Estimation of Safety of Four-Legged Unsignalized Intersections

CARL BÉLANGER

In this study, empirical Bayesien methods were applied to the estimation of the safety of four-legged unsignalized intersections. This application can be described as a two-step process. First, multivariate models were developed to estimate the number of accidents from various flow functions at these intersections. The best model was obtained from the product of major and minor flows, raised to a power. Attempts were made to develop models for specific patterns of collisions and to incorporate variables other than traffic flow functions to these models. The modeling results were then combined with the accident count of a four-legged unsignalized intersection to estimate its safety. Results were used to identify blackspot locations and to evaluate the effects of interventions more accurately.

In urban areas, more than half of all accidents occur at intersections, and the corresponding figure for rural environments is about one-quarter (*1*). Given the importance of safety problems at these locations and the relatively small portion of the highway network they represent, interventions aimed at improving the safety of intersections are desirable. To identify sites that have a potential for improvement, knowledge of their long-term mean number of accidents is required. This mean is defined as the "safety" and is denoted by m; its estimate is denoted by $\hat{m}$. Estimates of m are also needed to evaluate the effects of interventions adequately and to determine the success of such actions.

Several methods that have been developed to estimate m are based on the number of accidents observed at the site of interest in a relatively short period. Because of the rarity of accidents and annual variations in the accident count, these estimates are often inaccurate. Also, given that sites are generally chosen for treatment because of a recent poor accident record, $\hat{m}$ is often overestimated. In these situations, the count of accidents in the period after identification will generally revert toward its expected value even if no treatment is applied to the site. This phenomenon, which is called regression to the mean (RTM), introduces significant bias to the conclusions of safety studies (*2–4*). Using longer periods of analysis does not solve the problem because many factors that influence m change over time. Consequently, new methods of analysis were sought. In the past decade, an approach based on empirical Bayesian (EB) methods emerged as a better way of estimating safety. More recently, multivariate statistical analysis was proposed to enhance the benefits of EB methods. We were interested in using these techniques to estimate the safety of unsignalized intersections. To ensure a higher level of homogeneity, a subclass was chosen. Only four-legged intersections that are signed with two stops on minor approaches and have one lane in each direction were selected.

Ministry of Transportation of Québec, 700 René Lévesque Blvd. East, 22nd floor, Québec, Qc., Canada G1R 5H1.

DATA

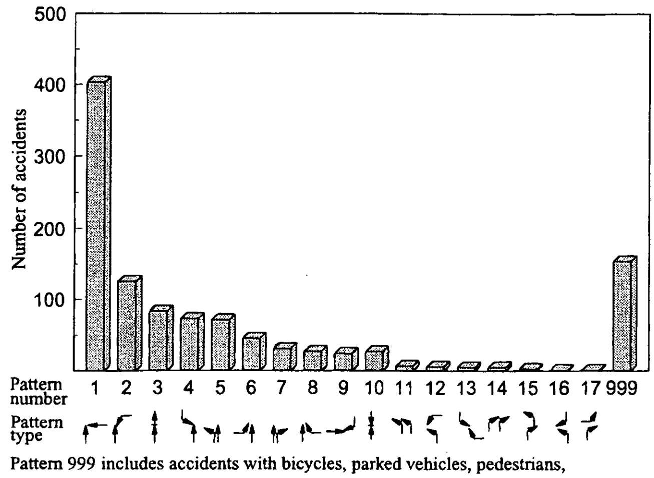
The detailed information about accidents, traffic flow, and geometric characteristics needed for this project was only available for a few sites that had been the object of a safety analysis in recent years. Because these analyses are generally motivated by requests from elected officials to improve sites that are perceived as hazardous, RTM problems are likely to be present. To improve the accuracy of the models, accident data relating to events preceding each site's identification were not used. Only the period following the demand (the number of accidents in the after period is not subject to RTM bias) was considered. With this decision, the establishment of a constant period of analysis became impossible because a sufficient number of accidents could not be gathered during any fixed period. Instead, specific period lengths were determined for each site; consequently, this analysis is based on the number of accidents per day. The sample for this project consists of 149 intersections located in eastern Quebec.

Accidents were considered pertinent if they occurred within 30 m of the intersection or were intersection related. A total of 1084 accidents fulfilled these criteria. Of these, more than 85 percent involved two vehicles. The determination of each pattern of collision and each combination of flows that caused the accident was required, but this information, as coded in the accident file, is unreliable. The list of patterns provided in the accident report form is not exhaustive, and codes are often missing or inconsistent. However, by analyzing the microfilm of each accident report, several of these problems were corrected. Eighteen patterns of accidents were identified (Figure 1). Pattern "999" is a miscellaneous category that includes single-vehicle accidents, accidents with pedestrians, bicyclists, parked vehicles, and reversing vehicles. Right-angle collisions represent 42 percent of the accidents with two or more vehicles.

For each site, a 12-hr count was available, which provided estimates of flows for each of the 12 possible maneuvers at a four-legged intersection: left turn, through, and right turn on each approach. On the basis of data collected from permanent traffic counters, these estimates were converted to daily flow estimates that are representative of an average day, month, and year of the period of analysis. The distribution of flows is shown in Figure 2. It ranges from 388 to 15,942 vpd.

METHODOLOGY

To reduce the regression-to-the-mean bias, EB estimates use not only information from the intersection analyzed but also information from a group of intersections having similar characteristics (called the reference population). The weight attributed to the reference population is a function of its homogeneity. A major diffi-



Pattern 999 includes accidents with bicycles, parked vehicles, pedestrians, vehicles reversing and single vehicle accidents.

FIGURE 1 Number of accidents, by pattern.

culty associated with the use of EB methods consists of defining a reference population that is sufficiently homogeneous to be reliable and yet large enough to improve the estimation. To alleviate this problem, the multivariate approach, as recently proposed by Hauer (*5*) was used. Regression models were first developed to estimate the moments E(m) and VAR(m) that describe the distribution of ms in an imaginary group of intersections having the same characteristics as the site under analysis. Once E(m) and VAR(m) become available, they are combined to the accident history (x) at the intersection of interest to obtain the updated estimate of safety [denoted E($m|x$)] and its variance [denoted VAR($m|x$)].

$$\mathrm{E}(m|x) = a\mathrm{E}(m) + (1 - a)x$$

$$\mathrm{VAR}(m|x) = a(1 - a)\mathrm{E}(m) + (1 - a)^2 x$$

$$\text{with } a = \frac{\mathrm{E}(m)}{\mathrm{E}(m) + \mathrm{VAR}(m)} \quad (1)$$

Thus, the major task consisted of developing multivariate models to estimate E(m) and VAR(m). In this project, modeling was undertaken in three stages: (a) development of models relating the

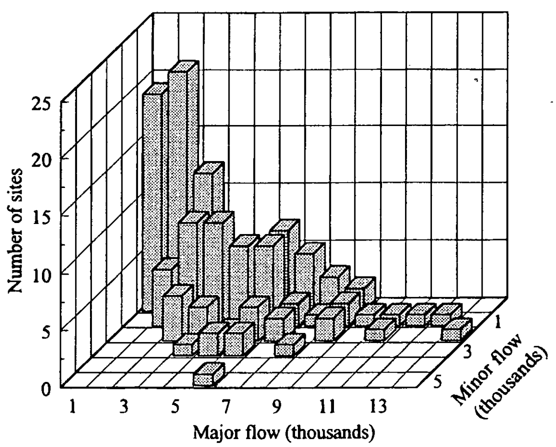


FIGURE 2 Distribution of traffic flow per site.

total number of accidents to various flow functions, (b) development of models relating accidents of a specific type to various flow functions, and (c) evaluation of the effect of variables other than traffic flow.

Most of the regression theory is based on the assumption that the error structure is normal with mean equal to 0 and a constant variance (σ^2); however, this hypothesis is not valid in road safety analysis because residuals tend to increase with larger fitted values. A number of recent studies have concluded that a negative binomial type of error is more appropriate to describe the variations in the number of accidents at several sites. This choice is based on the assumptions that the variations in the number of accidents (x) at any particular location can be described by a Poisson process and that the variations in the levels of safety (m) in a group of similar intersections can be fitted by a gamma distribution (*6*). The GLIM software (*7*) was selected to estimate the coefficients of our models using a negative binomial error structure consistent with the data.

The estimation of E(m) is straightforward because it is obtained directly from the models, but the estimation of VAR(m) from multivariate techniques is less common. It is only recently that a method has been proposed to estimate VAR(m) from the regression results, using the following empirical relationship (*8*):

$$\mathrm{VAR}(m) = \mathrm{E}(m)^2/k \quad (2)$$

The appropriateness of this relationship was confirmed with the data. As both VAR(m) and k need to be estimated, the process must be iterative, as explained elsewhere (*9*).

GOODNESS OF FIT

In ordinary least-square regression, the coefficient of determination, R^2, is frequently used to express the goodness of fit of a model. It represents the proportion of variation in the observation that is explained by the model and can be calculated in two ways

$$R^2 = 1 - \left(\frac{\text{Unexplained variation}}{\text{Total variation}}\right) \text{ or}$$

$$R^2 = \frac{(\text{Explained variation})}{(\text{Total variation})} \quad (3)$$

However, when the variance is not constant (as with the negative binomial distribution), both forms of this equation do not yield identical results, and the R^2 statistic does not constitute a precise estimator of the goodness of fit. Nevertheless, two values of "Pseudo R^2" have been calculated from Equation 3; they constitute a possible range of R^2 values. The difficulty arises in that no equivalent measure of goodness of fit has yet been developed and widely accepted when the error structure is other than normal.

McCullagh and Nelder (*10*) proposed to evaluate the discrepancy of a fit based on the deviance or on the generalized Pearson X^2 statistic. Maycock and Hall (*6*) determined that the expected value of the scaled deviance for a good model having a negative binomial type of error follows a χ^2 distribution with $(n - p)$ degrees of freedom as long as the fitted values are generally larger than 0.5; n is the number of observations, and p is the number of estimated parameters. Larger than expected values of scaled deviance indicate model deficiencies. The appropriateness of adding parameters to a model can be evaluated by comparing decreases in scaled deviance versus decreases in number of degrees of freedom between two models. A decrease in scaled deviance that exceeds the decrease in the number of degrees of freedom justifies the additional complexity of a model. However, when many fitted values are smaller than 0.5, the expected value of the scaled deviance is considerably less than 1 and the χ^2 comparison cannot be used to evaluate the goodness of fit of a model. The Pearson X^2 statistic is calculated from

$$\text{Pearson } X^2 = \sum_i \frac{[x_i - E(m_i)]^2}{VAR(x_i)} \tag{4}$$

Miaou et al. (*11*), Bonneson and McCoy (*12*), and Persaud and Dzbik (*13*) evaluated their models on the basis of this statistic, which also follows a χ^2 distribution. McCullagh and Nelder (*10*) mentioned that it may not provide adequate results for limited amounts of data. Hauer (*5*) based model evaluations on the maximum value of the k parameter of Equation 2. Given the relationship between $E(m)$ and $VAR(m)$, models with larger values of k provide a better overall fit because they have a smaller variance. In this research, the evaluation of the adequacy of our models was based on the average behavior of these four indicators: k, scaled deviance, Pearson X^2, and pseudo R^2.

RESULTS

Models for Total Intersection Accidents

In the first stage of regression modeling, relationships between the total number of accidents and various traffic flow functions were explored. To choose functional forms that were coherent with the data, the appropriateness of the selected relationships was verified. The procedure is illustrated with the simple model of the sum of entering vehicles (Q1). A graph of the number of accidents per day versus Q1 was prepared (Figure 3). Sites were ordered in increasing values of Q1 and assembled into groups. Each square on the graph represents an average of 15 sites. The relationship between these two variables is evident: it is almost linear with a hint of downward bend. Regression has been evaluated with the more general functional form of Q1 raised to a power. The best model is

$$\text{Acc/Day} = 3.65 \times 10^{-6} * Q1^{.86} \tag{5}$$

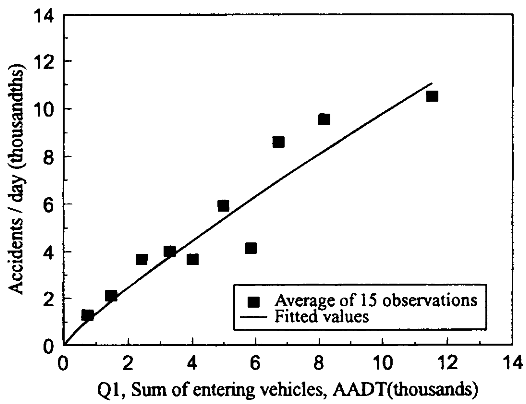


FIGURE 3 Model of total intersection accidents/day versus sum of entering vehicles.

The fitted curve is also shown in Figure 3. A similar approach has been used to develop models that estimate the number of accidents from the product of major and minor flows (F1, F2), the sum of products of conflicting flows (Q2), and the sum of weighted products of conflicting flows (Q3). Although various definitions have been proposed to describe the notion of conflicting flows, no agreement has yet been reached as to its best representation (*14,15*). In the determination of a conflict index, the concern was to ensure that the proposed function provides an adequate representation of accident occurrence in the population of sites. Accordingly, the conflict index consists of the sum of the products of each combination of flows that is involved in ten accident patterns responsible for 95 percent of all collisions involving at least two vehicles. Q2 is calculated from

$$Q2(l) = \sum_{i=1}^{10} \sum_{j=1}^{p} \left[F1(i, j, l) * F2(i, j, l) \right] \tag{6}$$

where

i = pattern number,
j = approach number,
l = intersection number, and
p = number of occurrences of the same pattern at each intersection.

Because different patterns of accidents have different probabilities of occurrence, a model that takes into consideration the relative risk of a maneuver is likely to provide a better fit than a model allocating the same weight to all products of conflicting flows. To test this hypothesis, weighting indexes (WI) have been calculated. These weighed indexes are obtained by dividing the total number of accidents of a given pattern by the corresponding sum of products of contributing flows

$$WI(i) = \frac{\sum_{j=1}^{p} \sum_{l=1}^{149} acc(i, j, l)}{\sum_{j=1}^{p} \sum_{l=1}^{149} F1(i, j, l) * F2(i, j, l)} \tag{7}$$

Values of the weighting indexes are shown in Table 1. They range from 0.08 for rear-end collisions to 4.19 for right-angle collisions. The flow function Q3 becomes

$$Q3(l) = \sum_{i=1}^{10} \sum_{j=1}^{p} \frac{WI(i)}{p} * [F1(i, j, l) * F2(i, j, l)] \tag{8}$$

Results of total intersection accident models are summarized in Table 2. Although ranking obtained from each goodness-of-fit indicator is unique, the product of major and minor flows is generally identified as the best functional form.

Modeling by Type of Accidents

A logical approach to modeling consists of relating accidents to the traffic flows that cause the impact. In our population of sites, right-angle collisions (Pattern 1) account for 42 percent of all collisions involving two or more vehicles, and a specific model has been developed for this accident pattern. The pattern second in importance is angle collisions between a through vehicle and a left-

TABLE 1 Weighting Factors, by Patterns

Pattern Number	Pattern Type	Number of accidents	Weighting indexes
1		403	4.19
2		125	1.08
3		83	0.08
4		73	1.00
5		71	0.91
6		45	0.49
7		31	0.25
8		27	0.35
9		24	0.21
10		21	0.46
Total		903	

turning vehicle; 125 collisions of this type have been coded. As causes of these collisions differ depending upon whether the left-turning vehicle is located on the minor or the major approach, these accidents were subdivided into two groups. The resulting subsets were too small to allow the determination of logical relationships based on observed trends of the data, and the goodness of fit was reduced. Instead of proposing several "intuitive models" that would present a poor fit for more than half the data, only right-angle collisions were analyzed in detail, and all remaining collisions were grouped into one aggregate model. The resulting tool to evaluate the safety of an intersection consists of two models: a right-angle model and a remaining patterns model.

Estimation of Contribution of Additional Features

To assess whether variables other than traffic could make a significant contribution to the explanation of accident occurrence, factors were added to the models; these are dummy variables that take distinct integer numbers for each specific value of a variable. For example, to evaluate the influence of flashing beacons on the observed number of accidents, a two-level factor is created: 0 for intersections with flashers and 1 for intersections without flashers. If the coefficients of each value of the factor are different, it means that the factor has an influence on the total number of accidents. The following functional form was used to examine the effect of flashing beacons, sight distance, turning lanes, and speed:

$$\text{Acc/day} = b_0 * F1^{b1} * F2^{b2} * e^{(\text{Factor})} \tag{9}$$

For example, the value of the "flasher factor" for intersections that are equipped with this warning device is 0.17, which indicates that at the same flow these junctions are expected to have 19 percent more accidents than intersections without such a device. However, given the magnitude of the standard error of this coefficient (0.15), the effect is uncertain. As shown in Table 3, similar results were obtained for sight distance, turning lanes, and speed. It should be remembered that regression equations provide relationships that are associative and not causative. That intersections with flashers have on average more accidents does not necessarily mean that beacons reduce the safety. Instead, it could be that they are generally installed at intersections that have a poorer safety performance and that they do not succeed in making these junctions as safe as other similar sites.

Whenever feasible, it is better to assess the effect of a variable by the development of distinct models for each level of a factor. With this data, it was possible to do so for the maximum posted speed at intersection approaches. Models were developed for the 50 and 90 km/hr speed limits. Results are summarized in Table 4.

TABLE 3 Effect of Causal Factors

Characteristic	Factor	Value	Standard error	k
Flashing beacon	1: no 2: yes	0.00 0.17	- 0.15	3.0
Sight distance	1: <100m 2: 100-200m 3: 200-300m 4: >300m	0.00 0.41 0.17 0.45	- 0.24 0.25 0.23	3.3
Turning lanes	1: 2 lanes 2: 2 + RT 3: 2 + LT 4: 2 + LT + RT	0.00 0.10 0.25 0.21	- 0.16 0.20 0.22	3.1
Speed limit	1: 50 Km/hr 2: 90 Km/hr	0.00 0.17	- 0.15	3.6

TABLE 2 Models for "Total Intersection Accidents"

Functional form	k	Deviance (d.f.)	X^2	R^2
$\text{Acc/Day} = 3.65 * 10^{-6} * Q1^{.86}$	2.50	168.45 (147)	144.01	.42, .50
$\text{Acc/Day} = 5.59 * 10^{-6} * F1^{.42} * F2^{.51}$	2.95	164.64 (146)	135.88	.47, .56
$\text{Acc/Day} = 4.41 * 10^{-5} * Q2^{.36}$	2.05	164.75 (147)	146.32	.33, .53
$\text{Acc/Day} = 3.14 * 10^{-5} * Q3^{.45}$	2.80	166.14 (147)	147.04	.50, .50

TABLE 4 Summary of Results

	Main road speed limit 50 km/hr	Main road speed limit 90 km/hr	All speeds limits
Total intersection models			
a1) Acc/Day=b_0 * Q^{b1}	b_0=4.59E-6 b1=0.83 k =2.70	b_0=4.42E-6 b1=0.83 k =2.70	b_0=3.65E-6 b1=0.86 k =2.50
a2) Acc/Day=b_0 * $F1^{b1}$ * $F2^{b2}$	b_0=1.07E-5 b1=0.34 b2=0.49 k =3.10	b_0=3.37E-6 b1=0.41 b2=0.59 k =5.10	b_0=5.29E-6 b1=0.42 b2=0.51 k =2.95
Pattern models			
b1) Pattern 1 (Right Angle): Acc/Day=b_0 * EXP(b_1 * F1) * $F1^{b2}$ * $F2^{b3}$ (50 km/hr and "all speeds") Acc/Day=b_0 * $F1^{b1}$ * $F2^{b2}$ (90 km/hr)	b_0=2.05E-6 b1=-3.69E-4 b2=0.57 b3=0.46 k =1.95	b_0=6.14E-6 b1=0.32 b2=0.43 k =1.40	b_0=1.09E-5 b1=-6.52E-5 b2=0.26 b3=0.46 k =1.50
b2) Remaining Patterns: Acc/Day=b_0 * $F1^{b1}$ * $F2^{b2}$	b_0=1.97E-6 b1=0.59 b2=0.36 k =3.30	b_0=1.57E-6 b1=0.57 b2=0.45 k =6.20	b_0=1.42E-6 b1=0.65 b2=0.35 k =2.80

ESTIMATION OF E($m|x$) AND VAR($m|x$)

With these models it is now possible to estimate the safety of a four-legged unsignalized intersection when its traffic flow and accident history are known. This estimation consists of two steps

1. Estimation of E(m) and VAR(m) with the multivariate models, and
2. Estimation of E($m|x$) and VAR($m|x$) from Equation 1.

The application of the method is shown with the following example. Suppose that the traffic flows and accident count at a four-legged unsignalized intersection are as indicated in Figure 4. Depending on the availability of the data, different models can be selected to estimate its safety. When the number of collisions by pattern and traffic flow estimates per movement are available, Models b1 and b2 of Table 4 can be used. If instead only the major and minor flows are known, Model a2 should be used. In this case, the calculation is as follows:

Step 1: Estimation of E(m) and VAR(m)

$$\begin{aligned} E(m) &= (1.07 * 10^{-5} * 4500^{.34} * 2000^{.49}) * 1095 \\ &= 7.74 * 10^{-3} * 1095 \\ &= 8.48 \text{ acc/3 years} \\ VAR(m) &= [(7.74 * 10^{-3})^2/3.10] * 1095^2 \\ &= 23.17 \text{ (acc/3 years)}^2 \end{aligned}$$

Step 2: Estimation of E($m|x$) and VAR($m|x$)

$$\begin{aligned} E(m|x) &= aE(m) + (1 - a)x \\ &\quad \text{with } a = E(m)/[E(m) + Var(m)] \\ &\quad = 8.48/(8.48 + 23.17) \\ &\quad = 0.27 \\ &= (0.27 * 8.48) + [(1 - 0.27) * 15] \\ &= 13.24 \text{ acc/3 years} \\ VAR(m|x) &= a(1 - a)E(m) + (1 - a)^2x \\ &= (0.27 * (1 - 0.27) * 8.48) + [(1 - 0.27)^2 * 15] \\ &= 9.67 \text{ (acc/3 years)}^2 \end{aligned}$$

In this example, the estimate of safety is reduced from 15 to 13.24 acc/3 years, which corresponds to a RTM correction of 12 percent. The larger the difference between the number of accidents at the site and the expected value of the reference population, the larger the correction. Once these estimates are made available, two major tasks can be accomplished: identification of entities that require intervention and evaluation of the effects of road safety interventions.

IDENTIFICATION OF DEVIANT SITES

A site is selected when the difference between its safety and the safety of sites having similar characteristics is unacceptable.

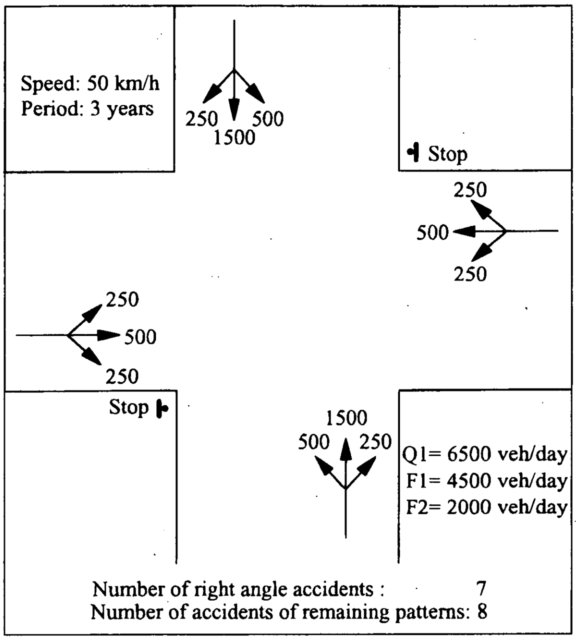


FIGURE 4 Numerical example.

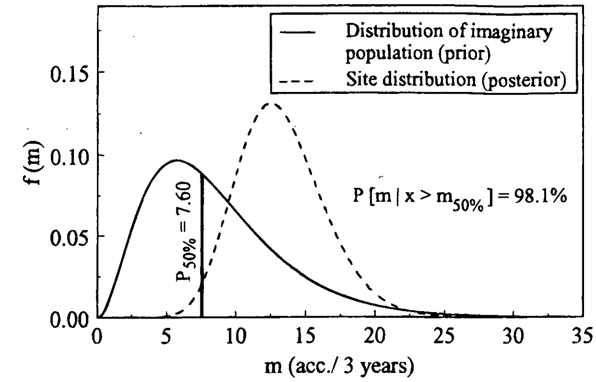


FIGURE 5 Gamma distributions, prior and posterior estimates of safety.

The determination of what is unacceptable should be a function of the resources allocated to the correction of deviant sites. The process is as follows:

- Estimate E(m) and VAR(m) from the multivariate models and plot the probability density function (pdf) of the reference population (gamma distribution).
- On the basis of this pdf, determine the value of m to be used as a point of comparison; the use of the median of the reference population ($P_{50\%}$) is recommended.
- Estimate E($m|x$) and VAR($m|x$) from Equation 1 and plot the pdf of the intersection of interest (gamma distribution).
- On the basis of this pdf, calculate the probability of the $m|x$ of this intersection being larger than the median of the reference population and decide whether the site is deviant.

The previous example is continued to illustrate this method. It has been estimated that E(m) = 8.48 acc/3 years and VAR(m) = 23.17 (acc/3 years)^2. When the density function is plotted one finds that the median of the distribution of ms for an imaginary group of four-legged unsignalized intersections having a major flow of 4500 vpd and a minor flow of 2000 vpd is 7.60 acc/3 years. It was also estimated that intersections with this flow combination that had 15 accidents in the last 3 years have E($m|x$) = 13.24 acc/3 years and VAR($m|x$) = 9.67 (acc/3 years)^2. When the corresponding pdf is plotted one finds that there is only a 1.9 percent probability for this intersection to have m smaller than 7.60 acc/3 years. In other words, there is a 98.1 percent chance that this intersection is less safe than 50 percent of intersections having similar characteristics; consequently, it is selected for treatment. The result is illustrated in Figure 5.

To facilitate the use of the method, a computer program has been developed that estimates the safety of these intersections and identifies blackspots. Information on this program is available from the author.

EVALUATION OF EFFECTS OF INTERVENTIONS

To estimate an intervention's effect on safety, an index of effectiveness (IE) must be calculated. It corresponds to the following ratio:

$$\text{IE} = \frac{\text{Safety in the after period}}{\text{Safety that would have been after, without intervention}} \tag{10}$$

The numerator and denominator of this equation need to be estimated adequately. When the number of accidents is large enough to minimize the effect of random variations, the count of accidents in the after period is a good estimate of the safety after treatment. The estimate of what would have been the safety of the entity in the after period if the intervention had not been implemented is more difficult to obtain because it corresponds to a quantity that cannot be observed directly. A commonly used estimator of the denominator of Equation 10 is the observed number of accidents in the period preceding the intervention, but it often leads to an overestimation of the benefits of our actions. To improve the accuracy of the denominator, two questions must be answered.

1. What was the safety of the entity before treatment?
2. How would the estimate of safety in the before period have changed between the before and after period if the intervention had not been implemented?

The safety before treatment at four-legged unsignalized intersections should be estimated from the multivariate models and the knowledge of the number of accidents at the site, as shown in the previous example. Between the before and the after period, several factors are likely to have changed and to have modified the level of safety at the site: traffic, weather, economy, and so forth. The influence of some factors is unknown and cannot be estimated, but it is important to calculate the effect of factors whose influence is known. For example, the impact of modifications in traffic flows can be estimated from the models. The previous example is continued to illustrate the method.

At the same intersection, 11 accidents have been recorded in a 3-year period following its treatment. In the same period, the average daily traffic increased from 4500 to 5000 vehicles/day on the major street and from 2000 to 2500 vehicles/day on the minor street. The best estimate of safety in the after period is 11 acc/3 years. Earlier it was found that the estimate of the safety of the intersection before treatment is 13.24 acc/3 years. To correct for changes in traffic, use is made of the Model a2 (Table 4). With the after flows, one would expect 9.81 acc/3 years, which represents an increase of 16 percent compared with the original level of traffic. Assuming that only traffic flow changes can be taken into consideration, the estimate of what would have been the safety of the entity in the after period without intervention is equal to $(13.24 * 1.16) = 15.36$ acc/3 years. Accordingly, the index of effectiveness is

$$\begin{aligned} IE &= 11/15.36 \\ &= 0.72 \end{aligned}$$

In other words, the treatment at this intersection is estimated to be associated with a 28 percent reduction in accidents. However, results of similar interventions at several intersections are required to increase the accuracy of this estimated effect.

SUMMARY AND CONCLUSIONS

In this study, multivariate models have been developed that can be used to estimate the safety of four-legged unsignalized intersections in an EB framework. Multivariate models are used to estimate the moments $E(m)$ and $VAR(m)$ of an "average" intersection; this information is then combined with the count of accidents (x) at a specific intersection to calculate its updated estimate of safety, as expressed by $E(m|x)$ and $VAR(m|x)$.

This study confirms the applicability of methodological elements proposed in recent research. The negative binomial error structure was shown to be consistent with the data. Also confirmed by the data is the useful empirical relationship between $E(m)$ and $VAR(m)$; that is $VAR(m) = [E(m)]^2/k$.

Total intersection models and pattern models for three categories of speed were developed: 50 km/hr, 90 km/hr, and all speeds. The 50 and 90 km/hr models are more precise than the all speeds models and should be used whenever possible. When only the total number of accidents and entering vehicles on each approach is known, total intersection models must be used. However, when accidents by pattern and traffic volumes by movements are available, the use of pattern models is preferred. They constitute a more powerful tool of analysis because they can provide a detailed identification of abnormal situations. For example, a site could have a total number of accidents not significantly higher than the average total for similar sites but show an abnormal frequency of right-angle collisions.

In practice, both the total intersection models and pattern models are of interest. Given that it is unlikely that accidents by pattern and detailed traffic flow estimates will be available on a large scale in the near future, total accident intersection models could be used as a first sieve. Data requirement is not as extensive as with pattern models and allows for a wider number of intersections to be considered initially. Detailed information could then be collected on the reduced sample to make possible the use of more precise pattern models.

ACKNOWLEDGMENTS

This research has been made possible by the support of the Ministry of Transportation of Quebec. We are grateful for the collaboration of Ezra Hauer, who supervised the work.

REFERENCES

1. Hodge, G. A., and A. J. Richardson. The Role of Accident Exposure in Transport System Safety Evaluations I: Intersection and Link Site Exposure. *Journal of Advanced Transportation*, Vol. 19, No. 2, 1985, pp. 179–213.
2. Abbess, C., D. Jarrett, and C. C. Wright. Accidents at Blackspots: Estimating the Effectiveness of Remedial Treatment, with Special Reference to the "Regression-to-Mean" Effect. *Traffic Engineering and Control*, Vol. 22, No. 10, 1981, pp. 535–542.
3. Brüde, U., and J. Larsson. *The "Regression-to-Mean" Effect. Some Empirical Examples Concerning Accidents at Road Junctions.* VTI Report 240, 1982.
4. Hauer, E., and B. N. Persaud. Common Bias in Before and After Accident Comparisons and its Elimination. In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 164–174.
5. Hauer, E. Empirical Bayes Approach to the Estimation of "Unsafety": the Multivariate Regression Method. *Accident Analysis and Prevention*, Vol. 24, No. 5, 1992, pp. 457–477.
6. Maycock, G., and R. D. Hall. *Accidents at 4-Arm Roundabouts.* TRRL Laboratory Report 1120. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
7. Baker, R. J., and J. A. Nelder. *The GLIM System—Release 3.* Rothamsted Experimental Station, Harpenden, U. K., 1978.
8. Hauer, E., and B. N. Persaud. How to Estimate the Safety of Rail-Highway Grade Crossings and the Safety Effect of Warning Devices. In *Transportation Research Record 1114*, TRB, National Research Council, Washington, D.C., 1988, pp. 131–140.
9. Hauer, E., J. C. N. Ng, and J. Lovel. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
10. McCullagh, P., and J. A. Nelder. *Generalized Linear Models.* The University Press, New York, 1983.
11. Miaou, S. P., P. S. Hu, T. Wright, A. K. Rathi, and S. C. Davis. Relationship Between Truck Accidents and Highway Geometric Design: a Poisson Regression Approach. In *Transportation Research Board 1376*, TRB, National Research Council, Washington, D.C., 1992, pp. 10–18.
12. Bonneson, J. A., and P. T. McCoy. Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways. In *Transportation Research Record 1401*, TRB, National Research Council, Washington, D.C., 1993, pp. 83–89.
13. Persaud, B., and L. Dzbik. Accident Prediction Models for Freeways. In *Transportation Research Record 1401*, TRB, National Research Council, Washington, D.C., 1993, pp. 55–60.
14. Surti, V. H. Accident Exposure for At-Grade Intersections. *Traffic Engineering*, Vol. 36, No. 3, 1965, pp. 26–27.
15. Peleg, M. Evaluation of the Conflict Hazard of Uncontrolled Junctions. *Traffic Engineering and Control*, Vol. 9, No. 7, 1967, pp. 346–347.

Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.