KAYE, D. H. (1987). Hypothesis tests in the courtroom. In *Contributions to the Theory and Application of Statistics* (A. Gelfand, ed.). Academic, Orlando, Fla.

LYKKEN, D. (1987). The validity of tests: Caveat emptor. *Jurimetrics J.* **27** 263–270.

SAKS, M. and KIDD, R. (1980–1981). Human information processing and adjudication: Trial by heuristics. *Law and Society* **15** 123–160.

SHAFER, G. (1986). The construction of probability arguments. *Boston Univ. Law Rev.* **66** 799–816.

Comment: Base Rates and the Statistical Precision of Polygraph Tests in Various Applications

John C. Kircher and David C. Raskin

In his analysis of the precision of medical screening procedures, Gastwirth discussed the effects of low base rates on the accuracy and utility of test data. The problem of low base rates has been discussed for many years in the psychologic literature (Meehl and Rosen, 1955). In general, when the prevalence of a characteristic such as AIDS or deception in the population is low, it is difficult for a test to improve upon the accuracy that would be obtained if only information about the base rate were used to make diagnoses. If the base rate of a disease is only 0.1%, then diagnosing all patients as disease-free would produce a diagnosis accuracy of 99.9%. To improve upon the accuracy attainable using only base rate information, the accuracy of a test to detect the disease would have to exceed 99.9%. Unfortunately, tests with that level of diagnostic accuracy are extremely rare, and populations with extreme base rates such as those encountered in screening situations are not uncommon.

In the polygraph literature, Raskin (1984) first called attention to the problem of low base rates in hearings before the Committee on Armed Services of the United States Senate on the proposed Department of Defense counterintelligence polygraph program. The Department of Defense was considering widespread testing of federal employees and defense contractors concerning unauthorized disclosures of sensitive information. The base rate issue was particularly important in that context because the vast majority of federal employees and contractors do not make unauthorized disclosures of sensitive information. The base rate of guilt in that population is probably less than 1 in 1000. As discussed by Raskin (1984, 1986) and as Gastwirth's analyses clearly confirm, deceptive polygraph outcomes under those circumstances would be considerably less than 50% correct, even if it is assumed that the polygraph is 90–95% accurate on populations with equal base rates of truthful and deceptive individuals.

Gastwirth focused on a different but related problem. His work reveals that the sampling error of estimates of test accuracy increases as the incidence of the trait in the tested population departs from 50%. In addition to reducing confidence in test outcomes, skewed base rates increase the error in estimating test validity. This makes an already bad situation worse. To our knowledge, this important issue has not been addressed in the polygraph literature, nor has it been discussed in the broader literature on psychologic assessment.

Although we agree with the statistical conclusions drawn by Gastwirth, the implications of his work for applications of polygraph techniques merit further comment. Polygraph tests are used in many different contexts. Law enforcement and private polygraph examiners administer polygraph tests to suspects, defendants and witnesses during criminal investigations. Businesses make extensive use of polygraph tests to screen job applicants and to test employees periodically. Government agencies use polygraphs in criminal investigations and in cases involving risks to national security. The base rate of deception and the costs associated with false positive decision errors are more problematic in some contexts than in others.

Analyses of data from the United States Secret Service for a 2-yr period suggest that the base rate of guilt is about 45% in their criminal investigations (Raskin, 1986). Raskin also reported findings from 292 polygraph tests that he had conducted over a 12-yr period on a confidential basis for defense attor-

John C. Kircher is Assistant Professor of Educational Psychology and David C. Raskin is Professor of Psychology at the University of Utah, Salt Lake City, Utah 84112.

neys. The estimated base rate of guilt in that sample was 60%. If those base rates are representative of the populations of criminal suspects and defendants who agree to take polygraph examinations from law enforcement and private polygraph examiners, then the standard error of estimated polygraph accuracy in those applications would be small. According to Gastwirth's calculations it would be on the order of 4–5%.

We are not too concerned about a sampling error of 5% since the variability of results in the polygraph literature is considerably greater than that (Office of Technology Assessment, 1983). Indeed, we found that sampling error accounted for less than 24% of the variance in levels of diagnostic accuracy in 14 laboratory studies of polygraph techniques (Kircher, Horowitz and Raskin, 1987). Thus, Gastwirth's view that scientists and consumers of polygraph test data should be circumspect in the conclusions they draw from the detection rates obtained in any given study is justified for more than the reasons he stated.

Although sampling error may not play a significant role in estimating the validity of polygraph techniques in criminal investigations, Gastwirth has shown that it is important in screening situations. When only a small percentage of tested individuals are deceptive, the probability that a deceptive polygraph outcome is correct may be less than 50%, and in some cases it may be far less. On the other hand, the probability that a truthful outcome is correct approaches unity. Thus, in most screening situations, one can place little confidence in a deceptive polygraph outcome and much greater confidence in a truthful one.

In the private sector, applications of polygraph techniques to screen job applicants and to test employees periodically are subject to the statistical problems and misinterpretations discussed by Gastwirth. In addition to the high rate of false positive outcomes and substantial sampling error, Gastwirth noted that deceptive results are not confidential and the individual has limited legal recourse when arbitrarily dismissed or denied employment because he or she has failed a polygraph test. It should also be noted that such tests are not taken voluntarily. A person who refuses to take a preemployment or periodic polygraph test may be denied employment or terminated. In contrast, a criminal suspect or defendant may refuse to take a polygraph test without prejudice.

A major problem not discussed by Gastwirth concerns the lack of empirical research on the validity of polygraph techniques in screening situations. We are unaware of any scientific studies that support routine use of polygraph techniques for preemployment screening and periodic testing. Moreover, Gastwirth adopted accuracy rates for his analyses from Raskin (1986). Those accuracies were obtained from laboratory experiments in which half of the subjects committed a mock crime. There is considerable controversy over whether the accuracies observed in mock crime experiments are representative of those obtained in actual criminal investigations (Lykken, 1979; Raskin and Podlesny, 1979). Critics argue that the laboratory data tell us nothing about the accuracy of forensic polygraph examinations, whereas we have taken the position that well-conducted laboratory experiments are essential and do provide an important source of converging evidence with which to assess the accuracy of polygraph techniques as they are or could be applied *in criminal investigations* (Podlesny and Raskin, 1977; Kircher, Horowitz and Raskin, 1987).

It is a serious mistake to assume that the accuracies of test results obtained from subjects who participate in mock crime experiments provide adequate estimates of test accuracy on individuals who undergo preemployment and periodic polygraph tests at the request of employers. The assumptions, procedures and goals of forensic polygraph tests differ in many important respects from those conducted as a condition of employment. Commercial polygraph examiners usually receive inadequate training and frequently conduct tests that fail to meet minimum standards for any serious application (Raskin, 1986; 1987). If Gastwirth overestimated the accuracy of screening tests in employment situations, as we suspect he did, then the problem of sampling error is even more serious than his analyses indicate. In any case, Gastwirth has provided additional support for the position that the polygraph should not be used in the private sector to decide who should be hired and who should be denied employment.

Our final comment concerns the use of polygraph techniques in national security screening programs. This application is also characterized by low base rates of deception, high rates of false positive errors and large standard errors of estimated accuracy. However, in contrast to the screening tests administered in the private sector, some federal agencies are beginning to use test formats that are similar to those used in criminal investigations. Because of the similarity between the two types of tests, the accuracies borrowed from Raskin (1986) may provide reasonable estimates of the accuracy of the screening test if it were used with a population of equal numbers of truthful and deceptive individuals. Of course, this does not mitigate the problems of false positive outcomes and sampling error when the base rate of deception is low. In fact, Gastwirth's assumptions concerning the sensitivity and selectivity of the screening test are more appropriate for this particular application of polygraph techniques than any other.

Recognizing the consequences of using an otherwise diagnostic test in situations where the base rate of deception is low, we would simply draw attention to a

possible justification for using a screening test in spite of these problems. The recent, highly publicized Walker spy case is but one example of several recent disasters in our national security system. The consequences of failing to detect leaks of secret information to foreign governments may be severe. A polygraph test that correctly identifies 88% of deceptive individuals tested, misclassifies only 3% and yields 9% inconclusive outcomes could be relied upon to identify most security risks. However, since the base rate of deception in this population is so low, most of the individuals who would fail the test would in fact be truthful. If a deceptive polygraph outcome is more often wrong than it is correct, it is clear that it should not be the sole basis for concluding that a person is a spy, for denying individuals access to secure information or for taking other action against them.

On the other hand, if the screening test is used only to eliminate from further consideration all those who pass the test, then the number of potential security risks would be reduced by a factor of approximately 10 (Raskin and Kircher, 1987). Extensive field investigations would then be required on a much smaller number of individuals with a somewhat higher base rate of deception than in the original sample. With this "successive hurdles" approach (Meehl and Rosen, 1955), polygraph screening tests could be used in the vast majority of cases in lieu of costly field investigations. The required follow-up investigations of those who fail the initial screening test would minimize the risk of false positive errors and probably identify the individuals who are guilty of compromising our national security.

ADDITIONAL REFERENCES

KIRCHER, J. C., HOROWITZ, S. W. and RASKIN, D. C. (1987). Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior*. To appear.

LYKKEN, D. T. (1979). The detection of deception. *Psychol. Bull.* **86** 47–53.

MEEHL, P. and ROSEN, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.* **52** 194–214.

PODLESNY, J. A. and RASKIN, D. C. (1977). Physiological measures and the detection of deception. *Psychol. Bull.* **84** 782–799.

RASKIN, D. C. (1984, March). Proposed use of polygraphs in the department of defense. Statement before the Committee on Armed Services, U.S. Senate.

RASKIN, D. C. (1987). Does science support polygraph testing? In *The Polygraph Test: Truth, Lies and Science* (A. Gale, ed.). Sage, London. To appear.

RASKIN, D. C. and KIRCHER, J. C. (1987). The validity of Lykken's criticisms: Fact or fancy? *Jurimetrics J.* **27** 271–277.

RASKIN, D. C. and PODLESNY, J. A. (1979). Truth and deception: A reply to Lykken. *Psychol. Bull.* **86** 54–59.

Comment

Janet Wittes

Professor Gastwirth's most interesting paper, coupled with my craving for poppy seed bagels and my passion for our Fourth Amendment right to privacy, has led me to a new appreciation of the importance of specificity θ in medical screening. My work with Dr. Goldberg (Goldberg and Wittes, 1978, 1981) has focused on the sensitivity η; the inverse symmetry of Dr. Gastwirth's equations (3.1) and (3.3) point to diametrically opposed criteria for optimality depending on whether one is interested primarily in the predicted value positive (PVP) or the predicted value negative (PVN). In the former case, Gastwirth shows that θ should be estimated most precisely; in the latter, the emphasis should be placed on η.

Janet Wittes is Chief, Biostatistics Research Branch, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892.

The context of the screening determines whether the sensitivity or the specificity is more important. For the cases that Dr. Goldberg and I have considered in the past, screening was performed for the benefit of the screenee. A woman elects to participate in a breast cancer screening, for example, because she is seeking an early diagnosis of a disease for which early diagnosis can translate to her own lengthened survival (Shapiro, Strax, Venet and Venet, 1973). Hence, from her, the consumer's point of view, a screening program consisting of a highly sensitive test, followed by a highly specific test if she is positive, is a sensible course of action. Consideration of the PVP is then secondary to the needs of the consumer. When, however, the consumer is not the screenee, but the society at large, and when that society assumes an implicitly adversarial position with respect to the screenee, Gastwirth's emphasis on the primacy of the specificity