guished editorial consultants of *Chest*, it is hoped the Illustrative Echocardiograms department can serve as one forum for communication and evolution in this new field. We hope to provide rapid transmission and assessment of important developments and re-evaluation of classic concepts by publishing especially educational material. We wish to provide examples of high quality echocardiograms against which one may judge his own work. The recognition of records of insufficient quality for diagnostic use is as important as the recognition of a new diagnostic pattern. Selected articles submitted directly for this department, as well as some chosen from the general submission to *Chest*, will be published as Illustrative Echocardiograms. In addition we will invite short reviews or critiques of echocardiographic principles and techniques from established authorities in the field.

Through the Illustrative Echocardiograms Department we hope to help echocardiographers understand and advance modern standards. By this type of "secondary education" we hope to further the proper growth of echocardiography, with its unique problems and promise. We welcome your help.

*Richard L. Popp, M.D.*
Stanford, California

*Assistant Professor of Medicine, Director, Cardiac Non-Invasive Laboratory, Cardiology Division, Stanford University School of Medicine.

**REFERENCE**

1 Feigenbaum H: Educational problems in echocardiography. Am J Cardiol 34:741-742, 1974

**Radiology and the Receiver Operating Characteristic (ROC) Curve**

Psychoptics is the unlikely source of a methodology which permits quantitative evaluation of the usefulness of a given medical test or of the ability of a radiologist to interpret a roentgenogram. Highly mathematical in orientation, it has been discussed by Lusted in connection with medical tests in general and radiology in particular. The purpose of this commentary is to try to clarify the Lusted adaptation and apply it to roentgenographic interpretation, to show how to compare the accuracy of inter- and intra-observer readings. The former variation is widely known but personally suppressed. The latter is virtually de-nied and seldom studied (especially by radiologists).

Start by assuming a score, *x*, to be the result of any medical test obtained from a patient. Continue in this abstract fashion and agree that *x* increases for patients who are more likely to be "sick," i.e., to have the disease being tested for. But since no test is infallible, sometimes a "well" patient will have a higher score than some "sick" patients. Now consider a retrospective study and plot separate histograms of the scores of "sick" and "well" patients: the average for "sick" patients will be larger, but the histograms will overlap.

In clinical practice, it is necessary to bite the bullet and define a criterion score, *x_c*. If a patient has a score higher than *x_c*, the test is considered positive, but if his score is less than *x_c*, the test is negative. In addition to true positive and true negative reports, for which the test result turns out to be an accurate predictor of the patient's condition, there are two kinds of false reports. One is the false positive, in which a "well" patient receives a positive report; the number of false positives is equal to the number of patients in the "well" histogram with scores *x* greater than *x_c*. The patients are subjected to the worry, expense, and possible risks of follow-up procedures which turn out to be unnecessary. A more dangerous situation is represented by the false negative, in which a "sick" patient receives a negative report. If the test is for cancer, for example, a false negative may doom the patient. The number of false negatives is equal to the number of patients from the "sick" histogram who have scores less than *x_c*. If a large enough number of patients are tested, there will be some from each histogram with very large and very small scores; thus, neither false negatives nor false positives can be entirely eliminated.

To many of you who have read thus far, the objection will be raised that many tests do not yield a numerical *x*; roentgenographic interpretation is an obvious example. In fact the interpreter looks at the film and implicitly rates it in terms of the apparent probability that it is positive; only if this probability exceeds some value will the report be positive. In effect, then, we can think of a numerical score as having been generated or computed. A certain number of false reports of each sort is again inevitable as every clinician knows and every radiologist fears.

What is actually at our disposal is the criterion score, *x_c*, defined above, which separates positive from negative reports. By setting this criterion score high, we reduce false positives; by setting it low, we reduce false negatives. By setting it at zero, we...
eliminate false negatives, but alas, we also eliminate the utility of the test, since every report is positive! Thus $x_c$ has to be chosen as a compromise to balance the risks, costs, and numbers of false positives against the risks and numbers of false negatives.

It is simple to create a graph with the percentage of false positives plotted along the horizontal axis and the percentage of true positives plotted along the vertical axis, for various values of $x_c$. (Raising the value of $x_c$ lowers both the number of true positives and the number of false positives, since it decreases the number from each histogram with scores greater than $x_c$; lowering $x_c$ has the opposite effect.) Such a curve is called a receiver operating characteristic (ROC) curve. ROC curves are often drawn on normal deviates graph paper, for which equal divisions along each axis represent equal increments of $x_c$. They are thus unequally spaced in terms of percentages of true and false positives, but these percentages are specified on the graph paper. On such paper, the ROC curve is a straight line.

Note that ROC curves can be drawn without knowledge of the underlying histogram, which is totally convenient if they are to be used for tests in which an explicit value of $x$ is not generated. All that is needed is enough follow-up information so that each report can be identified as having been "true positive," "false positive," "true negative," or "false negative." The slope and intercept of the straight line defined by an ROC curve cannot be altered by interpreting the test more or less conservatively; they are characteristic of the test itself. The position along this line, however, can be varied by simply changing the criterion. In the case of roentgenology, for example, a low criterion could be represented by a policy under which positive reports were issued when the reader suggested even the possible presence of a disease. A high criterion would be in effect if such a reading resulted in a negative report, with positive reports reserved for cases in which the evidence for the disease seemed fairly definite.

A physician can consciously modify his position on the ROC curve, but remain on the same straight line. This allows the practicing radiologist to score himself on a series of standardized roentgenograms. It would permit evaluation of diagnostic criteria, not of one's skill. Improved radiologic interpretation is the objective.

Progress in a reader's training or skill would be represented only by his ability to generate a new, more favorable curve of his own. The conclusions are clear even if the preceding explanation is not. Receiver operating characteristic curves offer: (1) an excellent way to help us all face the issue of variations in roentgenographic interpretations, from time to time by the same and different observers; (2) a stimulating technique to practice self-assessment; (3) an objective method to quantitate the learning rate of the resident in radiology; and (4) a striking approach to the establishment of diagnostic criteria in roentgenographic disease.

Should the American College of Chest Physicians lead the way again?

W. Scott Andrus, Ph.D.* and Kenneth T. Bird, M.D., F.C.C.P.**

Boston


*Consultant, Telemedicine Project, Massachusetts General Hospital.

**Associate Physician, Massachusetts General Hospital, and Associate Clinical Professor of Medicine, Harvard Medical School.

Reprint requests: Dr. Bird, 275 Charles Street, Boston 02114

REFERENCE


Nomograms for Pulmonary Ventilatory Function: Should Coronary Heart Disease Patients Be Included?

In today's modern and advanced medicine, uncertainty still remains concerning the normal values for pulmonary ventilatory functions (PVF). Admittedly, this uncertainty is more marked in the older age group than in our younger patients. Most of our knowledge about these values is derived from studies done some years ago, when the normal values were based on small groups of subjects in whom smoking and coronary heart disease (CHD) were not considered as factors.

The first nomogram for vital capacity was described by Hutchinson1 128 years ago. In this work it was already pointed out that vital capacity is directly related to height, and inversely proportional to age. Even though the criteria for "pulmonary normals" in Hutchinson's work are not clearly defined, it may be considered the pioneer investigation in this field. Since then, various other