CHEST

Official publication of the American College of Chest Physicians

Remote interpretation of chest roentgenograms

WS Andrus, CH Hunter and KT Bird

Chest 1975;67;463-468
DOI 10.1378/chest.67.4.463

SPECIAL COMMUNICATION

Remote Interpretation of Chest Roentgenograms*

*W. Scott Andrus, Ph.D.,*** *Charles H. Hunter, M.D.*† *and Kenneth T. Bird, M.D., F.C.C.P.*‡

A series of 98 chest films was interpreted by two physicians on the basis of monitor display of the transmitted television signal representing the roentgenographic image. The transmission path was 14 miles long, and included one active repeater station. Receiver operating characteristic curves were drawn to compare interpretations rendered on television view of the image with classic, direct view interpretations of the same films. Performance in these two viewing modes was found to be quite similar. When films containing only hazy densities lacking internal structure or sharp margins, were removed from the sample, interpretation of the remaining films was essentially identical via the two modes. Since hazy densities are visible on retrospective examination, interpretation of roentgenograms at a distance via television appears to be a feasible route for delivery of radiologic services.

It has been estimated that by the end of this decade the total number of medical roentgenograms taken will approach ten billion.[1] The burden placed upon the corps of radiologists who must interpret these films is compounded by the fact that a substantial number will be exposed at facilities other than teaching hospitals. The problem can be ameliorated if films made at extra-hospital facilities can be promptly interpreted without the need for the radiologist to waste time traveling. A promising approach is teleradiology: the use of television to transport the roentgenographic image to a central site for interpretation. This report is an assessment of the accuracy with which chest roentgenograms can be interpreted via television.

For editorial comment, see page 378

The Massachusetts General Hospital (MGH) participates in two systems for the practice of telemedicine, or medicine at a distance. These systems transmit and receive audio and video on the 12 GHz microwave band. One links the hospital to its medical station at Logan International Airport, and is used for remote diagnosis of somatic and psychiatric conditions as well as several other health, education and social welfare community activities.[2] The other system links MGH to the Veterans Administration Hospital (VAH), Bedford, Massachusetts, as part of a total health information system.[3] Figure 1 shows the essential features of this type of telemedicine system. The camera viewing the patient is under the remote control of the physician, via subcarrier tone channels which enable him to pan, tilt, zoom, and focus the camera, and to vary the lens opening. The patient can see the physician on a monitor placed just above the camera and can talk to him normally via the audio channel through discretely placed fixed microphones. The image is made more distinct by an electronic image enhancer which sharpens edges that may have been blurred by high frequency losses in transmission.

The system is well adapted for tests of teleradiology. The ability of the physician to control the camera remotely is essential, because at least 700 lines of resolution would be needed to transmit in a single view the information contained on a standard roentgenogram. Standard transmitted television images permit about half this resolving power, but by zooming in for an enlarged view and scanning each portion of the film sequentially, the physician can extract virtually all of the diagnostic information on the film. Although the objection has been raised that this requires him to scan the film in an artificial

*Supported in part by the Veterans Administration, Washington, D.C., Certificate of Award Nos. EMI-69-001-01 and EMI-69-001-02 (titled Teleconsultation: A New Health Information Exchange System), Certificate of Award No. EMI-70-001-03 (titled Teleconsultation: A New Health Information Exchange System), and Certificate of Award Nos. EMI-72C-001-01 and EMI-74C-011-03 (titled Telemedicine: A New Health Information Exchange System).

**Consultant, Telemedicine Project, Massachusetts General Hospital, Boston.

†Chief of Radiology, Veterans Administration Hospital, Bedford, Massachusetts.

‡Associate Physician, Massachusetts General Hospital; Associate Clinical Professor of Medicine, Harvard Medical School, Boston.

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

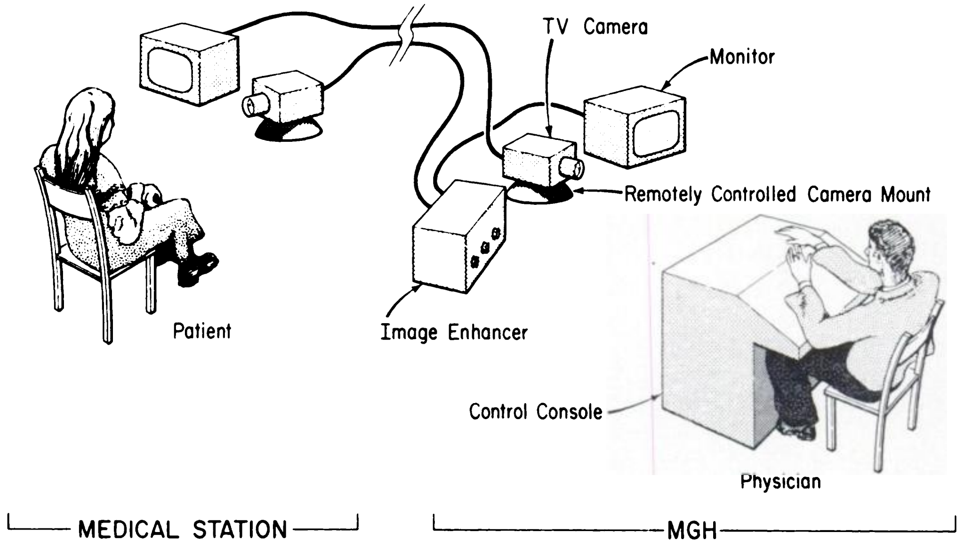

FIGURE 1. Conceptual schematic diagram of the telediagnosis system. Both television cameras are controlled from the physician's console.

manner, the fact is that scanning is required even in direct roentgenographic interpretation, because the film is too large to be included in the foveal field of view in a single eye fixation at the normal reading distance.[4] The more systematic scanning pattern required in teleradiology may in fact be beneficial.

MATERIALS AND METHODS

A previous study of remote roentgenologic interpretation[5] indicated that, if there was degradation of the image in television transmission, it was small and was masked by normal inter-observer variation.[6] Because of the importance of this point, a further study has now been completed of the accuracy of remote roentgenographic interpretation, using a series of 98 chest films. All films either were normal or were of patients with pulmonary tuberculosis or other pulmonary conditions, such as cancer or pulmonary fibrosis of unspecified etiology, resembling tuberculosis in roentgenographic appearance. Table 1 lists the conditions represented. The chest roentgenograms were chosen from hospital records on the basis of discharge diagnosis, without reference either to the film or to the roentgenographic interpretations.

Table 1—*Hospital Discharge Diagnoses and Radiologic Findings of Patients in Study*

Discharge Diagnosis	No. of Cases	Suspected by Radiologist* Direct View	T.V. View
Normal	11	11	11
Minimal tuberculosis	47	46**	42**
Moderately advanced Tuberculosis	10	10**	9**
Moderately advanced Tuberculosis *and* cancer	1	1**	1**
Cancer	13	8†	7†
Pulmonary fibrosis	8	—	—
Silicosis	2	—	—
Pneumonitis	2	—	—
Pneumonia	1	—	—
Asthma	1	—	—
Empyema	1	—	—
Pulmonary abscess	1	—	—

*Blanks represent situations where diagnosis is not in terms used by the radiologist.

**Figures given are number of patients suspected as tuberculosis.

†Figures given are number of patients suspected as cancer.

The physician (reader A) who had made the interpretations of films contained in these records was a participant in our study. The image of each film was shown to him at MGH via the microwave link from Bedford 14 air miles away with a simple interposed active-repeater microwave station. The images were also shown to a radiologist (reader B) at the VAH, Bedford, via the link from MGH. In addition, reader B reviewed each film directly, several weeks before or after seeing it by television. In each case, only a single chest film was shown. Each reader was asked to interpret the films in his usual fashion and dictate reports. Reader B at Bedford reported all significant abnormalities that he could see; reader A, interpreting at MGH, also reported his impression as to the most likely etiology of any pathology noted, as he had done in his original, direct view of each film years earlier.

METHODS OF DATA ANALYSIS

Although roentgenograms are not diagnostic, the ability of a radiologist to identify pulmonary tuberculosis depends on the quality of the image he sees.[7] It can happen that a roentgenogram is recognized as abnormal even though not all lesions demonstrated are noticed; in such a case, the probability of a correct diagnosis is reduced. Thus, even though tuberculosis often mimics or is mimicked by other conditions, a comparison of a radiologist's success in recognizing tuberculosis on television view and on direct view of the same roentgenogram is a very sensitive test of the potential accuracy of remote interpretation. One approach employed in determining whether the interpretations based on television view were of acceptable accuracy was a study of reader A's ability to perceive the radiologic densities and to equate the findings with the often characteristic patterns of pulmonary tuberculosis. His ability to perform this task on television view of the films was compared to his performance on direct view. A reader, while correctly suspecting a condition which in fact is present, may have doubts solely as a result of the use of an unfamiliar medium. We have therefore used receiver operating characteristic (ROC) curves[8] to determine whether the two modes are equally powerful for this purpose. A description of this technique follows.

Most medical tasks can be considered devices for assigning to the patient a number, x, which is proportional to the probability that the patient has the disease to which the test is sensitive. If the procedure is carried out for a large number of patients and their scores are recorded, the distribution of test results will resemble the binormal distribution shown in Figure 2A. If the clinical course of each patient is subsequently followed and an eventual determination finally made as to which patients did indeed have the disease in question, it becomes possible to prepare a graph similar to the one shown in Figure 2B. It is clear that, on the average, "sick" patients—those who do have the disease—have higher scores on the test. The scores for both "sick" and "well" patients vary over a considerable range, however, and in fact a "well" patient will occasionally have a higher score than some "sick" patients. This means that the test, like almost all medical tests, is a less than perfect discriminator between "sick" and "well" patients.

One aspect of the practice of medicine is the determination of a criterion: a value, x_c of the test results such that any patient whose score exceeds x_c is reported "sick." But any choice of x_c will result in some false reports, since the distributions in Figure 1B overlap.

There are four types of result possible for any medical test:

True positive The test result is positive, but the patient is "sick."

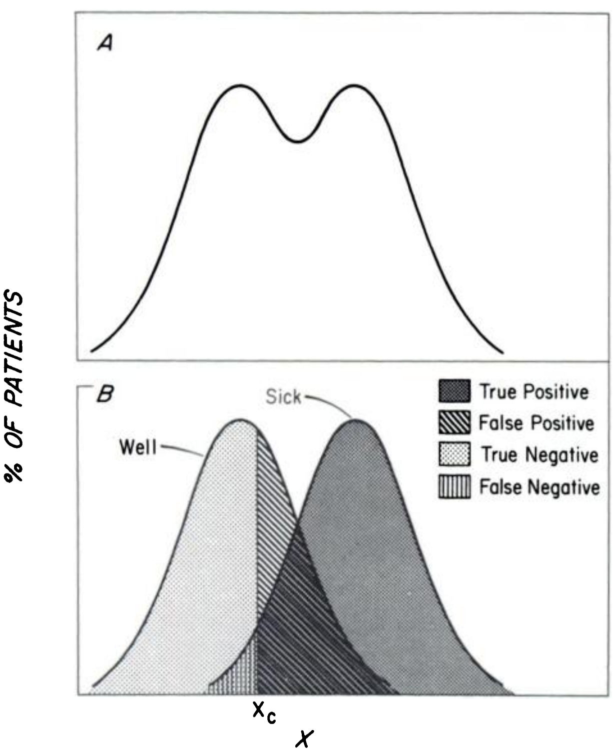

FIGURE 2. A. The bi-normal distribution of results expected in a medical test. B. The separate normal distributions generated with clinical knowledge of each patient's true condition. The line marked X_c is the criterion; all patients with scores greater than X_c receive positive reports. The numbers of true positive, false positive, true negative, and false negative reports are proportional to the areas of the regions of corresponding cross hatching, respectively, as shown.

False positive The test result is positive, but the patient is actually "well."

True negative The test result is negative and the patient is "well."

False negative The test result is negative, but the patient is in fact "sick."

Both false positive and false negative results have unfortunate sequelae. False positive results subject the patient to unnecessary and sometimes risky follow-up procedures, waste of time, expense, and worry. If x_c is chosen sufficiently high, the number of false positives can be made very small. In this case, however, the number of false negatives will increase. A false negative test result for a serious progressive disease such as cancer can doom the patient. False negatives can be avoided by choosing x_c very low, but again the number of false positives will increase.

Figure 2B illustrates these statements in graphic form. It is clear that decreasing x_c will decrease the size of the vertically crosshatched area which represents false negatives, but at the cost of increasing the size of the diagonally crosshatched area which represents false positives. Thus the choice of x_c represents the physician's judgment as to the relative costs and dangers of the two types of false report.

A plot can be made of the percentage of false positives versus the percentage of false negatives for various values of x_c. It is more common, in fact, to plot the percentage of true positives versus the percentage of false positives. The two are equivalent because the percentage of true positives is simply 100 percent minus the percentage of false negatives. A plot of the percentage of true positives versus the percentage of false positives for various values of x_c is called a receiver operating characteristic (ROC) curve. This plot is not a straight line when drawn on ordinary graph paper, but there is a type of graph paper, called normal deviates paper, on which ROC curves are expected to be linear. Normal deviates paper has a specific unequal spacing in the divisions along each axis. If the ROC curve is drawn on normal deviates paper as in Figure 3, using a relatively small number of points or even just two, the expected numbers of true positives and false positives for any value of x_c can be determined from the curve formed by connecting the points with a straight line.

The great value of ROC curves for analysis of radiologic

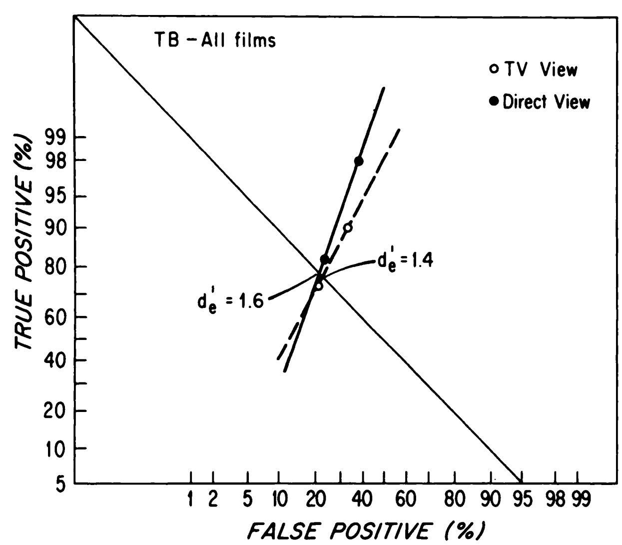

FIGURE 3. ROC curves for roentgenographic identification of tuberculosis from chest films, by direct inspection and by television visualization. The data are plotted on normal deviates paper.

interpretations is that factors relating to the radiologist's attitude are divorced from factors relating to his ability to perceive suspicious features in the film. The radiologist's optimism or pessimism merely defines the value of x_c at which he operates. It is in fact possible, as Lusted[8] has shown, for a radiologist to vary the position on the ROC curve at which he operates. He merely interprets first optimistically by assuming that all shadows which may be suspicious but are not clearly so, can be overlooked, and then pessimistically by assuming that all suspicious shadows are positive. Changing the criterion in this way will alter the position along the ROC curve at which the radiologist operates, but the line itself is not changed. The position of the line is an indicator of the sensitivity and specificity of the test. Lines which fall toward the top left of the graph represent more sensitive interpretation; lines which fall toward the bottom right represent less sensitive interpretation.

Note that the construction of an ROC curve does not depend on knowledge of the underlying distributions of Figure 2; it is only necessary to know the percentages of true positives and of false positives for at least two operating points of the test. In fact, the ROC curve actually provides enough information for the underlying distributions to be calculated. They are assumed to have the standard normal shape. The "sick" and "well" curves are so normalized that the area under each is the same: 100 percent. The ratio of their variances, or widths squared, is equal to the slope of the linear ROC curve. Then it is only necessary to know the distance between the peak of the "well" curve and the peak of the "sick" curve.

This parameter can be read directly from the ROC curve. In Figure 3, a straight line has been drawn from the upper left to the lower right corner. This is the negative diagonal: the locus of points for which the sum of the true positive percentage and the false positive percentage is 100 percent. The intersection of the ROC curve with the negative diagonal is the point of interest. The false positive percentage at this point is read from the graph. The associated normal deviate can be found in a standard table; this quantity is equal to half the separation between the peaks in the underlying distributions, in units of the variance of either curve if they have the same variance. Otherwise, the units are given by the average of the two variances.

Thus, even though an explicit value of x is not specified in a radiologic report, a distribution function in x can be determined from such reports by drawing the ROC curve and computing the underlying distributions as described. The meaning of this distribution can be understood from a conceptual model of roentgenographic interpretation in which the radiologist assigns a point score to each roentgenographic sign. Suppose, for example, he is attempting to assess the probability that tuberculosis is present. He might feel that a hazy density at the apex was worth ten points, that a blunted costophrenic angle was worth five points, that fullness of the hilar shadow was worth two points, and so forth (the actual point values do not matter; only their relative magnitudes are important).

The score, x, for a roentgenogram is then the sum of the point values for all signs seen on the film; a plot of scores for a number of films would look like Figure 2A. Retrospective separation of the patients into a "sick"—*ie* tuberculous—population and a "well"—nontuberculous, no matter what other conditions might be present—population would make it possible to draw curves like those of Figure 2B. A positive report results whenever the score exceeds whatever criterion score, x_c, has been selected. Performing this computation for several values of x_c makes possible determination of a number of pairs of values of true positive percentage and false positive percentage. From these pairs of values, an ROC curve can be drawn.

Naturally this is not the way roentgenograms are interpreted in practice. But from ROC curves drawn on the basis of interpretations actually rendered, it is possible to determine the distribution of scores, x, that would have resulted if the radiologist had explicitly recorded the score which, in fact, he implicitly associates with various signs.

Results

To return to the ROC curves of Figure 3, it can be seen that curves have been drawn for television interpretation and for direct view interpretation.

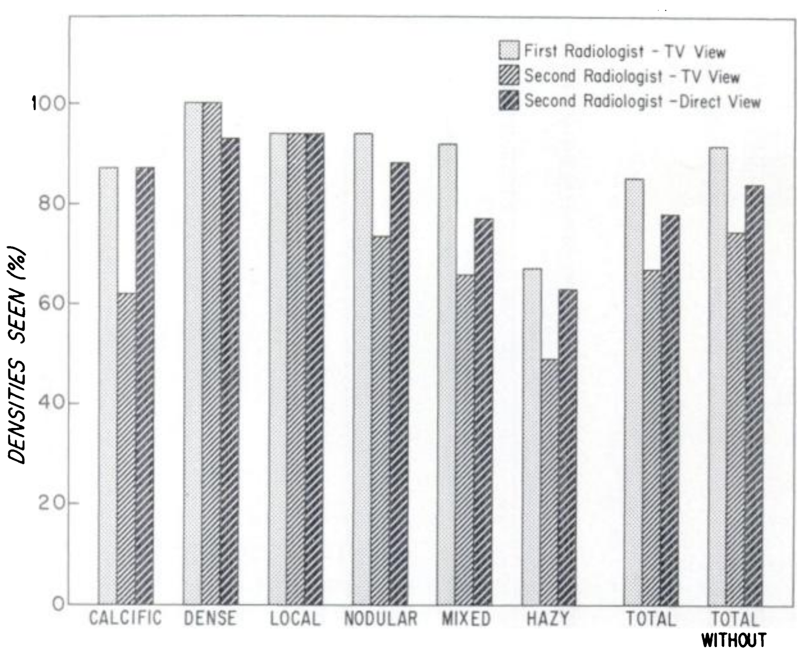

FIGURE 4. Probability of densities of various classifications, reported in the original roentgenographic interpretations, being reported in subsequent interpretations.

Each curve is drawn through two points, representing pairs of values of true positive percentage and false positive percentage, for *optimistic* and *pessimistic* interpretation. Optimistic interpretation (high value of x_c) means that the report was considered positive only if the reader said the findings were consistent with tuberculosis and mentioned no alternative possibility. In other words, "When in doubt, call the film negative (for tuberculosis)." Pessimistic interpretation (low value of x_c) means that the report was considered positive even if tuberculosis was only mentioned as a possibility or as one of several possibilities: "When in doubt, call the film positive."

Both the television view and direct view ROC curves intersect the negative diagonal at about the same point. Thus, the peaks of the distributions would be in about the same place for both modes. The direct view curve is more steeply sloping, however, showing that television interpretation gives a slightly broader distribution of effective scores for "sick" patients. This could well result from the reader's relative inexperience with interpretation based on television view.

Figure 4 is a graph of the probability that a density of each of several classifications will be reported by a radiologist. The original hospital interpretations, as found in the records, were arbitrarily taken as the standard. Each density reported has been sorted as to position and classification. In each subsequent report (including one via television by reader A, and one via television and one via direct view by reader B), a determination has been made as to whether a density was reported in that position.

Six arbitrary types of density were considered: calcific; dense; local or sharply bounded; nodular; mixed; and hazy. *Hazy* density means any shadow which is light in density and which does not have a clear-cut internal structure or sharp margins. Hazy densities can, however, have some linear structure, but may be difficult to distinguish from normal vascular shadows. Thus, a hazy density has to be recognized by a small difference in radiopacity, without reference to other features. In each group in the figure, the left hand bar gives the percentage of such densities seen by reader A on television view, the middle bar gives the percentage seen by reader B on television view, and the right hand bar gives the percentage seen by reader B on direct view. The major difference is in hazy densities: shadows of low density without internal structure or sharp margins. When all films whose intrepretation depended solely on perception of hazy densities were removed from the sample, and the reports on the remaining 84 films were analyzed in the way described above,

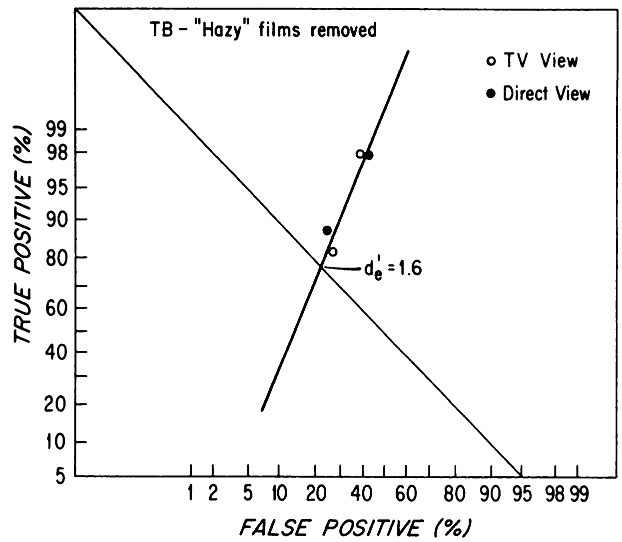

FIGURE 5. ROC curve for identification of tuberculosis, with films containing only hazy densities deleted from the sample.

Figure 5 resulted. A single straight line ROC curve comes very close to all points for direct and television interpretation, indicating that the two modes are about equally effective.

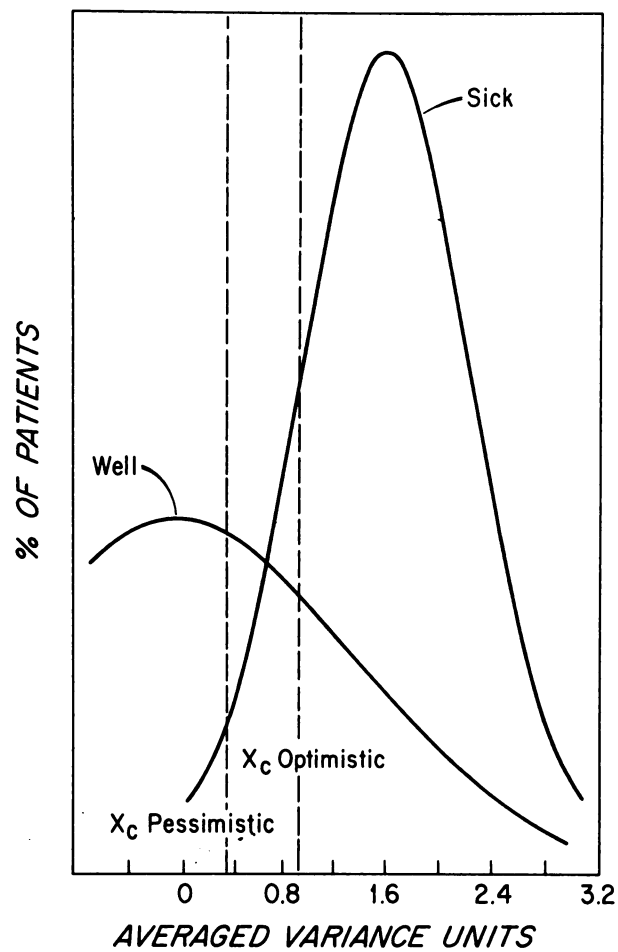

FIGURE 6. The underlying distributions of test results for "well" and "sick" patients, inferred from the ROC curve of Figure 5. The values of the criterion, X_c, for optimistic and for pessimistic interpretation are indicated by vertical lines.

The underlying distributions have been computed from this curve and are shown in Figure 6. The steep slope of the ROC curve indicates that the "well" distribution will be considerably broader than the "sick" distribution. The peak of the "sick" distribution is correspondingly higher, as the areas under the two distributions must be equal.

Discussion

The ability to detect hazy poorly defined densities can be improved by the following techniques:

Alert the teleradiologist to the fact that such densities are hard to see and should be carefully sought. The viewing monitor should be carefully adjusted so as to emphasize hazy densities by appropriate choice of contrast and brightness.

Provide a photographic step wedge on the viewbox so that absolute density levels can be identified.

Carefully eliminate electronic noise; small density variations are screened by noise. The electronic image enhancer must be carefully adjusted with this constraint in mind.

Scan the image with normal signal polarity and with the polarity reversed to provide a negative image; hazy densities are sometimes perceived more clearly in this negative image.

The fact that hazy densities missed on television view are visible in retrospective television examination indicates that they do not represent an insuperable obstacle; it is merely necessary to be aware of the problem and to take simple steps to deal with it. The fact that hazy densities were frequently missed by reader B on direct view indicates that such densities may have been missed by reader A in the original reports; no check on this possibility has been attempted in this study.

The deliberate choice of a test which is very difficult—detection of tuberculosis from a single chest film—is the reason for the high overlap in the distributions of Figure 6 and consequent low sensitivity which has been found. This choice assured that the comparisons of television interpretation and direct view interpretation would be an exacting one. If, for example, the reader had simply been asked to say whether any active disease was present, he would have been correct on almost every film. It would therefore have been difficult to say whether the television image of the film was really as good as the direct view image. Quantitatively, the peaks in the distributions would have been further apart, and the ROC curve would have been displaced upward and to the left.

The ROC curves in Figure 3 fall close together. The slightly less steep slope for television view suggests that results for tuberculous patients show more variation via this mode; this could very well be an artifact resulting from the reader's relative lack of experience with the television appearance of a chest film. When films containing only hazy densities are removed from the sample, the points fall close to the single ROC curve of Figure 5 for both television and direct view interpretation of the remaining films. In view of the sensitivity of the comparison, this is strong evidence that teleradiology can provide a thoroughly satisfactory alternative to classic, direct view, roentgenographic interpretation.

References

1 Andrus WS, Bird KT: Teleradiology: Remote interpretation of roentgenograms. Photographic Science and Engineering in Medicine (Proceedings of a tutorial seminar sponsored by the Society of Photographic Engineers and Scientists), 1972

2 Murphy RLH, Cohen GL, Bird KT, et al: Telediagnosis: A new community health resource. Massachusetts General Hospital, Boston, 1970

3 Bird KT: Teleconsultation: A new health information exchange system. Massachusetts General Hospital, Boston, 1971

4 Andrus WS, Bird KT: Teleradiology—Evolution through bias to reality (Editorial). Chest 62:655-657, 1972

5 Murphy RLH, Barber D, Bird KT, et al: Microwave transmission of chest roentgenograms. Am Rev Resp Dis 102:771-777, 1970

6 Yerushalmy J: The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Radiol Clin North Am 7:381-392, 1969

7 Murray RS. *In* Perry KMA, Selors TH: Chest Diseases (vol 1). London, Butterworth & Co, 1963, pp 76

8 Lusted LB: Introduction to Medical Decision Making. Springfield (Ill), Charles C Thomas, 1968