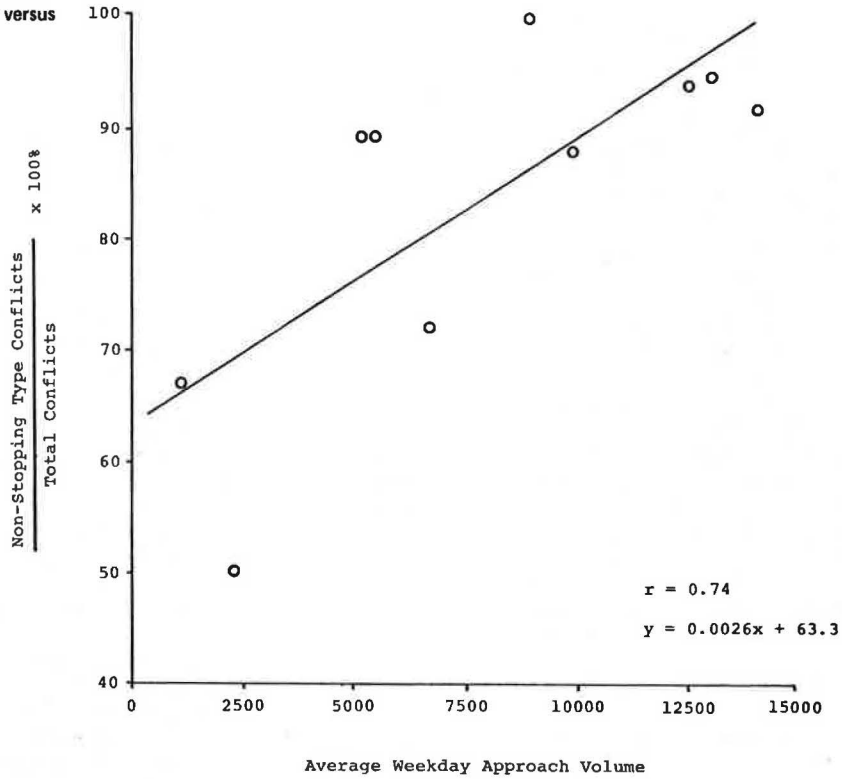
Figure 7. Percentage of nonstopping conflicts versus approach volume.



lemma zone and reinforce the usefulness of the study method.

Since no studies were made at a location where the dilemma zone was known not to be a contributor to the accident history, a comparison was not available. However, it appears that the traffic conflict study is of value in evaluating the timing of clearance intervals and stopping behavior in heavy traffic.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.

Design Considerations of Traffic Conflict Surveys

E. Hauer, Department of Civil Engineering, University of Toronto

The traffic conflicts technique is a device for measuring safety indirectly. It requires, at present, a field count of conflict occurrences, which gives the basis on which the rate at which conflicts occur is estimated. This report deals with the accuracy of such estimating and its dependence on the design of the field survey. Current practices in conflict-count duration are reviewed, and the relationship between count duration and estimation accuracy is examined. Using data from several sources, the daily variability of conflict counts is described. It is concluded that the expected conflict rate varies from day to day. Use of negative binomial distribution is suggested as appropriate for representing the distribution of sample means obtained from conflict studies. On this basis, confidence limits and probabilities of type I and type II errors in hypothesis testing are obtained and tabulated. Their use in study design is illustrated by numerical examples. The marginal increase in estimation accuracy diminishes rapidly as conflict-counting time increases. Thus, there is little to be gained by counting longer than 3 d. This establishes a practical limit to the accuracy with which expected daily conflict rates can be estimated.

The traffic conflicts technique is a device for indirectly measuring safety. Its early history may be traced (1, 2, 3, 4), and its recent applications have been described (5, 6, 7, 8, 9). There are also state-of-the-art surveys now available (9, 10, 11).

The traffic conflicts technique is applicable to a variety of situations. It can be used to assess changes in safety through before-and-after studies and by comparison with control sites; to investigate effectiveness of devices, layouts, design, and procedures; and to identify and diagnose hazards.

All such uses require a field study to observe and count the occurrences of conflicts and thus estimate their occurrence rate. The purpose of this paper is to

examine available empirical evidence in order to provide guidance for conducting conflict surveys.

The discussion will center on conflict-rate estimation accuracy, survey duration, and sample size selection. On the basis of information from several sources, the variability of daily conflict counts will be characterized and a model suggested for this counting distribution. Various aspects of the conflict study design will be illustrated by numerical examples. Tables and graphs will be supplied for use in survey design.

Present practice in conflict-count duration is summarized in Table 1.

Glennon and others (10) recently raised grave questions about the validity of present practice. They conclude: "For all three potential uses of conflict counts, existing relationships do not allow practical sample sizes." The conclusion, if true, would have far-reaching consequences. Not only does present practice in the conduct of conflict studies seem inadequate, there also seems to be little hope that the sample sizes required by Glennon and others would leave much interest in applying the traffic conflicts technique under any conditions. As this very important conclusion has been reached on the basis of limited empirical evidence, careful reexamination is in order.

EXPECTED CONFLICT RATE

The aim of a conflict survey is to obtain satisfactory estimates of the expected conflict rate. This is not a simple concept and requires delineation. In intuitive terms, the concept of expectation is closely associated with the notion of average, in the long run. We tend to believe that just as throws of a die will, in the long run, average 3.5, so would repeated conflict counts reveal a permanent characteristic of the site. The analogy, however, is incomplete. Unlike the die, the site changes its average property. There is little reason to assume that the expected conflict rate is the same during peak and off peak, Sundays and Mondays, winter and summer. It is essential, therefore, to specify which expected rate is subject to estimation.

We will proceed on the assumption that it is the expected weekday conflict rate that is of interest. That is, the average number of conflicts occurring per unit of time during a specified period of observation characterizing any weekday during a certain season of the year. There are two reasons for this choice. First, surveys are designed in terms of team days. Thus, it makes little sense to be concerned much about, say, hourly variations. Second, the traffic conflicts technique is used principally in comparisons between sites, devices, and treatments that are usually performed in a relatively short period of time. This eliminates the need to consider seasonal variation.

COUNTING, ESTIMATION, AND ACCURACY

In Figure 1, circles represent the number of conflicts counted on 19 consecutive weekdays during 7:00-10:00 a.m. and 3:00-6:00 p.m. at the intersection of St. Clair and Keele Streets in Toronto. The bars in the figure represent the estimation of expected conflict rate obtained by averaging the first 1 . . . 19 daily counts.

This simple graph illustrates all major features of the problem at hand.

First, the tangible evidence is the daily conflict count, which is subject to considerable variations. It is this random fluctuation that is the root source of difficulties in estimation.

Second, the fluctuating daily conflict counts are used to obtain an estimate of the expected conflict rate. In the present case, the simple average is used for estimation. Unlike the daily counts, the estimate of the expected rate is characterized by pronounced stability.

Third, the accuracy of the estimate increases with the number of daily counts. At first, every added count increases accuracy markedly. Beyond a certain point, not much accuracy is gained by counting more.

The qualitative observations made on the basis of all illustrative examples need to be quantified. This will be done, on the basis of data from several sources, with mathematical statistical methods. These allow measurement of variability in daily counts, characterization of accuracy of estimation, and so forth. However, how accurate the estimate should be cannot be determined by using statistics only. Standards of accuracy should depend on the circumstances of the survey and on the use to which its results will be put.

Figure 1. Daily conflict counts and their running averages.

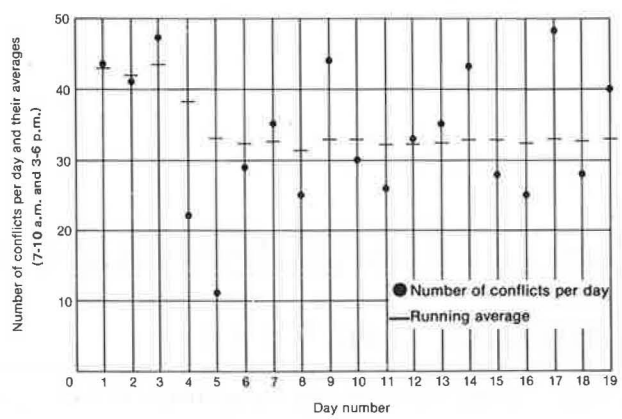


Table 1. Summary of conflict counts.

Source	Country	Intersections Surveyed	Duration	Counting Type	Team
Amundsen (9, 1974)	Norway	31 low volume, unsignalized	1 d, 7 h/d		Two- or three-person (one for conflicts)
Baker (3, 1972)	United States	392 before, 193 after modification	1 d	15-min samples on each approach with breaks between counts	Two-person
Cooper (12, 1973)	Canada	59 unsignalized, 51 urban, 5 rural	2 d, 14 h/d	Continuous	Five-person (three for conflicts)
Hydén (6, 1975)	Sweden	50 urban, varied control	2 d, 9.5 h/d	Continuous	One- or two-person
Malaterre and Muhlrad (8, 1976)	France	8 urban	1 d, 7:00 a.m.-12:00 midnight	Sample varies 13-57 percent of 1 h	Two-person
Perkins and Harris (1, 1963)	United States	30 signalized and unsignalized	3 d, 12 h/d	Alternate sampling per direction	
Spicer (4, 1973)	Great Britain	6 rural	1 d, 10 h/d	Continuous plus 16-mm time-lapse photography	Three- or four-person

This is unfortunate. One usually prefers to have firm and explicit standards for guidance. Conducting an elaborate decision analysis study in every case is impractical. There is a strong temptation, therefore, to simply adopt commonly used significance levels, because their use in medicine, sociology, quality control, psychology, and other fields lends them sufficient authority.

A word of caution is in order. It is wrong to blindly transfer levels of significance well suited for, say, concrete quality control directly to management of safety. The costs of conducting experiments are different; the implications of error are not the same. For most safety countermeasures, statistically conclusive evidence of effectiveness is unobtainable practically. Recognizing this as a fact of life, does one proceed and recommend that such things as driver licensing and vehicle inspecting be discontinued because neither can be shown effective at the 5 percent significance level? "The research community should consider carefully before doing so" (13). Uncritical use of high significance levels "can lead to the erroneous rejection of effective programs" (14). Rather, we need to adopt standards of accuracy that reflect both the benefit lost by not implementing measures that are likely but undemonstrably beneficial and the cost of implementing measures that are not effective.

DATA BASE

Investigating the distribution of daily conflict counts requires information on the number of conflicts at several sites over a relatively long period of time. Such information is not easy to come by. Daily counts are discarded or difficult to obtain, partly because conflicts are rarely counted for more than a couple of days and partly because usually only the average number is retained and archived.

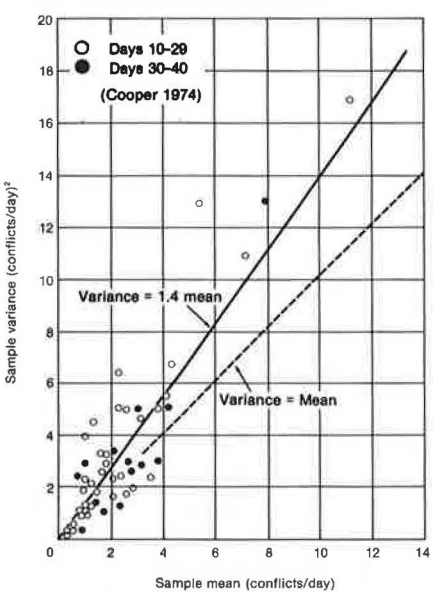
Fortunately, in the course of a study on the effectiveness of law enforcement (5), a good data base for the present purpose was generated. Conflicts were counted for 39 weekdays at each of seven urban intersections during 7:00-10:00 a.m. and 3:00-6:00 p.m.

The first 2 weeks (10 survey days) with normal police activity were followed by 4 weeks (19 survey days) with increased enforcement and another 2 weeks (10 survey days) with normal enforcement again. At each location seven conflict types were recorded. In the analysis below, the initial 10 d of the survey will not be used because "Initially, all observers tended to overcount drastically... stabilization of these counts did not proceed as quickly as anticipated and in most cases could not realistically have occurred by the beginning of increased enforcement" (5).

Thus, data from seven locations, seven conflict types, and two sequences of 19 and 10 d are used. It should be noted that the variations in counting will of necessity be overestimated because some will have been generated by the changes in police activity even within the phases that are analyzed separately.

To avoid reliance on one source of data, however extensive, two additional smaller sets of data were used. The first, conflict counts from 20 sites, was published by Hydén (6). In this case, 2 d of observation for each site are available. The second data set was collected by the Transport and Road Research Laboratory at four rural intersections; for three of them only 2 d of conflict counts exist, while for the remaining intersection 3 d of counts are available.

Figure 2. Sample mean and variance for homogeneous conflict classes.



VARIATIONS IN DAILY CONFLICT COUNTS

Variability is usually measured by sample variance. In Figure 2, sample variance is plotted against sample mean. Each point in Figure 2 represents the sample mean and variance of a homogeneous conflict class (cross traffic, rear end, etc.) at seven intersections for two phases of the study (5) with different enforcement levels. I have made the following observations.

First, the variance of the conflict count increases with the mean count, which is to be expected. Obviously, when the mean count is zero, so is the variance. Thus the origin is the starting point of the curve describing the relationship. In the range of mean conflict counts for which data are available, it is simplest to represent the relationship by a line through the origin. When conflicts for each homogeneous class are counted separately, the average variance-to-mean ratio is 1.4. When the sum of all conflict classes is of interest, the average variance-to-mean ratio is 2.2.

In their paper on evaluating the traffic conflicts technique (10), Glennon and others concluded that daily conflict counts are characterized by a constant variance of 530 (conflicts per day) (5) irrespective of the daily conflict rate. It is on the basis of this variance that they derive the number of days needed for conflict surveys. Our data do not confirm the assumption of constant variance; on the contrary, as in most known counting distributions, variance is found to increase with the mean. Nor can one find support for the high value of the variance used by Glennon and his coworkers.

The second observation to be made on the basis of Figure 2 pertains to the Poisson hypothesis. In the absence of empirical evidence to the contrary, it is usually assumed that rare events with a constant mean follow the Poisson distribution. Were this so, one would expect the dotted line labeled variance = mean to fit the data. It is apparent that the Poisson hypothesis does not hold. This may be so because the expected rate of conflict occurrence at intersections changes from day to day because of changes in weather, vehicle flow, pedestrian volumes, etc. In addition, some variability is introduced by the subjectivity of the observers identifying conflicts.

Third, there is no assurance that the same distribution describes the conflict-counting process irrespective

of the conflict type counted, the counting procedure used, the definition of the conflict event, or the specific circumstances of the site. Hydén's results (6), for example, suggest a smaller variability than the rest of the data. When specific information about count variability is not available, use of the average values obtained in this paper is recommended.

THE MODEL

To facilitate survey design and analysis in customary statistical terms, one has to adopt a model probability distribution that is simple to use, fits the data, and represents a process that bears a reasonable resemblance to our perception of reality.

The negative binomial distribution has been used for similar purposes in the past (15). It is founded on the assumption that the daily expected conflict rate follows a gamma distribution and the actual daily conflict counts a Poisson distribution with the aforementioned daily expected conflict rate as a mean.

By adopting the negative binomial distribution, it is shown that the distribution of the sample mean ($\overline{X}$) obtained from a count over j days is given by

$$P(X = n/j) - (-1)^n \binom{-\nu j}{n} p^n q^{\nu j}, \quad n = 1, 2, \ldots$$

with

$$E\{\overline{X}\} = \nu/a$$
$$VAR\{\overline{X}\} = E\{\overline{X}\}(1 + a)/(aj) \tag{1}$$

where

$\overline{X}$ = sum of j daily conflict counts divided by j,
$p = 1/(1 + a)$,
$q = 1 - p$,
ν = a (expected daily conflict rate), and
$a = \begin{cases} 2.5 \text{ for homogeneous conflict classes or} \\ 0.83 \text{ for the sum of several conflict classes.} \end{cases}$

What follows is a calculation that shows the origin of Equation 1 and provides useful information about characteristic functions, moments, and estimation.

Let X be the number of conflicts counted on a day during a given period of time and λ be the expected number of conflicts that day. If the count of that day obeys the Poisson distribution, then

$$P(X = k|\lambda) = (\lambda^k/k!)e^{-\lambda} \tag{2}$$

Regarding λ as a continuous random variable that assumes different values on different days, the distribution of X over many days is given by

$$P(X = k) = \int_0^\infty P(X = k|\lambda) f(\lambda) d\lambda \tag{3}$$

When λ obeys the gamma distribution

$$f(\lambda) = \begin{cases} [a^\nu/\Gamma(\nu)]\lambda^{\nu-1} e^{-a\lambda} & \text{for } \lambda > 0 \\ 0 & \text{for } \lambda \leq 0 \end{cases} \tag{4}$$

with $\nu > 0$ and $a > 0$.

Substituting Equation 4 into Equation 3 and integrating, we obtain

$$P(X = k) = (-1)^k \binom{-\nu}{k} p^k q^\nu$$
$$k = 0, 1, 2, \ldots \tag{5}$$

where

$p = 1/(1 + a)$,
$q = 1 - p$,
$\binom{-\nu}{k} = (-\nu)(-\nu - 1) \ldots (-\nu - k + 1)/k!$

The probability distribution defined by Equation 5 is the negative binomial distribution. Its characteristic function is given by

$$\phi_X(t) = q^\nu(1 - pe^{it})^{-\nu} \tag{6}$$

Ordinary moments of order r are given by

$$m_r = \sum_{i=0}^{r-1} (-1)^{r-1} \binom{r-1}{i} (p/q)^{r-i} (-\nu)_{r-i} \tag{7}$$

and the two central moments by

$$\mu_1 = \nu(p/q) = \nu/a \tag{8}$$

and

$$\mu_2 = \mu_1(1 + p/q) = \nu(1 + a)/a^2 \tag{9}$$

Thus, the negative binomial distribution is completely specified by the two parameters ν and a. To estimate their value from Equations 8 and 9

$$a = 1/[(\mu_2/\mu_1) - 1] \tag{10}$$

Also from Equation 8 and using Equation 10 we obtain

$$\nu = \mu_1/[(\mu_2/\mu_1) - 1] \tag{11}$$

Thus, when conflicts belong to a homogeneous class, $a \cong 1/(1.4 - 1) = 2.5$, and, when the sum of all conflict classes is of interest, $a \cong 1/(2.2 - 1) = 0.83$. The variance-to-mean ratios 1.4 and 2.2 were obtained earlier.

In the final account we are interested in the distribution of the average conflict count obtained from counting j days. Denote daily counts by $x_1, x_2, \ldots, x_i, \ldots x_j$ and

$$\overline{X} = (1/j) \sum_1^j X_i$$

As counts on successive days are statistically independent, using Equation 6,

$$\phi_X(t) = [\phi_X(t/j)]^j = q^{\nu j}(1 - pe^{it/j})^{-\nu j} \tag{12}$$

The characteristic function in Equation 12 belongs to the modified negative exponential distribution

$$P(\overline{X} = n/j) = (-1)^n \binom{-\nu j}{n} p^n q^{\nu j}$$
$$n = 0, 1, 2, \ldots$$

with

$$E\{\overline{X}\} = \mu_1$$
$$VAR\{\overline{X}\} = \mu_2/j \tag{13}$$

ACCURACY, ERRORS, AND DECISIONS

By using the data on daily variability in conflict counts and the suggested probability model, one can tabulate the probability distribution of conflict counts. (Tables of the probability distribution, confidence intervals, and type I and II errors for expected daily conflict rates between 2 and 20 are available from the Transport and Road Research Laboratory.) This is the basic information needed for statistical considerations of any kind.

The various uses of results obtained so far are best discussed within the framework of illustrative examples, or sample problems.

Example 1. Confidence Limits

From a 2-d survey of cross-traffic conflicts, an average daily count of 10.0 conflicts has been obtained. Find the 50 and 90 percent confidence intervals. The solution is from Figure 3, where the 50 and 90 percent confidence intervals are 4 and 11 conflicts per day.

For high confidence levels, the interval is large. Thus, one must either adopt modest standards of accuracy or invest in longer counts.

Example 2. Survey Design for Specified Confidence Limits

How many count days are needed to obtain a 90 percent confidence interval of 4 or less under the conditions of the previous example? The solutions, given in the table below for the 90 percent confidence limits, are

No. of Survey Days	90 Percent Confidence Limits
2	5.5-14.1
3	6.3-13.4
4	6.8-12.9
5	7.2-12.7
6	7.3-12.3

Figure 3. Confidence intervals for homogeneous conflict classes.

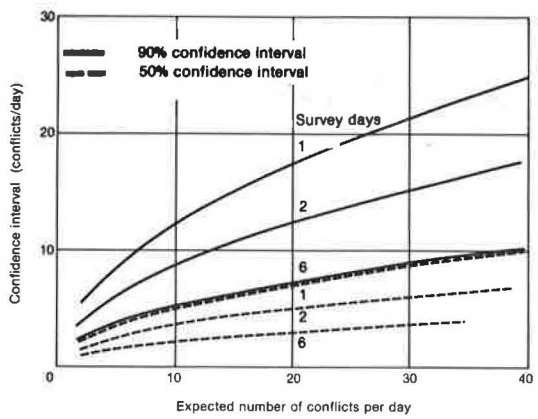
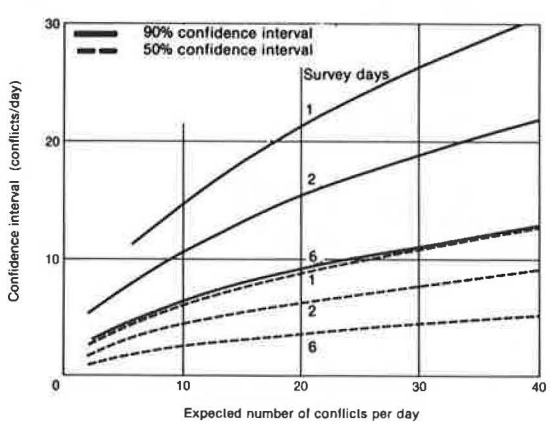


Figure 4. Confidence intervals for all conflict classes.



It appears that attainment of the specified accuracy is difficult. A count duration in excess of 10 d would be needed. As may be seen, the reduction in the confidence interval by counting longer diminishes. One should therefore weigh the increase in accuracy against the cost of prolonging the survey.

Example 3. Confidence Limits for Daily Conflict Rates Larger Than 20

Anticipating a daily conflict rate of 40 (sum of all conflict classes), how many days are needed for counting so that a 75 percent confidence interval is less than 10 conflicts per day? The answer is that, as the conflict rate exceed 20, tables of the normal probability distribution may be used. For the sum of all conflict classes, a = 0.83 (Equation 1). Also from Equation 1, VAR $\{\overline{X}\}$ = $40 \times (1 + 0.83)/(0.83j)$. To determine the 75 percent confidence limits given below, multiply $(\text{VAR}\{\overline{X}\})^{1/2}$ by the 1.15 obtained from tables of the normal probability distribution.

No. of Survey Days	$(\text{VAR}\{\overline{X}\})^{1/2}$	75 Percent Confidence Limits
1	9.4	±10.0
2	6.6	±7.6
3	5.4	±6.2
4	4.7	±5.4
5	4.2	±4.8
6	3.8	±4.4

For the prescribed accuracy, 5 d of counting are needed. For 50, 60, 70, 80, 90, and 95 percent confidence limits, 0.67, 0.84, 1.04, 1.28, 1.64, and 1.96 are the appropriate multipliers.

Determination of confidence limits for conflict counts is relatively easy; their interpretation is straightforward. This is therefore likely to be the best basis on which to make intuitive decisions about survey duration. After all, one can readily assess the gain in accuracy from prolonging the survey one more day and the cost of doing so.

Figures 3 and 4 give the size of confidence intervals in terms of survey duration and expected conflict rate. This should prove to be an effective guide in many circumstances.

In some situations, the probability distribution of count averages and the associated confidence intervals may not be deemed sufficient for conflict survey design and survey result analysis, notably when treatment effectiveness is the main concern. This is most common in so-called before-and-after studies.

In this context, one usually wishes to ascertain whether some treatment is effective in reducing the number of conflicts. A positive answer may lead to the modification of design standards, installation of new equipment, reconstruction of inferior site features, etc.

While the question of whether the treatment is effective is simple, the answer is not. Indeed a straightforward response is not forthcoming. It is for this reason that use of confidence limits may be preferred as the device least given to misinterpretation.

To provide an answer of a sort, the problem of treatment effectiveness needs first to be recast in terms of testing a statistical hypothesis. The outcome of such a test is not a statement that the hypothesis is true or false; it is merely a statement specifying the chance of error, should one decide on the basis of conflict studies to accept or reject the hypothesis.

It is customary (for little good reason) to test the hypothesis that the treatment is not effective. If so, con-

cluding on the basis of data that the treatment is effective when in fact it is useless constitutes a type I error. Conversely, maintaining on the basis of empirical evidence that the treatment has no effect, while in fact it is useful, is an error of type II.

These errors depend on the decision rule that determines acceptance or rejection of the hypothesis. While essential for formal decision analysis, this information is not understood intuitively and is therefore apt to be misinterpreted. It invites, therefore, use of arbitrary significance levels that are not derived from the reality of the situation at hand. As statistical hypothesis testing is deeply ingrained in present practice, the two types of error have been tabulated by the Transport and Road Research Laboratory. To guard against the possibility of misinterpretation, they are described in as clear a language as possible and their use is illustrated by numerical examples.

Example 4. Failure to Observe Improvement When Improvement Exists

Example 4 deals with the chance of failing to detect (through a reduction in the count of conflicts) a real decrease in the expected rate at which conflicts occur. The subsequent examples focus on the probability of observing a reduction in conflict counts when in fact there has been no change in the expected conflict rate.

Due to a successful treatment, the expected daily rate of cross-traffic conflicts has been reduced from 12 to 8. Thus, in the long run, the number of such conflicts is reduced by 33 percent. Determine the probability of observing no reduction in the average conflict count if 2 d of before counts are compared to 2 d of after counts. Such a result—failure to observe reduction in conflicts when improvement does exist—might lead one to conclude erroneously that the treatment had no beneficial effect.

The solution can be seen in Figure 5. For this example, the probability is 0.156. To aid interpretation, consider 100 sites at which treatment reduces the expected daily conflict rate from 12 to 8. Counting conflicts for 2 d before and after treatment, approximately 84 sites will show a reduction in the average number of conflicts; at the remaining 16 sites, no reduction will be observed.

It is natural to ask now how important it is to reduce the probability of failing to observe a reduction in the average conflict count. The answer depends on the specific objectives and circumstances of each survey. If the effect of the treatment is examined at 10 sites, a reduction in conflicts should be obtained on the average of

Figure 5. Probability of the before mean being equal to or smaller than the after mean when improvement exists: homogeneous conflict classes.

m_1	j	2	4	6	8	10	12	14	16	18	20
2.	1	0.595									
2.	2	0.562									
2.	3	0.547									
2.	4	0.543									
2.	5	0.538									
2.	6	0.534									
4.	1	0.301	0.562								
4.	2	0.192	0.543								
4.	3	0.132	0.534								
4.	4	0.094	0.529								
4.	5	0.068	0.526								
4.	6	0.056	0.526								
6.	1	0.138	0.344	0.549							
6.	2	0.050	0.252	0.534							
6.	3	0.020	0.196	0.528							
6.	4	0.008	0.156	0.524							
6.	5	0.003	0.126	0.521							
6.	6	0.001	0.103	0.519							
8.	1	0.059	0.192	0.367	0.543						
8.	2	0.011	0.094	0.287	0.529						
8.	3	0.002	0.050	0.236	0.524						
8.	4	0.000	0.027	0.197	0.520						
8.	5	0.000	0.015	0.168	0.518						
8.	6	0.000	0.009	0.144	0.516						
10.	1	0.024	0.100	0.227	0.382	0.538					
10.	2	0.002	0.029	0.129	0.310	0.526					
10.	3	0.000	0.010	0.078	0.263	0.521					
10.	4	0.000	0.003	0.045	0.227	0.518					
10.	5	0.000	0.001	0.031	0.198	0.516					
10.	6	0.000	0.000	0.020	0.174	0.515					
12.	1	0.010	0.050	0.132	0.252	0.393	0.534				
12.	2	0.000	0.008	0.050	0.156	0.327	0.524				
12.	3	0.000	0.001	0.020	0.103	0.283	0.519				
12.	4	0.000	0.000	0.009	0.070	0.249	0.516				
12.	5	0.000	0.000	0.004	0.049	0.222	0.515				
12.	6	0.000	0.000	0.002	0.034	0.198	0.513				
14.	1	0.004	0.023	0.073	0.158	0.272	0.402	0.531			
14.	2	0.000	0.002	0.017	0.069	0.179	0.340	0.522			
14.	3	0.000	0.000	0.004	0.033	0.125	0.299	0.518			
14.	4	0.000	0.000	0.001	0.016	0.090	0.267	0.515			
14.	5	0.000	0.000	0.000	0.008	0.065	0.240	0.513			
14.	6	0.000	0.000	0.000	0.004	0.048	0.218	0.512			
16.	1	0.001	0.011	0.039	0.094	0.179	0.287	0.408	0.529		
16.	2	0.000	0.000	0.005	0.027	0.087	0.197	0.351	0.520		
16.	3	0.000	0.000	0.001	0.009	0.046	0.144	0.312	0.516		
16.	4	0.000	0.000	0.000	0.003	0.023	0.107	0.281	0.514		
16.	5	0.000	0.000	0.000	0.001	0.014	0.081	0.256	0.512		
16.	6	0.000	0.000	0.000	0.000	0.008	0.062	0.234	0.511		
18.	1	0.001	0.005	0.020	0.054	0.112	0.196	0.300	0.414	0.528	
18.	2	0.000	0.000	0.001	0.010	0.038	0.103	0.213	0.360	0.519	
18.	3	0.000	0.000	0.000	0.002	0.014	0.058	0.160	0.322	0.515	
18.	4	0.000	0.000	0.000	0.000	0.005	0.034	0.123	0.293	0.513	
18.	5	0.000	0.000	0.000	0.000	0.002	0.020	0.096	0.269	0.512	
18.	6	0.000	0.000	0.000	0.000	0.001	0.012	0.075	0.248	0.511	
20.	1	0.000	0.002	0.010	0.029	0.068	0.129	0.211	0.310	0.418	0.526
20.	2	0.000	0.000	0.000	0.003	0.015	0.049	0.118	0.227	0.367	0.518
20.	3	0.000	0.000	0.000	0.000	0.004	0.020	0.071	0.174	0.331	0.515
20.	4	0.000	0.000	0.000	0.000	0.001	0.009	0.044	0.137	0.303	0.512
20.	5	0.000	0.000	0.000	0.000	0.000	0.004	0.027	0.109	0.280	0.511
20.	6	0.000	0.000	0.000	0.000	0.000	0.002	0.017	0.088	0.260	0.510

m_1 Expected daily conflict rate 'before'
m_2 Expected daily conflict rate 'after'
j Number of survey days

eight sites by counting conflicts for 2 d. It can happen, of course, that a reduction will be observed at five sites or less. The probability of this event is less than 2 percent (using the binomial distribution with p = 0.156). In this case, a 0.156 level of significance might offer sufficient insurance against the possibility that a reduction in conflicts will not be observed at most sites.

If, however, the survey is carried out at four sites only, the probability of obtaining a reduction in conflict count at two sites or less (when counting 2 d before and 2 d after) is 12 percent. By counting 3 d in this case, one can reduce the chance of not obtaining a reduction in conflicts at the majority of sites to 6 percent.

Example 5. Failure to Observe Improvement

The expected daily conflict rates (sum of all classes) are 35 before and 30 after. What is the probability that the average after count will not be less than the average before count when counting 1, 2, . . . , 6 days? This is a repetition of example 4 but with daily conflict rates above 20 to illustrate the use of the normal approximation.

The difference between the before and after sample means is approximately normally distributed with a mean that equals 35 minus 30, or 5 conflicts per day, and a variance that equals the sum of VAR $\{\bar{X}\}$ before and after, which is $[(35 + 30)(1.83)]/0.83j$ (see Equation 1 and example 3). The standard normal variable in this case is $5/\{[(65)(1.83)]/0.83j\}^{1/2}$. The probability that the difference between the counts will be negative is listed in the table below of the normal probability distribution.

No. of Survey Days (j)	Standard Normal Variable	Probability of No Reduction in Conflict Count
1	0.42	0.34
2	0.59	0.28
3	0.72	0.24
4	0.84	0.20
5	0.93	0.17
6	1.02	0.15

The value of the standard normal variable for any two expected daily conflict rates is given by the difference between the expected daily conflict rates, or

$$[\text{sum of expected conflict rates } (1 + a)/aj]^{1/2} \quad (14)$$

where a and j are defined by Equation 1.

Example 6. Critique of Results by Glennon and Others

Glennon and his group (10) use the probability of failure to observe a reduction in the count of conflicts as their criterion for survey duration determination. It is natural at this point, therefore, to discuss their results in the context of the following example.

Determine the duration of conflict survey needed to assure that the probability of obtaining a reduction in conflict count is 0.025, 0.005, . . . , 0.40, given that the expected daily conflict rate before treatment is 50 and that the reduction after treatment is 5, 10, . . . , 25 percent.

The solution, found by using the normal approximation as in example 5, is given in Table 2, where the results obtained by Glennon and others (10) are given in the first column. It is on this basis that they conclude that the required survey duration is not practical. The survey durations, according to our analysis, are approximately five times shorter for the same probability of failure to obtain reduction in counts (see column three). The discrepancy between the two results stems from the difference in the assumed variability of daily counts. While Glennon and others assume a constant variance of 530, our calculations are based on a variance-to-mean ratio of 2.2 as obtained from the data described above.

On the basis of Table 2, it appears that, as long as the difference between the before-and-after expected conflict rates is large, surveys of modest duration guard sufficiently against the probability of not observing a reduction in counts. When the difference between the expected conflict rates is small, even very long surveys do not offer protection against the chance that the after count will be larger than the before count.

With 4 d as a largest practical survey duration, the solid line in Table 2 is the boundary of combinations of expected conflict-rate reductions and probabilities of failure to observe a reduction in conflict counts. If the probability is to be as low as 0.05 (10), a conflict survey seems practical only when the before and after rates differ by more than 25 percent. If a 0.3 probability of not observing a reduction in conflict count is still acceptable, differences as low as 15 percent can be measured. It must be remembered that, if the effectiveness of the treatment is tested at, say, 20 sites, then, with a probability of 0.3 pertaining to each site, the chance of not obtaining a reduction at the majority of sites is less than 5 percent.

In summary, there is nothing sacred about a significance level of 0.05. In many circumstances, lower levels may be regarded as satisfactory. However, small differences between expected daily conflict rates cannot be measured even with very modest significance levels. This limitation is inherent in every estimation based on random variables with large variances.

In spite of this limitation, one needs to retain the proper perspective. At present, safety can be measured by using accident records or conflict counts. Accident records fluctuate no less than conflict counts. If a site has, on the average, 50 conflicts per day, then, for a 10 percent confidence level in a 25 percent reduction in the rates, one needs to count conflicts for 3 d. If the same site has 10 accidents per year, then, for a similar accuracy, accident records for 15 years (before and after treatment) need to be collected. Thus, the very real limitations on the conflict method of measuring safety discussed above are even more severe when accident data are used for the same purpose.

Table 2. Number of days to attain probabilities of failure to observe a reduction in conflicts (expected daily conflict rate before treatment -50).

Reduction in Expected Conflict Rate (%)	Probability of No Reduction in Conflict Count After Treatment						
	Glennon and Others (0.05)	0.025	0.05	0.10	0.20	0.30	0.40
25	26	5	4	3	2	1	1
20	41	8	6	5	3	2	1
15	72	14	12	9	6	4	2
10	162	32	27	21	14	8	4
5	650	132	110	86	56	35	17

This comparison is not quite fair, in that it disregards the question of proportionality between conflicts and accidents. It serves, however, to illustrate the main attraction of measuring safety via conflicts and the accelerated collection of information. The argument for indirect safety measurement, for instance by conflict studies, cannot be based on a claim of great estimation accuracy, which is ordinarily not attainable. It is based on the simple fact that in some circumstances indirect safety measurement is more accurate than any other method at our disposal.

So far we have been concerned about the possibility of not being able to show, through a reduction in the count of conflicts, a real reduction in the expected conflict rate. This may be thought of as the danger of not recommending a treatment that in fact is effective when implemented. The converse, of course, must be also of concern. It is quite possible (in fact it is very likely) to obtain a reduction in the count of conflicts in spite of there being no change in the expected conflict rate. This error is associated with the danger of implementing treatments, because of reduction in conflict counts, that are without effect. Such practice is wasteful of resources that could be spent more effectively elsewhere.

Example 7. Observing Improvement When No Improvement Exists

The expected daily rear-end conflict rate is 16 and remains so after treatment. What is the probability of obtaining a reduction in the average daily conflict count of 2, 4, 6, or more in a 1-d before-and-after survey?

The solution, from Figure 6, shows that the corresponding probabilities are 0.41, 0.30, and 0.20.

Note that even fairly large reductions are not unlikely, in spite of there being no real change in the rate at which conflicts occur. Out of 100 such sites to which an ineffective treatment has been applied, in a 1-d survey, some 30 will show a reduction of 4 or more in the average daily conflict rate. Conversely, if only treatments that reduce the daily conflict count in a 1-d survey by 4 or more are implemented, then 30 percent of all useless treatments under consideration will be implemented.

Also note that if the daily rate of the sum of all conflict classes is involved, Figure 7 and not Figure 6 should be used. As expected, when counting longer, the probability of obtaining a reduction exceeding a specified magnitude diminishes. This is illustrated graphically in Figures 6 and 7.

Example 8. Distribution of the Difference Between Count Averages

For large expected conflict rates, the normal approximation may be used. To illustrate, find the probability of the difference between the average counts (sum of all conflict classes) of 2-d surveys to exceed 10 if the expected daily conflict rate both before and after is 30.

The solution is that the difference is approximately normally distributed with a mean of 0 and variance that equals $[(2)(30)(1.83)]/0.83j$, or 66 (see Equation 1 and examples 3 and 5). The standard normal variate is $10/(66)^{1/2} = 1.23$. From tables of the normal distribution, the probability that the difference will exceed 10 is 0.11.

In general, the standard normal variate is given by

$$\text{Difference between conflict count averages} \div \{[2(\text{expected daily number of conflicts})(1 + a)]/(aj)\}^{1/2} \qquad (15)$$

where a and j are as defined in Equation 1.

SUMMARY AND DISCUSSION

Counts of the number of conflicts occurring per day are characterized by considerable variation. After observing available data, it appears that, for homogeneous classes, the variance-to-mean ratio is 1.4; for the sum of several (seven) conflict classes, the variance-to-mean ratio is 2.2. Accordingly, conflict counts do not follow a simple Poisson distribution. It is convenient to assume that the expected conflict rate varies from day to day. The negative binomial model is invoked to account for this variation.

Using this model in conjunction with the aforementioned empirically derived variance-to-mean ratios, the probability distribution has been tabulated, and confidence intervals have been derived and probabilities of types I and II errors computed. Their use is introduced through eight illustrative examples substantiated by graphs and tables.

Interest centers on questions of result accuracy and survey duration. Result accuracy is characterized through confidence limits and probabilities or error in testing hypotheses with respect to treatment effectiveness. It is suggested that, unless coupled with formal decision analysis, the framework of hypothesis testing is given to misinterpretation. Thus, for a judgmental decision to be made on survey design and standards of accuracy, confidence limits may be preferred.

As illustrated, accuracy increases with survey duration. However, the increase in accuracy per additional survey day diminishes rapidly. In general, there is not much to be gained by counting longer than 3 d. Thus,

Figure 6. Probability of reduction in average daily conflict count to equal or exceed r when no improvement exists (homogeneous classes).

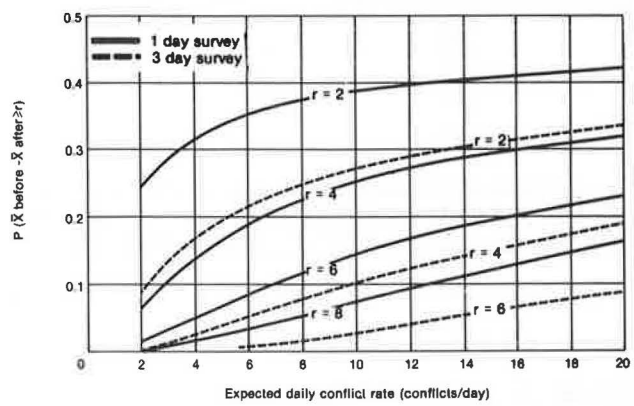
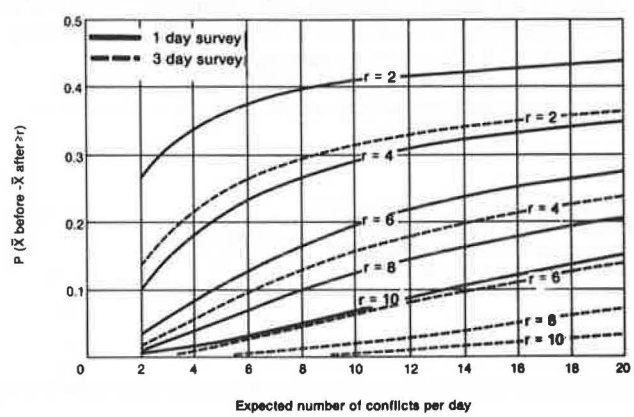


Figure 7. Probability of reduction in average daily conflict count to equal or exceed r when no improvement exists (all classes).



there is a practical limit to the accuracy with which the value of the expected daily conflict rate can be estimated. Existence of this practical limit on estimation accuracy must be considered when investigating treatment effectiveness. Conflict rate differences of 15 percent or less will prove difficult to demonstrate through conflict studies.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Research Council of Canada and by the Transport and Road Research Laboratory.

REFERENCES

1. S. R. Perkins and J. I. Harris. Traffic Conflict Characteristics—Accident Potential at Intersections. HRB, Highway Research Record 225, 1968, pp. 35-43.
2. S. R. Perkins. Traffic Conflicts Technique Procedures Manual. Research Laboratories, General Motors Corp., Warren, MI, GMR-896, Aug. 1969.
3. W. T. Baker. An Evaluation of the Traffic Conflicts Technique. HRB, Highway Research Record 384, 1972, pp. 1-8.
4. B. R. Spicer. A Study of Traffic Conflicts at Six Intersections. Transport and Road Research Laboratory, Crowthorne, England, TRRL Rept. 551, 1973.
5. P. J. Cooper. Effectiveness of Traffic Law Enforcement. Traffic Canada, Road Safety, Ottawa, 1974.
6. C. Hydén. Relation Between Conflicts and Traffic Accidents. Department of Traffic Planning and Engineering, Univ. of Lund, Sweden, mimeo, 1975.
7. V. A. Guttinger and J. H. Kraay. Development of a Conflict Observation Technique. Paper contributed to the Organization for Economic Cooperation and Development, Special Research Group on Pedestrian Safety, Netherlands Institute for Preventive Medicine (TNO), Nov. 1976.
8. G. Malaterra and N. Muhlrad. Intérêt et limite du concept de conflit. Jan. 1976 (ONSER).
9. F. H. Amundsen. Near Accidents. Development of a Recording Method and an Analysis of the Correlation With Traffic Accidents. Norges Teknisk-Naturvitenskapelige Forskningråd, Oslo, TOI-4413, May 1974.
10. J. C. Glennon, W. D. Glauz, M. C. Sharp, and B. A. Thorson. Critique of the Traffic Conflicts Technique. TRB, Transportation Research Record 630, 1977, pp. 32-38.
11. S. Oppe. Traffic Conflict Analysis, a Road Safety Research Technique. Institute for Road Safety Research (SWOV), Voorburg, The Netherlands, Dec. 1975.
12. P. J. Cooper. Predicting Intersection Accidents. Transport Canada Road Safety, Sept. 1973.
13. P. F. Waller. Challenging the Status Quo in Driver Licensing. Traffic Safety, Vol. 76, No. 10, Oct. 1976, pp. 20-21.
14. R. C. Peck. Towards a Dynamic System of Driver Improvement Program Evaluation. Human Factors, Vol. 18, No. 5, 1976, pp. 493-506.
15. M. Greenwood and G. U. Yule. An Enquiry Into the Nature of Frequency Distributions Representative of Multiple Happenings With Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents. Journal of the Statistical Society, Vol. 83, 1920, pp. 255-279.

Discussion

William D. Glauz, Midwest Research Institute, Kansas City, Missouri

A major concern with any data collection effort, and particularly when traffic safety data are involved, is how much to collect. Mr. Hauer has nicely reviewed prior work on this question as it relates to the traffic conflicts technique.

Prior assumptions regarding the statistical properties of traffic conflict counts have ranged from a presumed Poisson's distribution, which is intuitively enticing and the usual hypothesis for accident data, to the worrisome finding of Glennon and myself and others (10) that the variance may be so large as to make the technique impractical.

Mr. Hauer is to be congratulated for his careful study, which indicates that variances are more modest and within the range of practicality.

Probably the major, although not obvious, conclusion that can be reached through critical analysis of Mr. Hauer's and other studies is that the prime factor in determining the usefulness of the technique may be the experience and training of the observers. For example, the data we used were collected by several observers as a small segment of a much broader research project. There was a very large interobserver variance that can be attributed to minimal formal training. However, the Hauer paper itself furnishes a convincing argument for proper training, as I shall now demonstrate.

First of all, he states that the first 10 d of observations were not used in the analysis because of observer error. The total experiment for the Glennon data lasted less than 10 d. Thus, a certain amount of variance was arbitrarily eliminated. Even after deleting these 10 d, however, the effect of observer error is still present. This is clearly illustrated by the calculations in the table below on the effect of training on observer consistency.

No. of Days	Conflicts		
	Mean	Variance	Variance-to-Mean Ratio
4	38.7	87.0	2.25
5	33.2	192.6	5.80
6	32.5	163.0	5.02
7	32.9	140.7	4.28
8	31.9	129.8	4.07
9	33.2	129.8	3.91
10	32.9	117.9	3.58

The data in this table are derived from Hauer's Figure 1, which shows total daily conflict counts at one specific site (after the first 10 d of observations had been discarded). Note that, although the mean count became well stabilized after 5 d or more, the variance continued to decline substantially thereafter. Within 1 week's observation, the variance-to-mean ratio was 5.80; it dropped to 3.58 after 2 weeks.

An even more vivid demonstration is to compare this ratio by using all the data given by the author for this intersection (again, discarding the first 10 d) with that obtained by deleting yet another week. The 19-d ratio is 2.9; it is only 1.8 based on the last 14 d. Clearly, traffic conflict counts made by any but the most thoroughly trained observers are, statistically, a nonstationary process.

This illustration suggests that Mr. Hauer's final results, although very encouraging as a guide to conflict counting needs, are still too pessimistic. It also suggests that the overall average variance-to-mean ratio is

not as great as 2.2 but, if examined by the method of the illustration, is on the order of 1.5. In fact, I suspect that an approach to 1.0—the Poisson distribution—is theoretically possible.

He uses his statistical findings to carry out examples and concludes that "there is not much to be gained by counting longer than 3 d." The user must be cautioned that such a conclusion depends not only upon the variance-to-mean ratio (thus, observer training and experience) but also upon the definitions of traffic conflicts used (formal as well as operational interpretations) and upon the expected mean daily count.

Mr. Hauer notes that the expected conflict rate is dependent upon time of day, day of week, and season; I agree. Therefore, the user must be careful in interpreting his or her data in terms of the time frame in which they were collected. They need not be obtained as a daily count; a partial day may be more practical.

In the latter case it is important to note the particular time, if an after count is subsequently to be made. Also, for some locations, evening or weekend counts may be of more interest than weekday counts. Further, the time between before and after counts that are to be used to evaluate improvement effectiveness may well span many months in actual practice, so the possibility of seasonal variations must be considered.

Mr. Hauer states that the negative binomial distribution is appropriate for traffic conflict counts, a fact we had also noted (10). However, it is questionable whether λ, the expected number of conflicts on a day, should itself be a random variable. Moreover, based on our earlier comments, the value of a (or p) is really not yet known (nor is it likely to be a constant). Therefore, the tables and figures presented by Mr. Hauer are illustrative at best and should not be used indiscriminately.

In summary, the state of the art is that traffic conflicts and accidents are undoubtedly related (positively correlated), but the strength of the relationship is uncertain. The important need at this time is to improve and simplify the traffic conflict technique so that it can be applied easily yet uniformly by all interested agencies. In the author's words, we need to optimize the estimation of the "expected conflict rate."

Author's Closure

It is most gratifying to be congratulated for a careful study. Even more gratifying is Mr. Glauz's apparent concurrence with the main conclusions of the study. These squarely contradict their earlier findings (10), which appear to have been based on data collected by observers, as he says, with "minimal formal training" in the course of a "small segment of a much broader research project." Some points raised by Mr. Glauz warrant comment.

THE IMPORTANCE OF OBSERVER TRAINING

Mr. Glauz argues that observer training and experience are of prime importance, that observer reliability improves even after weeks of experience, and that with very experienced observers the variance-to-mean ratio may not be as large as 2.2. All these observations are most likely correct.

However, it is possibly somewhat hasty to draw conclusions about the manner in which the variance-to-mean ratio diminishes with observer training (and about its possible limiting value) on the basis of partial data obtained at a single intersection. (Complete data for all seven intersections are readily available.)

The issue cannot be resolved by using data presented for illustration and requires careful experimental design and analysis. Therefore, for the time being it seems prudent to follow the recommendation made in the paper that "when specific information about count variability is not available, use of the average values obtained in this paper is recommended." Admittedly, this may be conservative.

To keep the issue in proper perspective a glance at Figures 3 and 4 is useful. In Figure 4, the variance-to-mean ratio 2.2 has been used; for Figure 3 the ratio is 1.4. It is easy to see that the size of the corresponding confidence intervals is not all that different. Certainly, the difference is not large enough to drastically alter decisions about survey duration. It seems, therefore, that the results are not too sensitive to the variance-to-mean ratio used.

Finally, the tables available from the Transport and Road Research Laboratory can be requested specifically for the variance-to-mean ratio tables. Thus, corresponding tables will be generated.

SURVEY DURATION

Mr. Glauz takes exception to the conclusion that there is not much to be gained by counting longer than 3 d.

My entire paper is an attempt to provide the user with convenient tools for the statistical design of a conflict survey—equations, graphs, and tables. Also, the paper contains examples and illustrations demonstrating the use of such tools.

The user can, therefore, balance costs against accuracy and need not apply rules of thumb. However, the practitioner needs and appreciates useful generalizations that apply to most situations of interest.

Possibly the most useful generalization that emerges from the paper is that the increase in estimate accuracy after the third survey day is rather small. This is clear from inspection of Figures 3 and 4. The conclusion will not change for any reasonable conflict definition, expected daily conflict count, or variance-to-mean ratio. Therefore, Mr. Glauz's caution to the user is unnecessary.

PARAMETERS OF THE NEGATIVE BINOMIAL DISTRIBUTION

Mr. Glauz expresses some unspecified doubt about whether the expected number of daily conflicts (λ) should itself be a random variable. His basis for this doubt is difficult to understand. He seems to subscribe to the use of the negative binomial model, saying that its appropriateness has already been noted (10), in a note I could not find. The negative binomial model arises precisely when λ is a random variable. He also explicitly states a few lines earlier that λ "is dependent on the time of day, day of week," etc.

In summary, the first part of the discussion by Mr. Glauz raises the issue of observer training, experience, and conflict count variability. It is hoped that future research will shed more light on this problem. The second part of the discussion, unfortunately, is less constructive. It exhorts the user to be cautious, "to be careful in interpreting," to regard results as "illustrative," not to use them "indiscriminately," but without apparent reason.

Publication of this paper sponsored by Committee on Methodology for Evaluating Highway Improvements.