

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:

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

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

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

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

and the probability of a false negative,

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

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

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

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

Clearly, if we were diligent and clever enough, we could find from $p(\theta, \eta, \pi | d)$ the joint posterior density of τ and ρ , say $p(\tau, \rho | d)$. 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

