

## THE VALIDITY OF POLICE REPORTED ACCIDENT DATA†

DAVID SHINAR‡

Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer Sheva  
84120, Israel

JOHN R. TREAT

Highway Safety Research Institute, University of Michigan, Ann Arbor, MI 48109, U.S.A.

and

STEPHEN T. McDONALD

Institute for Research in Public Safety, School of Public and Environmental Affairs, Indiana University,  
Bloomington, IN 47405, U.S.A.

**Abstract**—Information theory and signal detection theory techniques were used to assess the validity of police reported traffic accident data. The validity criteria were the data and conclusions of multi-disciplinary accident investigation teams who investigated the same traffic accidents. The results indicated that the accident level variables reported by the police with least reliability were vertical road character, accident severity, and road surface composition. The most reliably reported data were those concerned with the accident location, date, and number of drivers, passengers, and vehicles. The informativeness of the police reports with respect to driver/vehicle characteristics was practically nil, with the exception of driver age, sex and vehicle model for which the police were correct most of the time (but not errorless). It was also found that police reports provided very little information regarding the presence of different human conditions and states, vehicle defects and environmental/road deficiencies. The sensitivity of police investigators to all accident causes was low. When causes were categorized into human direct, human indirect (conditions and states) vehicle, and environmental, police were the most reliable with respect to human direct causes and the least reliable with respect to environmental and human indirect causes. Implications for improvement and use of police data are discussed.

Police accident reports are probably the most ubiquitous source of traffic accident data analysis. While the primary purpose of such reports is to provide both summary descriptive statistics on accidents and information that might later be used for litigation purposes, very often data from these reports are taken at face value for inferential analysis, most notably in the area of traffic safety improvement programs. Thus, many safety programs are evaluated on the basis of whether or not they yield a reduction in accidents as reported by the police. In conducting such analyses, one must be aware that at least as far as rigorous scientific procedures are concerned, this approach is questionable. This is because in any scientific data gathering effort, the nature of the data collection process is often dependent on the objectives of the program. In the case of analyzing police data, however, the objectives of the researcher may be totally different from those of the policeman who is collecting the data at the scene. Thus, while police reports may be a useful source of information for the evaluation of various safety improvement programs, they are, as indicated by the research described below, often biased and/or incomplete.

Various studies have demonstrated that even at the level of reporting accident frequencies, sources other than police reports may be more complete. With the possible exception of alcohol related accidents, driver self-reports typically reveal more accidents than police reports [McGuire, 1973, 1976]. Accident reporting is also less complete the less severe the accident. Comparison of police records with hospital records in England showed that over 30% of injury-producing accidents were not reported to the police at all [Bull and Roberts, 1973;

†This work was conducted as part of a "Tri Level Study of the Causes of Traffic Accidents Modification, Task 3: Validity Assessment of Police Reported Accident Data" performed by the Indiana University Institute for Research in Public Safety for the National Highway Traffic Safety Administration, Contract DOT-HS-034-3-535. Thanks are due to Nicholas S. Tumbas, who was responsible for the mapping of accident data from IRPS to police categories and Ronald W. Drahos who reviewed the accident cases.

‡Requests for reprints should be sent to David Shinar.

Nicholl, 1981]. Similar results were obtained in Sweden [Thorson and Sande, 1969]. Probable reasons for the incompleteness of police accident data is fear of litigation by the drivers, reluctance to get involved in bureaucratic red tape, as well as the reluctance of police officers to file accident reports for accidents involving low levels of property damage only.

The same factors probably operate to influence the accuracy of details of each accident, once it is reported. Thus, when attempting to tease out factors such as location of accident, cause of accident and driver characteristics, errors in the data source are likely to lead to inappropriate conclusions concerning appropriate improvement programs.

Nonetheless, since police reports are so readily available, it is extremely important to gain more in-depth knowledge about the inaccuracy for purposes of accident data analysis. If the indisputable facts concerning each accident were known, then the police reports could be compared against them in order to assess the validity of police reports for different accident-related data.

It would be desirable to have a validity criterion for accident evaluation that would be independent of human judgment. Short of this goal, however, the most we can strive for is the careful analysis of accidents that would involve the most sophisticated techniques of information-gathering available, combined with expert skills of the accident investigators. An approximation to this more realistic goal is provided by the "in-depth" accident analysis conducted by Indiana University's Institute for Research in Public Safety (IRPS). The accident collection procedures involved in this data collection effort have been detailed elsewhere [Treat *et al.*, 1977; Treat and Shinar, 1976; Shinar, 1978], and will be described here only briefly. In that data collection effort, a relatively representative sample of 420 motor vehicle accidents were analyzed in depth by multi-disciplinary accident investigation (MDAI) teams and 2258 accidents were analyzed by on-site technicians. The analysis involved both a detailed description of the driver-vehicular-environmental context within which the accident took place, as well as an analysis of the causal factors involved in each accident.

In particular, the in-depth analysis was based on the data obtained by the on-site investigators as well as on an independent evaluation of the vehicle (by an automotive engineer), the driver (by a psychologist) and the roadway environment (by an accident reconstruction specialist).

The primary purpose of this study was to evaluate the validity of the police reported data relative to that of IRPS, with respect to four types of variables: accident characteristics, vehicle characteristics, driver characteristics, and accident causes<sup>†</sup>. In order to use IRPS data and conclusions as a criterion for the accuracy of the police report, the validity of the IRPS data must first be demonstrated.

### *The issue of relative validity*

The assumption that the validity of the police reported data and conclusions can be tested by comparing them to IRPS data and conclusions must be qualified. Accident investigation, by its nature, involves a *post hoc* analysis of events (i.e. the accident has already occurred). And, like any *post hoc* accident investigation effort, the IRPS investigation is likely to be to some extent erroneous. In the absence of an independent "true" criterion, the validity of the IRPS data and conclusions are very difficult to assess. Nonetheless, a strong case can be made for the higher accuracy (and, therefore, relative validity) of IRPS-collected data and conclusions over police-collected data and report for the following reasons:

(1) The time delay between the occurrence of the accident and the initiation of the IRPS investigation was as short as that of the police, but the professional time spent in investigating each accident, by both the on-site and the in-depth teams, was much longer than that available to the police (approximately 40 professional man hours).

(2) The IRPS teams consisted of professional accident investigators, each with his/her own area of expertise in either the vehicular, environmental, or human area. Accident analyses performed by IRPS was based on careful measurements taken by the automotive engineer and

<sup>†</sup>An additional objective of the project was to compare the relative accuracy of the three police agencies working in the area. Municipal, County and State. Some differences among the three agencies were found, the State police being slightly more accurate overall. These analyses are detailed in the original report by Shinar and Treat [1977].

the environmental specialist, and extensive testing and interviews conducted by the human factors specialist.

(3) IRPS reports were based on composite opinions of four or more experts, whereas police reports were often based on the opinions of a single investigating officer who did not have available to him/her any quality control or feedback mechanisms.

(4) The IRPS investigators disassociated themselves from the legal system, and the information provided by the drivers was perceived by them as confidential. This was especially helpful in providing cause-related data that might incriminate one or both of the drivers.

(5) The IRPS data was subjected to quality control checks, both within the team, by the project director, and NHTSA personnel as well as by statistical consistency tests (see Treat *et al.*, 1977, Vol. 1). Furthermore, in the case of causal assessment, multiple sources of evidence were considered in attributing causes.

In summary, the amount of time, expertise and money spent in investigating each accident by the multidisciplinary team; the quick response feature of the on-site investigators; the confidentiality of the driver reports; and the converging quality control operations; all of these make the IRPS accident investigation effort probably the most intense and accurate to be conducted in recent years on a representative sample of traffic accidents. And for these reasons it is probably as good a criterion for the validity of police reports as any other presently available.

## METHOD

### *Technical approach*

In the present study a random sample of 124 accidents, involving 207 drivers, was drawn from the 420 accidents investigated in depth. All the accidents occurred in Monroe Co., Indiana between May 1971 and June 1975. For each of these drivers and accidents, the police records were obtained, and comparisons between the police-reported data and conclusions and IRPS-reported data and conclusions were made. Each of these accidents was investigated by one—and only one—of the three different police agencies operating in the area: State (15 accidents involving 21 drivers), County (36 accidents involving 52 drivers), and Municipal (73 accidents involving 134 drivers). Although one objective of this study was to evaluate the relative accuracy of the three agencies, these analyses and results will not be discussed in detail here. They are fully documented in Shinar and Treat (1977).

The sample size was limited to 124 accidents because of the expense involved in the manual comparisons between the police record and the IRPS files. The mapping procedures (described below) always required that all the information be read by an experienced accident investigator (familiar with both IRPS terminology and the police terminology).

### *Mapping procedure for accident causation data*

Since IRPS data forms were designed with the specific objective of providing accident cause data as detailed as possible, the amount of information available for each case was much greater in the IRPS files than in the police files. This necessitated the derivation of mapping strategies from one file to the other. Because the police file was the one with the fewer categories, the mapping for most variables was from one of several IRPS categories to one police category. Furthermore, depending on the situation, a given IRPS category could be mapped into different police categories. To illustrate, depending on the specific circumstances any one of the following IRPS accident causes could be mapped into the police category "passed stop sign": "Inattention (to stop sign)", "Delay in perception of road signs", "Improper driving technique (braking later than should have or stopping too far out in road)". This required a case-by-case reanalysis of all the accidents to assign the relevant police accident cause labels to the IRPS team conclusions concerning the accident causes. This reanalysis imposed the major restriction on the sample size.

### *Presence versus cause*

A distinction was made between assessments of presence of a potential accident cause (e.g. an assessment that a driver is under the influence of alcohol) and assessments of actual cause (e.g. an assessment that the driver's being under the influence made a difference in whether or

not the accident occurred). While both assessments may involve large elements of judgment, the latter clearly requires an additional level of influence of that factor with additional opportunity for error. This distinction between presence items and causal items is useful since it relates to two types of information, the first being purely associative information, the kind that could be associated or not associated with the accident involvement, while the latter are the kind that definitely could be described as "causes" of accidents. Furthermore, it allows tabulating the less judgmental presence information for associative comparisons, while still making the alternative "clinical assessment" information readily available.

ANALYTICAL APPROACH

Two different statistical procedures were used to evaluate the accuracy of the police-reported data. The first procedure involved the derivation of an information metric which provides a way of describing the proportion of information that the police can transmit on each one of those items, given the amount of uncertainty that exists beforehand. This metric is based on the information theory model of communication [Shannon and Weaver, 1949]. The second technique involved the use of decision theory statistics used in the mathematical procedure developed in signal detection theory (SDT) by Swets, Tanner and Birdsall [see Green and Swets, 1966]. This technique separates the "sensitivity" of the police in terms of their ability to detect information from their response bias in terms of their readiness to state an accident cause in their report.

The use of these rather uncommon statistics is due to the nature of the data involved. For most presence variables and all causal variables, the data were at the nominal scale level, most often dichotomous, and usually not evenly distributed. The advantages of the information transmission metric and the SDT statistics can best be illustrated with an actual examination of the data on two causal factors: "Failure to Yield Right of Way" and "Fatigue".

*The shortcomings of some standard measures of association*

Table 1 contains the two frequency matrices that served as the data base for evaluating the reliability of the police data on two accident causes: "Failure to Yield Right of Way" and "Fatigue".

Table 1. A comparison of different measures of association on causal assessment and the extent to which they are affected by marginal distributions (numbers in parentheses are percentages)

FAILURE TO YIELD RIGHT-OF-WAY						
IRPS						
	NO	YES	TOTAL	STATISTICS		
POLICE	NO	166 (80.2)	1 (.5)	167 (80.7)	% Agreement:	95.7
	YES	8 (3.9)	32 (15.5)	40 (19.4)	Phi Coefficient:	.86
	TOTAL	174 (84.1)	33 (15.9)	207 (100)	Contingency Coefficient:	.65
				Uncertainty:	.71	

FATIGUE						
IRPS						
	NO	YES	TOTAL	STATISTICS		
POLICE	NO	204 (98.6)	2 (1.0)	206 (99.6)	% Agreement:	98.6
	YES	1 (.5)	0 (0)	1 (.5)	Phi Coefficient:	.007
	TOTAL	205 (99.0)	2 (1.0)	207 (100)	Contingency Coefficient:	.007
				Uncertainty:	.001	

Note that failure to yield right-of-way was identified as a cause by IRPS approximately 16% of the time, while the police identified it as a cause approximately 19% of the time. In the second example, fatigue was identified as a cause twice (or 1% of the time), while the police identified it as a cause only once (or 0.5% of the time). Thus, a significant difference between the two causes is that the marginal distributions for both the police and IRPS assessments are extremely different. Accepting IRPS as reflecting the best approximation to the true state of events, it appears that failure to yield right-of-way was an accident cause 16% of the time, whereas fatigue was an accident cause only 1% of the time. Now, to derive one commonly-used measure of association—percent agreement between the IRPS investigators and the police—we simply have to add the percent of times that both the investigators either agreed that these were the accident causes or agreed that these causes were not involved. In the case of failure to yield right-of-way, we obtain an agreement of approximately 96%, whereas in the case of fatigue, we obtain an agreement of approximately 99%. Thus, the high percent of agreement obtained for fatigue is mostly due to the fact that both IRPS and the police failed to cite this factor whether it existed or not. In fact, if the police were never to identify the factor of fatigue, we would still obtain the same 98.6% agreement!

In general, in the total absence of any police citations, the lower the probability of occurrence of a cause (or the more specific it is), the higher the expected percent agreement. Therefore, it can be easily concluded that percent agreement is not a very useful statistic in all cases, since the marginal probability of a cause being identified or not being identified is not the same for all causes.

Some traditional statistics, such as the Phi ( $\phi$ ) coefficient and the contingency ( $C$ ) coefficient, do account for the variation in marginal frequencies. Accordingly, in both cases, the Phi coefficient and contingency coefficient are higher for the failure to yield right-of-way than for the fatigue factor. The contingency coefficient, derived from the  $\chi^2$  distribution, is applicable to nominal variables of two or more categories. It is defined as

$$C = \sqrt{\chi^2/(\chi^2 + N)},$$

where  $\chi^2$  is Pearson's chi-square statistic and  $N$  is the sample size.

A serious shortcoming of the contingency coefficient is that its upper limit depends on the number of categories, making comparisons across variables with a different number of categories difficult to interpret. Also, the contingency coefficient is inappropriate when some of the cell-values approach zero.

The Phi coefficient is defined as

$$\phi = \sqrt{\chi^2/N},$$

where  $\chi^2$  and  $N$  are defined above.

The values obtained for the Phi coefficient are identical to the Pearson  $r$  correlations for a two-category variable. Its major shortcoming is that it is applicable to dichotomous variables only, and when the number of categories of some of the variables compared is greater than two it cannot be calculated for all the variables.

#### *The information metric*

The basic approach to this analysis assumes that some uncertainty exists with respect to the occurrence of accident-related variables, and that the purpose of the police investigation is to reduce such uncertainty. We further assume that the IRPS data reflect the true frequency of occurrence of various events, and then examine the degree to which knowledge of the police report reduces the uncertainty. Since the amount of pre-existing uncertainty depends on the prior probability of occurrence of the various events, we can adjust our measure to reflect the proportion of uncertainty reduction. The quantitative measure used for this purpose is the Uncertainty coefficient ( $U_c$ ), which is defined by Nie *et al.* [1975, p. 226] as follows:

$$\text{Uncertainty coefficient } (U_c) = \frac{U(Y) - U(Y|X)}{U(Y)},$$

where  $U(Y)$  is the uncertainty associated with the in-depth team citing by IRPS and  $U(Y|X)$  is the uncertainty associated with guessing the IRPS assessment given information obtained by the police.

The calculation of the uncertainty measures is based on the mathematical information theory where uncertainty of variable  $X$ ,  $U(X) = -\sum p(X_i) \log_2 p(X_i)$  and  $p(X_i)$  is the probability of category  $i$  in  $X$ .

The  $U_c$  can vary from 0.0 (where the association is random) to 1.0, where the correlation between the two data sources is perfect. This measure is preferable to the contingency coefficient since the expected value for some of the cells is small or zero. The advantage of the information metric over the Phi correlation is that the information metric can be used for any number of categories and is not limited to the case of the  $2 \times 2$  matrix. Thus, of the three measures above ( $C$ ,  $\phi$  and  $\{U_c\}$ ), it is the only measure that can provide useful information, based on a single mathematical formula, for all the IRPS–Police comparisons.

To illustrate the calculation of  $U_c$ , in the case of assessing the police performance in noting Failure to Yield Right of Way, from the cell frequencies presented in Table 1:

$$\begin{aligned}
 U(Y) &= -\sum p(Y_i) \log_2 p(Y_i) = -(0.423 + 0.211) = -0.634 \\
 U(Y|X) &= -\sum p(Y_i, X_j) \log p(Y_i|X_j) \\
 &= -\left[ \frac{166}{207} \log \frac{166}{167} + \frac{32}{207} \log \frac{32}{40} + \frac{8}{207} \log \frac{8}{40} + \frac{1}{207} \log \frac{1}{167} \right] = -0.184.
 \end{aligned}$$

Inserting the obtained values for  $U(Y)$  and  $U(Y|X)$  in the equation for  $U_c$ , we get  $U_c = 0.714$ . Thus, the police's performance on this variable reduces our uncertainty about its involvement by 71% (where 100% reduction in uncertainty indicates perfect performance).

For the dichotomous accident causes the  $U_c$  correlates highly with both Phi ( $r = 0.94$ ) and the Signal Detection Theory statistic  $d'$  ( $r = 0.98$ ) (discussed below) and so Phi will be included in the accident cause tables, to provide a better "feel" of the IRPS/Police correspondence for those familiar with the Pearson  $r$  correlations.

#### *Signal detection theory (SDT) statistics*

A decision theory approach to evaluating the police assessment was used in which the IRPS assessment is taken to reflect the true state of the world. A methodology typically associated with Signal Detection Theory (SDT) was then used to determine the  $\alpha$  and  $\beta$  error levels of the police, and indices based on these error rates were derived. The SDT approach will be briefly described below; a more extensive treatment of the SDT analytical approach and rationale is available in Green and Swets [1966].

According to SDT, when an event (signal) occurs in the outside world, it gives rise to a change in the person exposed to it. Whether this change in the situation will be detected or not is, however, a function of two different phenomena: (a) the extent to which the signal is stronger than the general "noise" in the system and (b) the bias or risk-taking level that the person has with respect to stating the signal is there when in fact it is not (a type  $I$  error). Each of the above phenomena can be quantified, as will be illustrated below.

For the purpose of this illustration, let us examine the police performance in correctly identifying failure to yield right-of-way (FYRW). We can then depict the factor detection process, as shown in Fig. 1. In this figure, the left curve is the frequency distribution ( $f$ ) of the "strength of evidence" or intensity of FYRW cues, when it is *not* a causal factor, i.e. the intensity of those cues when generated by "noise" alone. The right curve is the frequency distribution of the same cues when FYRW is a causal factor. Typically, the two curves will overlap, and the investigator then is assumed to have (not necessarily consciously) a critical cue intensity  $\beta$  (see Fig. 1) so that whenever the signal exceeds this intensity, he identifies FYRW as a causal factor; and whenever the signal intensity is less, he decides that the "signal" is not present, i.e. FYRW is not a factor. While the critical cue intensity itself will not determine the overall error rate, it does represent the bias the investigator has in terms of the relative proportion of times a factor is not cited when it is causal (misses), and the number of times a factor is cited when in fact it should not be cited (false alarms). Of interest here are the

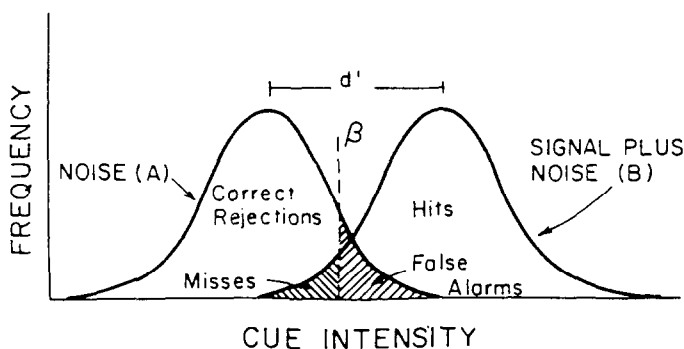


Fig. 1. A signal detection theory (SDT) representation of causal factors (see text for explanations).

conditional probabilities of hits (factor being cited, given signal plus noise, i.e. when it is in fact there) and false alarms. The probabilities of hits and false alarms can be readily calculated from the frequency table in Table 1 according to the following formula from Bayes Theorem:

$$P(\text{Hits}) = P(\text{Police citing of Factor given IRPS citing}) \\ = \frac{P(\text{citing by both Police and IRPS})}{P(\text{citing by IRPS regardless of Police citing})}$$

$$P(\text{False Alarms}) = P(\text{Police citing of Factor given IRPS not citing}) \\ = \frac{P(\text{Police citing while IRPS does not cite})}{P(\text{IRPS does not cite regardless of Police})}$$

In the case of FYRW,  $P(\text{Hit}) = 0.97$  and  $P(\text{False Alarm}) = 0.05$ . Note that for the area marked "misses" in Fig. 1,  $P(\text{Miss}) = 1 - P(\text{Hit})$ .

Obviously, it would be most desirable to both maximize the hits and minimize the false alarms. Since—short of increasing the investigator's sensitivity—this cannot be done, an alternative objective is to maximize the quantity  $P(\text{Hit}) - \beta P(\text{False Alarm})$ , where  $\beta$  is a constant. A decision rule that maximizes this quantity is to cite the presence of a factor (in this case FYRW) if and only if the likelihood ratio (LR, defined below) is greater than  $\beta$ :

$$LR = \frac{f(\text{Hits})}{f(\text{False Alarms})} = \frac{f(\text{critical cue intensity}|\text{signal} + \text{noise})}{f(\text{critical cue intensity}|\text{noise})} > \beta,$$

where  $f(\text{Hits})$  is the value of the ordinate of curve B in Fig. 1 at the critical cue intensity, and  $f(\text{False Alarms})$  is the value of the ordinate of curve A in Fig. 1 at the same point.

The LR is a statistic that enables us to evaluate the police performance in terms of both hits and false alarms. An ideal detector can optimize the criterion  $\beta$  so that  $\beta = 1$  whenever the value of a hit and the cost of a false alarm are identical, or when the *a priori* probability of a signal is 0.5;  $\beta > 1$  whenever the cost of a false alarm is greater than the value of a hit, or the probability of a signal is less than 0.5;  $\beta < 1$  whenever the cost of a false alarm is less than the value of a hit, or the probability of a signal is more than 0.5. In practice,  $\beta$  is calculated from the empirical data used to derive the likelihood ratio which corresponds to the decision criterion [Green and Swets, 1966]. In the case of FYRW,  $\beta = 0.66$ . Therefore, in the case of FYRW, the police were hedging in favor of *false alarms* rather than misses.

The likelihood ratio should also reflect the values and costs associated with hits and false alarms, and when these can be quantified, a procedure to adjust LR is available [Green and Swets, 1966, p. 211].

In the analysis of the police ability to detect causal factors, another variable is the distance between the "noise" and "signal plus noise" distributions (A and B in Fig. 1). This distance, labeled  $d'$ , denotes the discriminability of the signal, the "obviousness" of the factor (when it is present), or the discriminating capacity of the police independently of where the criterion  $\beta$  is. If we assume that both signal and noise are normally distributed and have equal variance, then

from the  $P(\text{Hit})$  and  $P(\text{False Alarm})$  we can determine the distance of the critical cue intensity from the means of the two distributions and, hence, the distance between the two distribution means (in standard scores). The greater the  $d'$ , the less the overlap between the two distributions and the more detectable the factor is. To calculate  $d'$  the standard  $Z$  scores for the values of hits and false alarms are subtracted from each other according to the following formula and normal distribution tables

$$d' = Z_p(\text{False Alarms}) - Z_p(\text{Hits})$$

in the case of FYRW we obtain

$$d' = Z_{0.05} - Z_{0.97} = 1.65 - (-1.88) = 3.53.$$

The above results also imply that if the police would give equal value to misses and false alarms—shifting their criterion  $\beta$  to 1—then the probability of either error would be  $P(Z > d'/2) = 0.04$ , i.e. any reduction in the rate of misses would be costly in terms of the increase in false alarms. In any case, assuming equal-variance distributions, the lowest error rate possible, given  $d' = 3.53$ , is 4% of each of the error types (false alarms and misses). In the discussion below a factor will be considered as adequately discriminable by the police whenever  $d' > 2.0$ , i.e. whenever the sum of  $P(\text{misses})$  and  $P(\text{false alarms}) < 0.32$ ; the total area under the noise curve  $1\sigma$  above the mean and the area under the signal + noise curve up to  $1\sigma$  below the mean.

The use of conditional, rather than unconditional, probabilities is helpful in reducing effects caused by variations in the actual frequencies of the factor.

In interpreting the results obtained by the SDT procedures, care should be taken to separate the appropriateness of the statistical procedure from the appropriateness of the underlying signal detection theory. The above discussion was primarily addressed to the appropriateness and implications of the procedure rather than the psychological theory. Whether or not it is appropriate to describe the detection of causal factors in terms of a "cue intensity" variable—and accordingly interpret  $\beta$  and  $d'$ —remains an open question. While the application is intuitively appealing, it has no precedence in accident causation research (though it has been applied successfully to quality control; Fox, 1973). Given the potential promise of this analytical technique, it is recommended that experiments be designed to test its appropriateness.

## RESULTS AND DISCUSSION

In the discussion that follows, a distinction is made among accident variables, traffic unit variables, and accident causes. These three variable categories can be distinguished on the basis of the ease of getting at the information. Accident variables describe the accident environment regardless of the specific vehicles and drivers. For the most part they require no more than observation at the scene after the accident. On the other hand, traffic unit variables involve determination of both the driver and vehicle condition that precipitated the accident—though these may not necessarily have been causally relevant. Finally, causal factors are those variables which are deemed to be responsible for the occurrence of the accident; the assumption being that had these events, behaviors, or conditions not existed, the accident would not have occurred.

The data base, on the bases of which the police reports were evaluated, consisted of the two types of agreements and three types of disagreements with the IRPS conclusions. For each accident variable the total percent agreements was the sum of the times that the event or cause was cited by both, and the sum of the times that it was cited by neither, i.e. the sum of correct identifications and correct rejections. Disagreements could arise from commission errors, omission errors, or misidentifications. A commission error was cited whenever the police identified a variable which IRPS did not, and an omission error was cited whenever the police failed to identify a variable cited by IRPS. Misidentification was cited whenever the police noted the presence of a variable but misidentified its level (e.g. severity level of accident, vehicle color, date, etc.), either due to an error in reasoning or coding.

Although interagency comparisons were made, the number of cases evaluated for each



agency was relatively small and a complete discussion of the results would be beyond the scope of this paper. Still, in general the differences between the agencies for most of the variables examined were rather small—and where they existed they could be attributed at least in part to differences in their areas of control. Pearson  $r$  correlations calculated on the uncertainty coefficient indicated that the correlations among the three agencies for the accident variables ranged from 0.68 to 0.90; for the traffic unit variables they ranged from 0.53 to 0.93; and for the causal variables they ranged from 0.42 to 0.71. Across all variables in each of the above categories, the highest average uncertainty coefficient was obtained for the state police ( $Uc = 0.57$ ) while the average uncertainty coefficient for the County and City police were 0.45 and 0.48, respectively.

#### Accident variables

Nineteen variables that together provide a description of the scenario for each accident were identified for this analysis. The variable names and the agreements and disagreements on their occurrence are provided in Table 2. The variables included in this analysis are of such a nature that commission errors on the part of the police are impossible. Therefore, the only two kinds of disagreements possible for these analyses were misidentifications (noting the wrong answer for that variable) or omissions (simply failing to make an entry for that variable).

It appears that the police are highly reliable in observing the correct location and date and may be considered to be sufficiently reliable in noting the day of week, number of drivers, passengers, and vehicles involved in each accident ( $Uc \geq 0.88$ , in Table 2).

Converging trajectories, which are important to crash data analysis, are also fairly well reported by the police ( $Uc = 0.80$ ). For this particular variable, of the 14 misidentifications, 3 were because headons were misclassified as opposing-oblique; 9 were because opposing-oblique, right-angle and acute-oblique were not distinguished properly; 3 because rearend and acute-oblique were confused; and 1 because a rearend was misclassified as a collision while backing.

Police performance begins to deteriorate when they note the ambient road (dry, wet, or covered with snow or ice), light and weather conditions ( $0.70 < Uc < 0.80$ ). Here, most of the police errors are probably due to misunderstanding of the coding procedures and confusion between weather conditions and road conditions (for example, snow was sometimes coded under both road condition and weather—even if it was not snowing at the time of the accident).

Police accuracy is poorest in noting the vertical curvature, i.e. grade of the road ( $Uc = 0.17$ ). This is an important variable since there is evidence suggesting that, at least in curves, vertical

Table 2. Agreements and disagreements between IRPS and police reports on accident variables ( $N = 124$  accidents)

Variable	Disagreements				Total		Agreements		Uncertainty Coefficient $u$
	Misidentifications $n$	%	Police Omissions <sup>a</sup> $n$	%	$n$	%	$N$	%	
Month	1	0.8	0	0.0	1	0.8	123	99.2	.99
Day of Month	6	4.8	0	0.0	6	4.8	113	95.2	.97
Year	1	0.8	0	0.0	1	0.8	123	99.2	.98
Day of Week	6	4.8	2	1.6	8	6.5	116	93.5	.88
* of Traffic Units	2	1.6	0	0.0	2	1.6	122	98.4	.90
* of Passenger Units	3	2.4	0	0.0	3	2.4	121	97.6	.88
* of Trucks (GVW $\neq$ 8,000 lbs.)	1	0.8	0	0.0	1	0.8	123	99.2	.91
* of Motorcycles	0	0.0	0	0.0	0	0.0	124	100.0	1.00
* of Parked Units	1	0.8	0	0.0	1	0.8	123	99.2	.90
Accident Severity	38	30.6	0	0.0	38	30.6	86	69.4	.25
Converging Trajectories	14	11.3	0	0.0	14	11.3	110	88.7	.80
Speed Limit	28	22.6	21	16.9	49	39.5	75	60.5	.59
Horizontal Character	9	7.3	4	3.2	13	10.5	111	89.5	.68
Vertical Character	38	30.6	13	10.5	51	41.1	73	58.9	.17
Surface Composition	13	10.4	1	0.8	14	11.3	110	88.7	.37
Road Ambience	7	5.6	0	0.0	7	5.6	117	94.4	.71
Weather Ambience	7	5.6	1	0.8	8	6.5	116	93.5	.72
Light Ambience	6	4.8	1	0.8	7	5.6	117	94.4	.78
Location	1	0.8	0	0.0	1	0.8	123	99.2	N/A

<sup>a</sup> Police omissions are equivalent to missing values (no entry on police reports), and are therefore not included in the computation of the uncertainty coefficients.

curvature may be related to accident propensity of a road section [Shinar, 1977]. In this case, the lower accuracy of the police may be due either to confusion concerning accidents occurring at intersections or to poor judgment and measurement capabilities. Of the 38 misidentifications 14 were because accidents on level road were misclassified as being on grades, 22 times the reverse occurred, and twice accidents on grades were misclassified as occurring on hill-crests (IRPS classified a road as level whenever the vertical curvature was less than 2%. This high precision level may account for police misclassification of grades as level, but then it underestimates the number of police errors for level roads).

The second variable for which the police data are definitely inadequate is the accident severity ( $Uc = 0.25$ ). Whereas the police report is typically filed immediately after the accident, the IRPS in-depth report is based on data collected both immediately following the accident, as well as on follow-up data collected up to a month after the accident. This allows the IRPS investigators to get more reliable information concerning the injuries to all drivers/passengers involved. Of the 38 misidentifications, all were because personal injury accidents were misclassified as property damage only. Obviously, the use of police data to code severity would therefore yield very conservative estimates of injury, and could be extremely misleading in various cost-benefit analyses of safety improvement programs.

Other variables where police accuracy is low are surface composition ( $Uc = 0.37$ ), speed limit ( $Uc = 0.59$ ) and horizontal character of road ( $Uc = 0.68$ ). For road surface composition, all the misclassifications were caused by confusing concrete and blacktop road surfaces. For speed limit, of the 28 misidentifications the police were within 10 mph of the actual speed limit 19 times, and made mistakes outside that range 9 times. For horizontal character, the police misclassified straight and curved roads 9 times (a curve was defined in the State Police Manual as an "apparent change in direction regardless of degree").

*Traffic unit variables*

Traffic unit variables are those measures which reflect characteristics of each one of the participating drivers and vehicles. Thus, these characteristics must be specified separately for each one of the involved units. Twenty-two traffic unit variables defined by the police that were considered potentially relevant for accident statistics were included in this analysis, and they are listed in Table 3, along with the agreement-disagreement analysis between the police and IRPS.

Table 3. Agreements and disagreements between IRPS and police reports on driver/vehicle variables ( $N = 207$  driver/vehicle units)

Variable	Misidentifications		Disagreements				Agreements				Total	Uncert. <i>c/f</i> u.	Phi			
	n	%	Police Commissions n	%	Police Omissions n	%	Total n	%	Not Present n	%				Present n	%	
Age	24	11.6	N/A	N/A	1*	0.5	25	12.1	N/A	N/A	N/A	N/A	182	87.9	.91	N/A
Sex	1	0.5	N/A	N/A	0	0.0	1	0.5	N/A	N/A	N/A	N/A	206	99.5	.96	.99
Model Year	11	5.3	N/A	N/A	20*	9.7	31	15.0	N/A	N/A	N/A	N/A	176	85.0	.91	N/A
Drinking-Degree of Impairment	6	2.9	5	2.4	6*	2.9	17	8.2	N/A	N/A	N/A	N/A	190	91.8	.32	N/A
Brakes Defective	N/A	N/A	0	0.0	90*	43.5	90	43.5	113	54.6	4	1.9	117	56.5	.02	.15
Lights Defective	N/A	N/A	0	0.0	40	19.3	44	21.3	162	78.7	0	0	163	78.7	+	+
Steering Defective	N/A	N/A	0	0.0	76	36.7	76	36.7	131	63.3	0	0	131	63.3	+	+
Other Vehicle Defects	N/A	N/A	0	0.0	168	81.2	168	81.2	37	17.9	2	1.0	39	18.8	.00	.05
Attention Diverted	N/A	N/A	6	2.9	17	8.2	23	11.1	181	87.4	3	1.4	184	89.9	.03	.17
Drinking	N/A	N/A	3	1.4	7	3.4	10	4.8	195	94.2	2	1.0	197	95.2	.09	.28
Blindsight Defective	N/A	N/A	0	0.0	87	42.0	87	42.0	120	58.0	0	0	120	58.0	+	+
Hearing Defective	N/A	N/A	0	0.0	1	0.5	1	0.5	206	99.5	0	0	206	99.5	+	+
Illness	N/A	N/A	0	0.0	3	1.4	3	1.4	204	98.6	0	0	204	98.6	+	+
Fatigue	N/A	N/A	1	0.5	8	3.9	9	4.3	198	95.7	0	0	198	95.7	.00	.01
View Obstruction (Hill Crest)	N/A	N/A	2	1.0	5	2.4	7	3.4	197	95.2	3	1.4	200	96.6	.13	.46
View Obstruction (Embankment)	N/A	N/A	3	1.4	5	2.4	8	3.9	199	96.1	0	0	199	96.1	.00	.02
View Obstruction (Growth)	N/A	N/A	1	0.5	8	3.9	9	4.3	196	94.7	2	1.0	198	95.7	.11	.35
View Obstruction (Other)	N/A	N/A	7	3.4	15	7.2	22	10.6	176	85.0	9	4.3	185	89.4	.14	.40
Foreign Substance on Road	N/A	N/A	14	6.8	3	1.4	17	8.2	187	90.3	3	1.4	190	91.8	.14	.26
Shoulder Defective	N/A	N/A	1	0.5	9	4.3	10	4.8	194	93.7	3	1.4	197	95.2	.15	.42
Other Road Defects	N/A	N/A	0	0.0	17	8.2	17	8.2	189	91.3	1	0.5	190	91.8	.04	.23
Make of Vehicle	2	1.0	N/A	N/A	4*	1.9	6	2.9	N/A	N/A	N/A	N/A	201	97.1	N/A	N/A

\* These omissions are equivalent to missing values (no entry on police report), and are therefore not included in the computation of the uncertainty coefficient.  
 + No statistics computed, because of insufficient data.

In this analysis, we begin to see large discrepancies between the percent agreements and the uncertainty coefficient statistics for the presence of either a driver deficiency, a vehicle deficiency, or a road-related problem. The poor level of agreement is due, for the most part, to a tendency by the police not to cite these variables, i.e. to make omission errors. The fact that commission errors are rare suggests that conservatism/nonreporting on the part of the police is indeed their underlying characteristic. Note also that for all but the first four variables, misidentification does not apply. Thus, the police can either cite or not cite defective brakes, but there is no opportunity for them to misidentify defective brakes as something else.

In light of the results presented in Table 3, the use of police data for evaluating the frequency and type of vehicle defects in accidents is very questionable. For all the vehicle defect categories evaluated here, the uncertainty coefficients are practically zero, i.e. no information is conveyed at all. Phrased in another way, the uncertainty is not reduced at all by the police reports. The situation is fairly similar with respect to driver mental characteristics, including the police report of whether a driver was drinking or not.

Finally, again with respect to road-related characteristics, the police report may be viewed as transmitting very little information. Whereas for vehicle and driver characteristics, the police tend to make omission errors significantly more than commission errors, in the case of road-related defects, the police are approximately just as likely to make a commission error as they are to make an omission error. As in the other two areas, the high percent agreement is mostly based on the lack of any detected deficiencies.

#### *Accident causes*

In many ways, the determination of an accident cause is the ultimate goal of an accident investigation. Nonetheless, as has been mentioned above, the definition of an accident cause is very different for the police investigator than for the IRPS investigation. Part of the policemen's role is to determine the most legally culpable driver in an accident. Thus, *a priori*, his orientation is to find some fault with one or both of the drivers. On the other hand, the IRPS investigators attempted to identify cause-and-effect relationships which led to the accident regardless of the legal culpability involved. Thus, discrepancies between the police and the IRPS investigations are as likely to be a result of: (1) differences in the focus of attention and the definition of the accident cause, and (2) the relative accuracy of the police investigations. Unfortunately, no statistical analysis can separate these two issues and determine the accuracy of the police on each, independent of the other. However, since the underlying issue here is the validity of the police-reported data for highway safety research, analysis, and development progress, the results of the comparison are still valid because they indicate the extent to which the police are accurate in reporting accident causes as defined by a research-oriented, multi-disciplinary accident investigation team (IRPS).

Twenty-three different accident causes, corresponding to preexisting police categories, were identified by IRPS for the purpose of this analysis. These causes are listed in Table 4, along with the results of the agreement/disagreement analysis. Since errors of misidentification were not applicable here, they are not listed for this table. The causes are grouped into the vehicular, human direct, human indirect, and environmental causes.

The analyses of the agreements and disagreements between the IRPS evaluations and the police evaluations again indicate a very high value for the total percent agreements between IRPS and the police, but an extremely variable relationship based on the uncertainty coefficient and Phi correlation. To illustrate, for 17 out of the 22 variables listed, the percent agreement between IRPS and the police exceeds 85. Thus, there is a high percent of agreement on the driver sex, whether or not the driver was drinking, fatigued or had his attention diverted and the presence of absence of view obstructions. Yet except for sex, where the uncertainty coefficient is 0.91, for all other variables  $0.00 \leq Uc \leq 0.14$  and  $0.01 \leq Phi \leq 0.46$ . The discrepancies between the measures stem from the number of correct noncitations (correct rejections) by the police. When the number is high there is a high percent of total agreements between IRPS and the Police. Furthermore, as has been noted above, the more detailed the description of the cause is likely to be, the higher the percent of noncitations by the Police. Therefore, the uncertainty coefficient is a more realistic measure of the accuracy of the police reporting procedures than the percent agreements or disagreements.

Table 4. Agreements and disagreements between IRPS and police reports on accident causes ( $N = 207$  driver/vehicle units)

Variable	Disagreements						Agreements						Uncert. coeff.	Phi
	Police Commissions		Police Omissions		Total		Cause Not Present		Cause Present		Total			
	n	%	n	%	n	%	n	%	n	%	n	%		
<b>Vehicular Causes</b>	2	1.0	9	4.3	11	5.3	191	92.3	5	2.4	196	94.7	.20	.48
Inadequate Brakes	2	1.0	3	1.4	5	2.4	198	95.7	4	1.9	202	97.6	.5*	.60
Tire Problems	0	0.0	7	3.4	7	3.4	200	96.6	0	0	200	96.6	*	*
Other Vehicle Causes	1	0.5	1	0.5	2	1.0	205	99.0	0	0	205	99.0	.00	<.01
<b>Direct Human Causes</b>	4	1.9	33	15.9	37	17.9	73	35.3	97	46.9	170	82.1	.40	.67
Speed Too Fast	8	3.9	8	3.9	16	7.7	181	87.4	10	4.8	191	92.3	.26	.51
Failed to Yield Right-of-Way	8	3.9	1	0.5	9	4.3	166	80.2	32	15.5	198	95.7	.71	.86
Drove Left of Center	7	3.4	3	1.4	10	4.8	193	93.2	4	1.9	197	95.2	.26	.43
Improper Overtaking	6	2.9	2	1.0	8	3.9	193	93.2	6	2.9	199	96.1	.43	.59
Passed Stop Sign	1	0.5	1	0.5	2	1.0	201	97.1	4	1.9	205	99.0	.63	.80
Followed Too Closely	5	2.4	0	0.0	5	2.4	202	97.6	0	0	202	97.6	*	*
Made Improper Turn	1	0.5	2	1.0	3	1.4	200	96.6	4	1.0	204	98.6	.49	.72
Other Improper Driving	8	3.9	56	27.1	64	30.9	121	58.5	22	10.6	143	69.1	.07*	.30
<b>Indirect Human Causes</b>	3	1.4	10	4.8	13	6.3	192	92.8	2	1.0	194	93.7	.06	.23
Had Been Drinking	2	1.0	1	0.5	3	1.4	202	97.6	2	1.0	204	98.6	.42	.57
Fatigue	1	0.5	2	1.0	3	1.4	204	98.6	0	0	204	98.6	.00	.01
Driver Inexperience	0	0.0	2	1.0	2	1.0	205	99.0	0	0	205	99.0	*	*
Other Indirect Human Causes	0	0.0	8	3.9	8	3.9	199	96.1	0	0	199	96.1	*	*
<b>Environmental Causes</b>	1	0.5	60	29.0	61	29.5	135	65.2	11	5.3	146	70.5	.07*	.30
Slick Roads	2	1.0	12	5.8	14	6.8	189	91.3	4	1.9	193	93.2	.13	.38
View Obstructions	0	0.0	30	14.5	30	14.5	176	85.0	1	.5	177	85.5	.02	.17
Other Highway-Related Causes	0	0.0	17	8.2	17	8.2	189	91.3	1	.5	190	91.8	.04	.23
<b>Ambience-Related Causes</b>	1	0.5	12	5.8	13	6.3	190	91.8	4	1.9	194	93.7	.15	.43
All Causal Factors	52	14.7	168	47.5	220	62.1	---	---	---	---	134	37.9	.35	*

\* No statistics computed.

While the uncertainty coefficient can provide us with a single measure of the police's accuracy, the SDT statistics, presented in Table 5, are useful in interpreting the reduced accuracy. Recall that  $d'$  is a "pure" measure of the investigator's sensitivity, while the likelihood ratio reflects the degree of conservatism or reluctance to make false alarm errors. Since these two measures reflect two different human information processes, they are susceptible to improvement by different methods. Thus, knowledge of performance along these two measures can be used in the design, development and improvement of police investigation procedures.

In evaluating the actual results, a note of caution is in order. Due to the extremely low *a priori* probabilities of some of the accident causes, and the relatively small sample of accidents studies, the cell probabilities on which these statistics are based may not be very stable. This is particularly true with respect to estimates of the likelihood ratio.

Looking at all the causal factors together, the SDT statistics support the notion that the police in fact are conservative in their attribution of causes since the probability of false alarms is extremely low for all causal factors, and as a result the likelihood ratio is extremely high. The true sensitivity of the police investigators is reflected in  $d'$  (average  $d' = 1.83$ ), which suggests that for all causes together the police are fairly insensitive. For those factors for which no police citations at all were available,  $d'$  and the likelihood ratio could not be calculated, even though the uncertainty coefficient could be calculated.

A comparison between the different categories of causal factors—vehicular, human direct, human indirect, and environmental causes—indicates that direct human causes are the ones that are best detected by the law enforcement agencies ( $d'$  is highest), while the human indirect and environmental causes are the ones that present the most difficulty ( $d'$  is lowest). Also, in the human direct cause category the police are willing to commit a slightly higher rate of false alarms than in indirect human, vehicular, or environmental causes. This is probably due to the police orientation to search for culpability in terms of inappropriate driver behaviors. Indirect human causes are simply difficult to detect within the short amount of time and level of expertise available to the police, and environmental causes often require careful measurements by an accident reconstruction specialist—something beyond the scope of the police capabilities in terms of time, cost, and possibly, expertise.

In the domain of vehicular causes, the only cause that the police are marginally successful at detecting correctly is that of inadequate braking ( $Uc = 0.37$ ,  $d' = 2.5$ ). One way of possibly

Table 5. Signal detection analysis of police performance in identifying accident causes

Causal Factor	P(Hit)	P(False Alarm)	LR	$d'$
<u>Vehicular Causes</u>	.56	.01	14.05	1.97
Inadequate Brakes	.57	.01	14.75	2.50
Tire Problems	---	---	---	---
<u>Other Vehicle Causes</u>	.08	.05	---	---
<u>Direct Human Causes</u>	.75	.05	3.08	2.32
Speed Too Fast	.56	.04	4.58	1.90
Failed to Yield Right-of-Way	.97	.05	.66	3.53
Drove Left of Center	.57	.04	4.56	1.93
Improper Overtaking	.75	.03	4.67	2.56
Passed Stop Sign	.80	.01	10.52	3.17
Followed Too Closely	---	---	---	---
Made Improper Turn	.67	.01	13.60	2.77
<u>Other Improper Driving</u>	.28	.06	2.83	.97
<u>Indirect Human Causes</u>	.17	.02	5.23	1.10
Had Been Drinking	.67	.01	13.60	2.77
Fatigue	.00	.01	---	---
Driver Inexperience	---	---	---	---
<u>Other Indirect Human Causes</u>	---	---	---	---
<u>Environmental Causes</u>	.16	.01	9.14	1.33
Slick Roads	.25	.01	11.94	1.65
View Obstructions	.03	.00	---	---
<u>Other Highway Related Causes</u>	.06	.00	---	---
<u>Grand Mean</u>	.48	.03	5.86	1.83

improving the police's detection rate of vehicular causes would be to be less conservative, and risk increasing the false alarm rate. An increase of from 0.01 to 0.05 would probably greatly increase the probability of hits without involving too great a "cost" in terms of permissible level of accuracy. It is most likely that the major limitation here is that of time—the police simply do not have the time (or do not consider the spending of such time appropriate) to actually remove a wheel and examine the brakes. In fact, they usually do not even have the time to drive the car themselves.

Human direct causes are perhaps the best identified by the police. Of these, failure to yield right-of-way and failure to stop at a sign are the best identified ( $U_c = 0.71$  and  $0.63$ , respectively, and  $d' = 3.53$  and  $3.17$ , respectively). The level of false alarms that the police are willing to tolerate here is much greater than it is for vehicular or environmental causes and appears to be appropriate. Note that failure to stop at a stop sign is associated with both a high probability of hits and a low probability of false alarms, indicating high sensitivity to this cause. Direct human causes with respect to which the police's assessment can be considered unreliable are speeding, driving left of the center of the road, and any other improper driving behaviors. The problem with identification of speeding and driving left of center is not one of poor criterion ( $\beta$ ), but actually one of the investigator's sensitivity to these factors. It may be that, given the stress that the police are under, it is impossible for them to actually determine whether a driver was speeding or driving left of the center line. Obviously, the driver himself/herself, viewing the police officer as a representative of the law would be reluctant to volunteer this type of information. Transient environmental evidence to this effect (tire markings on the pavement) were perhaps too time-consuming to be properly assessed by the police. Nonetheless, since both speeding and driving left of the center line are clear-cut violations of the law it may be

advantageous to try to improve the overall detection capability of the police ( $d'$ ) by providing them with short workshops that could give them additional cues to look for and some rules of thumb that they can use to calculate speed and paths prior to impact. Also, with respect to speeding, it is likely that the police use different criteria from those used by IRPS. This is because IRPS' evaluation of speeding was in essence "driving too fast for conditions", while the police citing of this variable may have been restricted to "above the speed limit". In any case when the speed limit was exceeded the police tended to cite this variable while IRPS cited it only when it was judged as causal.

The police assessment of "other improper driving" behaviors is practically useless, as indicated by both the uncertainty coefficient and  $d'$ , which reflect chance-level performance. This is probably because more subtle human errors escape the police's notice (especially if they do not have a specific code for these behaviors on the form), which is suggested by the high rate of omission errors in this category (27%).

Of all the indirect human causes, only drinking can be evaluated since for the three other categories—fatigue, driver inexperience, and other indirect human errors—no correct identifications were made at all (a negative reflection on police performance). The assessment of drinking may be considered adequate ( $U_c = 0.57$  and  $d' = 2.77$ ), especially since perhaps the only way to increase the number of current identifications may be to shift the criterion to increase the percent of false alarms. Since this is a type of cause in which the police would want to be conservative in their estimate, it is likely that short of increasing pressure on the police to give alcohol test prior to citing for driving while intoxicated, no improvement can be expected†.

In the assessment of environmental causes the police performance is also not very reliable. For none of the factors cited does the police performance exceed the chance level. This shortcoming is particularly critical if police reports are to be used as data sources for highway improvement programs. As to the reason for the poor performance, the near zero false alarm rate for the three environmental causes is a clue suggesting that the police in fact simply disregard or overlook this category. This can become a bad habit relevant to any causal factor that is relatively rare. This may be a particularly difficult problem to solve since it appears that subjective probabilities for (objectively) low probability events are often zero [Näätänen and Koskinen, 1975]. To the extent that failure to check this category is due to oversight it can be eliminated in part by proper formatting of the police forms, so that the investigating police officer will be forced to scale all relevant alternatives. Slick roads as a causal factor may be an exception since it is checked at some above-chance frequency, as is indicated by the probability of a hit = 0.25.

#### SUMMARY CONCLUSIONS AND RECOMMENDATIONS

##### *Summary of objectives and methodology*

In the present study, a random sample of 124 accidents involving 207 drivers was investigated by both multi-disciplinary accident investigation (MDAI) teams and by police. The MDAI team investigating an accident consisted of an accident reconstruction specialist, an automotive engineer, and a psychologist. The representatives of the three relevant disciplines each investigated the accident from his/her own viewpoint and then together, through a formal process of accident analysis, formulated conclusions concerning the characteristics of the accident and the relevant causal factors. In the absence of an external criterion for accident description and cause, the MDAI report was assumed to reflect the true state of events, and the validity of the police data was then evaluated relative to the MDAI report. Comparisons were made on three types of accident variables: (1) accident descriptors, including date and time of accident, number of traffic units involved, converging trajectories of the vehicle(s), accident severity, roadway characteristics, ambience, weather, etc.; (2) driver/vehicle descriptors, including driver age, sex, presence of alcohol, presence of hearing and visual deficiencies, reports of fatigue, vehicle condition, vehicle make and year, etc.; and (3) accident causes, including

†The assessment of drinking must be considered very tentative because in the total sample even IRPS cited alcohol as a cause only in three accidents. To rely on a less accurate but larger data base additional analysis was made in the 2258 accidents investigated by the on-site investigators. The results of these analyses are included in the original report by Shinar and Treat [1977].

vehicular factors (e.g. inadequate brakes and tires), human direct causes (e.g. speeding, failing to yield right-of-way, driving left of center, passing a stop sign, making improper driving maneuvers, following too closely), human indirect causes (e.g. alcohol intoxication, fatigue, driver inexperience), and environmental causes (e.g. slick roads, view obstructions).

The nature of the data (nominal categories, unevenly distributed) precluded the use of standard parametric statistical procedures. Instead, measures derived from information theory and signal detection theory were used.

In general the most valid police reported data were those concerned with accident descriptors and the least reliable were driver/vehicle variables. The ability of the police to accurately attribute accident causes varied considerably across the different causes. The main conclusions can be summarized as follows:

Among *accident-descriptive data* reported by the police, it was found that the police data were the most reliable for the following six variables: location, date, day of week and numbers of drivers, passengers and vehicles in each accident. At the other extreme, the least reliable police data concerned vertical road character, accident severity and road surface composition. Of the vertical road character errors, the biggest problem was misidentifying accidents which occurred on grades as occurring on level roads; out of 38 total misidentifications (30.6% of the cases) this error occurred 22 times. In 14 additional cases accidents on level roads were misclassified as being on grades. Accident severity was underestimated by the police in over 30% of the cases, and in all of them personal injury accidents were misclassified as involving property damage only. Under road surface composition, the 13 misidentification errors (10.4% of cases) all involved confusion of concrete and asphalt surfaces. Reliability was also inadequate for speed limit and horizontal character of roadway. The police improperly identified the speed limit in 28 of the 124 accidents (22.6%), and failed to indicate the speed limit in another 21 accidents (16.9%). Of the 20 misidentifications, the police were within 16 km (10 mph) of the actual speed limit 19 times. For horizontal character, the police misclassified straight vs curved roadway sections in nine accidents (7.3%).

The police reports analysed provided very little information regarding the *presence of driver factors, human conditions and states, and vehicular and environmental/roadway factors and deficiencies*. For example, the police misclassified driver age for 24 of the accident drivers (11.6%) and misclassified vehicle model year for 11 vehicles (5.3%), with model year not stated for an additional 20 vehicles (9.7%). For vehicle and driver characteristics, the police tended to make omission errors significantly more often than commission errors (i.e. the police often failed to provide any information on the report, rather than to identify a factor—such as a defective brake component—as being present when in fact it was not). However, in the case of road-related defects, the police were approximately just as likely to make omission as commission errors.

The sensitivity of police investigators to *accident causes* was also generally low. Police often failed to cite factors which in fact should have been cited, although they rarely cited factors which were not in fact involved (i.e. the false alarm rate was low). In terms of the identification of the general categories of causal factors (human, vehicle, environment), the police performed most reliably in detection of human direct causes followed by vehicular, environmental, and human indirect causes. In the area of human direct causes, police performance was relatively good in identifying “failure to yield” and “failure to stop at a stop sign”, and was relatively poor with respect to “speeding”, “driving left-of-center” and “other improper driving”. For vehicle factors, the police were marginally successful in detecting the role of inadequate braking, but performed inadequately with respect to all other vehicle factors.

With respect to environmental factors, police performance did not exceed the chance level for any of the factors cited. A particular problem exists with respect to police identification of view obstructions—the most frequent environmental cause identified by the MDAI teams, and a factor which police record systems could perform an important service in correctly identifying. The police failed to implicate view obstructions as causes in 30 accidents (14.5%) in which the in-depth team indicated this factor should have been cited. Overall, the police correctly implicated view obstructions in only 3% of the accidents.

Among human indirect causes, police performance was adequate only for the “had been drinking” involvement assessment. However, this conclusion must be considered very tentative, since it is based on only three accidents in which IRPS cited drinking as an accident cause.

### Recommendations

The prevalent use of police records for various non-police needs such as research, policy making and highway improvements, provides a strong justification for improving the validity of police reported data—or at the very least bringing the lack of validity to the awareness of the different data users. An important implication of the results obtained in the preceding analyses is the need to reevaluate police-reported data in the proposed National Accident Sampling System use of that data for accident statistics purposes.

Because it is very likely that police reports will remain a popular source for various traffic accident statistics in the foreseeable future, some steps should be taken to monitor the quality of police-reported accident data and where possible improve its accuracy. Specifically, three recommendations are made:

(1) The generally poor police performance indicated by this assessment provides a strong argument for improving the training and motivation of police officers in traffic accident reconstruction and investigation. Significantly, many of the errors were recorded for factors which clearly do not require high levels of expertise to correctly assess. For example, as important as driver age is to the use of a police record system for problem identification, the police improperly identified driver age in 11.6% of the accidents considered. In other instances, data simply were not entered. For example, whereas vehicle model year was improperly stated in 5.3% of these accidents, it was simply not provided in 9.7%. In addition to better motivating police officers through informing them of the importance of accurate records (and then demonstrating this importance through actual use of these record files), in some instances—such as vehicle model year—it may indicate a problem in the availability of needed reference information. Perhaps vehicle model year should be more clearly indicated on the vehicle (although it is on the registration card which is supposed to be with the vehicle).

(2) This assessment also demonstrates a need to periodically monitor and report the accuracy of police agencies. Such evaluation can be of benefit both in motivating law enforcement personnel, and through helpful feedback, in better informing them as to problem areas or errors they may be making. For example, the frequent misidentification of asphalt and concrete surfaces could reflect a procedural problem, such as the completion of police reports only upon return to the station, with large elements of guessing then taking place for certain items which are perceived as being of lesser importance. Where possible, such evaluations might be conducted either by supervisors within the agency itself, or by state personnel, to reduce the potential political impact and sensitivity of such assessments.

(3) Some of the problems detected emphasize the need for improved design of accident report forms. For example, the extreme lack of sensitivity in recording the presence of environmental problems may reflect a habit of simply failing to consider such factors or failing to address this section of the reporting form, due to the low frequency with which environmental factors are clearly involved. A change in the structure of the recording form might ensure that such relatively rare items are properly considered. In addition, police agencies should also monitor the rate of missing information, and take corrective action when missing value rates exceed reasonable levels.

### REFERENCES

- Bull J. P. and Roberts B. J., Road accident statistics—A comparison of police and hospital information. *Accid. Anal. & Prev.* 5, 45–53, 1973.
- Fox J. G., Recent human factors contribution to enhancing industrial quality control. *Behaviorometrics* 3, 99–118, 1973.
- Green D. M. and Swets J. A., *Signal Detection Theory and Psychophysics*. Wiley, New York, 1966.
- McGuire F. L., The nature of bias in official accident violation records. *J. Appl. Psychology* 57, 300–305, 1973.
- McGuire F. L., The validity of accident and violation criteria in the study of drinking drivers. *J. Safety Res.* 7, 46–47, 1976.
- Näätänen R. and Koskinen P., Simple reaction time with very small imperative stimulus probabilities. *Acta Psychologica* 39, 43–50, 1975.
- Nicholl J. P., *The use of hospital in-patient data in the analysis of the injuries sustained by road accident casualties*. Report No. SR268. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1981.
- Nie N. H., Hull C. H., Jenkins J. G., Steinbrenner K. and Bent D. H., *Statistical Package for the Social Sciences* (2nd Edn.). McGraw-Hill, New York, 1975.
- Shannon C. E. and Weaver W., *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Ill., 1949.



- Shinar D., *Psychology on the Road: The Human Factor in Traffic Safety*. Wiley, New York, 1978.
- Shinar D. and Treat J. R., *Tri-level study-modification. Task 3: validity assessment of police-reported accident data*. Report No. DOT-HS-034-3-535 (MOD.T-3). Indiana University, June 1977.
- Thorsen J. and Sande J., Hospital statistics on road traffic accidents. *Proc. 3rd Triennial Congress of the Int. Assoc. for Accident and Traffic Medicine*, University of Michigan, Highway Safety Research Institute, 1969.
- Treat J. R. and Shinar D., A methodology for assessing and classifying traffic accident causes. *Proc. Motor Vehicle Collision Investigation Symp.* Buffalo, New York, 1976.
- Treat J. R., Tumbas N. S., McDonald S. T., Shinar D., Hume R. D., Mayer R. E., Stansifer R. L. and Castellan N. J., *Tri-level study of the causes of traffic accidents*. Report No. DOT-HS-034-3-535-77 (TAC), Indiana University, March 1977.