Determining Transit Drug Test Accuracy: The Multidrug Case

DAROLD T. BARNUM AND JOHN M. GLEASON

The accuracy of simultaneous testing for two or more drugs of abuse is analyzed. Probability theory and drug-testing accuracy concepts applicable to the testing for multiple drugs are reviewed, these concepts are applied to laboratory proficiency and transportation drug usage data, and accuracy levels are estimated that could occur in transit agency drug testing programs that simultaneously test for five different abused substances. The finding of this analysis is that as the number of drugs tested for increases, the probability that a positive test result is erroneous (the false accusation rate) increases significantly. For example, the false accusation rate when testing for five drugs is about 4 times the false accusation rate when testing for one drug. Therefore, it is suggested that if transit system decision makers wish to obtain certain maximum false accusation rates at their own organizations, they must adapt laboratory sensitivity and specificity rates for the number of drugs actually being tested for.

In U.S. workplaces in general, and in transportation organizations in particular, the testing of job applicants and employees for drugs of abuse continues to play a central role in the battle to eliminate the use of illegal drugs (*1–10*). However, if drug tests are to be an acceptable method of detecting the use of unlawful substances, they must accurately discriminate between those who are using drugs and those who are not. If the tests are not sensitive enough to identify most of those taking drugs, they will neither discourage drug use nor eliminate abusers from the workforce.

More importantly, the tests must be sufficiently specific to classify most nonusers as such. Otherwise, many who are innocent of drug use will be falsely accused. The U.S. systems of justice and of workplace jurisprudence both require that workers must be presumed innocent of infractions until proved guilty with compelling evidence (*11–16*). That is, if drug tests are to be an acceptable way of identifying drug users, they must not classify many people who do not abuse drugs as drug users. That is, there should be a low probability that someone who tests positive is a nonuser, which is referred to as a low false accusation rate.

Recently, Barnum and Gleason (*17,18*) discussed methods by which transit decision makers could set the maximum false accusation rates that would occur in their organizations. Thus, if a decision maker wanted no more than one false positive out of every 1,000 people testing positive, representing a maximum false accusation rate of 0.001, then the methods proposed would permit this goal to be achieved.

This earlier work, consistent with research on testing accuracy (*11,19,20*), developed accuracy estimates based on formulas that implicitly assume that a single test is conducted for the presence or absence of drugs. That is, they assume that each specimen is subjected to one test, and the results of this test will be positive if drugs are present and negative if drugs are absent. In fact, often multiple tests are conducted simultaneously on a specimen, because different substances are used to identify the presence of each drug of interest.

The purpose of this paper is to examine appropriate procedures for dealing with simultaneous tests for multiple drugs, because it is most typical for transit and other transportation modes to test for five illegal substances. That is, the U.S. Department of Transportation (DOT) mandates that most transportation modes simultaneously test for five illegal drugs: amphetamines, marijuana, cocaine, opiates, and phencyclidine (*21*). Although urban transit systems are not now covered by such regulations, most already test simultaneously for the same five drugs. Moreover, legislation has been proposed that would require all to do so by law (*6*).

The procedures discussed incorporate probability concepts that are appropriate under conditions of testing for multiple drugs, and issues are identified that have not been previously considered in the drug testing literature. The procedures may be incorporated into the processes suggested by Barnum and Gleason (*17,18*) to enable transit decision makers to obtain false accusation rates no greater than the level they find acceptable.

UNDERLYING PROBABILITY THEORY

Under proper drug testing protocol, a specimen is first screened for the presence of a drug. If the specimen tests negative for the drug on the screening test, then no more testing is done and it is assumed that the specimen does not contain the drug. However, if the specimen tests positive for the drug on the screen, then a confirmation test is conducted. If the specimen tests negative for the drug on this confirmation test, then it is assumed that the drug is not present; if it tests positive for the drug on this confirmation test, then it is assumed the drug is present. In other words, if a specimen screens negative, it is assumed to contain no drugs. Likewise, if a specimen screens positive but tests negative on the confirmation, it is assumed to contain no drugs. Only if the specimen tests positive on both the screen and confirmation, is it assumed to contain drugs. It is hereinafter assumed that the preceding protocol always has been used to reach test outcomes.

Carefully note that use of the words "one test" refers to one complete test for one drug, which has screening and confirmation components.

D. T. Barnum, College of Business Administration, University of Illinois at Chicago, Chicago, Ill. 60680–2451. J. M. Gleason, College of Business Administration, Creighton University, Omaha, Neb. 68178–0130.

Now, the situation when one test for one drug is being conducted can be reviewed. Excellent detailed descriptions of this process, from differing perspectives, have been provided (*11,19,20,22*).

When a specimen is tested for a given drug, one of four outcomes must occur. If the drug is not present in the specimen, the specimen may test negative, resulting in a true negative; or it may test positive, resulting in a false positive. Likewise, if the drug is present in the specimen, the specimen may test positive, resulting in a true positive; or it may test negative, resulting in a false negative. These four possible testing outcomes are referred to as individual test outcomes, because they refer to conducting one test of the specimen for one specific drug. (If the specimen is tested for several drugs, as is discussed later in this section, then there will be an individual test outcome for each of the drugs involved. At this point, however, it continues to be assumed that only one drug is being tested for.)

Three measures of drug testing accuracy are used in the health-related professions: sensitivity, specificity, and predictive value. In the following formulas, a positive test result is represented by a + sign, a negative test result is represented by a − sign, a specimen truly containing a drug is represented by a D, and a truly drug-free specimen is represented by an N.

Sensitivity is the probability that a specimen containing a drug will test positive for that drug. Thus, it is the probability of obtaining a true positive. The notation for the probability of a positive test result, given that the specimen contains the drug, is

$$\text{Sensitivity} = P(+|\text{D}). \tag{1}$$

Specificity is the probability that a specimen not containing a drug will test negative for that drug, or the probability of obtaining a true negative. The notation for the probability of a negative test result, given the specimen does not contain the drug, is

$$\text{Specificity} = P(-|\text{N}). \tag{2}$$

Thus, sensitivity measures the ability of the test to correctly report the presence of a drug, whereas specificity measures the ability of the test to correctly report the absence of a drug. Ideally, sensitivity and specificity would both be equal to 1.0, meaning that every drugged specimen tests positive and every nondrugged specimen tests negative.

Another important concept, although not used in studies measuring laboratory proficiency, is the predictive value of a test. For drug tests, the positive predictive value is the probability that the drug is present in a specimen, given that the test yielded a positive result for that drug.

$$\text{Positive predictive value} = P(\text{D}|+) \tag{3}$$

For example, if 90 out of every 100 people testing positive for a drug truly have the drug in their specimens, then the positive predictive value of the test would be 0.90. The probability that people with positive test results truly do not have the drug in their specimens would be 0.10. Therefore, if a drug test has a positive predictive value of X, then the probability is $(1 - X)$ that a person testing positive is free of that drug.

Thus, by maximizing the positive predictive value of a test, the probability is minimized that specimens testing positive are actually drug-free. Herein, this probability that specimens testing positive are drug-free is called the false accusation rate. That is,

$$\text{False accusation rate} = P(\text{N}|+) = 1 - \text{Positive predictive value} \tag{4}$$

This concept is key to determining whether a positive result on a drug test provides sufficient evidence of drug usage. If, for example, positive results on a test are known to be untrue in one out of every 10 cases (that is, the false accusation rate = 0.1), then a positive test probably would not be considered sufficient evidence to accuse a person of drug use. More importantly, if one wishes to protect the innocent from false accusation, then it is the positive predictive value of the test (or, analogously, the false accusation rate) that is of prime concern.

All of these concepts implicitly assume that one test for one drug is being conducted (with that test consisting of two parts when a specimen screens positive). That is, the concepts do not account for the situation where tests for several drugs are being conducted simultaneously. The situation can be extended to the case of testing for more than one drug, a topic that to our knowledge has not been addressed in any other publication.

Consider, therefore, a specimen that will be tested for multiple drugs. It either contains none of these drugs or contains one of them. Although it is easy to extend the theory to cases where more than one drug is present in a specimen, the necessary empirical data to make use of this extension are not available. Luckily, the final results would be little affected by the inclusion of more than one drug in a specimen, so the significance of specimens containing several drugs has little practical implication.

Further, only those cases are considered in which an error in the test for any one drug occurs independently of errors in the tests for any of the other drugs. That is, the errors are mutually independent, so the probability of any error is unrelated to the presence of other errors. More formally, if X_i represents a false test result for Drug i, then for n drugs,

$$P(X_1, X_2, X_3, \ldots, X_n) = P(X_1)P(X_2)P(X_3) \ldots P(X_n) \tag{5}$$

Thus, systematic errors such as the following are not included in this paper's analysis: cases where conditions causing specimens to test falsely positive for one drug increase the probability of the specimen's testing falsely positive for other drugs, and cases where conditions causing specimens to test falsely negative for one drug increase the probability of the specimens testing falsely negative for other drugs. Such systematic errors are important, and would increase the total error rate to a level higher than that caused by mutually independent errors alone; however, examination of systematic and other sources of error are beyond the scope of this paper.

In the multidrug case, a specimen either will contain no drugs or will contain one of the drugs being tested for. If it contains no drugs, one of two outcomes must occur: (1) it will test negative for all of the drugs, or (2) it will test positive for at least one of the drugs.

If the specimen contains one of the drugs of interest, one of four outcomes must occur: (1) the specimen will test positive for the drug it contains and negative for the other drugs, (2) the specimen will test positive for the drug it contains and positive for at least one of the other drugs, (3) the specimen will test negative for the drug it contains and positive for at least one of the other drugs, or (4) the specimen will test negative for all of the drugs. These results, which are presented in Table 1, are discussed in the following paragraphs.

First, consider the case where no drugs are present in the specimen. If it tests negative for all drugs, then there are true individual negatives for each of the drugs and a true group negative for the specimen. If, however, a drug-free specimen tests positive for at least one drug, then there are true individual negatives for drugs that had negative test results, false individual positives for drugs that had positive test results, and consequently a false group positive for the specimen (because the specimen itself will be declared to be positive when indeed it is not).

Now, consider the case where Drug *i* is present in the specimen, where Drug *i* could be any one of the substances being tested for. If the specimen tests positive for Drug *i* and negative for the others, then there is a true individual positive result and several true individual negative results, leading to a true group positive (and a specimen that is declared positive). If the specimen tests negative for all drugs, then there is a false individual negative result and several true individual negative results, leading to a false group negative (and a specimen that is declared negative incorrectly). These outcomes are fairly easy to interpret; the remaining two possibilities are more difficult.

Again assuming Drug *i* is present, consider the outcome in which the specimen tests positive both for Drug *i* and at least one of the absent drugs. Here, there is one true individual positive and at least one false individual positive, with the remainder of the outcomes being true individual negatives. The specimen would be declared positive; this outcome is classified as a true group positive, because the specimen indeed tests positive for the drug it contains.

Likewise, consider the outcome in which the specimen tests negative for Drug *i* and positive for at least one of the absent drugs. There is one false individual negative and at least one false individual positive, with the remainder of the outcomes being true individual negatives. The specimen would be declared positive. This outcome is classified as a true group positive, because a specimen containing a drug tests positive for a drug, even though it is the wrong drug.

TABLE 1 POSSIBLE OUTCOMES WHEN TESTING A SPECIMEN FOR MULTIPLE DRUGS

TRUE SPECIMEN STATE	POSSIBLE TEST OUTCOMES	SPECIMEN CLASSIFICATION
No Drugs Present	- for all drugs	true group negative
	+ for one or more drugs	false group positive
Drug i Present	+ for drug i, and - for absent drugs	true group positive
	+ for drug i, and + for one or more of absent drugs	true group positive
	- for drug i, and + for one or more of absent drugs	true group positive
	- for all drugs	false group negative

Note that a multiple-test outcome is classified herein as a true group positive whenever the sample tests positive for the drug it contains, regardless of the classification of the other drugs. The multiple outcome is also classified herein as a true group positive whenever the sample tests falsely negative for the drug it contains and falsely positive for one or more of the other drugs. Others may want to modify these two classifications, as they have more to do with what one believes is justice than with probability concepts. As seen later, however, both of these outcomes have such small probabilities of occurrence that their classification makes little practical difference.

Now, for the multiple drug case the individual probabilities can be developed for each of the six possible identified outcomes. As already noted, only mutually independent events are considered.

First, consider the two possible test outcomes for a specimen that contains no drugs. The probability that this specimen will test negative for all drugs is equal to the probability that it tests negative for the first drug, times the probability that it tests negative for the second drug, times the probability that it tests negative for the third drug, and so on.

Recall that the probability that a sample tests negative for a drug it does not contain, or a true negative, is the individual specificity rate. Assume that the individual specificity rates are equal for all of the drugs, that is, $(\text{individual specificity})_1 = (\text{individual specificity})_2 = \ldots = (\text{individual specificity})_n = (\text{individual specificity})_\mu$. The subscripted numbers identify the drug test to which the specificity applies, that is, the individual specificity of the test for Drug 1, the individual specificity of the test for Drug 2, and so on. The subscript μ indicates the mean specificity value for all of the tests. This assumption of equal individual specificities can be written

$$P(-_1|\mathrm{N}_1) = P(-_2|\mathrm{N}_2) = \ldots = P(-_n|\mathrm{N}_n) = P(-|\mathrm{N})_\mu \qquad (6)$$

Of course, specificity may vary by drug. But, because there is no good evidence that individual specificities differ, the insufficient reason approach to decision making suggests that individual specificities be assumed equal.

Now, assume tests are being conducted for n drugs (which means that n tests, one for each drug, must be conducted). The probability that the specimen will test negative for all drugs is the individual specificity rate raised to the nth power, because all individual specificities are identical, and, as elsewhere, errors are assumed to be random and mutually independent. This parameter is group specificity, as opposed to individual specificity that concerns a test for a single drug. Thus,

$$\text{Group Specificity} = [(\text{Individual Specificity})_\mu]^n = [P(-|\mathrm{N})_\mu]^n \qquad (7)$$

Consider a case in which the individual specificity for each drug is 0.999 (the individual false positive rate of each test for a drug is 0.001). This value means that for any one of the drugs, out of every 1,000 samples that do not contain that drug, 999 will test negative for that drug and 1 will test positive for that drug. If five different drugs are being tested for, then the group specificity for those five drugs would be $(0.999)^5 = 0.9950$.

Once the group specificity rate (the probability of all negative test results, given that none of the drugs are truly present) is estimated, this value can be used to identify the probability that this drugless sample will test positive for at least one drug. Of course, if the specimen tests positive for even one of the drugs, it will be declared positive. Because the specimen truly does not contain drugs, this result will be incorrect and hence will be a false positive. Because in this case the false positive applies to a specimen that has been tested for several drugs, this event is called a "group false positive." Recall that the individual false positive probability equals (1 − individual specificity); thus, the group false positive rate equals (1 − group specificity).

For example, recall that in the previous example the group specificity rate for five drugs was 0.9950. The probability that the specimen will test positive for at least one of the five drugs (the group false positive rate) is $(1 - 0.9950) = 0.0050$.

To summarize, the two possible outcomes of a drug test on a truly drug-free specimen have been examined. The specimen can test negative for all of the n drugs being tested for, or it can test positive for one or more of the n drugs. The probability that it will test negative for all of the drugs is the group specificity rate, which is simply the mean individual specificity rate of the drugs being tested for, raised to the nth power. The probability that the specimen will register at least one positive is the group false positive rate, which is 1 minus the group specificity rate. Finally, the probability that a person will falsely test positive for at least one drug increases as the number of drugs being tested for increases.

Now consider the case of tests for multiple drugs, where the specimen indeed contains one of these drugs. One of four outcomes must occur, as presented in Table 1. In order to develop the probabilities of these four outcomes, it is again necessary to start with the individual sensitivity and specificity rates. As before, on the basis of the principle of insufficient reason, the tests for all drugs are assumed to have identical individual sensitivities and identical individual specificities.

First, consider the case in which a total of n drugs are being tested for and the specimen correctly tests positive for the drug it contains and negative for the $n - 1$ drugs that it does not contain. The probability of this occurring equals the individual sensitivity rate for the drug in the specimen, times the individual specificity rate for the remaining drugs raised to the $n - 1$ power. That is, in testing for n drugs, if only Drug i is in the sample, then the probability of a positive test result on Drug i and negative test results for the other drugs is

$$P(+_i, -_j|\mathrm{D}_i, \mathrm{N}_j) = P(+_i|\mathrm{D}_i) * [P(-|\mathrm{N})_\mu]^{n-1} \quad [j = 1, 2, \ldots, n; j \neq i] \qquad (8)$$

where $+_i$ indicates a positive test result on Drug i, D_i represents the presence of Drug i in the specimen, N_j indicates Drug j is not present in the specimen, and $-_j$ indicates a negative test for Drug j. Note that when a probability has a μ for the subscript, then the probability is equal for all drugs, and therefore the particular drug involved is immaterial.

In other words, *Group Sensitivity Part One* = (*Individual Sensitivity*) * (*Individual Specificity*)$^{n-1}$. Because all drugs are assumed to have identical true individual sensitivities and identical true individual specificities, the mean values are simply estimated from empirical laboratory proficiency studies, and do not need to be obtained for each drug. The result is only a partial measure of group sensitivity because, as is described later, other events also contribute to total group sensitivity.

Assume, for example, that a specimen is being tested for five drugs, that their individual sensitivities are all 0.9, and that their individual specificities are all 0.999. Assume also that the specimen contains one of the drugs. The Group Sensitivity Part One, or the probability that the sample will test positive for the drug it contains and negative for the four others, is $(0.9) * (0.999)^4 = 0.8964$.

Second, consider the probability that a specimen containing a drug will test positive for that drug, and will test positive for one or more of the drugs that it does not contain. This probability is the probability that there will be a true positive for the drug that the specimen contains, and at least one false positive for the other drugs. The probability of a true positive for the drug in the specimen will again be the individual sensitivity rate. The probability of at least one false positive for the $n - 1$ remaining drugs will be 1 minus the probability of obtaining all true negatives, that is, [1 − (Individual Specificity)$^{n-1}$]. Thus,

$$P(+_i, + \text{ on one or more others } |\mathrm{D}_i, \mathrm{N}_j) = P(+_i|\mathrm{D}_i)*\{1 - [P(-|\mathrm{N})_\mu]^{n-1}\} \quad [j = 1, 2, \ldots, n; j \neq i] \qquad (9)$$

Because this group contains both true and false positives, it cannot unambiguously be classified into either a true or false category. However, because the objective of the test is to identify drug users, and this event does in fact identify a drug user, it is classified as a true group positive result. Hence, Equation 9, as Equation 8, is a partial measure of group sensitivity. The equation can be written in words as *Group Sensitivity Part Two* = (*Individual Sensitivity*) * [1 − (*Individual Specificity*)$^{n-1}$].

Consider, for example, the case of testing for five drugs with the same individual specificity and sensitivity as before. Then the Group Sensitivity Part Two would be $(0.9) * (1 - 0.999^4) = 0.0036$.

Third, the probability of obtaining a false negative for the drug in the specimen and one or more false positives for the other drugs must be calculated. This probability is the product of two terms: (1 − Individual Sensitivity), and [1 − (Individual Specificity)$^{n-1}$], that is,

$$P(-_i, + \text{ on one or more others } |\mathrm{D}_i, \mathrm{N}_j) = P(-_i|\mathrm{D}_i) * \{1 - [P(-|\mathrm{N})_\mu]^{n-1}\} \quad [j = 1, 2, \ldots, n; j \neq i] \qquad (10)$$

This event includes false negatives, false positives, and potentially one or more true negatives, making its correct classification even more difficult than the previous one. Again assuming that the main purpose of testing is to identify drug users, and because this result does indeed identify drug users, it will be classified as a true group positive. Because it contributes to total group sensitivity, it can be written in words as *Group Sensitivity Part Three* = (*1 − Individual Sensitivity*) * [*1 − (Individual Specificity)*$^{n-1}$].

Using the same assumptions as previously, the probability of this event is $(0.1) * (1 - 0.999^4) = 0.0004$.

The total group sensitivity probability is the sum of the probabilities of these three events occurring, or, in other words, the sum of Group Sensitivity Part One, Part Two, and Part Three. That is,

$$\text{Group Sensitivity} = P(\text{at least one } +|\text{at least one D})$$

Note that this definition is consistent with the meaning of individual sensitivity. Just as individual sensitivity is the probability of a positive test result, given a drug is being used, group sensitivity is the probability of one or more positive test results, given that a drug is being used. For the examples given, the probability of a true group positive, or group sensitivity, is 0.8964 + 0.0036 + 0.0004 = 0.9004. The formulas for determining group specificity and sensitivity are presented in Table 2.

For completeness, the probability that a specimen with one drug will test negative for all drugs is calculated. This is the probability that there will be a false negative for the drug in question and true negatives for the remaining drugs of interest. The probability of a false negative is one minus the probability of a true positive, or, in other words, one minus the sensitivity rate; and the probability of a true negative is the specificity rate. Thus, the group probability is the probability of a false individual negative, times the probability of a true individual negative raised to the $n - 1$ power, where n drugs are being tested for.

$$P(-_i, -_j|\mathrm{D}_i, \mathrm{N}_j) = P(-_i|\mathrm{D}_i) * [P(-|\mathrm{N})_\mu]^{n-1} \quad [j = 1, 2, \ldots, n; j \neq i] \qquad (11)$$

Assume, as before, that the individual sensitivity rate is 0.9, that the individual specificity rate is 0.999, that five drugs are being tested for, and that the specimen contains a drug. Then, the probability that the sample will test negative for that drug and negative for the four others is $(0.1) * (0.999)^4 = 0.0996$. Given that a drug is in the specimen, four events could occur, so the total probability of these four events must equal 1.0. When this last probability is added to the previous three (0.0996 + 0.9004), the total is indeed 1.0.

Finally, consider the false accusation rate for the multidrug case. On the basis of the classification scheme used herein, the group false accusation rate is the probability of not being on *any* drug, given one or more positive test results:

$$\text{Group false accusation rate} = P(\mathrm{N}_i|\text{at least one } +_i) \quad [i = 1, 2, \ldots, n] \qquad (12)$$

Thus, the more drugs being tested for, the more chances that one of them will test positive. The exact relationship between

TABLE 2 FORMULAS FOR COMPUTING GROUP SENSITIVITY AND SPECIFICITY

ACCURACY MEASURE	FORMULA
GROUP SPECIFICITY	$(\text{Individual Specificity})^n$
GROUP SENSITIVITY	
= Grp Sensitivity I	$(\text{Indiv. Sensitivity})*(\text{Indiv. Specificity})^{n-1}$
+Grp Sensitivity II	$(\text{Indiv. Sensitivity})*[1-(\text{Indiv. Specificity})^{n-1}]$
+Grp Sensitivity III	$[1-(\text{Indiv. Sensitivity})]*[1-(\text{Indiv. Specificity})^{n-1}]$

Note: It is assumed that the tests for all drugs have the same individual specificities, that the tests for all drugs have the same individual sensitivities, and that errors are mutually independent and random.

the number of drugs and the probability of a false positive specimen is illustrated later with transportation data. That is, the theory is used to estimate the potential real-world differences caused by testing for various numbers of drugs.

DRUG ABUSE BY TRANSPORTATION WORKERS

Transportation employee drug use rates vary greatly, depending on such factors as whether the tests are conducted randomly, postaccident, or for cause, and on the basis of such factors as age, gender, the drugs being tested for, and so on. Usually, an organization's positive rate will be lowest for its random tests and highest for its for-cause testing. Rates tend to be much lower for females, workers over 35, and workers in certain regions of the country. Also, because usage rates have declined over the years, more recent estimates tend to be lower than older ones (*23,24*).

In order to identify some appropriate ranges for transit properties in the early 1990s, a few recent rates from transit and other transportation organizations are reported. Averages for the transit industry as a whole, or for any particular transit agency or a subset of its employees, may be higher or lower than the rates presented.

Drug usage rates based on the random testing of Class I railroad employees during 1990 have ranged from 0.3 to 10 percent, with an average of 3 to 4 percent, according to the Federal Railroad Administration (*25*). One urban transit property in an area with very high drug use conducted random testing during January 1990 and found a drug usage rate of 2.7 percent.

Random drug tests of 65,000 current transportation employees during 1990 in a variety of transportation industries by Smith-Kline Beecham Clinical Laboratories showed a positive rate of 3.1 percent. This percentage, according to DOT officials, did not take into account those taking legal prescription drugs or the large variety of legitimate conditions that can result in false positive test results that normally are screened out in medical reviews of the tests. Thus, DOT estimated that the true positive rate was only about half of the 3.1 percent figure, or 1.55 percent (*26*).

Finally, DOT administered 9,941 tests to its own employees, all but 29 random, between 30 September 1989 and 31 March 1990. Of these, 26 were positive, for an overall rate of 0.26 percent (*27*). The drug usage rate of DOT employees is probably lower than this. Given that the 29 nonrandom tests were probably tests for cause, and given that DOT's experience is about half of for-cause tests are positive, about 14 positive results would be expected from the 29 nonrandom tests. This leaves $26 - 14 = 12$ positive test results coming from the $9{,}941 - 29 = 9{,}912$ random tests, for a rate of $12/9{,}912 = 0.0012$, or 0.12 percent.

ACCURACY OF DRUG TESTING

Because of the concerns over whether testing correctly identifies the presence or absence of drugs, a number of laboratory proficiency studies have been conducted. Prepared samples (called "challenges") are sent to laboratories to determine their testing accuracy. When the laboratories cannot distinguish the challenges from normal specimens received for routine testing, the study is a "blind" one. When the laboratories know which specimens are challenges, the study is called "open." Herein, the concern is with how laboratories perform under normal operating conditions, so only blind studies are relevant.

Rather than review all laboratory proficiency studies, estimates from a 1988 article published in the *Journal of the American Medical Association* (JAMA) are used (*28*). The study on which this article was based is considered to be the most relevant to date, because it is recent, it followed a valid research design, and it truly was a blind study.

A more recent study was conducted by the American Association for Clinical Chemistry (*29*). This was a well-conducted study, but it was not a truly blind study. Ten per-

cent of the laboratories in this study knew exactly which client would be submitting the challenges. An additional percentage, the magnitude of which was not identified in the article, knew that one of two clients would be submitting the challenges. Thus, although the laboratories did not know precisely which specimens were challenges, many of the laboratories did know precisely which clients would be submitting the challenges. While the study does indicate what the best laboratories in the country can do when they know they are being tested, it does not necessarily indicate what the laboratories might do under routine day-to-day testing of regular specimens from normal clients. Comprehensive discussions of the applicability to transit of other laboratory proficiency studies has been provided by Barnum and Gleason (*17,18*) and by Allen (*30*). Those concerned with quality assurance issues and false result rates are referred to these three publications.

On the basis of calculations from data presented for the blind phase of the JAMA laboratory proficiency study (*28*), the mean individual false positive rate (number of false positives divided by number of negative challenges) was 0.0016, representing findings on the proportion of drugless specimens where drugs were incorrectly reported to be present. (A "negative challenge" is a specimen that does not contain a drug being tested for by the study.) This statistic is an estimate of $P(+|N)_{\mu}$, and is equivalent to a mean individual specificity level of 99.84 percent. There is no evidence that the presence of a false positive for one drug is related to the presence of false positives for the other drugs.

The mean individual false negative rate (number of false negatives divided by number of positive challenges) was 0.3114 in the JAMA study (*28*). (A "positive challenge" is a specimen that does contain a drug being tested for by the study.) This statistic estimates $P(-|D)_{\mu}$, and reflects a mean individual sensitivity level of 68.86 percent. There is no evidence that the presence of a false negative for one drug is related to the presence of false negatives for other drugs.

Although the 1988 JAMA results (*28*) are used for illustrative purposes, these accuracy levels are not necessarily the averages one would find in the transit industry or at any given property. Accuracy for the industry as a whole or for an individual property may be higher or lower than the JAMA results.

As of July 1991, DOT regulations on drug testing still exclude transit. So, the rigorous uniform standards and National Institute on Drug Abuse (NIDA) laboratory certification procedures that DOT mandates for many transportation modes do not apply to transit agencies, and many properties do not follow such procedures. For example, a 1990 survey of 203 transit agencies from 44 states found that only 70 percent said they always confirmed positive test results, and only 37 percent said that they permitted employees to submit another sample from the original specimen to a laboratory of the employee's choice in the event of a positive test result (*1*). Moreover, not all properties use NIDA-certified laboratories (*18*).

If uniform transit standards eventually were required by DOT, as has been proposed (*6*), most transit properties would be expected to suffer from much lower false positive and false negative rates than were present in the JAMA study. But, even though DOT procedures often would result in extremely high accuracy, both false positives and false negatives undoubtedly still would occur. These could be the result of random and mutually independent laboratory errors, or could be caused by problems in the chain of custody, systematic laboratory errors, or in the misinterpretation of laboratory results by the medical review officer (MRO).

For example, there have been reports of serious divergence between laboratory procedures theoretically required by DOT and actual practice, as discussed in detail in two recent GAO reports (*21,31*). Moreover, there have been a number of false positives reported. In one such case, an MRO discovered the false result, and the following investigation uncovered a number of other incorrect positives that previously had gone undetected (*32*). In another report, the false results were only discovered when a worker, removed from his job several months earlier as the result of the test, filed a grievance through his union. The investigation of the grievance showed that the test was in error (*33*). In this situation also, the uncovering of this one case led to many more that had remained undetected by MROs and had not been challenged by the falsely accused workers. These cases are examples of faulty gas chromatography/mass spectrometry (GC/MS) procedures, and the resulting systematic errors are much easier to discover than the mutually independent laboratory errors that are analyzed here. Although such systematic errors probably represent a very small percentage of the total tests conducted, they show that undetected false positives can and do occur under DOT regulations. Furthermore, although their discussion is beyond the scope of this article, the possibilities for errors beyond the laboratory, caused by factors such as logistic and chain-of-custody problems, are also significant.

Likewise, although DOT wisely included provisions for requiring interpretation of results by MROs in its mandated procedures, and such analyses clearly are very valuable in reducing false positive errors, MRO interpretations are also subject to error (*34*).

In summary, false positives and false negatives undoubtedly occur under current transit industry practices, and undoubtedly would occur even if the uniform industry procedures were mandated by DOT. The actual laboratory error rates are expected to be much lower if the procedures are subject to DOT regulation, but such errors still would occur. The actual error rates in 1990, a period when the transit industry was not regulated by DOT, may be lower or higher than the ones reported by the JAMA study (*28*).

Again, JAMA results are used for illustration only, and the results do not necessarily represent the average for the industry as a whole or the results to be expected at any specific transit property.

Moreover, no matter how high the mutually independent random error rates actually are, false accusation rates can be lowered to acceptable levels by using sufficient reconfirmations of the results with currently available technology.

FALSE ACCUSATION RATES IN TESTING FOR MULTIPLE DRUGS

Consider the impact on the false accusation rate of testing for from 1 to 10 drugs, in which the drug usage rate for the target workforce is 3.0 percent, and in which the individual specificity and individual sensitivity are based on the JAMA (*28*) results, as presented in Table 3.

TABLE 3 GROUP FALSE ACCUSATION RATES BY NUMBER OF DRUGS TESTED

[1] # OF TESTS	[2] STATE	[3] P(State)	[4] Group Sensitivity, & Group False Pos Rate	[5] P(GS)*P(S), & P(GFPR)*P(S)	[6] P(S|+....)
1	Drugs	0.03	0.688600	0.020658	0.930
	No Drugs	0.97	0.001600	0.001552	0.070
	Total	1.00		P(+) = 0.022210	1.000
2	Drugs	0.03	0.689098	0.020673	0.870
	No Drugs	0.97	0.003197	0.003102	0.130
	Total	1.00		P(+) = 0.023774	1.000
3	Drugs	0.03	0.689596	0.020688	0.817
	No Drugs	0.97	0.004792	0.004649	0.183
	Total	1.00		P(+) = 0.025336	1.000
4	Drugs	0.03	0.690092	0.020703	0.770
	No Drugs	0.97	0.006385	0.006193	0.230
	Total	1.00		P(+) = 0.026896	1.000
5	Drugs	0.03	0.690588	0.020718	0.728
	No Drugs	0.97	0.007974	0.007735	0.272
	Total	1.00		P(+) = 0.028453	1.000
6	Drugs	0.03	0.691083	0.020732	0.601
	No Drugs	0.97	0.009562	0.009275	0.309
	Total	1.00		P(+) = 0.030007	1.000
7	Drugs	0.03	0.691578	0.020747	0.657
	No Drugs	0.97	0.011146	0.010812	0.343
	Total	1.00		P(+) = 0.031559	1.000
8	Drugs	0.03	0.692071	0.020762	0.627
	No Drugs	0.97	0.012729	0.012347	0.373
	Total	1.00		P(+) = 0.033109	1.000
9	Drugs	0.03	0.692564	0.020777	0.600
	No Drugs	0.97	0.014308	0.013879	0.400
	Total	1.00		P(+) = 0.034656	1.000
10	Drugs	0.03	0.693056	0.020792	0.574
	No Drugs	0.97	0.15885	0.015409	0.426
	Total	1.00		P(+) = 0.036200	1.000

In the first case in the table, one drug is being tested for. As indicated in Column 2, a urine specimen must be in one of two states: either it contains a drug or it contains no drugs. The specimen has a 0.03 probability of being in the first state, and a 0.97 probability of being in the second, as indicated in Column 3. These probabilities imply that 3.0 percent of the target population uses drugs. The next column, Column 4, identifies the probability of the urine specimen's testing positive for the drug when the drug truly is present (0.6886), and when there truly are no drugs in the sample (0.0016). That is, $P(+|\mathrm{N}) = 0.0016$, and $P(+|\mathrm{D}) = 0.6886$. (The rates in Column 4 are calculated using the formulas in Table 2, and are based on the fact that the false positive rate = 1 − specificity.)

The numbers in Column 5 are the products of the numbers in Columns 3 and 4. That is, for the population being tested, the probability that a person truly is on drugs and tests positive for drugs is 0.020658, whereas the probability that a person truly is not on drugs and tests positive for drugs is 0.001552. The sum of these two probabilities, denoted by $P(+)$ and equal to 0.022210, is the probability of a positive test result.

Dividing each of the numbers in Column 5 by $P(+)$ yields the numbers in Column 6, which are the probabilities of being in the particular states, given a positive test result. Thus, the probability that specimens that test positive will contain a drug is 0.930, meaning the test has a positive predictive value of 93.0 percent. The probability that specimens that test positive will truly contain no drugs is 0.070, meaning the test has a false accusation rate of 7.0 percent. That is, $P(\mathrm{D}|+) = 0.930$, and $P(\mathrm{N}|+) = 0.070$, with the two probabilities totaling 1.0.

The false accusation rates presented are based on illustrative rates for drug usage, individual sensitivity, and individual specificity, and are not necessarily applicable to any particular transit agency. But, all the illustrative rates are ones that could

occur in some circumstances. Because of the extremely serious consequences of false accusations of drug use, an employer would be wise to ensure for itself that such estimated rates, or similar rates, do not apply to its case.

In the first case discussed, as previously noted, it is assumed that only one drug is being tested for, similar to the assumptions that led to all of the outcomes discussed by Barnum and Gleason (*17,18*). However, the situation changes when the number of drugs being tested for increases. As indicated in Table 3, as the number of drugs tested for increases, the false accusation rate increases. As seen from Column 6, the one-drug false accusation rate of 7.0 percent increases to 27.2 percent (almost 4 times the one-drug rate) when 5 drugs are tested for, and the rate increases to 42.6 percent (over 6 times the one-drug rate) when tests for 10 drugs are involved.

It is easy to lower the false accusation rates to acceptable levels with automatic multiple confirmation testing, as discussed by Barnum and Gleason (*17,18*), but use of automatic multiple confirmation is not often required. Although employees are sometimes allowed the opportunity to request a second confirmation, the fact that they must request it makes it less likely to occur. The empirical evidence suggests that many falsely accused employees may not request second confirmations (*32,33*). This fact is not too surprising, because many have been told that the tests are foolproof. And, in many cases, the employees themselves have to pay for the second confirmation. Moreover, sometimes more than two confirmations may be necessary, a circumstance that is not typically provided for.

CONCLUSIONS AND IMPLICATIONS OF THE STUDY

Past research in drug testing accuracy (*17,18*) has highlighted the importance of identifying the percentage of those testing positive who are not on drugs, herein called the false accusation rate. However, the impact on this rate of the number of different drugs being tested for has never been addressed. As indicated in the examples, the false accusation rate increases at approximately the same rate as the number of drugs being tested for. That is, the false accusation rate for five drugs is about 4 times the false accusation rate for one drug. Because, in the past, false accusation rates have been based on the implicit assumption that one drug is being tested for, the rates presented have badly understated the true facts.

False accusation rates caused by random and mutually independent errors can be lowered to any desired level by several different means. But, to truly achieve the required rate, it is necessary to take into consideration the number of drugs being tested for.

In order to attain a false accusation rate below some desired maximum when testing for multiple drugs, the following steps must be taken.

1. Determine the maximum false accusation rate that is considered acceptable by the relevant decision makers (including union representatives, if applicable).
2. Estimate the lowest likely rate of drug usage by the workers to be tested.
3. Estimate the highest likely individual false positive and false negative rates for the organization involved.
4. Using the estimated individual false positive and false negative rates, the number of drugs to be tested for, and the pertinent formulas, estimate the group false positive and group false negative rates.
5. Assuming that one test includes a screen and confirmation, and using the expected drug usage rate and the group false positive and false negative rates, estimate the actual false accusation rate.
6. Compare the actual false accusation rate to the maximum acceptable rate. If the latter is higher, the procedure is sufficiently accurate. If the former is higher, the procedure is not sufficiently accurate; either additional confirmations of positive tests must be conducted, individual specificity or sensitivity must be increased, or fewer drugs must be tested for.

This process ensures that false accusations caused by random and mutually independent laboratory errors occur at a lower rate than the desired maximum. It does not address systematic errors or errors occurring outside of the laboratory. These too are important, and could make the total error rate substantially higher, but are beyond the scope of this paper.

In closing, we would like to make a few personal observations. We feel that DOT is working hard to establish rigorous and just testing standards, and the DOT-mandated procedures, once fully implemented, will likely result in substantial improvements in average test accuracy in the transportation modes to which they apply. These error rates can be expected to be substantially lower than those reported by JAMA (*28*).

In the authors' opinion however, there will never be perfect accuracy in drug testing, and there does not need to be. Some individuals will be falsely accused and convicted under all systems of justice. It is not by chance that both the American legal and workplace jurisprudence systems have identified levels of proof that range from preponderance of evidence, to clear and convincing evidence, to proof beyond a reasonable doubt (with the increasingly higher standards being applied as punishments become more severe). But, even for the most severe punishments, absolutely perfect proof is not required, and it is expected that there will be cases of error. The critical question is not how to avoid false accusations, which is impossible unless everyone is assumed innocent, but how to be sure that the errors that do occur will be at an acceptably low rate. One good way to achieve desired error rates is to control them by methods such as those identified in this paper.

REFERENCES

1. L. E. Henriksson. *Consequences of Drug Testing Programs in Urban Mass Transit*. Ph.D. dissertation. Graduate School of Business, Indiana University, Bloomington, Ind., 1991.
2. L. E. Henriksson. The Unconvincing Case for Drug Testing. *Canadian Public Policy*, (In press).
3. D. A. Lee. Employee Assistance Programs in the Public Transit Industry: The Experience of Connecticut Transit and Some Concerns for the Future. In *Transportation Research Record 1266*, TRB, National Research Council, Washington, D.C., 1990, pp. 3–9.

4. D. T. Bliss. Employee Drug Testing: Lessons to be Learned from the Transportation Initiative. *Federal Bar News & Journal*. Vol. 35, No. 6, July/Aug. 1988, pp. 280–285.
5. N. Eisner. Drug Testing: Regulatory and Legal Issues Confronted by the Department of Transportation. *Federal Bar News & Journal*, Vol. 35, No. 8, Oct. 1988, pp. 364–368.
6. Senate Committee Okays Bill Authorizing UMTA Screening. *The National Report on Substance Abuse*, Vol. 4, No. 17, Aug. 1, 1990, p. 2.
7. A. J. McBay. Drugs and Transportation Safety. *Journal of Forensic Sciences*, Vol. 35, No. 3, May 1990, pp. 523–529.
8. C. Zwerling, J. Ryan, and E. J. Orav. The Efficacy of Preemployment Drug Screening for Marijuana and Cocaine in Predicting Employment Outcome. *Journal of the American Medical Association*, Vol. 264, No. 20, Nov. 28, 1990, pp. 2639–2643.
9. R. Winslow. Study May Spur Job-Applicant Drug Screening. *The Wall Street Journal*, Nov. 28, 1990.
10. G. E. Dodge. OMB Issues Final Rules for Drug-Free Workplace Act. *Employment Testing: A Biweekly Reporter on Drug, Polygraph, AIDS, and Genetic Testing*. Vol. 4, No. 11, July 15, 1990, pp. 613–615.
11. R. P. DeCresce, M. S. Lifshitz, A. C. Mazura, J. E. Tilson, J. Ambre, and K. M. Cochran. *Drug Testing in the Workplace*. Bureau of National Affairs, Washington, D.C., 1989.
12. K. B. Zeese. *Drug Testing Legal Manual*. Clark Boardman, New York, 1989.
13. C. Morris, ed. *The Developing Labor Law*. Bureau of National Affairs, Washington, D.C., 1983.
14. M. F. Hill, Jr., and A. V. Sinicropi. *Evidence in Arbitration*. Bureau of National Affairs, Washington, D.C., 1987.
15. F. Elkouri and E. A. Elkouri. *How Arbitration Works*. Bureau of National Affairs, Washington, D.C., 1985.
16. K. W. Thornicroft. Arbitrators and Substance Abuse Discharge Grievances: An Empirical Assessment. *Labor Studies Journal*, Vol. 14, No. 4, Winter 1989, pp. 40–65.
17. D. T. Barnum and J. M. Gleason. Accuracy in Transit Drug Testing: A Probabilistic Analysis. In *Transportation Research Record 1266*, TRB, National Research Council, Washington, D.C., 1990, pp. 10–18.
18. D. T. Barnum and J. M. Gleason. Closure: Bayesian Approaches to Maintaining Proper Accuracy Standards in Transit Drug Testing. In *Transportation Research Record 1266*, TRB, National Research Council, Washington, D.C., 1990, pp. 20–22.
19. M. A. Rothstein. *Medical Screening*. Bureau of National Affairs, Washington, D.C., 1989.
20. R. V. Blanke. Quality Assurance in Drug-Use Testing. *Clinical Chemistry*, Vol. 33, No. 11(B), 1987, pp. 41B–45B.
21. *Drug Testing: Management Problems and Legal Challenges Facing DOT's Industry Programs*. GAO/RCED-90-31. U.S. General Accounting Office, Nov. 1989.
22. J. D. Osterloh and C. E. Becker. Chemical Dependency and Drug Testing in the Workplace. In Addiction Medicine [Special Issue]. *The Western Journal of Medicine*, Vol. 152, No. 5, May 1990, pp. 506–513.
23. A. Kopstein and J. Gfroerer. *Drug Use Patterns and Demographics of Employed Drug Users: Data from the 1988 National Household Survey on Drug Abuse*. Working Paper, Division of Epidemiology and Prevention Research, National Institute on Drug Abuse, Rockville, Md., 1990.
24. *National Household Survey on Drug Abuse: Population Estimates 1988*. National Institute on Drug Abuse, U.S. Department of Health and Human Services. U.S. Government Printing Office, 1989.
25. R. Watson. Short Lines Prepare for Random Drug Tests. *Modern Railroads: Short Lines and Regionals*. Vol. 45, No. 16, Sept./Oct. 1990, pp. 39–42.
26. 3.1% of Transport Workers Test Positive for Drug Use. *Omaha World Herald*, July 15, 1990.
27. DOT Leads Federal Agencies in Testing Programs, Data Show *National Report on Substance Abuse*, Vol. 4, No. 19, Sept. 12, 1990, pp. 8–9.
28. K. H. Davis, R. L. Hawks, and R. V. Blanke. Assessment of Laboratory Quality in Urine Drug Testing: A Proficiency Testing Pilot Study. *Journal of the American Medical Association*, Vol. 260, No. 12, Sept. 23/30, 1988, pp. 1749–1754.
29. C. S. Frings, D. J. Battaglia, and R. M. White. Status of Drugs-of-Abuse Testing in Urine Under Blind Conditions: An AACC Study. *Clinical Chemistry*, Vol. 35, No. 5, May 1989, pp. 891–894.
30. M. Allen. Discussion: Rebuttal of "Accuracy in Transit Drug Testing: A Probabilistic Analysis." In *Transportation Research Record 1266*, TRB, National Research Council, Washington, D.C., 1990, pp. 18–20.
31. *Employee Drug Testing: DOT's Laboratory Quality Assurance Program Not Fully Implemented*. GAO/GGD-89-80. U.S. General Accounting Office, Sept. 1989.
32. Lab Suspended Over False Positives; NIDA Will Assess Method's Reliability. *National Report on Substance Abuse*, Vol. 4, No. 23, 7 Nov. 1990, pp. 1, 8.
33. Federal Agency Finds Drug Test May Be Flawed. *Omaha World-Herald*, Oct. 25, 1990.
34. H. W. Clark. The Role of Physicians as Medical Review Officers in Workplace Drug Testing Programs: In Pursuit of the Last Nanogram. *Western Journal of Medicine*, Vol. 152, May 1990, pp. 514–524.

DISCUSSION

DONNA R. SMITH
Drug Enforcement and Program Compliance, U.S. Department of Transportation, 400 Seventh St., S.W., Washington, D.C. 20590.

The authors' conclusion that the false accusation rate (FAR) is an appropriate measure of drug testing accuracy is misleading. According to the authors' definition, the FAR is a function of the sensitivity and specificity of drug testing and the prevalence of drug use in the population being tested. Drug use prevalence does not affect laboratory accuracy. The application of the author's statistical methodologies to workplace drug testing is also based on experience from clinical diagnostic testing, which is not necessarily germane to analytical toxicology. Furthermore, the conclusions are based on an assumption that random testing errors occur at an equal rate for any combination of drugs being tested for. This assumption does not consider the procedures available to control for random errors in an independent serial testing protocol.

The procedures required in National Institute on Drug Abuse (NIDA) certified laboratories control for random laboratory error, in part, by the independent administration of two types of testing methodologies [immunoassay and gas chromatography/mass spectrometry (GC/MS)]. Specimens that are positive on the immunoassay are then tested using a separate aliquot subjected to GC/MS confirmation; it is not simply a continuation of the testing process using the same aliquot. Thus, the assumption that random laboratory error can be determined on the basis of dependent factors in the testing process is unsubstantiated. The authors use the insufficient reason approach to make their assumption that the individual specificity, sensitivity, and predictive value for each drug tested for are equal. In fact, sensitivity, specificity, and predictive values are very different for the various classes of drugs.

Although the authors emphasize that their many assumptions may not apply to DOT-mandated testing conducted in NIDA-certified laboratories, they claim that their conclusions

and recommendations are universally applicable to drug testing in the workplace. This progression is questionable. DOT does not assert that random laboratory errors are impossible. However, current procedures and NIDA laboratory protocols that respond to different sensitivity and specificity rates for individual drugs through controlled cutoff levels; independent serial testing procedures; and internal laboratory controls, standards, and calibration greatly minimize random laboratory error.

The authors discuss false positive rates, FAR values, and false conviction rates (the latter, undefined). False positive rates are based on laboratory findings (i.e., identifying a drug or drug metabolite when it is in fact not present) and are not based on whether or not the individual illicitly used a controlled substance. When a specimen contains an identified drug or metabolite, the determination of whether that metabolite got there as a result of prescribed medication or illicit drug use has no relationship to its being a false positive.

Blind proficiency data from DOT-mandated testing have yet to produce a false positive. There have been reported false negatives. The authors' observation of the lack of published data on the blind proficiency testing program for DOT-mandated testing is valid. The blind proficiency program is not a one-shot research design. It is a programmatic requirement and the only source of information for the results is from the individual employers participating in the ongoing proficiency testing program. There is no evidence to support that employers are withholding information concerning false positive events in the blind proficiency programs. The employer-supplied blind proficiency specimens, the NIDA proficiency testing program, and the laboratories' own internal open and blind proficiency testing programs combine to provide a comprehensive quality control program that monitors laboratory accuracy (and random laboratory error) on an ongoing basis.

Because of the factors discussed by the authors and the findings in the *Journal of the American Medical Association* (JAMA) articles cited by the authors, DOT and NIDA adopted the rigorous standards and procedures that currently exist in NIDA-certified laboratories. The GAO reports cited by the authors did not explore the issue of laboratory accuracy or random laboratory error. GAO's concern was the implementation of workplace drug testing policies and programs. The analytic framework used in the authors' paper is somewhat inappropriate for viewing the accuracy of a workplace drug testing program. It is more appropriate for use in medical diagnostic work. In medical diagnosis, the positive predictive value (PPV) has an individualistic interpretation (i.e., the probability that an individual has a disease given that they test positive on whatever indicative test is performed). In a medical diagnosis scenario, the prevalence of the disease plays a critical role in interpreting a test result. In assessing drug testing programs, the prevalence of drug use in the population should not necessarily play a role in the accuracy of the testing methodology. This should be done on the basis of the test (analytic process) itself (i.e., the false positive rate or false negative rate).

The FAR is given as the number of false positives divided by the number testing positive. The number testing positive is influenced by the prevalence rate. A larger prevalence rate should, therefore, have a smaller FAR than a program with a small prevalence rate—even though both may have the same number of false positives. The FAR does not therefore present an equitable evaluation across the universe of drug testing programs. Drug use prevalence is thus irrelevant to drug testing methodology accuracy.

False positive rates are derived from laboratory accuracy and reliability and are not affected by drug prevalence or incidence rates. The protection of the individual employee is paramount in any workplace drug testing program and that is why DOT believes that the rigorous procedures and standards imposed on NIDA-certified drug testing laboratories are essential. Hopefully, the transit industry, though not currently required to adopt DOT and NIDA guidelines, will pursue the use of such standards in their nonregulated programs.

Authors' Closure

Dr. Smith's remarks almost entirely concern DOT and NIDA procedures, yet, as she admits, these procedures currently are inapplicable to transit. Moreover, she does not address the topic of our paper: the relationship between the number of drugs tested for and the percentage of people who are falsely accused of drug use. That is, even if everything Dr. Smith said were true, it would have no bearing on transit and no bearing on the accuracy of testing for multiple drugs. Further, her discussion contains many incorrect and misleading statements. Unfortunately, treatment of all these issues is precluded by the space limit placed on our response.

The objective of our paper was to exhibit the impact that testing for several drugs, rather than for one drug, has on the accuracy of the drug-testing process. Because of the way in which sensitivity and specificity have been calculated heretofore, all prior drug testing analyses have implicitly assumed that a specimen was being tested for only one drug. That is, past analyses have failed to take into account the simple laws of probability that prevail in cases in which specimens are tested for several drugs. Our concern was to point out the probabilistic implications of the multiple-drug case, in order that potential problems could be dealt with in the design of accurate testing processes. Our conclusion was that an increase in the number of drugs tested for in a specimen also increases the probability that a given specimen will be falsely classified as positive. Dr. Smith never addresses this conclusion or the discussion justifying it.

Dr. Smith states "DOT does not assert that random laboratory errors are impossible. However, current procedures and NIDA laboratory protocols . . . greatly minimize random laboratory error." Thus, the discussant acknowledges that random errors can occur. It is notable, however, that DOT's blind proficiency sampling procedure has not been powerful enough to discover any false positives, although false positives have been reported by others (*1*,p.8).

We agree with Dr. Smith's statement that drug use prevalence does not affect laboratory accuracy—however, we never claimed it did. Our point, in the current paper and in previous papers (*2,3*), is that sensitivity and specificity rates should not be viewed in the absolute. Rather, the impact of sensitivity and specificity must be viewed from the perspective of their interaction with drug use prevalence rates, as determined by a Bayesian analysis. Clearly, as indicated by the results pro-

vided in our previous paper (*2*), which deals with the Bayesian concepts, the false accusation rates differ depending on prevalence of drug use. That is, for given sensitivity and specificity levels, the false accusation rate will be higher in groups with low drug use prevalence than in high-prevalence groups.

By focusing on laboratory error rates, Dr. Smith misses the critical point. Laboratory accuracy is only a means to an end, the end being to avoid false accusations. For example, the fact that a laboratory reports only one false positive out of every million drug-free samples is completely irrelevant if this still results in 9 out of every 10 positives being false accusations. A given laboratory accuracy level may result in very high or very low false accusation rates. It is important to start with acceptable false accusation rates and work backwards to the laboratory accuracy levels required for each target group. Our effort has been to suggest procedures to identify the potential impact of random errors on various target groups, and to suggest processes to deal with these potential errors.

We appreciate Dr. Smith's discussion, and the opportunity it has given us to reemphasize these issues.

REFERENCES

1. NIDA Suspends a Third Laboratory for False Positive Results. *The National Report on Substance Abuse*. Vol. 5, No. 4, Jan. 30, 1991.
2. D. T. Barnum and J. M. Gleason. Accuracy in Transit Drug Testing: A Probabilistic Analysis. In *Transportation Research Record 1266*, TRB, National Research Council, Washington, D.C., 1990, pp. 10–18.
3. D. T. Barnum and J. M. Gleason. Authors' Closure: Bayesian Approaches for Maintaining Accuracy Standards in Transit Drug Testing. In *Transportation Research Record 1266*, TRB, National Research Council, Washington, D.C., 1990, pp. 20–22.

Publication of this paper sponsored by Committee on Transit Management and Performance.