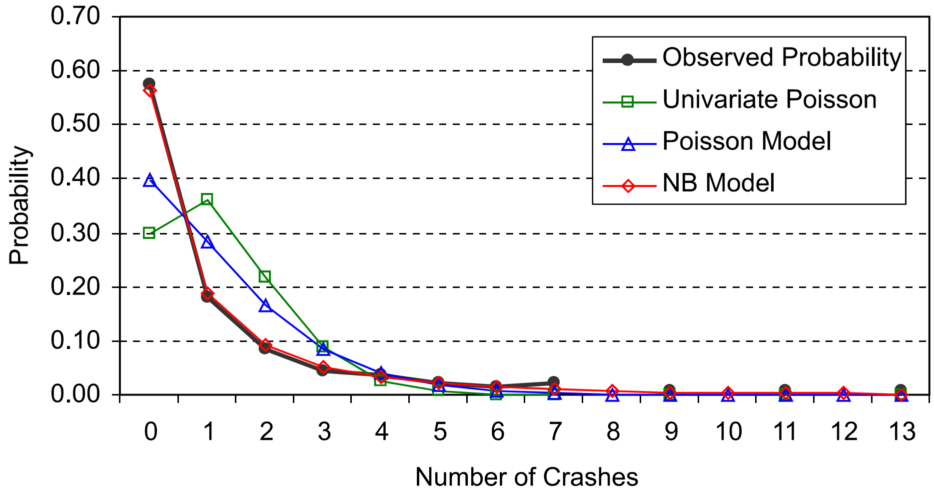


Figure 12. Mean Observed and Predicted Probabilities for Crash Pattern 1 in P.M. Peak

visual comparison of different predicted probabilities. The NB model resulted in the closest fit to the data in terms of the mean probabilities. The Poisson model and univariate Poisson overestimated the probability of no crash occurrences by about 40% and underestimated the probabilities of 1, 2, and 3 crashes by between 45% and 55%.

The final NB mean model (or SPF) and the final variance model are written as follows:

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\left(-9.2153 + 0.08012 \log Flow_{total,i} + 0.8731 PMPeakHours_i\right)$$

for the mean model with $\hat{k} = 1.6944$ for the mean model and

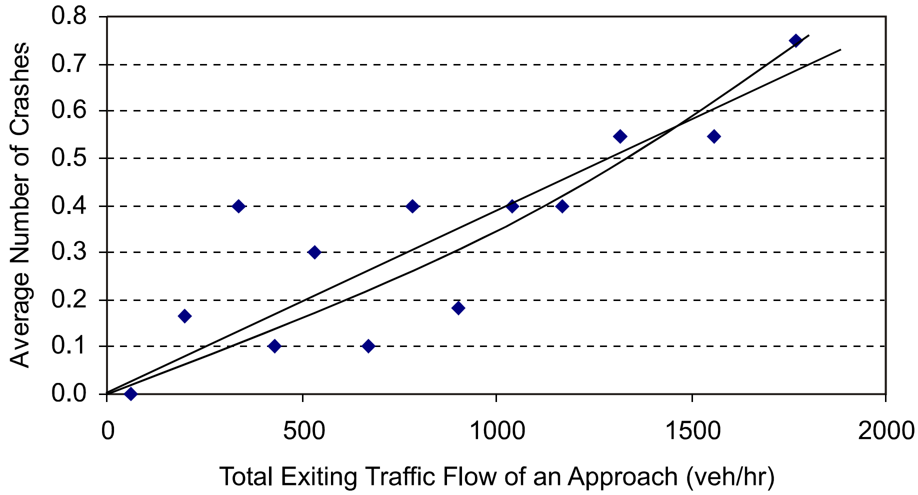
$$\hat{\sigma}^2_{\kappa_i} = \hat{V}(\kappa_i) = \frac{\hat{\mu}_i^2}{1.6529} \text{ for the variance model}$$

where $Flow_{total,i}$ is the total traffic flow on an approach *i* exiting from an intersection (veh/hr), and $PMPeakHours_i$ is the number of hours of the P.M. peak period.

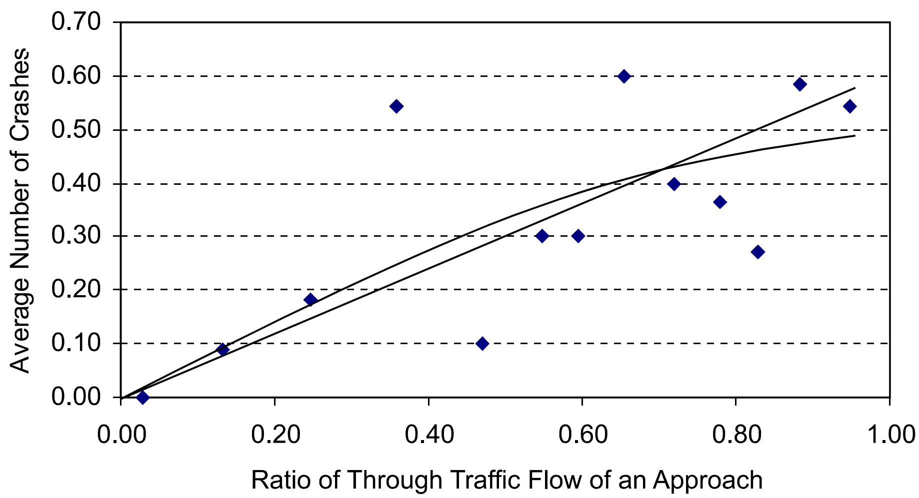
Crash Pattern 1 in Evening Off Peak

A total of 35 intersections were included, and 139 observations (or approaches) exclusive of one approach with missing information were used for the development of the final mean model. The result of the quick diagnostic check was indecisive, implying that there is a chance for the crash counts of this reference group to be overdispersed.

Figure 13 presents scatter plots for offering hints on appropriate functional forms for the mean model (or SPF). These forms were found to be acceptable by the plots of the cumulative sum of residuals. However, the estimation results (i.e., the statistical significance of the coefficients and the goodness of fit) using the forms suggest $\mu = \alpha \cdot Flow^{\beta}$ is the most suitable for this population reference group.



(a) Total traffic flow of an approach exiting from an intersection



(b) Ratio of straight-through traffic flow in total traffic flow of an approach exiting from an intersection

Figure 13. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 1 in Off Peak. Plot (a) suggests $\mu = \alpha \cdot Flow^{\beta}$ (with β being close to 1) and $\mu = \alpha \cdot Flow^{\beta+\gamma \log(Flow)}$, and plot (b) suggests $\mu = \alpha \cdot Ratio^{\beta}$ (with β being close to 1) and $\mu = \alpha \cdot Ratio^{\beta}$.

Using the estimated final Poisson and NB models, the four tests for overdispersion were performed. All four tests suggested that overdispersion was not present after some explanatory variables were added to the models. This test result implies that the Poisson model is appropriate for this reference group. For assessing the goodness of fit, two pseudo R-squared measures were computed as R_L^2=0.097 and R_C^2=0.077. In addition, Figure 14 provides a graphical comparison of the model fitness using mean predicted probabilities. Although the Poisson model is more appropriate than the NB model according to the results of the four overdispersion tests, the plot of mean probabilities looks slightly more favorable for the NB model than for the Poisson model. However, the difference across different average predictions appears negligible.

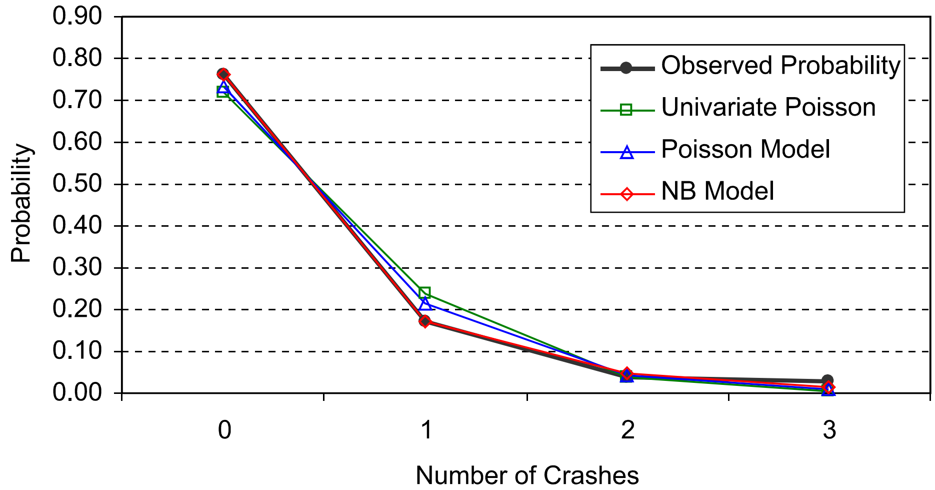


Figure 14. Mean Observed and Predicted Probabilities for Crash Pattern 1 in Off Peak

The final Poisson mean model (or SPF) is written as:

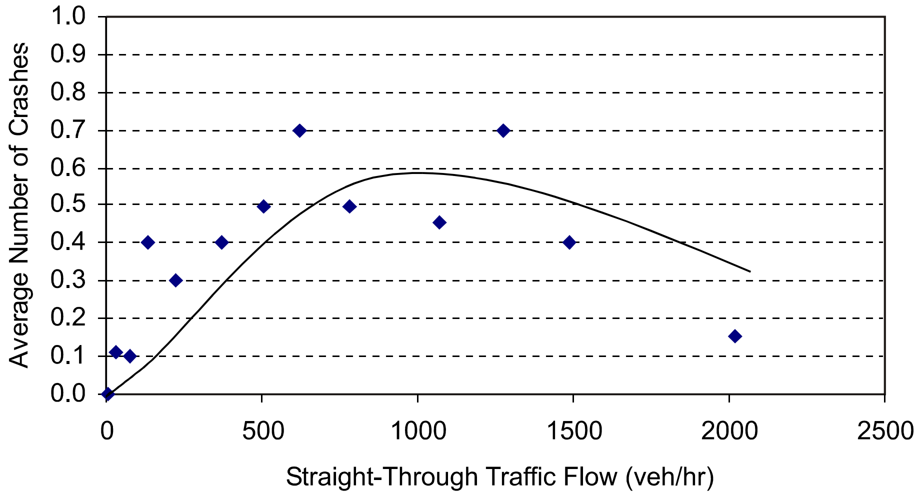
$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp(-8.8123 + 0.7337 \log Flow_{total,i} + 0.8731 OffPeakHours_i)$ where $OffPeakHours_i$ is the number of hours from the end of the P.M. peak to the start of the night "free operation" signal plan, and $Flow_{total,i}$ is the total traffic flow on an approach *i* exiting from an intersection (veh/hr) (see the SPF of crash pattern 1 in the A.M. peak for a detailed description of the total traffic flow). Because the Poisson model was selected over the NB model, it was not necessary to estimate a variance model.

Crash Pattern 6 in A.M. Peak

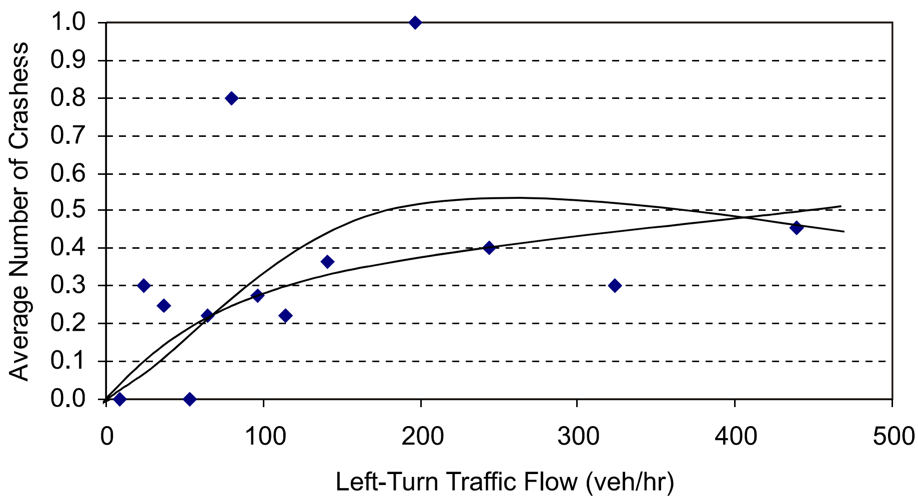
A total of 33 intersections were included for this reference group, and 131 (33 × 4) observations (or pairs of straight-through and opposing left-turn traffic flows) excluding 1 pair with missing information were used for model development. The quick diagnostic check for overdispersion resulted in an indecisive outcome; thus, further tests were required to understand the nature of the crash counts for this reference group.

Figure 15 offers visual clues on the functional forms of the SPF with respect to traffic flow variables. They were all confirmed as acceptable by plots of the cumulative sum of their residuals and used for the final models.

Using the estimates of the final Poisson and NB models, the four tests for overdispersion were performed. All four tests suggested no overdispersion; thus, the Poisson model was chosen for this reference group of crash pattern 6 in A.M. peak. For the goodness of fit, two pseudo R-squared measures were computed as R_L^2=0.136 and R_C^2=0.155, which appear high relative to those values for other reference groups. For visual evaluation of the model fit, mean probabilities of observed and predicted crash counts were computed and are displayed in Figure 16. The Poisson and NB models appear to be equally well fitted to the observed data.



(a) Straight-through traffic flow from an approach



(b) Left-turn traffic flow from an opposing approach

Figure 15. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 6 in A.M. Peak. Plot (a) suggests $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$, equivalently, $\exp(\alpha^* + \beta \log Flow + \gamma \cdot Flow)$ for straight-through flow, and plot (b) suggests $\mu = \alpha \cdot Flow^{\beta}$ or $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ for left-turn flow.

The final Poisson mean model (or SPF) is written as:

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\begin{pmatrix} -6.8155 + 0.7736 \log Flow_{through,i} - 0.0013 Flow_{through,i} \\ +0.3212 \log Flow_{left-turn,i} + 0.7209 PmPt_{left-turn,i} \end{pmatrix}$$

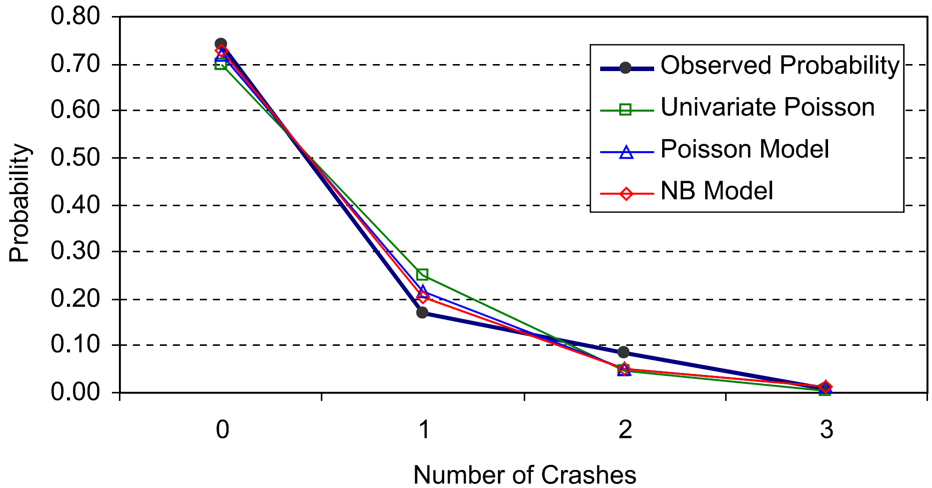


Figure 16. Mean Observed and Predicted Probabilities for Crash Pattern 6 in A.M. Peak

where

i is an index of a pair of straight-through traffic flow and opposing left-turn traffic flow

$Flow_{through,i}$ is a straight-through traffic flow from an approach (veh/hr)

$Flow_{left-turn,i}$ is a left-turn traffic flow from an approach opposite to the approach of the straight-through flow (veh/hr)

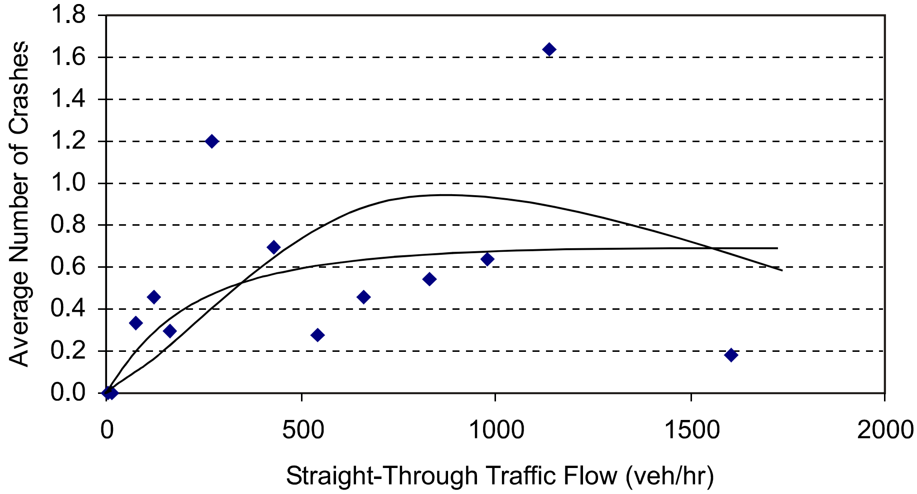
$PmPt_{left-turn,i}$ is an indicator of permissive and protected left-turn signal phase (1 if so, 0 if not).

A variance model was not needed for this reference group because the Poisson model was selected over the NB model.

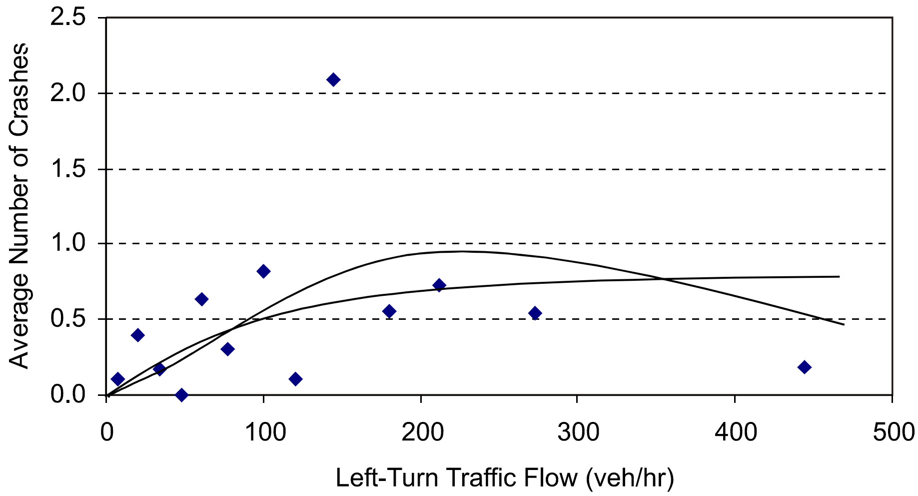
Crash Pattern 6 at Mid Day

A total of 35 intersections were included, and 137 observations (or pairs of straight-through and opposing left-turn traffic flows) excluding three pairs with missing information were used for developing final models. The quick overdispersion diagnostic check identified probable overdispersion in the crash data of this reference group. Figure 17 suggests appropriate functional forms for the SPF. They were all acceptable according to plots of the cumulative sum of their residuals. According to the statistical significance of the coefficients and the goodness of fit, $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ was best suited for both flow variables.

Using the final estimates of the Poisson and NB models, the four tests for overdispersion were performed. Their results confirmed the presence of overdispersion. Therefore, the NB model was selected for the reference group of crash pattern 6 in mid day.



(a) Straight-through traffic flow from an approach



(b) Left-turn traffic flow from an opposing approach

Figure 17. Average Aggregated Crash Counts for Crash Pattern 6 in Mid Day. Plots (a) and (b) suggest $\mu = \alpha \cdot Flow^{\beta}$ and $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ for both straight-through and left-turn flows.

For the goodness of fit, three pseudo R-squared measures were computed as R_L^2=0.158, R_C^2=0.092, and R_k^2=0.445. The dispersion parameter-based R-square shows a very high value compared to the other two R-squared measures. Figure 18 allows a graphical comparison among different predictions of mean probabilities. As expected, the NB model provided the best fit to the data.

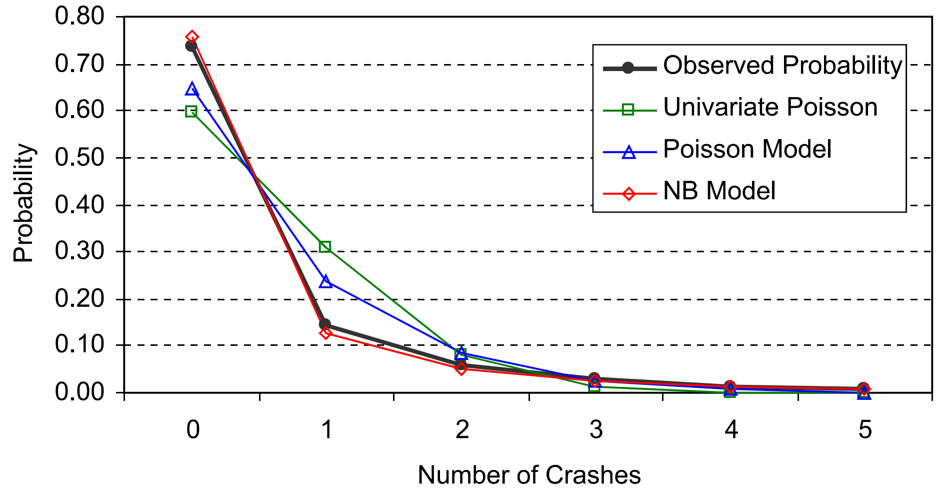


Figure 18. Mean Observed and Predicted Probabilities for Crash Pattern 6 in Mid Day

The final NB mean model (or SPF) and the final variance model were estimated and are written as follows:

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\begin{pmatrix} -14.9690 + 1.6388 \log Flow_{through,i} - 0.0026 Flow_{through,i} \\ +1.3497 \log Flow_{left-turn,i} - 0.0096 Flow_{left-turn,i} \\ +1.5125 Split_{left-turn,i} + 2.7704 Perm_{left-turn,i} + 1.2094 PmPt_{left-turn,i} \end{pmatrix} \text{ with}$$

$\hat{k} = 1.8137$ for the mean model and

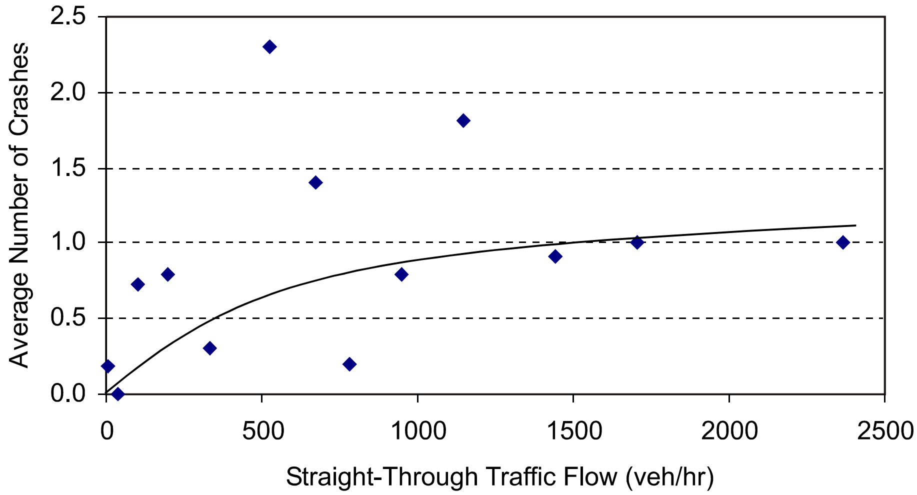
$$\hat{\sigma}^2_{\kappa_i} = \hat{V}(\kappa_i) = \frac{\hat{\mu}_i^2}{0.6739} \text{ for the variance model}$$

where $Split_{left-turn,i}$ and $Perm_{left-turn,i}$ are indicators for split and permissive signal phases for left-turn traffic, and the other parameters and variables are the same as described previously in crash pattern 6 in A.M. peak.

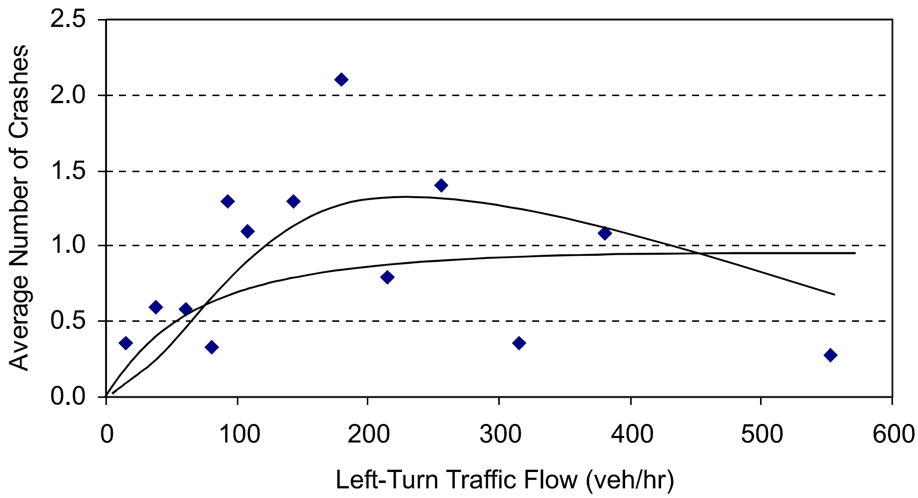
Crash Pattern 6 in P.M. Peak

A total of 34 intersections were included, and 135 observations (or pairs of straight-through and opposing left-turn traffic flows) excluding 1 pair with missing information were used for model development. The quick diagnostic check suggested potential overdispersion in that the unconditional variance of the crash counts (1.531) was greater than twice the unconditional mean of the crash counts (2 × 0.5182 = 1.0365).

Scatter plots of average aggregated crash counts by traffic flow group were created to gain insights into functional forms for the SPF and are displayed in Figure 19. The suggested forms were considered acceptable according to the plots of the cumulative sum of the residuals. The forms $\mu = \alpha \cdot Flow^{\beta}$ for straight-through flow and $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ were.



(a) Straight-through traffic flow from an approach



(b) Left-turn traffic flow from an opposing approach

Figure 19. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 6 in P.M. Peak. Plot (a) suggests $\mu = \alpha \cdot Flow^{\beta}$ for straight-through flow, and plot (b) suggests $\mu = \alpha \cdot Flow^{\beta}$ or $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ for left-turn flow.

selected for the final model development based on the statistical significance of the coefficients of the flow variables and the goodness of fit.

Using the final estimates of the Poisson and NB models, the four tests for overdispersion were performed. All four tests confirmed the presence of overdispersion. Therefore, the NB model was selected for the reference group of crash pattern 6 in P.M. peak. To evaluate the fit of the model, three pseudo R-squared measures were computed as R_L^2=0.044, R_C^2=0.031, and R_k^2=0.14, which were low compared to those for the other reference groups. Figure 20 allows a

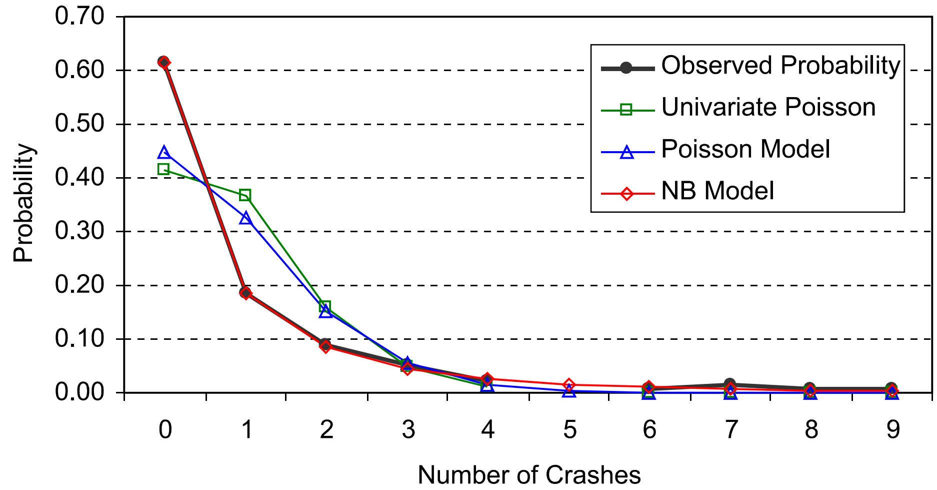


Figure 20. Mean Observed and Predicted Probabilities for Crash Pattern 6 in P.M. Peak

visual comparison of the mean probabilities predicted by different models. The predictions of the NB model seem to fit the data better than those of the Poisson models.

The final NB mean model (or SPF) and the final variance model were written as follows:

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\left(-2.1953 + 0.3309 \log Flow_{through,i}\right) \text{ with } \hat{k} = 1.9570 \text{ for the mean model,}$$

and

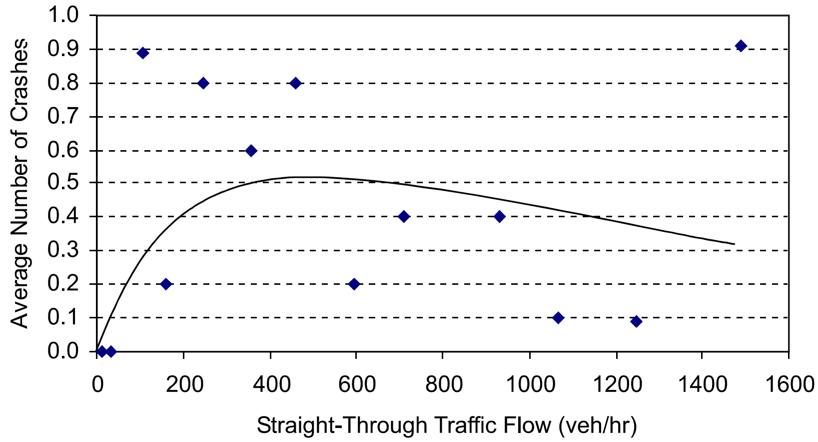
$$\hat{\sigma}^2_{\kappa_i} = \hat{V}(\kappa_i) = \frac{\hat{\mu}_i^2}{0.5561} \text{ for the variance model.}$$

Crash Pattern 6 in Evening Off Peak

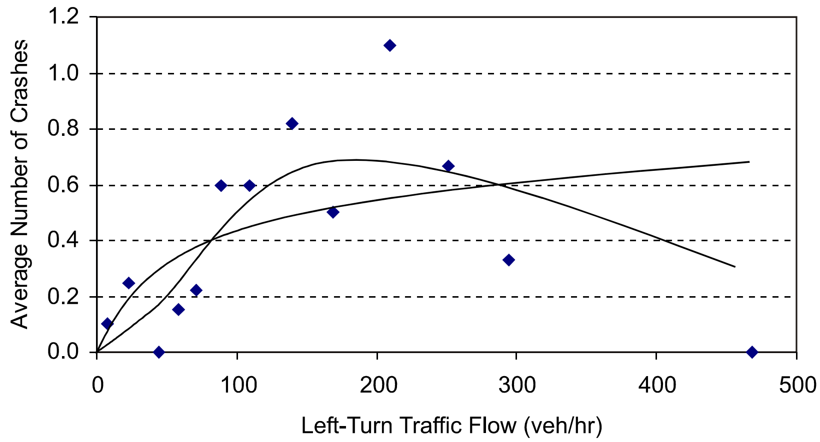
A total of 34 intersections were included in this reference group, and 133 observations (or pairs of straight-through and opposing left-turn traffic flows) excluding 3 pairs with missing information were used for model development. The result of the quick diagnostic check for overdispersion was indecisive.

Figure 21 provides clues to functional forms for the SPF with respect to traffic flows. The proposed forms were regarded as acceptable according to the plots of the cumulative sum of the residuals. Form $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ was selected for the final model development for both flows based on the statistical significance of the coefficients of the flow variables and the goodness of fit models.

The final estimates of the Poisson and NB models were obtained, and four tests for overdispersion using those estimates were performed. All four tests confirmed overdispersion conditional on the explanatory variables. Thus, the NB model was chosen for the reference group of crash pattern 6 in off peak. Three pseudo R-squared measures were computed for



(a) Straight-through traffic flow from an approach



(b) Left-turn traffic flow from an opposing approach

Figure 21. Average Aggregated Crash Counts Against Flow Variables for Crash Pattern 6 in Off Peak. Plot (a) suggests $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ for straight-through flow, and plot (b) suggests $\mu = \alpha \cdot Flow^{\beta}$ or $\mu = \alpha \cdot Flow^{\beta} \cdot \exp(\gamma \cdot Flow)$ for left-turn flow.

evaluating the goodness of fit, R_L^2=0.062, R_C^2=0.087, and R_k^2=0.351, which were somewhat low except for the dispersion parameter-based R-square. Figure 22 provides a graphical comparison among different models. As expected, the NB model outperformed the other models. The final Poisson mean model (or SPF) and the final variance model were estimated as follows:

$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\left(-7.1126 + 1.6355 \log Flow_{left-turn,i} - 0.0102 Flow_{left-turn,i}\right)$$

with $\hat{k} = 1.6200$ for the mean model, and

$$\hat{\sigma}_{\kappa_i}^2 = \hat{V}(\kappa_i) = \frac{\hat{\mu}_i^2}{0.3685} \text{ for the variance model.}$$

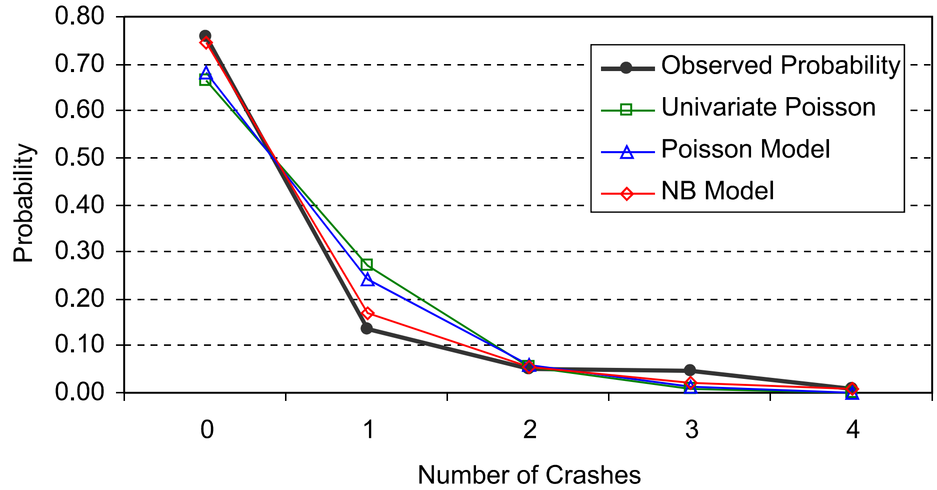


Figure 22. Mean Observed and Predicted Probabilities for Crash Pattern 6 in Off Peak

Crash Pattern 4 from A.M. Peak Through Off Peak

A total of 35 intersections were included in this reference group, and 140 observations (or pairs of two through traffic flows from two approaches perpendicular to each other) were used for model development. The quick diagnostic check suggested no overdispersion in the crash counts in this reference group.

Figure 23 provides hints about the proper functional forms for the SPF.

Using the final estimates of the Poisson and NB models, the four tests for overdispersion were performed. The tests suggested overdispersion; thus, the NB model was selected for the reference group of crash pattern 4 from A.M. peak through off peak. The results of the tests conflict with the result of the quick diagnostic check, but this is not a serious concern because the result of the quick diagnostic check is preliminary in its nature for assessing overdispersion.

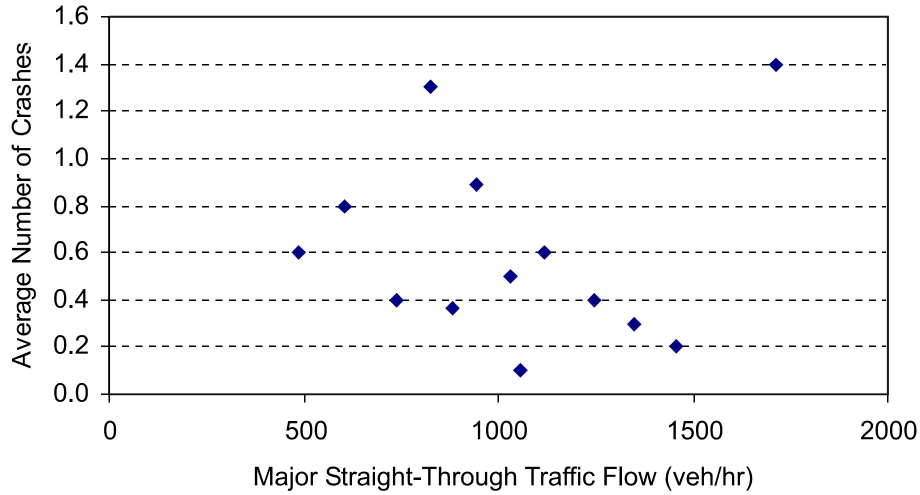
For the model of the goodness of fit, three pseudo R-squared were computed as R_L^2=0.104, R_C^2=0.111, and R_k^2=0.444, and the dispersion parameter-based R-square (R_k^2) was very high relative to the other two measures. Figure 24 provides a visual comparison. As expected, the NB model outperformed the other models in terms of mean predicted probability.

The final NB mean model (or SPF) and the final variance model were written as follows:

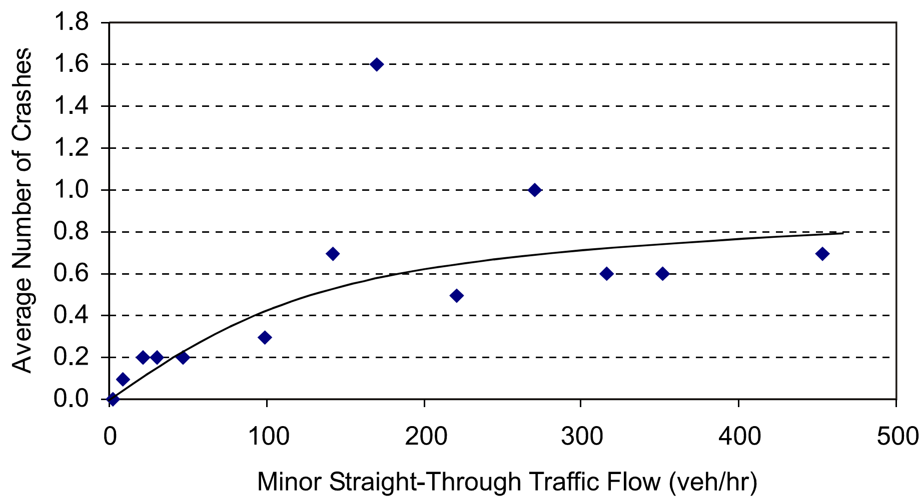
$$\hat{\mu}_i = \hat{E}(\kappa_i) = \exp\left(-3.4740 + 0.5780 \log Flow_{minor,i}\right)$$

with $\hat{k} = 0.8072$ for the mean model, and

$$\hat{\sigma}^2_{\kappa_i} = \hat{V}(\kappa_i) = \frac{\hat{\mu}_i^2}{2.4022} \text{ for the variance model}$$



(a) Major straight-through traffic flow from an approach



(b) Minor straight-through traffic flow from a perpendicular approach

Figure 23. Average aggregated crash counts for crash pattern 4 during hours from A.M. peak until off peak. As seen from plot (a), there seems to be no noticeable relationship between crash counts and major traffic flow. Plot (b) suggests $\mu = \alpha \cdot Flow^{\beta}$ for minor straight-through flow.

where i is an index of a pair of two through traffic flows from two perpendicular approaches and $Flow_{minor,i}$ is a smaller straight-through traffic flow (veh/hr). $Flow_{minor,i}$ is an average straight-through traffic volume over the hours from the start of the A.M. peak to the end of the off peak, which is the beginning of the night "free operation" signal phasing.

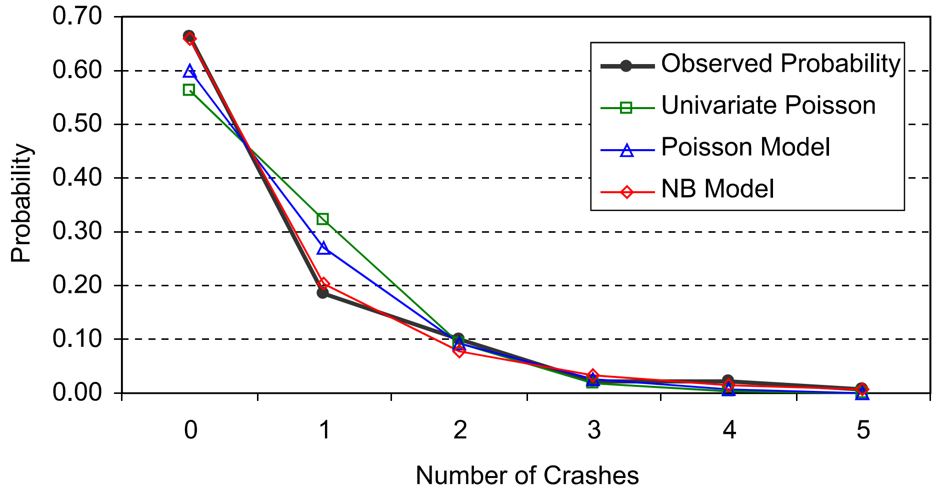


Figure 24. Mean Observed and Predicted Probabilities for Crash Pattern 4 During Period from A.M. Peak Until Off Peak

Summary of Procedure Development

For crash pattern 1 (i.e., collision between straight-through traffic flows after the intersection), the A.M. peak and evening off peak periods were different in the type of a mean model (i.e., Poisson versus NB model), and the four times of day were quite different in the coefficients of traffic flow variables in the mean models. The Poisson model was chosen for the A.M. peak and evening off peak periods, whereas the NB model was chosen for the mid day and P.M. peak periods. A total traffic flow summing straight-through, left-turning, and right-turning traffic volumes was included in the final mean models for all four periods. However, the estimates of their coefficients varied from 0.08 to 0.73 depending on the time of day.

For crash pattern 6 (i.e., collision between straight-through traffic flow and opposing left-turn traffic flow in the intersection), the A.M. peak period was different in the type of a mean model (i.e., Poisson versus NB models), and the four times of day had different traffic flow and/or left-turn signal phase variables in the mean models. The Poisson model was chosen for the A.M. peak period, whereas the NB model was chosen for the other three times of day. Straight-through and left-turn traffic flows were included in the final mean models for the A.M. peak and mid day periods. Only straight-through traffic flow was included in the final mean model for the P.M. peak period, whereas only left-turn traffic flow was included for the evening off peak period.

The permissive-plus-protected left-turn signal type was found to be statistically different from other signal types for the A.M. peak and mid day periods. However, different left-turn signal types were found to be statistically indistinguishable in terms of their influences on traffic safety for the P.M. peak and evening off peak periods. This suggests a more frequent crash occurrence of crash pattern 6 during the A.M. peak and mid day periods for the intersections with the permissive-plus-protected left-turn phase than for those with the protected left-turn phase. However, the crash frequency of crash pattern 6 during the P.M. peak and off peak periods was expected to remain the same across different left-turn signal types if traffic conditions are the same.

For crash pattern 4 (i.e., collision between straight-through traffic flows perpendicular to each other in the intersection), only the minor traffic flow variable was found to be statistically influential on crash occurrence. Thus, it was included in the final NB mean model.

CONCLUSIONS

- *The EB procedure developed in this study can be used by traffic engineers to evaluate the safety of a four-legged signalized intersection.* Traffic engineers can follow the procedure using field data and will obtain the expected crash frequency and its variance for different crash patterns and different times of day. By using fundamental statistical methods such as a confidence interval or a hypothesis test, traffic engineers can determine whether the intersection of interest is associated with an abnormally high crash risk.

- *Additional data do not need to be collected in order to apply the EB procedure.* Because all the data required for applying the EB procedure should be obtainable from VDOT's crash database and from Synchro input data that are already available to traffic engineers for traffic signal phase plans, the EB procedure is cost-effective and readily applicable.

- *The EB procedure is valid for use with only four-legged signalized intersections in VDOT's NOVA District within the valid input ranges.* The data used to develop the estimated mean and variance models in Table 6 in the full report were collected from four-legged signalized intersections in the district. If the intersection geometry, traffic patterns, and driver behavior were similar to those in VDOT's NOVA District, the EB procedure might be usable for other areas. However, the results for such areas might not be valid; a proper validation process using local data would be necessary to confirm the results.

- *The EB procedure may not be very useful for some of the nine crash population reference groups.* Traffic crashes for some crash population reference groups, such as reference group 9 (i.e., crash pattern 6 during the evening off peak period), were rare during the 4-year data period. The expected crash frequency for such reference groups would be less than 1 crash per 4 years over the entire range of input values. Thus, even 1 crash in 4 years is likely to lead to a conclusion that an intersection is associated with an abnormally high crash risk (e.g., reference groups 6 and 9, corresponding to crash pattern 6 in mid day and off peak periods, respectively, as shown in the EB case study in the full report and Appendix D).

- *An EB Spreadsheet, which aids in the application of the EB procedure, and a users' guide were developed.* For easier application of the EB procedure, an EB spreadsheet was developed using Microsoft Excel, and a users' guide was prepared. They are available from the author upon request.

RECOMMENDATIONS

1. *VDOT's Information Technology Division (IT Division), VTRC, and VDOT's NOVA District should facilitate the application of the developed EB procedure for the NOVA District.* Although the EB procedure is not difficult for traffic engineers to follow, it can be cumbersome and time-consuming for them to apply to the many intersections that would need to be evaluated for traffic safety. Thus, the IT Division, VTRC, and the NOVA District should collaborate to automate the application of the EB procedure to assess the safety of four-legged signalized intersections in the NOVA District.

    - *The IT Division should integrate data for calibration of the EB procedure and automate the application of the EB procedure.* The IT Division should extract the necessary data (i.e., traffic volumes, left-turn signal types, times of data, traffic crash characteristics and counts, and vehicle information) from Synchro files, time-based coordinate event sheets, and VDOT's crash database and integrate them into datasets in a format suitable for calibrating the EB procedure. After the calibration is done by VTRC, the IT Division should automate the application of the calibrated EB procedure for the NOVA District.

    - *VTRC should calibrate the EB procedure using the datasets that the IT Division integrates.* Using the datasets that the IT Division integrates, VTRC should calibrate all the model parameters embedded in the EB procedure. In addition, it should develop new models if necessary to enhance the reliability and accuracy of the EB procedure.

    - *The NOVA District should provide assistance to the IT Division and VTRC.* In the process of data integration and/or procedure calibration, practical insights and local information will more than likely be needed from the NOVA District.

2. *VTRC and VDOT's NOVA District should update the EB procedure when traffic characteristics of the four-legged signalized intersections change.* The EB procedure is based on the data believed to represent the prevailing traffic characteristics of the four-legged signalized intersections in the NOVA District during the years from 2001 through 2004. Intersection geometry, traffic patterns, and driver behaviors continue to change over time; as a consequence, the traffic characteristics influencing crash occurrence change. Thus, when the traffic characteristics of these intersections become significantly different from those used in this study, the EB procedure should be updated using newly collected data representing contemporary prevailing traffic characteristics. There are no established criteria for determining when the results should be updated. Engineers' judgment will certainly play a major role in such a determination.

COSTS AND BENEFITS ASSESSMENT

This study provided an explicit procedure whereby traffic engineers in VDOT's NOVA District can quickly evaluate the safety of four-legged signalized intersections. Such an intersection carrying a crash risk higher than normally expected can be identified by following the EB procedure with input data.

Using the EB procedure, traffic engineers can identify not only which intersections carry a high risk but also what traffic movements at the intersection and which time of day carry a high crash risk for the intersection. Thus, they can focus only on the identified movements and time of day to improve the safety of the intersection. In addition, when a site visit to the identified high-risk intersection is needed, the most appropriate time for the visit (e.g., A.M. peak or P.M. peak) can be identified using the results from the EB procedure.

Moreover, the EB procedure does not require additional data collection efforts as long as Synchro input data are available, which is common for signalized intersections in Virginia. Use of the EB procedure is likely to save a considerable amount of time and cost involved with field data collection whenever VDOT's NOVA District conducts a safety evaluation of its four-legged signalized intersections.

If the entire procedure from data preparation to application of the EB procedure were automated, traffic engineers could instantly assess the safety of the four-legged signalized intersections at any time just by choosing intersections of interest without manually entering the input. Moreover, calibrating the models and updating the results should be much more efficient and much less time-consuming. When the automated process is established, development and application of the EB procedure can be readily achieved by other VDOT districts as long as the appropriate data are provided.

ACKNOWLEDGMENTS

The author thanks Robert Jastrzebski and Hari Sripathi (NOVA District) for providing Synchro files and helping in selecting sample intersections for the study; Wayne Ferguson (VTRC) for his general guidance over the entire process of the project; John Miller (VTRC) for his efforts for initiating the project and insightful comments; John Jenkins III (NOVA District) for providing help with the time event sheets and helpful information; Robert Rasmussen (VDOT Central Office) for his help and information about VDOT's crash database; John Miller, Cheryl Lynn (VTRC), Matt Grimes (VTRC), Roger Howe (VTRC), and Hari Sripathi (NOVA District) for their reviews and helpful comments on the report; Roger Howe and Linda Evans (VTRC) for their excellent editorial assistance; Mike Griffith and Joe Bared (FHWA) for their useful information regarding past EB projects; Martin Parker (Wade Trim/Associates, Inc.) for providing tables of expected accident counts for Virginia; Ali Hadayeghi (Synectics Inc.) for sharing the EB spreadsheet developed for theOntario Ministry of Transportation in Canada; Meagan Powers (DKS Associates) for sharing the EB tutorial spreadsheet developed for the Montana DOT; the Transportation Safety Research Advisory Committee for supporting the project; and VDOT and FHWA for sponsoring this project.

REFERENCES

Ardekani, S., Hauer, E., and Namei, B. Traffic Impact Models. In *Traffic Flow Theory: A State-of-the-Art Report: Revised Monograph on Traffic Flow Theory,* Gartner, N., Messer, C.J., and Rathi, A.K., Eds. Transportation Research Board, Washington, DC, April 2002. www.tfhrc.gov/its/tft/tft.htm. Accessed October 2006.

Bauer, K.M., and Harwood, D.W. *Statistical Models of At-Grade Intersection Accidents: Addendum*. FHWA-RD-99-094. Federal Highway Administration, Washington, DC, 2000.

Belanger C. Estimation of Safety of Four-Legged Unsignalized Intersections. In *Transportation Research Record 1467.* Transportation Research Board, Washington, DC, 1994, pp. 23-29.

Cameron, A.C., and Trivedi, P.K. Regression-Based Tests for Overdispersion in the Poisson Model. *Journal of Econometrics*, Vol. 46, 1990, pp. 347-364.

Cameron, A.C., and Trivedi, P.K. *Regression Analysis of Count Data*. Cambridge University Press, Cambridge, MA, 1998.

Cameron, A.C., and Windmeijer A.G. An R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models. *Journal of Econometrics*, Vol. 77, No. 2, 1997, pp. 329-342.

Ceder, A., and Livneh, M. Further Evaluation of the Relationships Between Road Accidents and Average Daily Traffic. *Accident Analysis and Prevention*, Vol. 10, 1978, pp. 95-109.

Ceder, A., and Livneh, M. Relationships Between Road Accidents and Hourly Traffic Flow: I: Analyses and Interpretation. *Accident Analysis and Prevention*, Vol. 14, No. 1, 1982a, pp. 19-34.

Ceder, A., and Livneh, M. Relationships Between Road Accidents and Hourly Traffic Flow: II: Probabilistic Approach. *Accident Analysis and Prevention*, Vol. 14, No. 1, 1982b, pp. 35-44.

Demaris, A. *Regression with Social Data: Modeling Continuous and Limited Response Variables*. John Wiley & Sons, Inc., Hoboken, NJ, 2004.

Forbes, G., and Belluz, L. *Synthesis of Safety for Traffic Operations*. Final Report. Intus Road Safety Engineering Inc., Transport Canada, Ottawa, Ontario, March 2003.

Hadayeghi, A., Malone, B., and DeGannes, R. Development of New Crash Experience Warrants for Traffic Signals in Ontario, Canada. In *Transportation Research Record 1953.* Transportation Research Board, Washington, DC, 2006, pp. 120-127.

Hakkert, A.S., and Mahalel, D. Estimating the Number of Accidents at Intersections from a Knowledge of the Traffic Flows on the Approaches. *Accident Analysis and Prevention*, Vol. 10, 1978, pp. 69-79.

Harwood, D.W., Council, F.M., Hauer, E., Hughes, W.E., and Vogt, A. *Prediction of the Expected Safety Performance of Rural Two-lane Highways*. FHWA-RD-99-207. Federal Highway Administration, Washington DC, 2000.

Hauer, E. *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon Press, Tarrytown, NY, 1997.

Hauer, E., Ng, J.C.N., and Lovell, J. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185.* Transportation Research Board, Washington, DC, 1988, pp. 48-61.

Heinzl, H., and Mittlbock, M. Pseudo R-squared Measures for Poisson Regression Models with Over- or Underdispersion. *Computational Statistics & Data Analysis*, Vol. 44, 2003, pp. 253-271.

Kockelman, K.M., and Kweon, Y.-J. Driver Injury Severity and Vehicle Type: An Application of Ordered Probit Models. *Accident Analysis and Prevention*, Vol. 34, No. 3, 2002, pp. 313-321.

Kweon, Y.-J., and Kockelman, K.M. Effects of Speed Limit Changes Using Speed, Road Use, and Geometries. In *Transportation Research Record 1908*. Transportation Research Board, Washington, DC, 2005, pp. 148-158.

Lau, M.Y.-K. and May, A.D. *Accident Prediction Model Development for Unsignalized Intersections: Final Report*. Institute of Transportation Studies, University of California at Berkeley, 1989.

Lin, D.Y., Wei, L.J., and Ying, Z. Model-Checking Techniques Based on Cumulative Residuals. *Biometrics,* Vol. 58, 2002, pp. 1-12.

Long, J.S. *Regression Models for Categorical and Limited Dependent Variables*. SAGE Publication, Thousand Oaks, CA, 1997.

Lyon, C., Haq, A., Persaud, B., and Kodama, S.T. Safety Performance Functions for Signalized Intersections in Large Urban Areas: Development and Application to Evaluation of Left-Turn Priority Treatment. In *Transportation Research Record 1908.* Transportation Research Board, Washington, DC, 2005, pp. 165-171.

Matson T.M., Smith, W.S., and Hurd, F.W. *Traffic Engineering*. McGraw-Hill, New York, 1955.

McCullagh, P., and Nelder J.A. *Generalized Linear Models*, Second Edition. Chapman & Hall, New York, 1991.

McGee, H., Taori, S., and Persaud, B. *Crash Experience Warrant for Traffic Signals*. NCHRP Report 491. Transportation Research Board, Washington, DC, 2003.

Miaou, S.-P. *Measuring the Goodness-of-Fit of Accident Prediction Models*. FHWA-RD-96-040. Federal Highway Administration, Washington, DC, 1996.

Mountain L. and Fawaz, B. Estimating Accidents at Junctions Using Routinely-Available Input Data. *Traffic Engineering and Control*, Vol. 37, No. 11, 1996, pp. 624-628.

National Highway Traffic Safety Administration. *Traffic Safety Facts 2003*. DOT-HS-809-775. Washington, DC, 2005. www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2003F.pdf. Accessed October 5, 2005.

Persaud, B., and Nguyen, T. Disaggregate Safety Performance Models for Signalized Intersections on Ontario Provincial Roads. In *Transportation Research Record 1635.* Transportation Research Board, Washington, DC, 1998, pp. 113-120.

Poch, M., and Mannering, F. Negative Binomial Analysis of Intersection-Accident Frequencies. *Journal of Transportation Engineering*, Vol., No. 2, 1996, pp. 105-113.

Powers, M., and Carson, J. *Before-After Crash Analysis: A Primer for Using the Empirical Bayes Method Tutorial*. FHWA/MT-04-002-8117-21. Montana Department of Transportation, Helena, MT, 2004.

Regional Municipality of Durham, Public Works Department. *Safety Improvement Program and Software*. Ontario, Canada, 2001.

Regional Municipality of Halton, Planning and Public Works Department. *Safety Improvement Program and Software*. Ontario, Canada, April 2001.

Sayed, T., and Rodriguez, F. Accident Prediction Models for Urban Unsignalized Intersections in British Columbia. In *Transportation Research Record 1665.* Transportation Research Board, Washington, DC, 1999, pp. 93-99.

Schaechterle, K., Kurzak, H., Pfund, K., and Mensebach, W. Risk of Accidents and Capacity at At-grade Road Junctions. *Theme 2, 10th International Study Week in Traffic & Safety Engineering*, Rotterdam, The Netherlands, 1970.

Slymen, D.J., Ayala, G.X., Arredondo, E.M., and Elder, J.P. A Demonstration of Modeling Count Data with an Application to Physical Activity. *Epidemiologic Perspectives & Innovations,* Vol. 3, No. 3, 2006. www.epi-perspectives.com/content/3/1/3. Accessed May 10, 2006.

Virginia Department of Motor Vehicles. *Virginia 2005: Traffic Crash Facts*. Richmond, 2006. www.dmvnow.com/webdoc/pdf/vacrashfacts_05.pdf. Accessed November 15, 2006.

Virginia Department of Transportation, Traffic Engineering Division. *Expected Values for Accident Analysis at Intersections.* Richmond, 1991.

Vogt, A. *Crash Models for Rural Intersections: Four-lane by Two-lane Stop Controlled and Two-lane by Two-lane Signalized*. FHWA-RD-99-128. Federal Highway Administration, Office of Safety Research and Development, Washington, DC, 1999.

Vogt, A., and Bared, J. Accident Models for Two-lane Rural Segments and Intersections. In *Transportation Research Record 1635.* Transportation Research Board, Washington, DC, 1999, pp. 18-29.

Washington, S., Persaud, B., Lyon, C., and Oh, J. *Validation of Accident Models for Intersections*. FHWA-RD-03-037. Federal Highway Administration, Washington, DC, 2005.

APPENDIX A
INTERSECTIONS SELECTED FOR STUDY

Table A-1. The 49 Intersections Initially Selected for the Study

No.	Intersection	Note
1	Leesburg Pike (Rt. 7) @ Shreve Rd./Haycock Rd. (Rt. 703)	
2	Jefferson Davis Hwy. (Rt. 1) @ Optiz Blvd./Reddy Dr. (Rt. 2000)	
3	Church St. (Rt. 625) @ Sterling Blvd. (Rt. 846)	
4	Whiele Ave. (Rt. 828) @ Barron Cameron Ave. (Rt. 606)	
5	Little River Tnpk. (Rt. 236) @ Wakefield Chapel Rd./Shelley Ln. (Rt. 710)	
6	Chain Bridge Rd. (Rt. 123) @ Jermantown Rd./Blake Ln. (Rt. 655)	
7	Gallows Rd. (Rt. 650) @ Prosperity Ave. (Rt. 6066) /Park Tower Dr.	
8	Braddock Rd. (Rt. 620) @ Twinbrook Rd. (Rt. 652)	
9	Burke Centre Pkwy./Lee Chapel Rd. (Rt. 643) @ Burke Lake Rd. (Rt. 645)	X
10	Jefferson Davis Hwy. (Rt. 1) @ Cardinal Dr./Neabsco Rd. (Rt. 610)	
11	Lawyers Rd./Reston Pkwy. (Rt. 602) @ West Ox Rd. (Rt. 608)/Folkstone Dr. (Rt. 5640)	
12	Lee Hwy. (Rt. 29) @ Nutley St. (Rt. 243)	
13	Lee Hwy. (Rt. 29) @ Cedar Ln. (Rt. 698)	
14	Burke Lake Rd. (Rt. 645) @ Private Dr./Shiplett Blvd. (Rt. 5236)	
15	Fair Lakes Pkwy. (Rt. 7700) @ Fair Lakes Cir. (Rt. 7701)	
16	Waxpool Rd./Smith Switch Rd./Loudoun Co. Pkwy. (Rt. 607) @ Farmwell Rd. (640)	
17	Richmond Hwy. (Rt. 1) @ Southgate Dr. (Rt. 1779)	X
18	Old Bridge Rd. (641) @ Rolling Brook Dr. (1389)	
19	Old Keene Mill Rd. @ Greeley Blvd. /Bauer Dr.	
20	Arlington Blvd. (Rt. 50) @ Nutley St. (Rt. 10272)	
21	Leesburg Pike (Rt. 7) @ Towlston Rd. (Rt. 5020 & 676)	
22	Fairfax County Pkwy. (Rt. 7100) @ Rugby Rd.	X
23	Lee Jackson Mem. Hwy. (Rt. 50) @ Pleasant Valley Rd. (Rt. 609)/Dulles South Ct.	X
24	Centreville Rd. (Rt. 28) @ Green Trails Blvd. (Rt. 8024)/Old Mill Rd.	
25	Jefferson Davis Hwy. (Rt. 1) @ Joplin Dr./Fuller Rd. (Rt. 619)	X
26	Lee Hwy. (Rt. 29) @ James Madison Hwy. (Rt. 15)	X
27	Lee Jackson Hwy. (Rt. 50) @ Centreville Rd./Walney Rd. (Rt. 657)	
28	Leesburg Pike (Rt. 7) @ Patrick Henry Dr. (Rt. 2327)	
29	Old Keene Mill Rd. (Rt. 644) @ Lee Chapel Rd. (Rt. 643)	
30	Hayfield Rd. (Rt. 635) @ Kingstowne Village Pkwy. (Rt. 8690)	
31	Braddock Rd. (Rt. 620) @ Clifton Rd. (Rt. 645)	
32	Sideburn Rd. (Rt. 653) @ Zion Dr./Rd. (Rt. 654)	X
33	Hooes Rd. (Rt. 636) @ Silverbrook Rd. (Rt. 600)	X
34	Lawyers Rd. (Rt. 673) @ Hunter Mill Rd. (Rt. 674)	X
35	Spring Hill Rd. (Rt. 684) @ Old Dominion Dr. (Rt. 738)	X
36	Great Falls St. (Rt. 694) @ Idylwood Rd./Kirby Rd. (Rt. 695)	X
37	Arlington Blvd. (Rt. 50) @ Prosperity Ave. (Rt. 699)	
38	Lee Hwy. (Rt. 29) @ Prosperity Ave. (Rt. 699)	
39	Braddock Rd. (Rt. 620) @ Guinea Rd. (Rt. 651)	
40	Guinea Rd. (Rt. 651) @ Braeburn Rd. (Rt. 2430)	X
41	Harry Flood Byrd Hwy. (Rt. 7) @ Potomac View Rd. (Rt. 637)	
42	Charles Town Pike (Rt. 9) @ Berlin Turnpike (Rt. 2887)	X
43	Minnieville Rd. (Rt. 640) @ Dale Blvd. (Rt. 784)	
44	Jefferson Davis Hwy. (Rt. 1) @ Neabsco Mills Rd./Blackburn Rd. (Rt. 638)	
45	Lee Hwy. (Rt. 29) @ Stringfellow Rd./Clifton Rd.(Rt. 645)	
46	Lee Hwy. (Rt. 29) @ Gallows Rd. (Rt. 650)	
47	Hoadly Rd. (Rt. 642) @ Purcell Rd. (Rt. 643)/Dale Blvd. (Rt. 784)	X
48	Minnieville Rd. (Rt. 640) @Smoketown Rd. (Rt. 2000)	
49	Prince William Pkwy. (Rt. 3000) @ Minnieville Rd. (Rt. 640)	

Note: "X" indicates that the intersection was excluded from analysis for all crash population reference groups.

APPENDIX B
SYNCHRO DATA REPORT

Lanes, Volumes, Timings
1: Leesburg Pike & Haycock Road 11/15/2006

Lane Group	SEL	SET	SER	NWL	NWT	NWR	NEL	NET	NER	SWL	SWT	SWR
Lane Configurations												
Ideal Flow (vphpl)	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900	1900
Storage Length (ft)	235		1000	235		0	150		0	280		0
Storage Lanes	1		1	1		0	1		0	1		1
Total Lost Time (s)	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Leading Detector (ft)	35	5	5	35	5		35	35		35	35	35
Trailing Detector (ft)	-5	0	0	-5	0		-5	-5		-5	-5	-5
Turning Speed (mph)	15		9	15		9	15		9	15		9
Lane Util. Factor	1.00	0.95	1.00	1.00	0.95	0.95	1.00	0.95	0.95	1.00	1.00	1.00
Frt			0.850		0.989			0.985				0.850
Flt Protected	0.950			0.950			0.950			0.950		
Satd. Flow (prot)	1770	3539	1583	1770	3500	0	1770	3486	0	1770	1863	1583
Flt Permitted	0.052			0.208			0.339			0.197		
Satd. Flow (perm)	97	3539	1583	387	3500	0	631	3486	0	367	1863	1583
Right Turn on Red			Yes			No			No			Yes
Satd. Flow (RTOR)			32									260
Headway Factor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Link Speed (mph)		35			35			35			35	
Link Distance (ft)		1919			2160			1296			1096	
Travel Time (s)		37.4			42.1			25.2			21.4	
Volume (vph)	302	1160	32	68	1160	92	232	480	52	140	160	260
Peak Hour Factor	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Adj. Flow (vph)	302	1160	32	68	1160	92	232	480	52	140	160	260
Lane Group Flow (vph)	302	1160	32	68	1252	0	232	532	0	140	160	260
Turn Type	pm+pt		Perm	pm+pt			pm+pt			pm+pt		Perm
Protected Phases	5	2		1	6		3	8		7	4	
Permitted Phases	2		2	6			8			4		4
Detector Phases	5	2	2	1	6		3	8		7	4	4
Minimum Initial (s)	5.0	15.0	15.0	5.0	15.0		5.0	5.0		5.0	5.0	5.0
Minimum Split (s)	10.0	36.0	36.0	10.0	36.0		10.0	36.0		10.0	36.0	36.0
Total Split (s)	45.0	122.0	122.0	15.0	92.0	0.0	37.0	45.0	0.0	28.0	36.0	36.0
Total Split (%)	21.4%	58.1%	58.1%	7.1%	43.8%	0.0%	17.6%	21.4%	0.0%	13.3%	17.1%	17.1%
Maximum Green (s)	40.0	116.0	116.0	10.0	86.0		32.0	40.0		23.0	31.0	31.0
Yellow Time (s)	4.0	4.0	4.0	4.0	4.0		4.0	4.0		4.0	4.0	4.0
All-Red Time (s)	1.0	2.0	2.0	1.0	2.0		1.0	1.0		1.0	1.0	1.0
Lead/Lag	Lead	Lag	Lag	Lead	Lag		Lead	Lag		Lead	Lag	Lag
Lead-Lag Optimize?												
Vehicle Extension (s)	2.0	2.0	2.0	2.0	2.0		2.0	2.0		2.0	2.0	2.0
Recall Mode	None	C-Max	C-Max	None	C-Max		None	None		None	None	None
Walk Time (s)		7.0	7.0		7.0			7.0			7.0	7.0
Flash Dont Walk (s)		23.0	23.0		23.0			24.0			24.0	24.0
Pedestrian Calls (#/hr)		0	0		0			0			0	0
Act Effct Green (s)	143.0	130.6	130.6	113.9	105.4		59.0	36.3		46.1	27.3	27.3
Actuated g/C Ratio	0.68	0.62	0.62	0.54	0.50		0.28	0.17		0.22	0.13	0.13
v/c Ratio	0.91	0.53	0.03	0.26	0.71		0.71	0.88		0.68	0.66	0.60
Control Delay	76.4	24.6	5.2	17.6	45.9		70.4	93.2		67.1	94.5	12.4
Queue Delay	0.0	0.0	0.0	0.0	0.0		0.0	0.0		0.0	0.0	0.0
Total Delay	76.4	24.6	5.2	17.6	45.9		70.4	93.2		67.1	94.5	12.4
LOS	E	C	A	B	D		E	F		E	F	B

Rt 7B 8:00 am 5/17/2002 AM Retimed Synchro 6 Report
VDOT Page 1

Figure B-1. Example of Synchro report of intersection of Leesburg Pike at Shreve Rd./Haycock Rd.

APPENDIX C
TIME BASE COORDINATION EVENTS SHEET

Bank 1

< C + 0 + 9 = 0.1 >

Row	Time	Plan	Day of Week						
			S	M	T	W	T	F	S
0	06:00	1	0	1	1	1	1	1	0
1	09:30	2	0	1	1	1	1	1	0
2	15:00	3	0	1	1	1	1	1	0
3	19:30	4	0	1	1	1	1	1	0
4	22:00	62	1	1	1	1	1	1	1
5	09:00	5	1	0	0	0	0	0	1
6	00:00	62	1	1	1	1	1	1	1
7	00:00	0	0	0	0	0	0	0	0
8	00:00	0	0	0	0	0	0	0	0
9	00:00	0	0	0	0	0	0	0	0
A	00:00	0	0	0	0	0	0	0	0
B	00:00	0	0	0	0	0	0	0	0
C	00:00	0	0	0	0	0	0	0	0
D	00:00	0	0	0	0	0	0	0	0
E	00:00	0	0	0	0	0	0	0	0

Bank 2

< C + 0 + 9 = 0.2 >

Row	Time	Plan	Day of Week						
			S	M	T	W	T	F	S
0	00:00	0	0	0	0	0	0	0	0
1	00:00	0	0	0	0	0	0	0	0
2	00:00	0	0	0	0	0	0	0	0
3	00:00	0	0	0	0	0	0	0	0
4	00:00	0	0	0	0	0	0	0	0
5	00:00	0	0	0	0	0	0	0	0
6	00:00	0	0	0	0	0	0	0	0
7	00:00	0	0	0	0	0	0	0	0
8	00:00	0	0	0	0	0	0	0	0
9	00:00	0	0	0	0	0	0	0	0
A	00:00	0	0	0	0	0	0	0	0
B	00:00	0	0	0	0	0	0	0	0
C	00:00	0	0	0	0	0	0	0	0
D	00:00	0	0	0	0	0	0	0	0
E	00:00	0	0	0	0	0	0	0	0

Figure C-1. Example of signal timing plan of intersection of Leesburg Pike at Shreve Rd./Haycock Rd. (Screen from MIST Operator Interface).

APPENDIX D
EB CASE APPLICATION

Table D-1. Results of Empirical Bayes Application to Crash Pattern 6 in A.M. Peak, Mid Day, P.M. Peak, and Evening Off Peak

Note: "Unsafe" indicates being unsafe at the 95% confidence level.

Intersection No.	Bound		A.M. Peak			Mid Day			P.M. Peak			Off Peak		
	Through	Left-turn	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe
1	E	W	0	1.01E-02	No	1	9.73E-05	Yes	**1**	5.61E-01	**No**	0	7.16E-03	No
1	W	E	0	1.25E-02	No	1	8.14E-05	Yes	**1**	5.52E-01	**No**	0	2.11E-03	No
1	N	S	0	1.96E-02	No	0	1.31E-04	No	0	1.58E-01	No	0	7.92E-03	No
1	S	N	0	2.17E-02	No	0	7.96E-05	No	0	1.56E-01	No	0	6.55E-03	No
2	E	W	0	8.01E-03	No	0	2.12E-04	No	0	1.62E-01	No	0	7.92E-03	No
2	W	E	0	1.13E-02	No	0	2.06E-05	No	0	1.65E-01	No	0	1.00E-03	No
2	N	S	1	5.98E-03	Yes	1	4.86E-04	Yes	**1**	5.20E-01	**No**	0	7.92E-03	No
2	S	N	1	2.31E-02	Yes	2	3.83E-04	Yes	2	8.64E-01	Yes	0	3.32E-03	No
3	E	W	1	8.36E-03	Yes	0	6.18E-04	No	0	1.55E-01	No	1	2.80E-02	Yes
3	W	E	0	8.29E-03	No	0	6.67E-04	No	0	1.57E-01	No	0	5.37E-03	No
3	N	S	0	1.41E-02	No	0	7.95E-05	No	0	1.88E-01	No	0	4.89E-03	No
3	S	N	0	1.67E-02	No	0	1.40E-04	No	0	1.84E-01	No	0	7.86E-03	No
4	E	W	0	1.65E-02	No	1	5.13E-04	Yes	**1**	5.05E-01	**No**	1	2.62E-02	Yes
4	W	E	0	1.45E-02	No	0	1.56E-04	No	0	1.81E-01	No	0	7.86E-03	No
4	N	S	1	2.30E-02	Yes	0	1.45E-04	No	0	1.69E-01	No	0	7.87E-03	No
4	S	N	1	1.71E-02	Yes	4	1.76E-03	Yes	4	1.38E+00	Yes	0	5.63E-03	No
5	E	W	0	3.43E-03	No	0	3.42E-05	No	0	1.96E-01	No	0	4.75E-03	No
5	W	E	0	8.83E-03	No	0	3.54E-05	No	0	1.96E-01	No	0	2.08E-03	No
5	N	S	0	8.14E-03	No	0	4.94E-05	No	0	1.34E-01	No	0	7.49E-03	No
5	S	N	0	8.23E-03	No	0	5.91E-06	No	0	1.18E-01	No	0	1.30E-03	No
6	E	W	0	9.10E-03	No	0	1.81E-04	No	0	1.65E-01	No	0	2.47E-03	No
6	W	E	0	1.20E-02	No	0	5.91E-05	No	0	1.62E-01	No	0	2.32E-03	No
6	N	S	0	8.69E-03	No	1	2.28E-04	Yes	**1**	5.42E-01	**No**	2	3.96E-02	Yes
6	S	N	1	1.61E-02	Yes	3	4.82E-04	Yes	3	1.25E+00	Yes	1	2.19E-02	Yes
7	E	W	0	6.89E-03	No	0	5.15E-06	No	0	1.09E-01	No	0	7.49E-03	No
7	W	E	0	3.96E-03	No	0	8.82E-06	No	0	1.24E-01	No	0	6.92E-03	No
7	N	S	0	7.25E-03	No	0	6.83E-05	No	0	1.95E-01	No	0	3.26E-03	No
7	S	N	2	1.51E-02	Yes	3	7.40E-04	Yes	3	1.24E+00	Yes	0	2.08E-03	No

Table D-1. Results of Empirical Bayes Application to Crash Pattern 6 in A.M. Peak, Mid Day, P.M. Peak, and Evening Off Peak (Continued)

Intersection No.	Bound		A.M. Peak			mid day			P.M. Peak			Off Peak		
	Through	Left-turn	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe
8	E	W	1	7.96E-03	Yes	0	3.95E-05	No	0	1.94E-01	No	0	2.32E-03	No
8	W	E	2	7.71E-03	Yes	2	4.36E-04	Yes	2	8.92E-01	Yes	0	5.50E-03	No
8	N	S	0	8.56E-03	No	0	9.08E-05	No	0	1.14E-01	No	0	6.42E-03	No
8	S	N	0	1.86E-02	No	0	3.37E-05	No	0	1.28E-01	No	0	4.99E-03	No
10	E	W	0	8.52E-03	No	0	5.24E-05	No	0	1.31E-01	No	0	7.92E-03	No
10	W	E	0	4.95E-03	No	0	8.38E-06	No	0	1.46E-01	No	0	2.47E-03	No
10	N	S	0	8.36E-03	No	0	1.19E-04	No	0	1.87E-01	No	1	2.96E-02	Yes
10	S	N	1	1.71E-02	Yes	0	1.24E-04	No	0	1.87E-01	No	0	7.66E-03	No
11	E	W	0	6.63E-03	No	0	1.03E-04	No	0	1.23E-01	No	0	3.26E-03	No
11	W	E	1	8.44E-03	Yes	0	2.58E-04	No	0	1.34E-01	No	0	6.92E-03	No
11	N	S	0	7.04E-03	No	0	1.19E-04	No	0	1.81E-01	No	0	4.16E-03	No
11	S	N	2	1.67E-02	Yes	1	5.76E-04	Yes	**1**	5.14E-01	**No**	1	2.74E-02	Yes
12	E	W	0	7.73E-03	No	0	9.33E-05	No	0	1.82E-01	No	0	4.16E-03	No
12	W	E	0	2.33E-02	No	0	3.16E-05	No	0	1.82E-01	No	0	3.04E-03	No
12	N	S	0	0.01216	No	1	2.77E-05	Yes	**1**	4.90E-01	**No**	0	2.58E-04	No
12	S	N	2	8.69E-03	Yes	3	2.27E-03	Yes	3	1.12E+00	Yes	3	6.65E-02	Yes
13	E	W	0	5.38E-03	No	0	7.20E-05	No	0	1.94E-01	No	0	7.29E-03	No
13	W	E	1	1.88E-02	Yes	0	4.03E-05	No	0	1.95E-01	No	0	5.50E-03	No
13	N	S	0	9.29E-03	No	0	1.94E-04	No	0	1.59E-01	No	0	7.92E-03	No
13	S	N	0	8.47E-03	No	2	1.41E-03	Yes	2	7.23E-01	Yes	2	5.05E-02	Yes
14	E	W	0	2.95E-03	No	0	7.20E-05	No	0	1.09E-01	No	0	0.00744	No
14	W	E	0	2.11E-03	No	0	2.90E-05	No	0	1.09E-01	No	1	2.09E-03	Yes
14	N	S	0	9.34E-03	No	0	4.89E-05	No	0	1.84E-01	No	0	1.00E-03	No
14	S	N	0	5.96E-03	No	0	2.25E-04	No	0	1.85E-01	No	0	NA	NA
15	E	W	1	1.77E-02	Yes	0	2.35E-05	No	0	1.97E-01	No	0	5.63E-03	No
15	W	E	0	1.43E-02	No	0	4.25E-05	No	0	1.98E-01	No	0	7.77E-03	No
15	N	S	0	1.70E-02	No	0	1.23E-04	No	0	1.61E-01	No	0	7.38E-03	No
15	S	N	0	1.48E-02	No	0	1.05E-04	No	0	1.57E-01	No	0	5.76E-03	No

Table D-1. Results of Empirical Bayes Application to Crash Pattern 6 in A.M. Peak, Mid Day, P.M. Peak, and Evening Off Peak (Continued)

Intersection No.	Bound		A.M. Peak			Mid Day			P.M. Peak			Off Peak		
	Through	Left-turn	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe
16	E	W	1	4.75E-03	Yes	0	2.52E-05	No	0	1.98E-01	No	2	3.29E-02	Yes
16	W	E	0	1.01E-02	No	0	3.44E-05	No	0	1.98E-01	No	2	4.12E-02	Yes
16	N	S	0	7.28E-03	No	0	3.55E-05	No	0	1.24E-01	No	0	6.42E-03	No
16	S	N	0	5.26E-03	No	0	2.74E-05	No	0	1.18E-01	No	0	6.77E-03	No
18	E	W	1	0.000876	Yes	1	2.89E-04	Yes	**1**	5.44E-01	**No**	0	7.98E-03	No
18	W	E	1	NA	NA	0	9.83E-06	No	0	1.96E-01	No	0	8.05E-04	No
18	N	S	0	NA	NA	0	NA	NA	0	0.111	No	0	NA	NA
18	S	N	0	1.92E-03	No	0	6.82E-06	No	0	9.28E-02	No	0	5.58E-03	No
19	E	W	0	1.69E-03	No	0	2.46E-05	No	0	1.88E-01	No	0	4.89E-03	No
19	W	E	0	8.26E-03	No	0	3.25E-05	No	0	1.87E-01	No	0	7.97E-03	No
19	N	S	0	2.88E-03	No	0	7.91E-05	No	0	1.09E-01	No	0	7.65E-03	No
19	S	N	0	2.79E-03	No	0	4.19E-05	No	0	1.09E-01	No	0	4.21E-03	No
20	E	W	0	2.29E-03	No	0	2.09E-05	No	0	1.96E-01	No	0	3.26E-03	No
20	W	E	2	1.25E-02	Yes	0	3.76E-05	No	0	1.98E-01	No	3	5.03E-02	Yes
20	N	S	0	5.55E-03	No	0	8.48E-05	No	0	1.18E-01	No	0	4.87E-03	No
20	S	N	0	3.48E-03	No	0	5.82E-05	No	0	1.09E-01	No	0	7.65E-03	No
21	E	W	**1**	0.001051	**No**	0	1.28E-05	No	0	2.05E-01	No	0	7.92E-03	No
21	W	E	0	7.99E-03	No	0	1.19E-05	No	0	2.05E-01	No	0	4.89E-03	No
21	N	S	0	6.61E-03	No	0	3.06E-04	No	0	1.36E-01	No	0	7.65E-03	No
21	S	N	0	5.13E-03	No	0	2.09E-04	No	0	1.28E-01	No	0	6.42E-03	No
24	E	W	0	6.75E-03	No	0	2.32E-04	No	0	1.34E-01	No	0	7.65E-03	No
24	W	E	0	5.78E-03	No	0	3.37E-04	No	0	1.41E-01	No	0	6.42E-03	No
24	N	S	0	2.91E-03	No	0	2.45E-05	No	0	2.02E-01	No	0	3.72E-03	No
24	S	N	2	7.62E-03	Yes	1	7.45E-05	Yes	**1**	5.69E-01	**No**	0	7.65E-03	No
27	E	W	0	2.41E-03	No	0	1.82E-07	No	0	2.06E-01	No	0	3.32E-03	No
27	W	E	0	2.01E-03	No	0	1.11E-07	No	0	2.06E-01	No	0	6.37E-04	No
27	N	S	0	9.11E-03	No	0	2.43E-05	No	0	1.85E-01	No	0	5.37E-03	No
27	S	N	1	9.80E-03	Yes	0	4.23E-05	No	0	1.78E-01	No	0	7.53E-03	No

Table D-1. Results of Empirical Bayes Application to Crash Pattern 6 in A.M. Peak, Mid Day, P.M. Peak, and Evening Off Peak (Continued)

Intersection No.	Bound		A.M. Peak			Mid Day			P.M. Peak			Off Peak		
	Through	Left-turn	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe
28	E	W	0	1.01E-02	No	0	1.33E-04	No	0	1.57E-01	No	2	2.46E-02	Yes
28	W	E	3	9.82E-03	Yes	1	7.76E-04	Yes	**1**	4.46E-01	**No**	0	7.98E-03	No
28	N	S	0	6.08E-03	No	0	2.11E-05	No	0	2.02E-01	No	1	2.94E-02	Yes
28	S	N	0	1.04E-02	No	0	3.28E-05	No	0	2.00E-01	No	0	7.92E-03	No
29	W	E	0	2.24E-02	No	0	1.72E-04	No	0	1.73E-01	No	0	7.96E-03	No
29	N	S	0	1.11E-02	No	0	4.70E-06	No	0	1.69E-01	No	0	3.68E-04	No
29	S	N	0	8.51E-03	No	1	1.72E-03	Yes	**1**	4.68E-01	**No**	1	1.55E-02	Yes
30	E	W	0	1.11E-02	No	1	1.05E-03	Yes	1	4.23E-01	Yes	3	2.87E-02	Yes
30	W	E	0	8.18E-03	No	2	3.97E-03	Yes	2	7.17E-01	Yes	0	6.00E-03	No
30	N	S	0	1.51E-02	No	0	1.43E-04	No	0	1.82E-01	No	0	7.48E-03	No
30	S	N	0	1.51E-02	No	0	1.44E-04	No	0	1.80E-01	No	0	6.42E-03	No
31	E	W	0	5.95E-03	No	0	1.23E-04	No	0	1.87E-01	No	0	7.58E-03	No
31	W	E	**1**	0.02422	**No**	2	2.64E-04	Yes	2	8.61E-01	Yes	0	2.25E-03	No
31	N	S	1	1.60E-02	Yes	2	1.07E-03	Yes	2	7.28E-01	Yes	0	7.92E-03	No
31	S	N	0	1.56E-02	No	0	1.47E-04	No	0	1.59E-01	No	1	2.88E-02	Yes
37	E	W	1	0.001644	Yes	10	1.18E-03	Yes	10	3.76E+00	Yes	3	3.50E-02	Yes
37	W	E	1	1.16E-02	Yes	4	3.70E-04	Yes	4	1.62E+00	Yes	1	2.96E-02	Yes
37	N	S	0	1.01E-02	No	0	4.15E-05	No	0	1.67E-01	No	0	9.67E-04	No
37	S	N	0	8.20E-03	No	0	1.98E-04	No	0	1.59E-01	No	0	7.96E-03	No
38	E	W	0	9.83E-03	No	1	2.80E-04	Yes	**1**	5.30E-01	**No**	1	6.08E-03	Yes
38	W	E	1	1.67E-02	Yes	0	8.22E-05	No	0	1.92E-01	No	0	6.42E-03	No
38	N	S	1	1.36E-02	Yes	1	7.20E-04	Yes	**1**	4.88E-01	**No**	0	7.92E-03	No
38	S	N	1	2.18E-02	Yes	2	4.52E-04	Yes	2	7.74E-01	Yes	1	1.13E-02	Yes
39	E	W	0	5.34E-03	No	0	8.09E-06	No	0	1.93E-01	No	0	5.79E-04	No
39	W	E	0	3.69E-03	No	0	1.66E-05	No	0	1.96E-01	No	1	2.04E-02	Yes
39	N	S	0	6.13E-03	No	0	4.04E-05	No	0	1.64E-01	No	0	5.50E-03	No
39	S	N	0	9.18E-03	No	0	4.49E-05	No	0	1.60E-01	No	0	7.96E-03	No
41	E	W	0	1.69E-03	No	0	6.40E-06	No	0	2.02E-01	No	0	7.58E-03	No
41	W	E	0	3.32E-03	No	0	3.28E-06	No	0	2.06E-01	No	0	7.96E-03	No
41	N	S	0	0.009107	No	0	7.02E-06	No	0	1.64E-01	No	1	1.19E-02	Yes
41	S	N	0	8.05E-03	No	0	4.15E-05	No	0	1.59E-01	No	0	3.26E-03	No

Table D-1. Results of Empirical Bayes Application to Crash Pattern 6 in A.M. Peak, Mid Day, P.M. Peak, and Evening Off Peak (Continued)

Intersection No.	Bound		A.M. Peak			Mid Day			P.M. Peak			Off Peak		
	Through	Left-turn	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe	Crash	E(k|K)	Unsafe
43	E	W	0	2.88E-03	No	1	6.20E-05	Yes	1	5.35E-01	No	0	3.83E-03	No
43	W	E	1	1.05E-02	Yes	0	2.20E-05	No	0	1.89E-01	No	1	1.50E-02	Yes
43	N	S	0	1.20E-02	No	0	1.40E-05	No	0	1.80E-01	No	0	2.78E-03	No
43	S	N	0	8.56E-03	No	0	1.90E-05	No	0	1.79E-01	No	0	5.37E-03	No
44	E	W	0	7.72E-03	No	0	1.46E-04	No	0	1.49E-01	No	0	7.87E-03	No
44	W	E	0	6.61E-03	No	0	4.69E-05	No	0	1.24E-01	No	0	7.92E-03	No
44	N	S	1	5.47E-03	Yes	1	4.09E-04	Yes	1	5.29E-01	No	0	7.49E-03	No
44	S	N	0	2.08E-02	No	1	1.76E-04	Yes	1	5.20E-01	No	1	1.03E-02	Yes
45	E	W	0	4.36E-03	No	1	1.84E-04	Yes	1	5.40E-01	No	1	1.46E-02	Yes
45	W	E	2	2.38E-02	Yes	1	4.02E-05	Yes	1	5.39E-01	No	0	9.67E-04	No
45	N	S	2	2.00E-02	Yes	3	1.28E-03	Yes	3	1.03E+00	Yes	2	4.29E-02	Yes
45	S	N	0	1.98E-02	No	0	1.54E-04	No	0	1.63E-01	No	4	9.18E-02	Yes
46	E	W	1	1.22E-02	Yes	0	5.44E-05	No	0	1.90E-01	No	0	7.48E-03	No
46	W	E	0	1.62E-02	No	0	6.73E-05	No	0	1.88E-01	No	0	7.04E-03	No
46	N	S	2	9.77E-03	Yes	2	3.52E-04	Yes	2	8.75E-01	Yes	3	5.39E-02	Yes
46	S	N	2	1.94E-02	Yes	0	1.14E-05	No	0	1.85E-01	No	1	1.67E-02	Yes
48	E	W	2	4.50E-03	Yes	18	0.000854	Yes	8	0.272	Yes	3	3.60E-02	Yes
48	W	E	1	9.67E-03	Yes	5	7.45E-04	Yes	5	1.80E+00	Yes	1	2.19E-02	Yes
48	N	S	0	1.07E-02	No	0	4.23E-05	No	0	1.79E-01	No	0	6.80E-03	No
48	S	N	0	8.66E-03	No	1	1.10E-04	Yes	1	5.02E-01	No	2	1.54E-02	Yes
49	E	W	0	4.26E-03	No	0	2.89E-05	No	0	1.90E-01	No	0	7.86E-03	No
49	W	E	0	1.13E-02	No	0	1.91E-05	No	0	1.81E-01	No	0	6.42E-03	No
49	N	S	0	5.75E-03	No	0	8.43E-06	No	0	1.93E-01	No	0	2.70E-03	No
49	S	N	0	5.48E-03	No	0	1.03E-06	No	0	1.94E-01	No	0	4.03E-05	No