Estimation of Safety at Signalized Intersections

EZRA HAUER, JERRY C.N. NG, AND JANE LOVELL

Models to estimate the safety of a signalized intersection on the basis of information about its traffic flow and accident history are provided. They are based on data from 145 intersections in Metropolitan Toronto. Several insights were obtained during the development of the models. First, logically sound models require that the frequency of collisions be related to the traffic flows to which the colliding vehicles belong and not to the sum of the entering flows. Second, it is therefore necessary to categorize collisions by the movement of the vehicles before the collision and not by the initial impact type, as is customary. Third, the relationship between collision frequency and the related traffic flows is at times unexpected in form. For each of the 15 accident patterns, an equation is given to estimate the expected number of accidents and the variance using the relevant traffic flows. When data about past accidents are available, estimates based on traffic flow are revised with a simple equation. Several practical questions can now be answered. Given the traffic flow for a signalized intersection, one can predict how many and what kinds of accidents should be expected to occur on it; one can also show the probability density function (pdf) of the estimate. Knowledge of the pdf allows the determination of what an unusually high number of accidents would be on such an intersection. If the traffic flow of the intersection changes from year to year, one can estimate what changes in safety should be attributed to changes in flow. Also, one can correctly compare the safety of several intersections that have different flow patterns. Most important, one can estimate safety when both flows and accident history are given and, on this basis, judge whether an intersection is unusually hazardous. This method of estimation is recommended for accident warrants in the Manual on Uniform Traffic Control Devices.

In this paper we give equations to estimate the safety of a signalized intersection when the vehicular flows using it are known. This kind of estimate describes the safety of an "average" signalized intersection with given flows. How the safety of such an average signalized intersection depends on the details of the traffic flow may be regarded as basic knowledge, the raw material on which engineering design and decisions are or should be based. Ours is not the first attempt to explore this question, and earlier work will be reviewed first.

The ability to estimate intersection safety as a function of traffic flows may be useful when one has to judge whether the pattern of accidents and their number at some specific intersection are similar to what one might normally expect with such flows. Surely if one wishes to identify what is unduly hazardous and to diagnose what the reasons for such deviation might be, one has to have a good idea about what is normal. To judge what is normal and what is deviant, one has to know also what variability is found in the population of similar intersections. This is why an estimate of the variance of safety will also be provided.

Another circumstance in which one has to know how safety depends on traffic flows is when one has to judge whether some intervention has affected safety and what the extent of the effect is. In this case one has to separate those changes in safety that are due to the intervention from those that are due to the concurrent and inevitable changes in traffic flow.

When for a specific intersection the number of accidents is known (in addition to the traffic flow), this added information must also be reflected in the estimate of safety for this specific intersection. We will show how to combine the safety estimate based on flows with accident data. In our view, it is this combined estimate that should be the basis of the accident warrants used in the Manual on Uniform Traffic Control Devices.

We spoke of "safety" without stating what the word means. "Safety" is the property of some specific entity, in this case that of a signalized intersection. The "safety property" of an intersection is defined as the number of accidents and their adverse consequences expected to occur on it per unit of time. The term "expected" is equivalent to the average in the long run if it were possible to freeze all prevailing conditions that affect safety, such as traffic, weather, driver characteristics, and so on. The safety of some intersection will be denoted m. The mean of the m's in a population of intersections will be denoted $E\{m\}$ and their variance $\mathrm{Var}(m)$.

PREVIOUS WORK

Numerous relationships between accident frequencies and traffic flows have been suggested over the years. A comprehensive survey of these relationships has been given by Chapman (*1*) and Satterthwaite (*2*). The following summarizes findings that pertain to intersections.

Transport Safety Studies Group, Department of Civil Engineering, University of Toronto, Toronto, M5S 1A4, Canada.

Thorpe (*3*), Smith (*4*), and Worsey (*5*) suggest that the number of all accidents at an intersection is proportional to the sum of flows that enter the intersection. The merit of this approach is its simplicity. Its shortcoming is that it is logically unsatisfactory and not a suitable basis for the engineering analysis, which attempts to link cause and effect. One expects, for example, that the number of rear-end accidents at an intersection approach will strongly depend on the flow on approach A and depend only weakly on the flow on approaches B, C, and D. Similarly one should expect that collisions between vehicles from streams A and B moving at right angles to each other might be related to the product of flows A and B. To use the sum of flows A and B leads to the logical difficulty that one will predict accident occurrence even when one of the flows is zero; to use A + B + C + D just compounds two logical difficulties. When a model rests mainly on considerations of correlation and has obvious logical faults, its prediction performance is usually poor. To illustrate, suppose that apples and cherries are grown in the orchards of Ontario. The bushels of fruit produced in Ontario are strongly correlated with the acres of orchards in Ontario and could be estimated on that basis. However, most would agree that it would be better to estimate fruit production by first estimating bushels of cherries on the basis of the acres of cherry orchards and bushels of apples on the basis of the acres planted with apple trees and only then add the two numbers. Not only does one get more detailed information (about apples and cherries, rather than "fruit") but one rightly expects the result to be more reliable. This is the approach that we advocate.

Breunning and Bone (*6*), Surti (*7*), and Hakkert and Mahalel (*8*) relate accidents to the products of the conflicting flows. This is based on the speculation that were drivers blind, the number of collisions could be expected to be proportional to the product of vehicle flows. However, in other empirical research (*9, 10*) it has been found that the number of accidents is in fact not proportional to the product of flows. Rather, accidents were found to be related to the product of flows with each flow raised to a power of less than 1. Tanner (*11*) suggested that the square root of the product of flows would be sufficiently accurate as a rule of thumb.

Support for the "product-of-flows-to-power" relationship can be found in other circumstances as well. For example, when the expected number of accidents at a highway-rail grade crossing is calculated, train flow and highway traffic flow usually enter into the product with an exponent of 0.3 to 0.6 (*12–14*). Similarly, when estimating the expected number of "opposite direction" accidents on two-lane rural roads, Zegeer et al. (*15*) opt for the product-of-flows-to-power model.

Our effort to relate accidents at signalized intersections to traffic flow will be guided by the primacy of logical requirements. First, we will attempt to relate accidents to the traffic flows to which the colliding vehicles belong. This means that accidents between vehicles proceeding in the same direction have to be estimated separately from accidents between, say, vehicles turning left and those proceeding straight through the intersection. (This is analogous to the separation of apples and cherries.) Second, we will examine the data to see what functional relationship is indicated before we decide on the functional form.

It should be obvious that this kind of estimation requires information about turning flows. This puts a strain on practicality when only approach volumes are counted by automatic counters. Even in this case, the suggested estimation procedure can still be used except that turning flows have to be estimated first. Methods for doing so are easily available and commonly used [see, e.g., paper by Hauer et al. (*16*)].

THE DATA

To enhance the chances of success we have selected for analysis a set of signalized intersections that are similar in most respects except traffic flows and accident history. Thus, the data are for 145 four-legged, fixed-time, signalized intersections in Metropolitan Toronto that carry two-way traffic on all approaches and have no turn restrictions. Most are on straight, level sites with a speed limit of 60 km/hr (35 mph).

One-day vehicle counts were collected manually. Thus, for each approach we have details of turning flow and straight-through flow for the a.m. peak, p.m. peak, and off-peak. All vehicle counts are for weekday conditions, and the majority of counting was done during 1984.

The accident data are for 1982, 1983, and 1984. They were derived from the computerized version of the police accident report. A consistency check was performed with the computerized data, but the hard copy of the form was not consulted. To correspond to the available traffic count information, we used only "daytime" accidents, that is, those that occurred between 7:00 and 9:00 a.m. (the morning peak), those that occurred between 4:00 and 6:00 p.m. (the evening peak), and those in the time interval between 10:00 a.m. and 3:00 p.m. (the off-peak). The hours 9:00 to 10:00 a.m. and 3:00 to 4:00 p.m. were excluded from the analysis because this is when signal timing plans change. Accident data were divided into collisions involving pedestrians, single-vehicle accidents, collisions between two vehicles, and collisions involving more than two vehicles. The frequency with which these four classes arise is as follows:

Accident Type	*No.*
Single vehicle	54
Two vehicles	2,084
Multivehicle	248
Vehicle-pedestrian	187
Total	2,573

For the purpose of this analysis, only collisions involving two vehicles were examined. These accidents represent 81 percent of the total accidents.

In summary, the following analysis concerns weekday, daytime crashes of two vehicles at simple signalized intersections.

ANALYSIS OF ACCIDENTS BY PATTERN

To begin the statistical analysis from a logically satisfactory foundation, we thought it best to relate accidents to the flows to which the two colliding vehicles belong. The real interactions are perhaps more complex. However, as a point of departure and in view of the paucity of our data, it appears sensible to begin with the obvious.

The 15 patterns in which two vehicles at a four-legged intersection can collide are shown in Figure 1. Accidents in each pattern are defined by the maneuvers of the two vehicles before the collision (which are recorded on the police accident report as "turning left," "turning right," "going ahead," etc.). Thus, for example, in Pattern 1 the collision is between vehicles that are proceeding straight through the intersection, the crash occurring before the stop line; in Pattern 2 the same vehicles are involved but the crash is within the intersection; in Pattern 13, one of the vehicles is turning right.

We avoid using the more common categorization by initial impact type (such as rear-end, angle, turning movement, sideswipe, etc.) because of its ambiguity. It is indeed illuminating to examine the cross-tabulation of accidents classified by both criteria as is shown in Table 1.

Categorization by initial impact type is most common in the practice of intersection safety analysis. To demonstrate the imprecision of this type of analysis, Table 1 reveals that less than half of the angle collisions arise from Pattern 4; Patterns 5 through 12 all involve a left-turning vehicle, yet only 423 of 910 are classified as turning-movement accidents according to the initial impact type; most approaching initial impacts are not in Pattern 3 but

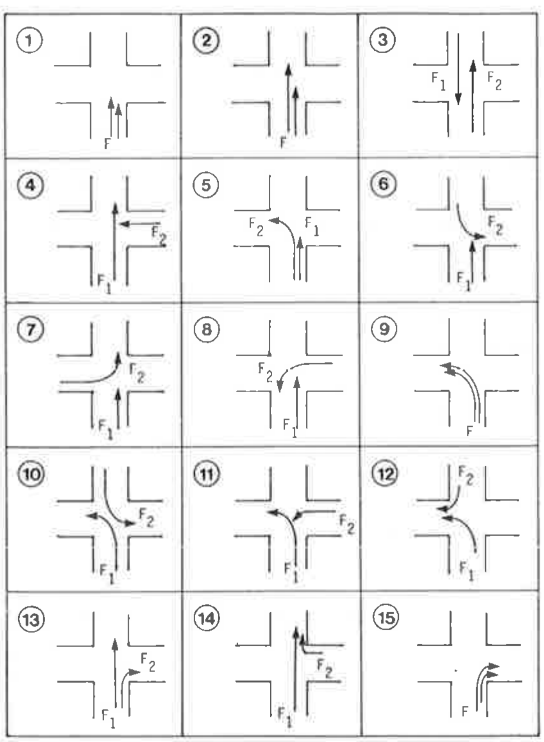

FIGURE 1 Accident patterns by vehicle streams.

in Pattern 6; and so on. We are led to the conclusion that (as many practicing traffic engineers have known all along) when accidents are categorized by initial impact type, their cause-and-effect relationship with traffic flow is weakened. To understand and analyze accidents at intersections, it is better to use the "vehicle maneuver" entry from the police accident report.

EXPLORATION: HOW ACCIDENTS DEPEND ON THE CONTRIBUTORY TRAFFIC FLOWS

To illustrate how the form of the model equations was chosen, consider the typical left-turn accident represented by Pattern 6. For each intersection this pattern arises four times. Therefore, the 145 intersections give 580 "sites," each with different values for the two "contributory" vehicle flows. Table 2 gives the average number of accidents per site in 3 years for five ranges of the left-turning flow versus five ranges of the straight-through flow (in vehicles per day). The cells give the average number of accidents per site in 3 years and, in parentheses, the number of Pattern 6 accidents and the number of sites with flows in the range for that particular cell. The rightmost column and the bottom row give the same information summed over the corresponding row or column. The somewhat irregular flow ranges were selected so that each row and column would have approximately one-fifth of all accidents.

In principle, one should seek a function form that fits the cell entries. However, in this case, the cell entries in Table 2 are based on too few accidents to allow for finesse. Therefore, useful clues are derived from examining the row and column totals. This is shown in Figure 2a and b.

There is a clear suggestion in Figure 2a that within the range of the data, accidents are proportional to the through traffic. Therefore, denoting the through flow in Pattern 6 F_1, we wish to select a model form such that the expected number of Pattern 6 accidents ($E\{m_6\}$) is proportional to F_1. The increase of accidents with the left-turning flow (F_2) appears to be nonlinear. The kind of increase indicated in Figure 2b can be captured by a function such as $F_2^{b_2}$, where b_2 is a coefficient smaller than 1. Therefore, for Pattern 6 accidents, a simple functional relationship that can closely match what we observe in Figure 2a and b is

$$E\{m_6\} = b_0 \times F_1 \times F_2^{b_2} \qquad (1)$$

No cause-and-effect arguments were used in selecting this functional form; the guiding principle was the wish to ensure a satisfactory fit with parsimony of parameters and without violation of the obvious logical requirements. We also explored alternative functional forms, which all increased the number of parameters. In no case did this seem worthwhile in terms of the improved fit to the data.

A similar exploration of how Pattern 1 and Pattern 2 accidents depend on traffic flow led to the expected conclusion: in this case, a straight-line fit seems satisfactory.

The examination of Pattern 4 proved more interesting. Because of the symmetry in the situation (see Figure 1) we

TABLE 1 CROSS-TABULATIONS OF ACCIDENTS BY VEHICLE MANEUVER PATTERN AND INITIAL IMPACT TYPE

PATTERN	NUMBER OF ACCIDENTS	INITIAL IMPACT TYPE					
		RE	An	Tn	Ss	Ap	Other
1	538	429	45	5	52	1	6
2	197	109	30	1	52	0	5
3	12	0	4	0	3	5	0
4	258	1	237	0	6	11	3
5	47	10	13	18	5	1	0
6	729	0	329	355	6	32	7
7	42	5	22	14	0	0	1
8	32	0	17	9	1	5	0
9	29	15	0	10	3	0	1
10	2	0	0	2	0	0	0
11	7	1	3	3	0	0	0
12	22	0	7	12	2	0	1
13	73	20	14	29	5	0	5
14	80	5	33	34	8	0	0
15	16	8	3	2	3	0	0
Total	2084	603	727	494	146	55	29

Initial Impact Type : RE : Rear-End
An : Angle
Tn : Turning Movement
Ss : Sideswipe
Ap : Approaching

TABLE 2 CROSS-TABULATIONS OF ACCIDENTS BY VEHICLE FLOWS FOR PATTERN 6

STRAIGHT-THROUGH FLOW (veh./day)	LEFT-TURNING FLOW (veh./day)					ROW TOTAL
	0-521	522-821	822-1033	1034-1408	1409-4038	
0-3525	0.49 (51/104)	0.60 (35/ 58)	0.63 (20/ 21)	1.57 (33/ 21)	0.56 (9/ 16)	0.64 (148/231)
3526-4825	0.61 (27/ 44)	1.00 (25/ 25)	1.91 (44/ 23)	1.24 (21/ 17)	1.75 (28/ 16)	1.16 (145/125)
4826-5941	1.27 (19/ 15)	1.22 (33/ 27)	1.44 (23/ 16)	1.36 (15/ 11)	2.79 (53/ 19)	1.63 (143/ 88)
5942-7771	1.31 (17/ 13)	1.40 (35/ 25)	3.10 (21/ 8)	2.20 (36/ 8)	1.40 (36/ 11)	1.78 (145/ 53)
7772-12091	1.67 (30/ 18)	2.75 (22/ 8)	2.63 (21/ 8)	4.50 (36/ 8)	3.27 (36/ 11)	2.73 (145/ 53)
COLUMN TOTAL	0.74 (144/194)	1.05 (150/143)	1.56 (139/ 89)	1.93 (149/ 77)	1.91 (147/ 77)	1.26 (729/580)

have chosen to distinguish the two flows by calling the larger flow F_1 and the smaller F_2. In Figure 3a and b we show the observed relationships between the average number of accidents per site for 3 years and the two flows. It appears that, for the range of flows for which data are available, the larger of the two flows exerts little influence on the number of accidents. The smaller of the two flows exerts a great deal of influence initially, but this tapers off as the flows become larger.

This is an unexpected and tantalizing finding. With hindsight one can offer speculative explanations. It is possible, for example, that accidents of this kind involve mostly platoon leaders; the number of vehicles arriving after the platoon leader would in this case be immaterial. Alternatively, one could think of the points in Figure 3a as being a continuation of the points in Figure 3b—the same functional relationship for both flows. Of course, speculation is not a substitute for explanation and does not amount to "understanding."

Thus, on the basis of an exploratory analysis, one can suggest functional forms for expressions that fit what has been observed in Patterns 1, 2, 4, and 6 with parsimony. There are not enough data for any of the remaining accident patterns to warrant similar exploratory analyses. For these we had to select model forms by analogy and judgment.

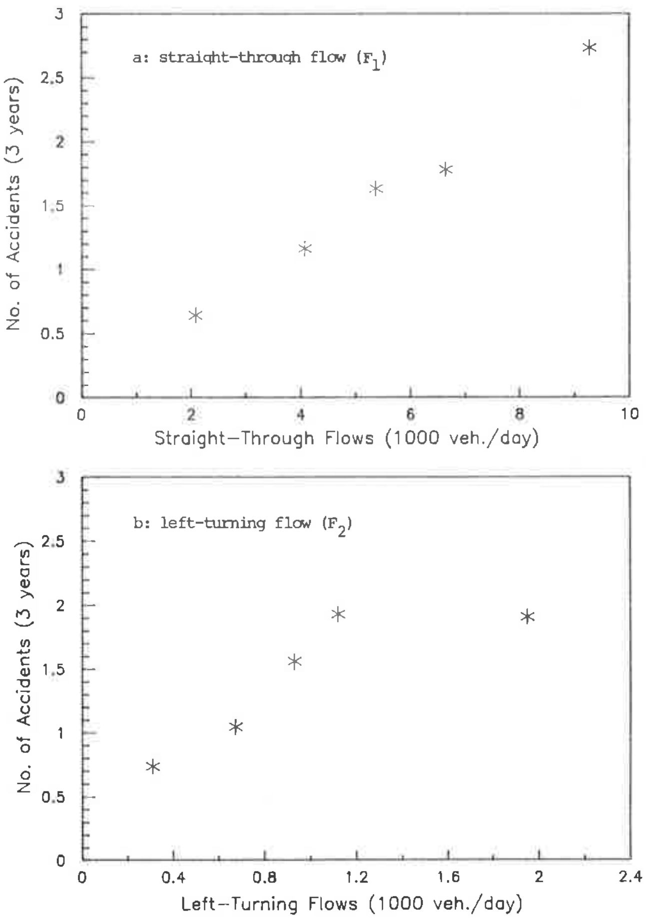

FIGURE 2 Relationship between accidents and contributory flows for Pattern 6.

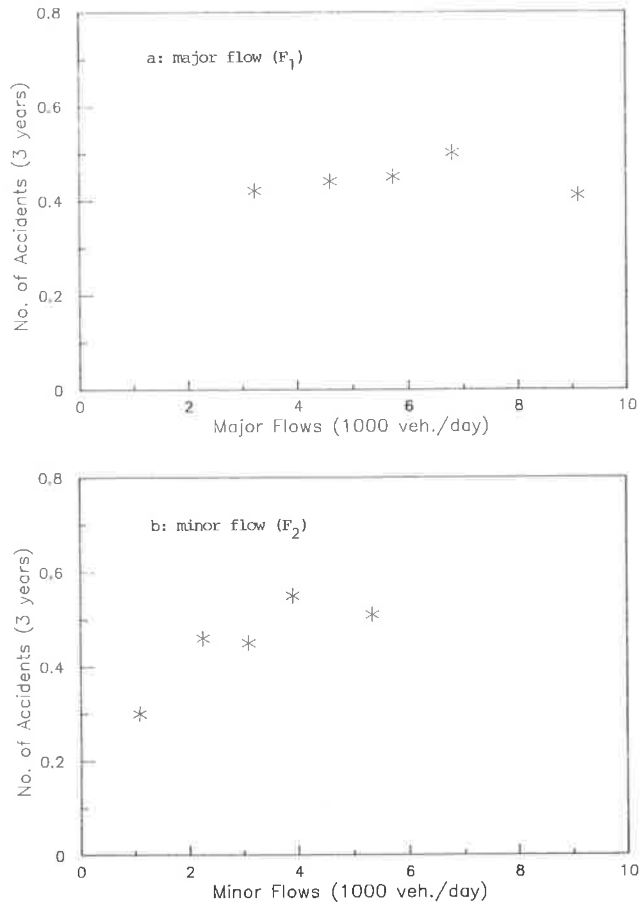

FIGURE 3 Relationship between accidents and contributory flows for Pattern 4.

On the basis of insight gained in the exploration stage we proceeded to the task of examining the statistical performance of the several model forms that appeared promising.

ESTIMATION OF COEFFICIENTS

Coefficient estimation is the domain of the professional statistician. It is a domain too frequently invaded by those who mistakenly believe that a statistical software package can be used by nonexperts. In this paper we do not attempt a comprehensive coverage of technical detail. Rather, we will point out some matters of method that seem important when it comes to the statistical treatment of accident data. It is best to do so with reference to the models at hand.

For accident Patterns 1 and 2 we have concluded that a simple model of the form $E\{m\} = b_0 * F$ seems sufficient. The reflex inclination is to "run a regression" in order to find an estimate $\hat{b}_0$ of parameter b_0. Although the estimate so obtained may be adequate, in principle to do this would be a mistake. To understand why, it is necessary to describe the conceptual framework within which this kind of parameter estimation takes place.

We have denoted m the safety of a specific intersection. Imagine a population of intersections that all have the same traffic flows. In this imaginary population, the m's would still vary from intersection to intersection because, although flows are identical, they involve different drivers in different parts of different cities, and so forth. Thus, one can speak of the mean of the m's ($E\{m\}$) in this imaginary population of intersections with identical traffic flows. This mean of m's is what describes the safety of a "representative" or "average" intersection for this imaginary population. Similarly, one can speak of the variance of the m's (Var$\{m\}$) in this imaginary population of intersections. To make a statement about the safety of a specific intersection of this population, it can be said that "intersections of this kind (e.g., signalized, with a specific pattern of traffic flow) have on the average $E\{m\}$ accidents and the variability of m in the population of intersections of this kind is Var$\{m\}$."

When fitting a model to accident data, we are trying to estimate $E\{m\}$ as a function of traffic flow (in this case). That is, we are trying to determine what the m is of some "average" or "representative" intersection and how it varies with traffic flow. However, the data used for estimation are not for "average" intersections. Each accident count we use is for one specific intersection from the imaginary population of intersections with the same flows. It follows that if $E\{m\}$ is what we wish to estimate, the accident count must be considered as a Poisson random variable that comes from a site with m as its expected value and that this m, in turn, is one of a distribution of m's characterized by $E\{m\}$ and $\mathrm{Var}\{m\}$.

Thus, the distribution of accident counts around $E\{m\}$ is one of a family of "compound Poisson distributions." In the special case in which the distribution of m's in these imaginary populations can be described by a gamma probability density function, the distribution of accident counts around the $E\{m\}$ must be taken as negative binomial. This is a radical departure from the assumptions on which the usual regression software is based. Therefore, a "least-squares" regression model should not be used without expert modification.

In this project we have estimated coefficients using the Generalized Linear Interactive Modeling (GLIM) software package (*17*). This software yields maximum likelihood estimates of coefficients and allows the user to specify the "error structure" that corresponds to the data used. In our case we specified the negative binomial error structure, following Maycock's lead (*18*).

It is already clear that to answer some practical questions about safety it is not sufficient to know what the safety ($E\{m\}$) is of an "average" intersection with given traffic flows. Ordinarily one wishes to make statements about the the safety (m) of some specific intersection. If the $\mathrm{Var}\{m\}$ is very large, knowledge of $E\{m\}$ tells us little about the m of a specific intersection, and vice versa. To make informative statements about the safety of specific intersections, one also needs to know the $\mathrm{Var}\{m\}$. From the methodological point of view, it is the approach to the estimation of $\mathrm{Var}\{m\}$ that is of the most interest.

It can be shown that if accident occurrence follows the Poisson probability law for each intersection in the aforementioned imaginary population, the variance of accident counts in such a population is given by

$$\mathrm{Var}\{\text{accident counts}\} = \mathrm{Var}\{m\} + E\{m\} \qquad (2)$$

It follows that one can estimate $\mathrm{Var}\{m\}$ if estimates of $E\{m\}$ and $\mathrm{Var}\{\text{accident counts}\}$ are available. An estimate of $E\{m\}$ is the direct product of coefficient estimation. Thus, once we have estimates of b_0, b_1, . . . for a certain accident pattern, we can estimate $E\{m\}$. We explain below how to get an estimate of $\mathrm{Var}\{\text{accident counts}\}$.

Consider one site in our data with its specific traffic flows. The squared difference between the accident count on that site and the corresponding $E\{m\}$ is an estimate of $\mathrm{Var}\{\text{accident counts}\}$ for that specific combination of flows. When we plot these squared differences (often called residuals) against $E\{m\}$, we find a relationship of the following form:

$$\hat{\mathrm{Var}}\{\text{accident counts}\} = \hat{E}\{m\} + [\hat{E}\{m\}]^2/k \qquad (3)$$

The same relationship was found in the analysis of accidents at grade crossings (*14*) and was suggested by others (*19*) earlier. Thus, taken together, the residuals behave in a sufficiently regular fashion to allow the estimation of the parameter k. Once an estimate of k is available, the estimate sought is given by

$$\hat{\mathrm{Var}}\{m\} = [\hat{E}\{m\}]^2/\hat{k} \qquad (4)$$

The process of coefficient estimation that we used is iterative. We begin by assuming a value for k and proceed to estimate the vector of b-coefficients with GLIM. Now we can calculate residuals. These serve as input into a program to obtain the maximum likelihood estimate of k. The new k is fed back into GLIM to obtain new estimates of b-coefficients, and the cycle is repeated until closure.

The most frequent accidents are of Patterns 1, 2, 4, and 6, accounting for 83 percent. Thus it was possible to study these patterns in considerable detail. It was found that the number of accidents varies with the time of day. Therefore, the b-coefficients were estimated separately for the a.m. peak, the p.m. peak, the off-peak, and average daytime (daily) conditions. Although we provide models and coefficient estimates for all other patterns as well, because of the limited data, these should be regarded as unreliable. The final models selected for the 15 accident patterns and their coefficients are given in Table 3.

For clarity, the dimensions of $\hat{E}\{m\}$ (the estimate of $E\{m\}$) are accidents per hour (a/h) and the dimensions of the traffic flows are correspondingly vehicles per hour (v/h). The convenience of this consistency will become evident in the next section.

Not all accidents are "reportable" and not all reportable accidents are in fact reported. During the 1982–1984 period, accidents that exceeded \$400 CDN in damages as well as accidents involving injury (visible or complained of) were reportable. These criteria as well as the completeness of reporting vary in time and place. The equations specified by Table 3 allow the estimation of total accidents in Metropolitan Toronto during 1982–1984. To increase the transferability of our findings, we have also provided estimates of the injury ratio (injury accidents/total accidents) for each accident pattern in Table 4.

In Figure 4 we show the ratio of observed accident count per $\hat{E}\{m\}$ for the total number of accidents. The 145 intersections were arranged in ascending order according to $\hat{E}\{m\}$. It is of course pleasing to see the ratios symmetrically distributed around 1. The upper and lower bounds (1.72, 0.37) in Figure 4 contain 90 percent of the ratios. The circles represent the averages of 20 ratios. This is not a great achievement because the same data used earlier for coefficient estimation are now used again to describe model performance.

In Figure 5a and b we show the same ratios (observed/estimated) for accident Patterns 1 and 6. It can be seen clearly that the variability of the ratios decreases as $\hat{E}\{m\}$ increases. The circles represent the averages of 20 ratios. As expected, these averages fall neatly around 1.

Finally, the reader is reminded that the equations in Table 3 are for weekday, daytime accidents at signalized intersections in which two vehicles collided.

TABLE 3 ACCIDENT PREDICTION MODELS

PATTERN	MODEL FORM	TIME	COEFFICIENT ESTIMATES			$\hat{k}$
			b_0	b_1	b_2	
1	$\hat{E}(m) = b_0 \times F$	AM	0.1655×10^{-6}			2.98
		PM	0.2178×10^{-6}			2.73
		off	0.2164×10^{-6}			3.54
		daily	0.2052×10^{-6}			4.59
2	$\hat{E}(m) = b_0 \times F$	AM	0.0987×10^{-6}			1.49
		PM	0.0933×10^{-6}			0.94
		off	0.1080×10^{-6}			4.15
		daily	0.1014×10^{-6}			1.97
4	$\hat{E}(m) = b_0 \times F_2^{b_2}$	AM	19.020×10^{-6}		0.1536	2.65
		PM	1.4127×10^{-6}		0.6044	2.33
		off	9.7329×10^{-6}		0.3860	3.38
		daily	8.1296×10^{-6}		0.3662	5.51
6	$\hat{E}(m) = b_0 \times F_1 \times F_2^{b_2}$	AM	0.0283×10^{-6}		0.5163	1.39
		PM	0.0940×10^{-6}		0.3091	2.70
		off	0.0718×10^{-6}		0.4127	2.20
		daily	0.0418×10^{-6}		0.4634	2.10
3	$\hat{E}(m) = b_0 \times F_2^{b_2}$	daily	8.6129×10^{-9}		1.0682	
5	$\hat{E}(m) = b_0 \times F_1^{b_1} \times F_2^{b_2}$	daily	0.3449×10^{-6}	0.1363	0.6013	
7	$\hat{E}(m) = b_0 \times F_1^{b_1} \times F_2^{b_2}$	daily	0.2113×10^{-6}	0.3468	0.4051	
8	$\hat{E}(m) = b_0 \times F_2^{b_2}$	daily	2.6792×10^{-6}		0.2476	
9	$\hat{E}(m) = b_0 \times F^{b_1}$	daily	6.9815×10^{-9}	1.4892		use $\hat{k}$ = 1, 2
10	$\hat{E}(m) = b_0 \times F_2^{b_2}$	daily	5.590×10^{-12}		2.7862	
11	$\hat{E}(m) = b_0 \times F_1^{b_1} \times F_2^{b_2}$	daily	1.3012×10^{-9}	1.1432	0.4353	
12	$\hat{E}(m) = b_0 \times F_1^{b_1} \times F_2^{b_2}$	daily	0.0106×10^{-6}	0.6135	0.7858	
13	$\hat{E}(m) = b_0 \times F_1^{b_1} \times F_2^{b_2}$	daily	0.4846×10^{-6}	0.2769	0.4479	
14	$\hat{E}(m) = b_0 \times F_1^{b_1} \times F_2^{b_2}$	daily	1.7741×10^{-9}	1.1121	0.5467	
15	$\hat{E}(m) = b_0 \times F^{b_1}$	daily	0.5355×10^{-6}	0.4610		

TABLE 4 INJURY RATIO BY PATTERN

PATTERN	MEAN RATIO	95% CONFIDENCE LIMIT
1	0.3885	0.34 - 0.45
2.	0.1929	0.14 - 0.27
4.	0.3256	0.23 - 0.40
6.	0.3169	0.28 - 0.36
other	0.1685	0.13 - 0.22
TOTAL	0.2989	0.28 - 0.33

NUMERICAL EXAMPLE

To illustrate the use of the equations in Table 3, we examine one intersection in detail. The a.m. peak traffic flows for the intersection are shown in Figure 6. To find the $\hat{E}\{m\}$ for a particular accident pattern, the flows in Figure 6 are substituted in the equations of Table 3. For example, to obtain the $\hat{E}\{m_6\}_{am}$ we use

$$\hat{E}\{m_6\}_{am} = \sum_{j=1}^{4} (0.0283 \times 10^{-6} \times F_{1j} \times F_{2j}^{0.5163})$$
$$= 7.96 \times 10^{-4} \text{ (a/h)}$$

and from Equation 4

$$\hat{\mathrm{Var}}\{m_6\}_{am} = \sum_{j=1}^{4} [\hat{E}_j\{m_6\}_{am}]^2/\hat{k}_{6,am}$$
$$= 1.18 \times 10^{-7} \text{ (a/h)}^2$$

where

F_1 = 450 (v/h)	F_2 = 120 (v/h)	for j = 1
= 986 (v/h)	= 41 (v/h)	= 2
= 850 (v/h)	= 96 (v/h)	= 3
= 869 (v/h)	= 59 (v/h)	= 4

To find $\hat{E}\{m_6\}$ for 2 a.m. peak hours over 261 weekdays in 3 years, multiply 7.96×10^{-4} by $2 \times 3 \times 261 = 1{,}566$, and multiply the $\hat{\mathrm{Var}}\{m_6\}_{am}$ by $(1{,}566)^2$. The equations in Table 3 are for estimating total accidents; therefore, to determine the number of injury accidents we multiply the estimate by the injury ratio from Table 4. Hence, the

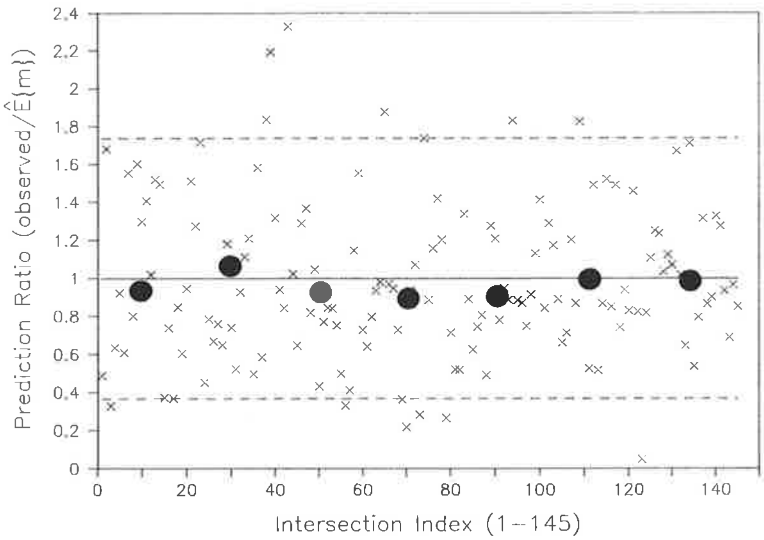

FIGURE 4 Prediction ratio for total number of accidents.

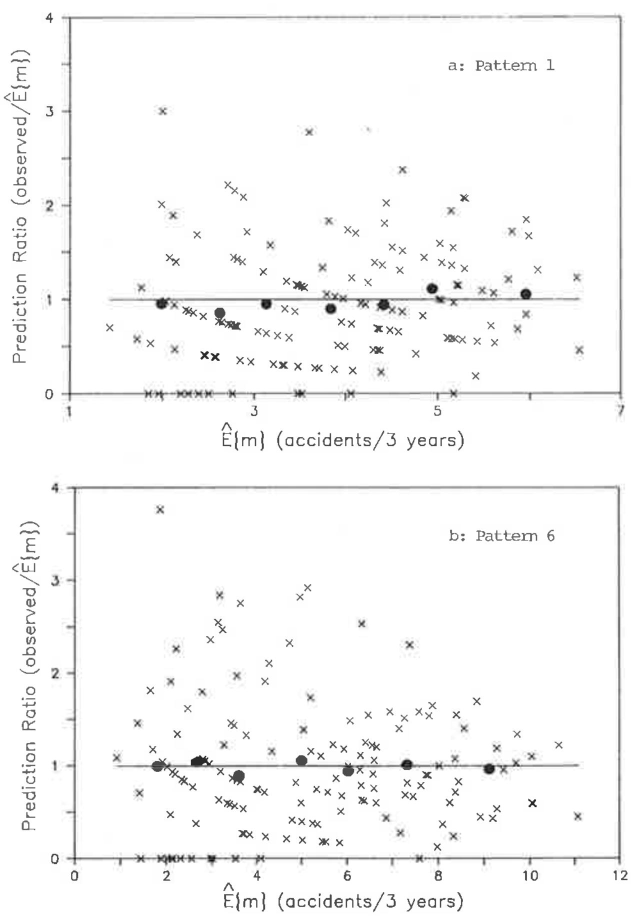

FIGURE 5 Prediction ratio for accident patterns 1 and 6.

estimated mean number of a.m.-peak, Pattern-6 injury accidents for 3 years is

$$\hat{E}\{m_{6,i}\}_{am} = 7.96 \times 10^{-4} \times 1{,}566 \times 0.3169$$
$$= 0.40 \text{ (injury accidents/3 years)}$$

In Table 5 we show the total observed counts, $\hat{E}\{m\}$'s, and $\hat{\text{Var}}\{m\}$'s for each accident pattern for 3 years at this

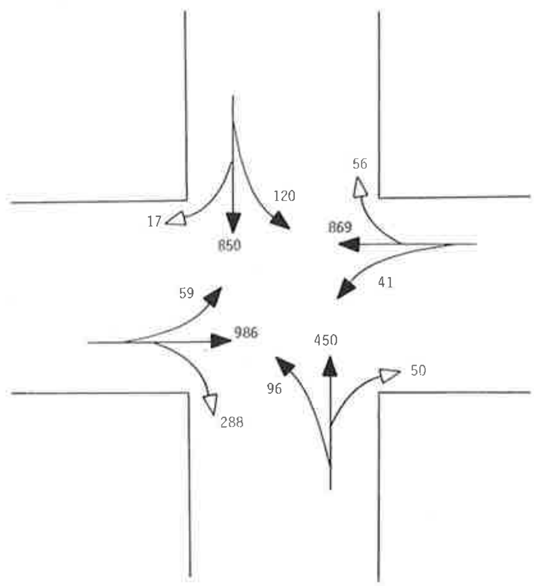

FIGURE 6 Hourly flows for a.m. peak.

intersection. The estimated means were calculated by using the average "daily" models in Table 3 multiplied by a factor of 7,047 (9 × 3 × 261). The reader is reminded that the flows are given in hourly numbers of vehicles. That is,

$$F_{day} = (2 * F_{am} + 2 * F_{pm} + 5 * F_{off})/9$$

The estimated mean and variance for total number of accidents on the intersection are shown on the last row of Table 5. It is the sum of the individual $\hat{E}\{m\}$'s and $\hat{\text{Var}}\{m\}$'s for each accident pattern.

ESTIMATION OF SAFETY WHEN ACCIDENT RECORD IS GIVEN

If someone provides us with the traffic flows on a signalized intersection and asks about its safety, we first calculate $\hat{E}\{m\}$ and $\hat{\text{Var}}\{m\}$ and then proceed to state: "Intersections

TABLE 5 CORRESPONDENCE BETWEEN OBSERVED AND ESTIMATED NUMBER OF ACCIDENTS

PATTERN	OBSERVED (1982-1984)	$\hat{E}\{m\}$ (3 years)	$\hat{VAR}\{m\}$ (3 years)
1	4	4.837	1.295
2	2	1.886	0.457
3	0	0.119	0.004
4	1	2.380	0.257
5	0	0.318	0.013
6	6	6.565	5.245
7	0	0.327	0.139
8	1	0.220	0.006
9	0	0.137	0.003
10	0	0.011	0.000
11	0	0.034	0.000
12	1	0.141	0.003
13	1	0.601	0.051
14	0	0.772	0.088
15	0	0.116	0.002
TOTAL	16	18.464	7.563

with such flows are estimated to have, on the average, $\hat{E}\{m\}$ accidents and the variability of m's among similar intersections with such flows is estimated to be $\hat{\mathrm{Var}}\{m\}$."

How does the statement change if, in addition to information about the traffic flows, we are also told that in the last n units of time, X accidents (of a certain pattern) have been recorded?

Before the question is answered, it is instructive to see that the conceptual framework specified earlier still continues to serve. In the section headed Estimation of Coefficients, we have imagined a subpopulation of all intersections such that all its members have the same pattern of traffic flow. This allowed us to make statements about the safety of intersections with a specific flow. The statement in the opening paragraph of this section is about one member of such a subpopulation. Now we are given more information. Not only do we know the flow of traffic for the intersection of interest, we also know its accident record. Thus, the earlier subpopulation (intersections with a certain traffic flow) can be further subdivided in our imagination into still more specific subsets; those with 0 accidents in n units of time, those with 1 accident in n units of time, and so on. The more information we have, the more specific is the subpopulation. For notational clarity we will use $E\{m \mid X,n\}$ and $\mathrm{Var}\{m \mid X,n\}$ to denote the mean and variance of the m's in the subpopulation of intersections with the given flow of traffic and a record of X accidents in the past n units of time.

It can be shown that if the distribution of m's in these imaginary populations can be described by a gamma probability density function (and accident occurrence on any entity follows the Poisson probability law), then

$$E\{m \mid X,n\} = (X + b)/(n + a) \quad (5)$$

$$\mathrm{Var}\{m \mid X,n\} = (X + b)/(n + a)^2 \quad (6)$$

in which,

$$a = E\{m\}/\mathrm{Var}\{m\} \quad (7)$$

$$b = (E\{m\})^2/\mathrm{Var}\{m\} \quad (8)$$

To illustrate this we continue with the earlier numerical example. We found that for the flows given in Figure 6, $\hat{E}\{m_6\} = 7.96 \times 10^{-4}$(a/h) and $\hat{\mathrm{Var}}\{m_6\} = 1.18 \times 10^{-7}$ (a/h)^2. What can be said about the safety of one such intersection on which two weekday, a.m.-peak, Pattern 6 accidents occurred in the last 3 years?

Using the estimates of $E\{m_6\}$ and $\mathrm{Var}\{m_6\}$ above, we find that $\hat{a} = 6{,}745.76$ and $\hat{b} = 5.37$. Therefore, from Equations 5 and 6 we get

$$\hat{E}\{m \mid 2{,}1566\} = 8.87 \times 10^{-4} \quad \text{(a/h)}$$

and

$$\mathrm{Var}\{m \mid 2{,}1566\} = 1.07 \times 10^{-7} \quad \text{(a/h)}^2$$

This can be made more tangible by visual representation. The gamma probability density function is given by

$$f(m) = a^b\, m^{b-1}\, e^{-am}/\Gamma(b) \quad \text{for } m > 0 \text{ and 0 otherwise} \quad (9)$$

Curve A in Figure 7 shows the $f(m_6)$ for intersections with flows given in Figure 6. The mean of m's for such intersections is 7.96×10^{-4} (a/h), and the variance is 1.18×10^{-7} (a/h)^2. By using $f(m)$ one can even say, for example, that 5 percent of intersections of this kind have m's in excess of 0.00146 (a/h).

Curve B in Figure 7 shows the $f(m_6 \mid X,n)$ for the same F_1's and F_2's when $X = 2$ accidents and $n = 1{,}566$ hr. To plot this curve we use Equation 9 again, except that a is replaced by $(n + a)$ and b is replaced by $(X + b)$. The mean of m's for intersections carrying these flows and with such an accident record is 8.87×10^{-4} (a/h), and the variance of m's is 1.07×10^{-7} (a/h)^2.

SUMMARY AND DISCUSSION

We have used accident data and information about intersection traffic flows to build models for the estimation of safety at signalized intersections.

During the course of model development we reached some useful insights. First, for a model to portray a relationship between cause (traffic flow) and effect (collisions between vehicles) we chose to relate accidents to the traffic flows to which the colliding vehicles belong. In our view, the logic of attempts to seek an aggregate relationship between accident frequency and some function of all flows (sum of entering flows, sum of products of flows, etc.) is unsatisfactory.

Second, it appears that the customary categorization of accidents by initial impact (rear end, angle, turning movement, sideswipe, etc.) is not very informative. One cannot assume, for example, that classification of an accident as an angle accident implies that the vehicles were traveling at right angles to each other or that most accidents involving left- or right-turning vehicles will be classified as turning accidents.

Third, a close examination of how the frequency of collisions depends on the traffic flows from which they arise reveals that preconceived notions are at times not borne out by empirical evidence. We find that the frequency of collisions between vehicles traveling in the same

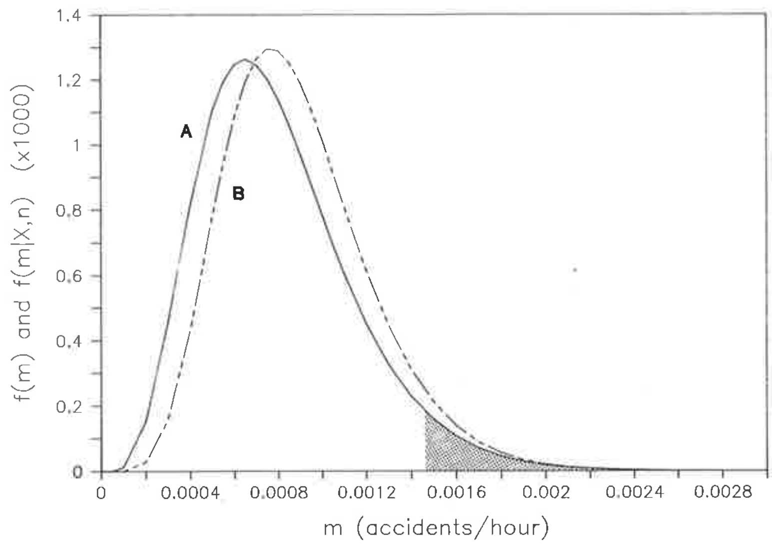

FIGURE 7 Gamma probability density function for *m*.

direction is proportional to the traffic flow in that direction, as one would expect. However, we also observe that the frequency of collisions between vehicles turning left and those proceeding straight through is proportional to the flow of through traffic but less than proportional to the flow of left-turning vehicles. In addition, the frequency of collisions between vehicles traveling at right angles to each other (~12 percent of all collisions) does not seem to depend at all on the larger of the two traffic flows; it increases with the smaller flow, but less than linearly.

Taken together, these observations lead to the conclusion that the popular assumption that intersection accidents are proportional to the sum of entering volumes is not in line with empirical evidence for several common accident types. Therefore, it cannot be true for the totality of intersection accidents. It follows that it is not correct to use intersection accident rates calculated on the basis of the sum of entering volumes to compare the safety of two different intersections nor is it proper to use the sum of entering volumes to correct for exposure in before-and-after studies.

On the basis of such exploratory analysis we have selected plausible model forms that would fit the data with parsimony of coefficients.

The following conceptual frame serves for both coefficient estimation and later for the model use. We think of a specific intersection as being a member of an imaginary population of intersections with similar features. Thus, for example, if we know the accident history and traffic flows for a certain signalized, four-legged intersection in metropolitan Toronto, we imagine a population of such intersections with the same flows and accident history. When making a statement about the safety, m, of a specific intersection, we say that intersections with similar features have on the average $E\{m\}$ accidents and that in this imaginary population, m's have a variance given by $\mathrm{Var}\{m\}$. Thus, the estimate of the population average $[E\{m\}]$ is our best estimate of the m for the specific intersection about which we speak and the estimate of $\mathrm{Var}\{m\}$ describes the uncertainty surrounding this estimate of m.

This sounds like academic hairsplitting and an unnecessary stretching of the imagination. However, the consequences of adopting this approach are immediate, practical, and far-reaching.

First, this approach implies that the run-of-the-mill least-squares regression software should not be used in the analysis of accident data. The error structure must be taken to be the compound Poisson kind. Second, the same approach allows one to use the residuals in order to obtain an estimate of $\mathrm{Var}\{m\}$. The knowledge of $\mathrm{Var}\{m\}$ is essential, as is shown; it allows us to combine data about accidents and information about traffic flows and use both for the estimation of safety. Knowledge of $\mathrm{Var}\{m\}$ is also the basis of all statements about the effect of any safety treatment or about what is "normal" or "unduly hazardous."

The results of our modeling and coefficient estimation efforts are summarized in Table 3. Here we give equations for each of 15 accident patterns to estimate $E\{m\}$ and $\mathrm{Var}\{m\}$ as a function of intersection flows. The use of these equations is demonstrated by a numerical example. We show that given the details of traffic flow at an intersection, we are now in a position to estimate the number of accidents expected to occur per unit of time in each of 15 accident patterns and also to describe the uncertainty surrounding this estimate.

Thus far the estimates have been "personalized" to account for the specific traffic flows at an intersection. The next step is to harness for estimation also the information about the accident history of a specific intersection. The conceptual framework set up earlier again stands in good stead. Simple equations allow the transition from the estimate of $E\{m\}$ and $\mathrm{Var}\{m\}$ based on traffic flows to the corresponding estimates, which are now based on the accident history as well. The underlying assumption is that the distribution of the m's in each imaginary population of intersections with similar features can be described by a gamma probability density function.

Relying on the same assumption, one can now obtain a complete description of the probability density of m's. We have shown this in another numerical example. Thus, we are now in a position to make several informative statements.

First, given the flows of a signalized intersection, we can say how many accidents and what type should be expected to occur on it. We can also specify the variance to be expected; in fact, the complete probability density function can be specified and plotted. On the basis of the probability density function we can decide what an unusually high m would be for accidents of a certain type on intersections with such flows.

Second, we can calculate how the number of accidents by type is expected to change when flows change. This allows us to separate changes in safety due to changes in traffic flow from changes due to other reasons. It also allows the correct cross-sectional comparison of the safety of several intersections instead of incorrect comparisons couched in terms of accidents per million entering vehicles.

Third, we can obtain an estimate of m for an intersection with a known accident history ($E\{m \mid X,n\}$) and compare it with what is expected at an average intersection with such flows ($E\{m\}$). It is on the basis of the magnitude $E\{m \mid X,n\}$ and its comparison with $E\{m\}$ (and also the two corresponding variances) that one should make decisions on what is deviant and where remedial action is warranted.

In particular, some warrants in the Manual on Uniform Traffic Control Devices (e.g., 2B-6 and 4C-8) refer to an "accident problem" and suggest that this is indicated by a certain "number of reported accidents of a type susceptible to correction" that occur in a 12-month period. In our view, such warrants should refer to estimates of $E\{m\}$ and $E\{m \mid X,n\}$ and their comparison and not to the count of reported accidents in a relatively short period of time.

ACKNOWLEDGMENT

Work on this subject has been made possible by the support of the National Science and Engineering Research Council of Canada, Transport Canada (Road Safety), and the Federal Highway Administration, U.S. Department of Transportation. The data for the analysis were provided by the Metropolitan Toronto Roads and Traffic Department.

REFERENCES

1. R. A. Chapman. The Concept of Exposure. *Accident Analysis and Prevention*, Vol. 5, 1972, pp. 95–110.
2. S. P. Satterthwaite. *A Survey of Research into Relationships Between Traffic Accidents and Traffic Volumes*. TRRL SP 692. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
3. J. D. Thorpe. Traffic Accident Measurements. *Journal of the Institution of Engineers*, Australia, Vol. 35, 1963.
4. W. L. Smith. Probability Study of High-Accident Locations in Kansas City, Missouri. *Traffic Engineering*, Vol. 40, No. 7, 1970, pp. 42–49.
5. G. Worsey. Predicting Urban Accident Rates from Road and Traffic Characteristics. *ITE Journal*, Dec. 1985.
6. S. M. Breunning and A. J. Bone. Interchange Accident Exposure. In *Highway Research Board Bulletin 240*, HRB, National Research Council, Washington, D.C., 1959, pp. 44–52.
7. V. H. Surti. Accident Exposure for At-Grade Intersections. *Traffic Engineering*, Vol. 36, No. 3, 1965, pp. 26–27, 53.
8. A. S. Hakkert and D. Mahalel. 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.
9. J. W. McDonald. Relationship Between Number of Accidents and Traffic Volumes at Divided Highway Intersections. In *Highway Research Board Bulletin 74*, HRB, National Research Council, Washington, D.C., 1966, pp. 7–17.
10. G. M. Webb. The Relationship Between Accidents and Traffic Volumes at Signalized Intersections. *Proc., Institute of Traffic Engineers*, 1955, pp. 149–160.
11. J. C. Tanner. Accidents at Rural Three Way Junctions. *Journal of the Institution of Highway Engineers*, Vol. 2, No. 11, 1953, pp. 56–67.
12. J. Coleman and G. R. Stewart. Investigation of Accident Data for Railroad-Highway Grade Crossings. In *Transportation Research Record 611*, TRB, National Research Council, Washington, D.C., 1976, pp. 60–67.
13. P. Mengert. *Rail-Highway Crossing Resource Allocation Procedure: User's Guide*, 2nd ed. Report FHWA-IP-86-11. FHWA, U.S. Department of Transportation, July 1986.
14. E. Hauer 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.
15. C. V. Zegeer, J. G. Mayes, and R. C. Deen. *Cost-Effectiveness of Lane and Shoulder Widening of Rural, Two-Lane Roads in Kentucky*. Bureau of Highways, Kentucky Department of Transportation, July 1979 (also published in *Transportation Research Record 806*, TRB, National Research Council, Washington, D.C., 1981).
16. E. Hauer, E. Pagitsas, and B.-Y. T. Shin. Estimation of Turning Flows from Automatic Counts. In *Transportation Research Record 795*, TRB, National Research Council, Washington, D.C., 1981, pp. 1–7.
17. R. J. Baker and J. A. Nelder. *The GLIM System: Release 3, Manual*. Rothhamsted Experimental Station, Harpenden, England, 1978.
18. G. Maycock and R. D. Hall. *Accidents at 4-Arm Roundabouts*. TRRL Laboratory Report 1120. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1984.
19. T. P. Hutchinson and A. J. Mayne. The Year-to-Year Variability in the Number of Road Accidents. *Traffic Engineering and Control*, Vol. 18, No. 9, Sept. 1977.

DISCUSSION

CARL N. MORRIS
Department of Mathematics and Statistical Sciences Center, University of Texas, Austin, Tex. 78712.

The idea by Hauer et al. of estimating traffic accident rates based on relevant flow rates certainly must be a welcome advance. But I will leave further assessment of that topic to others and concentrate on the statistical issues, about which I am most familiar.

The authors build on their weighty past contributions to accident analysis by again developing hierarchical and empirical Bayesian models for accident rates, this time further including regression model features. I will summarize their statistical model within a very useful paradigm, which I sometimes call the General Model for Statistics (*1*, *2*) because it contains the main branches of statistical thought. Without going through the general ideas, I immediately specialize that model to the Poisson-gamma model of Hauer et al., using their notation. The model has two completely equivalent renderings, a descriptive model that describes how we visualize distributions for the data and parameters, and an inferential model that rewrites the distributions in more convenient form for making inferences. Table 6 will help to clarify my questions.

The descriptive model is mathematically equivalent to the inferential model (Table 7), which specifies the marginal distribution for the data and the conditional distribution for the parameters.

Equation 14 in Table 7 declares the marginal distribution of X to be a negative binomial distribution, with usual parameters $b > 0$ and $p = n/(n + a)$. Expression 15 specifies the mean and variance of this negative binomial distribution inside the square brackets. Similarly, the square-bracket notation in Equation 18 indicates the mean and variance.

I now ask several questions that arise from consideration of the authors' methodology as structured in Equations 10–18.

TABLE 6 DESCRIPTIVE MODEL

Distribution for Observed Data:	$X \mid m \sim \text{Poisson}(mn)$	(10)
	X = accidents at a site,	
	m = true accident rate,	
	n = exposure (time).	
Distribution for Unobserved Parameters:	$m \mid b, a \sim \text{Gam}\left(b, \frac{1}{a}\right)$	(11)
	$Em = b/a = b_0 F_1^{b_1} F_2^{b_2} = \mu$,	(12)
	$\text{Var}(m) = b/a^2$.	(13)
	ϕ = unknown hyperparameters = (b, b_0, b_1, b_2) .	

TABLE 7 INFERENTIAL MODEL

Distribution for Observed Data:	$X \sim \text{NegBin}\left(b, p \equiv \frac{n}{n+a}\right)$	(14)
	$= \text{NegBin}\left[\frac{nb}{a}, \frac{n^2 b}{a^2}\right] = \text{NegBin}\left[\mu, \mu + \frac{\mu^2}{b}\right]$	(15)
	$\mu = b/a = b_0 F_1^{b_1} F_2^{b_2}$.	(16)
Distribution for Unobserved Parameters:	$m \mid X \sim \text{Gamma}\left(b + X, \frac{1}{n+a}\right)$	(17)
	$= \text{Gamma}\left[\frac{b+X}{n+a}, \frac{b+X}{(n+a)^2}\right]$	(18)

1. Parameters b and k: The authors have an expression similar to my Expression 15 for $\text{Var}(X) = \mu + \mu^2/b$, which is their Equation 3 of the section headed Estimation of Coefficients. But they use k for b. Isn't $k = b$? In Pattern 6 they estimate $k = 1.39$ in the morning (Table 3) but $b = 5.4$ (Figure 7) for the same data. Shouldn't these values agree?

2. The link function: The regression coefficients b_0, b_1, b_2 are estimated by the authors using GLIM, which requires, among other things, specification of the so-called "link function" $\eta = g(\mu)$. Here η, the linear form of the model, is

$$\eta = \log(b_0) + b_1 \log(F_1) + b_2 \log(F_2) \tag{19}$$

The mean structure therefore satisfies

$$\mu = EX = \frac{nb}{a} = b_0 F_1^{b_1} F_2^{b_2} = \exp(\eta) \tag{20}$$

and so, presumably, we are required to use

$$\eta = \log(\mu) \tag{21}$$

which is the "log link." Is this correct? Note that the "natural link" for the negative binomial family,

$$\eta = g(\mu) = \log\left(\frac{\mu}{b + \mu}\right) = \log\left(\frac{n}{n + a}\right) \tag{22}$$

would be fit by GLIM if no specification were made. But the natural link is not the link in Equation 21.

3. Variable exposures (n): Do the exposures n in Equation 10 vary from intersection to intersection? Assuming so, should that complicate the estimation of b and of the coefficients (b_0, b_1, b_2) through Equation 13?

4. Details about GLIM: Models like Equation 14 are quite difficult to fit. The authors seem to have found an ingenious way to estimate the "hyperparameters" (b, b_0, b_1, b_2) in the model. (Note that a can be defined in terms of the other paramters by using Equation 16.) More details, beyond the material surrounding the authors' Equation 4, are needed for an adequate understanding of this procedure. Most particularly, the subscripts that indicate the specifics of the intersection are not shown by Hauer et al., and so are also avoided in my rendering above. The clarification would reveal which values are intersection dependent and specify the assumed probabilistic independence of the data $\{X_i\}$ and of the parameters $\{m_i\}$.

5. Need for dependence of the mean on both F_1 and F_2: Although the points lie on a horizontal line in Figure 3a, the fitted curve must go through the origin (no flow implies no accidents). Thus, forms like $b_0 F_2 b_2$ used for Pattern 4 do not seem reasonable because they predict a substantial number of accidents when $F_1 = 0$.

6. Errors in variables: Substantial errors must be made in estimating (F_1, F_2). If so, this produced an errors-in-variables bias in estimation of b_1, b_2. Is there any way to determine how large this bias might be, or how accurately the flows are measured?

In conclusion, the paper offers a potentially very useful methodology that should further improve assessment of traffic hazards. The answers to the foregoing questions will help me to further understand the approach offered and its advantages.

ACKNOWLEDGMENT

This discussion was based on research supported by FHWA at the Texas Transportation Institute, Texas A&M University, College Station, Texas. Olga Pendleton is the principal investigator.

REFERENCES

1. C. N. Morris. Parametric Empirical Bayes Inference: Theory and Applications (with Discussion). *Journal of the American Statistical Association*, Vol. 78, No. 3, 1983, pp. 47–65.
2. C. N. Morris. Determining the Accuracy of Bayesian Empirical Bayes Estimators in the Familiar Exponential Families. In *Statistical Decision Theory and Related Topics*, IV, Vol. 1 (S. S. Gupta and J. O. Berger, eds.), Springer-Verlag, New York, 1988, pp. 251–263.

The opinions are those of the author alone, and have not been reviewed by FHWA.

AUTHORS' CLOSURE

It is most gratifying to have a prominent statistician review and comment on our work. We will attempt to respond to the questions raised.

PARAMETERS *b* AND *k*

There is no natural reason why in some population of intersections the *m*'s should fit a gamma distribution. However, if they do, the plot of squared residuals should be of the form specified by Equation 3. Having reassured ourselves that using the gamma assumption is in accord with our data, we estimated the value of k. Indeed, the estimate of k is also the estimate of b, as pointed out by the discussant.

As pointed out by the discussant also, for Pattern 6 (a.m. peak), $\hat{k} = 139$, whereas a few pages later, for the same pattern, we use $\hat{b} = 5.37$, an apparent discrepancy.

To see why the two values must differ and how, consider the characteristic function of a gamma distribution,

$$\phi(t) = 1/(1 - it/a)^b$$

If four independent random variables have the same gamma distribution, their sum will have a characteristic function,

$$\phi^*(t) = 1/(1 - it/a)^{4b}$$

Indeed, the 1.39 ($\hat{k}$) value in Table 3 refers to one gamma-distributed variable. The 5.37 ($\hat{b}$) used later pertains to a sum of four random variables (similar but not identical). Thus the usage does seem to be correct.

QUESTIONS ABOUT GLIM

The GLIM software has been used for this purpose by others before us [e.g., Baker and Nelder (*1*), Pickering et al. (*2*), and Hall (*3*)]. That is why we decided not to "attempt a comprehensive coverage of technical detail" in our paper (section on Estimation of Coefficients). Because the frame of reference has not been provided in the paper, it is difficult to give an intelligible discussion here.

In answer to Questions 3 and 4, accident count, duration of accident history, and traffic flow data are all intersection specific and are so represented for use in GLIM.

We have elected to measure μ in accidents per hour and F in vehicles per hour. Suppose now that we have at a certain intersection 3 years' worth of accident data for two morning peak hours during 250 weekdays a year. The n for this intersection would be $3 \times 2 \times 250$. For an intersection i,

$$n_i\mu_i = n_i \prod F_j^{b_j}$$
$$\eta_i = \ln(n_i) + \sum [b_j \times \ln(F_j)]$$

In GLIM, $\ln(n_i)$ is treated as an "offset."

NEED FOR F_1 AND F_2

The discussant notes that if $b_0F_2^{b_2}$ is used, one would predict $\mu > 0$ even when $F_1 = 0$. Recall that in the section of our paper on how accidents depend on contributory traffic flows, we called the larger flow F_1 and the lesser flow, F_2. Therefore, $F_1 = 0$ implies that $F_2 = 0$ and "no flow implies no accidents," as is required.

ERRORS IN VARIABLES

As noted by the discussant, F_1 and F_2, which are taken to apply to a 3-year period, are actually estimated from a 1-day volume count. Thus, although GLIM and other statistical software treat the independent variables as if they were measured accurately, in fact they are also subject to error. Some ramifications of the "error-in-variables" problem have been explored recently by Weed and Barros (*4*). Unfortunately, we still do not know what effect this might have on estimates of b_1 and b_2.

REFERENCES

1. R. J. Baker and J. A. Nelder. *The GLIM System: Release 3, Manual.* Rothhamsted Experimental Station, Harpenden, England, 1978.

2. D. Pickering, R. D. Hall, and M. Grimmer. *Accidents at Rural T-Junctions*. TRRL Research Report 65. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1986.
3. R. D. Hall. *Accidents at Four-Arm Single Carriageway Urban Traffic Signals*. TRRL Contractor Report 65. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1986.
4. R. M. Weed and R. T. Barros. Demonstration of Regression Analysis with Error in the Independent Variable. In *Transportation Research Record 1111*, TRB, National Research Council, Washington, D.C., 1987, pp. 48–54.