LETTERS

Environmental Analysis

In reading articles on environmental issues, I have repeatedly felt that three areas receive insufficient attention.

1) *Extrapolation.* "Everyone knows" that extrapolation is statistically unsound and that curves apply only to regions validated by data. In environmental issues extrapolation is routinely used, since no data exist in some important areas. The fallacy becomes obvious in looking at the case of trace elements in the human body. Copper and zinc, for example, are absolutely necessary to life. In larger doses they are dangerous poisons. No extrapolation could predict this reversal at low dosage. We need data, not bad guesses.

2) *Analysis.* Environmental risk assessment is a cost-benefit analysis, but not a very good one. Cost-benefit analysis was popularized by industrialists, for whom it was a good tool. It consists of stating a mathematical inequality, or sometimes a simultaneous series of inequalities, which may be compared to determine the course most favoring some objective. Industrialists state their data in dollar amounts, quite easy to contrast. But this is not the case in environmental matters. Consider cancer; in what unit do we state its risk? The probability of developing it? The probability of dying from it? The dollar loss due to disability? The amount of suffering? And can this last be quantified at all? How do we state as a number the value of saving the whale from extinction? Or the snail darter? Are these equal? There is no "unit of risk" as distinguished from the nature of the risk itself.

3) *Risk.* The "no risk" concept has no real existence. We live at risk from conception to death. Adding a specific risk may or may not be justified by circumstance. We save no lives by eliminating carcinogens from our environment. The mortality of those having and not having cancer is exactly the same—100 percent. The real issue, length and quality of life, is hard to evaluate. Consider insecticides; on one side is the risk of poisoning; on the other is the risk of malnutrition or starvation. Which is cost? Which benefit? In choosing a course that will minimize outcomes perceived as "bad" and maximize those perceived as "good" there will be risks on both sides. There is often no conclusive way to identify and select an optimum.

LEWIS G. JACOBS

1701 Vallejo Street,
San Francisco, California 94123

Assessing Diagnostic Technologies

The method offered by Swets *et al.* (24 Aug. 1979, p. 753) to describe the diagnostic efficiency of a test is an elegant way to graph the sensitivity and the specificity (Table 1) of a diagnostic test where observers are not or cannot be standardized to diagnostic criteria.

Unfortunately the "fundamental index, termed A_z," advocated by Swets *et al.* is less useful for comparing diagnostic methods than is visual inspection of their figure 4 or of the usual graphs such as their figure 5, which plots sensitivity against specificity on arithmetic paper.

The best diagnostic criterion is never determined by consideration of sensitivity or specificity alone (*1*). If there are no cost considerations relative to diagnosis and therapy and there is no disadvantage to falsely diagnosing someone as ill, one tries to attain 100 percent sensitivity even though the specificity of that diagnostic criterion may be low. If one is only concerned about screening for healthy individuals one tries to attain 100 percent specificity even at the expense of a low sensitivity. In practice one is usually somewhere between these extremes.

Sometimes the best criterion depends also upon the prevalence of the disease, as when a specific positive predictive value (Table 1) is desired. In that case the optimum combination of sensitivity and specificity for the method changes with prevalence and therefore the appropriate best criterion changes. A graph of sensitivity against specificity such as the graphs in Swets *et al.* is useful to determine the method that delivers the best combination of sensitivity and specificity for a particular diagnostic method given the objectives of the diagnosis (*2*).

However, the A_z index is never useful for any of these choices. The reason can best be understood if two methods (I and II) being compared have lines that cross when sensitivity is plotted on one axis against specificity on the other. At low specificities one method (I) has a higher sensitivity for a given specificity than has the other method (II); the opposite is true at high specificities. If one favors a criterion with a high sensitivity one will choose method I; method II is better when one favors a high specificity. The A_z index gives no information on this important matter.

Once the optimum method is chosen from a sensitivity-specificity plot, the statistical significance of the difference between the methods can be tested at the particular sensitivity or specificity chosen as appropriate for the intended use of the diagnoses. Neither the A_z index nor its variance gives any information about this statistical significance even when the sensitivity-specificity lines do not cross. In fact, comparison of A_z indices is meaningless unless there is a good likelihood that the sensitivity-specificity lines are parallel. This appears unlikely in the comparison of computer tomography and radionuclide scanning for the detection of brain lesions (figure 4 of Swets *et al.*), and unascertainable from the data presented for their other comparisons.

Even when the use of the A_z index is statistically permissible, it is, as noted above, not the logical measure of comparison when choosing a method for its cost-benefit in screening, for its precision in estimating prevalence, for its sensitivity in monitoring change, or for any other characteristic. Is there, then, any use for the A_z index in comparing medical diagnostic methods?

JEAN-PIERRE HABICHT

Savage Hall, Cornell University,
Ithaca, New York 14853

Table 1. Conventional terms used in describing the accuracy of diagnostic methods.

*A. "Truth table"**

Diagnosis by method used	"True" condition	
	Ill	Healthy
Ill	*TP*	*FP*
Healthy	*FN*	*TN*

B. Conventional terms

Sensitivity = $Se = TP/(TP + FN) = P(TP)\dagger = 1 - P(FN)\dagger$

Specificity = $Sp = TN/(TN + FP) = P(TN)\dagger = 1 - P(FP)\dagger$

Negative predictive value = $TN/(TN + FN)$

Positive predictive value = $TP/(TP + FP)$

C. Prevalence estimate of true disease

$Pr = (TP + FN)/(TP + FN + TN + FP)$

D. Interrelation between conventional terms and prevalence

Positive predictive value = $PrSe/[PrSe + (1 - Pr)(1 - Sp)]$

Negative predictive value = $PrSp/[PrSp + (1 - Pr)(1 - Se)]$

*T = true, F = false, P = positive, N = negative.
†Notation as per Swets *et al.*

References

1. R. S. Galen and S. R. Gambino, *The Predictive Value and Efficiency of Medical Diagnosis* (Wiley, New York, 1975), pp. 49-51.
2. J.-P. Habicht, *Am. J. Clin. Nutr.*, in press.

Although the article by Swets *et al.* is excellent methodologically in its demonstration of how to develop relative (or receiver) operating characteristic (ROC) curves for use in assessing diagnostic techniques, the authors distort the role of radionuclide imaging in the diagnosis of brain lesions. In the study "under-