

Comparing the quality of indicators of nutritional status by receiver operating characteristic analysis or by standardized differences^{1,2}

Jean-Pierre Habicht

The publication of the article by Rice et al (1) in this issue of the Journal is a good occasion to review the use of the receiver operating characteristic (ROC) method almost 30 y after it was introduced for medical diagnoses (reviewed in reference 2) and 20 y after it was introduced to readers of the Journal (3). In medical diagnosis, ROC analysis is used to select indicators that differentiate between diseased and nondiseased states (2). In nutrition, ROC analysis is used to select reflective indicators that differentiate between better and worse nutritional status (3) and to select risk indicators that predict health or survival outcomes based on nutritional variables (4). ROC analysis has also been used to identify response indicators (1, 5), measures that respond to nutritional influences. A response indicator should not be confused with a potential-to-benefit indicator, which predicts a future response (6).

ROC analysis requires a standard with which the indicator is compared. For reflective and risk indicators in nutrition, the nutrition or health standards are directly and logically equivalent to disease and nondisease. However, for the response indicator investigated by Rice et al (1), the standard was whether the subject was in a nutritional intervention group (equivalent to nondisease at the end of the study) or in a placebo group. Although the nondisease label is appropriate for the intervention group, the disease label for the placebo group is more problematic. The indicators measured in many of the subjects in the intervention group did not respond to treatment because the subjects were not vitamin A deficient; in other words, they were always nondiseased. A similar proportion of always nondiseased subjects were in the placebo group. The values of these always nondiseased persons in both groups overlapped and therefore decreased the indicator's discriminatory power at higher sensitivity values and lower specificity values in the placebo and intervention groups, respectively. This misclassification impaired the selection of the best response indicators.

All ROC parameters can be formulated as a standardized difference or as a *t* test between the means of the specificity (nondiseased) and the sensitivity (diseased) distributions (4). The indicator of choice is the one that produces the least overlap between the specificity and sensitivity distributions, resulting in the highest *t* test value and largest ROC parameter (area under the curve). ROC parameters can only be compared if the ratio of the specificity to sensitivity variances is the same between indicators, which results in parallel ROC lines if the sensitivity and specificity values are transformed to *z* scores (Gaussian probability

scale). If the *z* scored ROC lines are not parallel, the ROC comparison statistic is meaningless. For instance, 2 ROC lines can cross, as is the case in Table 4 of the article by Rice et al (1). In that case, the better indicator before the crossover is the poorer indicator after the crossover.

In most publications presenting ROC analyses, computer programs are used to perform the analyses. This is necessary when dealing with indicators that are not continuous variables, such as X-ray diagnoses (2). However, when continuous variables are used, as is usually the case in nutrition, calculations made without a computer program (4) are preferable because they force explicit consideration of the components of the ROC parameters—the variances of the specificity and sensitivity distributions and the difference between the means of the distributions. This method is certainly preferable to the method most cited in the literature for comparing indicators (7). The method described by Erdreich and Lee (7) does not take into account the correlations of the indicator variables among themselves and therefore can result in severely biased ROC statistics. The major limitation to using the method described by Brownie et al (4) is understanding the publication that describes it. The authors present much more information than is necessary. The reader is advised to start with the first 2 sections and then to turn to the example.


When the *z* scored ROC lines are parallel, ROC analyses are better by far than any other technique for describing and comparing indicators to be used for screening. The next step is choosing an appropriate cutoff (8, 9). When the conditions of parallelism are not met, one must select the cutoff points and then test their discriminatory powers by other less satisfactory means (10).

Screening with a reflective indicator to identify malnourished children or with a risk indicator to identify children at risk of bad outcomes makes sense because one can do something for these children. The population prevalences derived from such screens also make intuitive sense. The rationale for screening children with a response indicator to identify those who did or did not receive a past intervention is less obvious.

A response indicator is, however, the right indicator to use to measure influences on nutrition at the population level. For this

¹From the Division of Nutritional Sciences, Cornell University, Ithaca, NY.

²Address reprint requests to J-P Habicht, Savage Hall, Cornell University, Ithaca, NY 14853. E-mail: jh48@cornell.edu and bjs25@cornell.edu.

use, estimating the effect on nutrition with dichotomous prevalences is usually much less efficient statistically than using the means or ordinary least squares (OLS) regressions analyses (9). Selecting the best indicator for OLS analyses is best done by comparing standardized differences directly (1, 5) or indirectly through sample size estimates (11). The standardized difference is directly related to the t and F tests for statistical significance and therefore establishes the sample sizes needed to identify a nutritional determinant (5). The higher the standardized difference and the lower the required sample size, the better the indicator. The ranking of indicator quality is almost identical whether one uses the ROC or the standardized difference method (1), which is not surprising because the ROC is also based on standardized difference considerations (4). Therefore, using the standardized difference technique is most appropriate for selecting the best response indicator. 

I thank Gretel Pelto for helpful discussions and editing.

REFERENCES

1. Rice AL, Stoltzfus RJ, de Francisco A, Kjolhede CL. Evaluation of serum retinol, the modified-relative-dose-response ratio, and breast-milk vitamin A as indicators of response to postpartum maternal vitamin A supplementation. *Am J Clin Nutr* 2000;71:799–806.
2. Swets JA, Pickett RM. Evaluation of diagnostic systems. New York: Academic Press, 1982.
3. Trowbridge FL, Staehling N. Sensitivity and specificity of arm circumference indicators in identifying malnourished children. *Am J Clin Nutr* 1980;33:687–96.
4. Brownie C, Habicht J-P, Cogill B. Comparing indicators of health and nutritional status. *Am J Epidemiol* 1986;124:1031–44.
5. Stoltzfus RJ, Habicht J-P, Rasmussen KM, Hakimi M. Evaluation of indicators for use in vitamin A intervention trials targeted at women. *Int J Epidemiol* 1993;22:1111–8.
6. Ruel MT, Habicht J-P, Rasmussen KM, Martorell R. Screening for nutrition interventions: the risk or the differential-benefit approach? *Am J Clin Nutr* 1996;63:671–7.
7. Erdreich L, Lee ET. Use of relative operating characteristic analyses in epidemiology. *Am J Epidemiol* 1981;114:649–62.
8. Habicht J-P. Some characteristics of indicators of nutritional status for use in screening and surveillance. *Am J Clin Nutr* 1980;33:531–5.
9. Brownie C, Habicht J-P. Selecting a screening cut-off point or diagnostic criterion for comparing prevalences of disease. *Biometrics* 1984;40:675–84.
10. Galen RS, Gambino SR. Beyond normality: the predictive value and efficiency of medical diagnoses. New York: Wiley and Sons, 1975.
11. de Pee S, Yuniar Y, West CE, Muhilal. Evaluation of biochemical indicators of vitamin A status in breast-feeding and non-breast-feeding Indonesian women. *Am J Clin Nutr* 1997;66:160–7.

