but inaccurate, estimate will lead us astray in practice. Our inferences can be erroneous and lead us to poor policy decisions.

The large effects on estimation and inference that can be attributed to misclassification suggest that resources should be allocated to estimation of these error rates prior to the implementation of a mass screening program and on an ongoing basis for the duration of the program. The costs of classification errors are high to both individuals and society. The existence of a screening program itself may alter behavior of individuals, and the disease process may change from the intervention after screening and from improvements in both the screening method and therapy. These and other related issues in the evaluation of medical screening procedures are discussed in Goldberg and Wittes (1981).

ADDITIONAL REFERENCES

GOLDBERG, J. D. and WITTES, J. T. (1981). The evaluation of medical screening procedures. *Amer. Statist.* **35** 4–11.

SHAPIRO, S., STRAX, P., VENET, L. and VENET, W. (1973). Changes in five-year breast cancer mortality in a breast cancer screening program. *Seventh Nat. Cancer Conf. Proc.* 663–678.

VECCHIO, T. J. (1966). Predictive value of a single diagnostic test in unselected populations. *New England J. Med.* **274** 1171–1173.

Comment

Seymour Geisser

We are indebted to Professor Gastwirth for an enlightening discussion regarding the reliability of the results of screening tests in two rather important areas: AIDS and lie detectors. His main concern is with the conditional probabilities of correct classification and the sampling error of their frequentist estimators.

I would like to outline an approach that I believe might be more informative and illuminating for inferring the results of such screening tests. For the sake of simplicity, let us assume that there is a properly identified population and a single test (multiple tests and varying populations would only further serve to complicate the situation but not change the conceptual framework for handling such problems).

With the use of Prof. Gastwirth's notation, we have a table exhibiting the following probabilities:

	D	$\bar{D}$	
S	$\pi\eta$	$(1-\pi)(1-\theta)$	$\pi\eta + (1-\pi)(1-\theta)$
$\bar{S}$	$\pi(1-\eta)$	$(1-\pi)\theta$	$\pi(1-\eta) + (1-\pi)\theta$
	π	$1-\pi$	1

where, e.g., $P(D) = \pi$, $P(S \mid D) = \eta$, $P(\bar{S} \mid \bar{D}) = \theta$; $P(S) = \pi\eta + (1-\pi)(1-\theta) = p$. The critical so-called PVP,

$$P(D \mid S) = \frac{\pi\eta}{\pi\eta + (1-\pi)(1-\theta)} = \tau, \quad \text{say,}$$

and the probability of a false negative,

$$P(D \mid \bar{S}) = \frac{\pi(1-\eta)}{1 - \pi\eta - (1-\pi)(1-\theta)} = \rho, \quad \text{say,}$$

are functions of the three parameters, π, η and θ.

The type of sampling that Professor Gastwirth deals with in the paper presumably would yield a likelihood function for θ, η and π,

$$L(\theta, \eta, \pi) \propto \eta^{r_1}(1-\eta)^{n_1-r_1}\theta^{r_2}(1-\theta)^{n_2-r_2}\left(\frac{\pi\eta}{\tau}\right)^{t}\left(1 - \frac{\pi\eta}{\tau}\right)^{n-t},$$

recalling that τ is a function of θ, η and π. Suppose a joint prior for η, θ and π, $g(\eta, \theta, \pi)$ is available. Then the posterior density of θ, η and π is

$$p(\theta, \eta, \pi \mid d) \propto L(\theta, \eta, \pi)g(\eta, \theta, \pi),$$

where $d = (r_1, r_2, n_1, n_2, t, n)$.

Clearly, if we were diligent and clever enough, we could find from $p(\theta, \eta, \pi \mid d)$ the joint posterior density of τ and ρ, say $p(\tau, \rho \mid d)$. Ostensibly then for any set S on the unit square we could find

$$P[(\tau, \rho) \in S] = P,$$

or conversely for any fixed P we could find the "smallest" set S_P such that

$$P[(\tau, \rho) \in S_P] = P.$$

Similar results could be obtained marginally for either ρ or τ. This would be much more informative than the calculation of the approximate standard errors of the estimates $\hat{C}$ and $\hat{F}$. Of course this would require a good deal of heavy calculation involving numerical

Seymour Geisser is Professor and Director, School of Statistics, University of Minnesota, 270 Vincent Hall, 206 Church Street, S.E., Minneapolis, Minnesota 55455.

integration and approximation as well as some prior knowledge about θ, η, π which no doubt is available in the AIDS case, for example. Note that Professor Gastwirth indicates that π is probably around .01 in a particular low risk AIDS population.

For reasonably large data sets, uniform distributions for the three parameters would not in all likelihood induce posterior distributions that appreciably differed from ones based on somewhat more informative priors. Indeed one could also explore various conditional posterior probabilities of τ, say, given one or two of θ, η, π or functions of them.

The expectation of τ or ρ is of course the minimum mean squared error predictor of τ or ρ.

If in the future one could only observe individuals who would test positive for D and M among those actually are in class D, then τ is the limiting proportion, i.e.,

$$\tau = \lim_{N\to\infty} \frac{M}{N},$$

with distribution function $P(\tau \mid d)$, assuming stability of the population. Hence, τ and $P(\tau \mid d)$ so interpreted inform us about the potential fraction of the class S who are also D, when observation is limited to S. However, for a new individual or testee who has been classified as S and regards himself as exchangeable with the previous n testees, he would calculate his predictive probability of being D as

$$\int \tau \, dP(\tau \mid d')$$

where $P(\tau \mid d')$ is the posterior distribution of τ derived from the likelihood modified by adding 1 to t and n, respectively.

This is the probability that concerns him. What should concern the health authorities is developing screening tests that give rise to distributions for τ that are concentrated more and more closely about one.

Consider the situation wherein a hospital administrator will receive only individuals who are classified as S (the laboratory does not inform him how many were $\bar{S}$). These individuals will, within a fixed time period, either exhibit D or $\bar{D}$ and those that exhibit D will require a bed. The administrator would like to compute the number M of beds he will need given that he will be assigned K individuals classified as S. In such situations where there is a censoring of $\bar{S}$ we can compute

$$P[M = m] = \binom{K}{m} \int \tau^m (1-\tau)^{K-m} \, dP(\tau \mid d).$$

Note that

$$\lim_{K\to\infty} \frac{M}{K} = \tau$$

with $P(\tau \mid d)$ as the probability distribution of τ. For sufficiently large K we could use

$$P(\tau \leqq y \mid d)$$

as an approximation for

$$P(MK^{-1} \leqq y)$$

if in fact this calculation becomes too burdensome to execute.

Suppose, on the other hand, the total sample of individuals J screened in the laboratory was known to the administrator. Assuming these individuals were exchangeable with the original n testees then the calculation follows along these lines:

$$P(M = m, K = k) = \binom{k}{m}\binom{J}{k} \int \tau^m (1-\tau)^{k-m} \cdot \left(\frac{\pi\eta}{\tau}\right)^k \left(1 - \frac{\pi\eta}{\tau}\right)^{J-k} dP(\pi, \eta, \theta \mid d),$$

$$P(K = k) = \binom{J}{k} \int \left(\frac{\pi n}{\tau}\right)^k \left(1 - \frac{\pi n}{\tau}\right)^{J-k} dP(\pi, n, \theta \mid d),$$

and

$$P(M = m \mid K = k) = \frac{P(M = m, K = k)}{P(K = k)},$$

a computation more complex but more informative as more is known.

Thus the predictive approach outlined facilitates inferring probabilities of interest to (a) the testee, (b) some hospital, clinic or institution involving a finite number of testees or (c) a public health authority or commissioner concerned with a very large number (close enough to infinity so as to make no appreciable difference between $P[MN^{-1} \leqq y]$ and $P[\tau \leqq y]$) of potential testees. These predictive probabilities, combined with other ingredients, would be the basis for forming conclusions, making decisions and taking whatever actions are indicated.

ACKNOWLEDGMENT

This work was supported by National Science Foundation Grant DMS-8601314 and National Institutes of Health Grant GM 25271.