Psychophysical Studies of Detection Errors in Chest Radiology[1]

George Revesz, Ph.D., and Harold L. Kundel, M.D.

In 62 of 124 cases analyzed, there occurred a failure to detect a pulmonary nodule which was retrospectively noted only after detection at a subsequent examination. The characteristics of the lesions and their surrounds were measured, and a computer analysis used, to identify that combination of parameters (the conspicuity, *K*) which was best capable of separating the group into populations of missed and detected lesions.

INDEX TERMS: Lung neoplasms, diagnosis • (Lungs, error in diagnosis, 6[0].940) • Radiographs, image analysis • Radiographs, interpretation • Radiology and radiologists

Radiology 123:559–562, June 1977

COMPETENT OBSERVERS, at the first reading of radiographs, miss about 30% of those pulmonary nodules present, even though in retrospect they are clearly visible (2). The missing of these lesions may result in delayed or inappropriate treatment not only for patients being screened for lung cancer, but also for patients with known cancers elsewhere who are being staged or followed for evidence of metastases. In the absence of a reasonable alternative to chest radiography, there are two possible ways to decrease these errors: improving either the quality of the image or the reader performance. Neither approach has thus far been successful. Image quality can be improved by better equipment or technique, but developments of the past twenty years have not yielded improved detection accuracy. Nor is there evidence in the published literature that image processing has resulted in decreased error rates. Except for multiple reading of radiographs, the cost-effectiveness of which is questionable, no effective changes in radiograph reading have been suggested. Reader performance is so poorly understood that it is hazardous to suggest changes in either the method of reading or in the way new readers might be taught.

We have been conducting studies of the psychophysics of radiograph reading in order to link psychological factors, *i.e.,* the ability of readers to detect nodules on chest radiographs, with the physical properties of the image (5–7). These physical properties include nodule properties such as size, shape, and edge gradient; background properties such as overall density and complexity; and joint properties of nodule and surround such as contrast and overlap with other structures. An understanding of these factors may aid in the selection of optimal radiological techniques, or may lead to suggestions for image processing techniques that will improve observer performance.

The quantitative relationship between the ability of observers to detect a pulmonary nodule and the physical properties of the nodule-surround complex is expressed in terms of p, the probability of detecting a nodule, and a quantity K which we call conspicuity and which is calculated from measurements made on the radiographs. The probability *p* of detecting a lesion is thus a function of K:

$$p = f(K) \qquad (1).$$

As a first approximation, conspicuity was assumed to be proportional to the contrast and inversely proportional to the complexity of the surround, expressed as the rate of change of density in the area immediately surrounding the nodule. With these assumptions, *p* was found to be proportional to log K with good accuracy (7).

These relationships were obtained with artificially made pulmonary nodules because they are easier to work with in a laboratory setting. We attempted to verify the laboratory model in clinical cases.

MATERIALS AND METHODS

We wanted to compare the physical properties of lesions that were missed in clinical practice with those that were detected. A retrospective search of the radiographic files for missed lung lesions was therefore undertaken. First a search was done using the Temple University Tumor Registry and the Diagnostic Radiology Index files, which are based on the American College of Radiology-Index for Roentgen Diagnosis. A computer search to match cases entered in the tumor registry as primary or secondary lung cancer against chest radiographs indexed as either positive or negative for lung lesions in the diagnostic radiology index was done for the years 1969 and 1970, and yielded 632 matches. These cases were retrieved from the files and were reviewed by two radiologists, who identified 47

[1] From the Diagnostic Radiology Research Laboratory, Temple University School of Medicine, Philadelphia, Pa. Accepted for publication in January, 1977.
This work was supported in part by Grant CA 16491, National Cancer Inst., United States Public Health Service. ss

cases in which a pulmonary nodule was correctly reported on at least one examination of the chest *and* there was a prior chest examination in which the nodule could be seen in retrospect but was not reported. These 47 cases (a total of 124 examinations, since some cases had more than two examinations) contained 62 lesion pairs. These 62 radiograph pairs (one with a detected nodule and one missed) are the clinical material used in this study.

PHYSICAL PARAMETERS

We measured 4 parameters of each of the 62 nodule pairs: contrast, surround complexity, size, and edge index. Density measurements were performed with a Macbeth Densitometer TD102 having a 1-mm resolution. Multiple measurements were made just inside and just outside the edge of the nodule. The inside measurements were used to calculate mean nodule density and the outside measurements the mean surround density. The contrast, ΔD, is the difference between these two mean values. The rate of change of the density measurements around the nodule border was used to calculate the complexity of the area immediately surrounding the nodule. This approximately amounts to forming the Laplacian, ∇^2, of the density surrounding the nodule. These measurements have been described in detail in a previous paper (7).

Since many of the nodules were not round, but irregular, a simple measurement of diameter was not sufficient to express the size, and this parameter *S* was defined as the geometric mean of the longest diameter d_{max} and the shortest diameter, d_{min}:

$$S = \sqrt{d_{max} \cdot d_{min}} \qquad (2).$$

The characteristics of the nodule edge were the most difficult nodule property to quantify. At least two parameters are considered important: the steepness of the edge gradient (the sharpness of the border perpendicular to the edge) and the smoothness or regularity of the border around the perimeter of the nodule. Even though the edges may be sharp, a ragged irregular nodule is more difficult to detect than a smooth regular shape.

A visual rating scale of 1 to 5 was used for each parameter. For the gradient, 1 was steep and 5 was shallow; for the contour, 1 was smooth and 5 was irregular. The two ratings were then added, giving values between 2 (steep gradient, smooth contour) and 10 (low gradient, ragged edges). This sum was divided by 6 to give an edge index, EI:

$$EI = \frac{1}{6}(\text{Gradient} + \text{Contour}) \qquad (3).$$

Three observers independently rated each nodule, and their results were in good agreement.

RESULTS

The studies done with artificial nodules (5–7) used a test series and multiple observations to establish the probability of detecting each nodule. Although this could have been done here, we were more concerned with using nodules actually missed in clinical practice. We therefore studied the ability of the descriptors such as size and contrast to

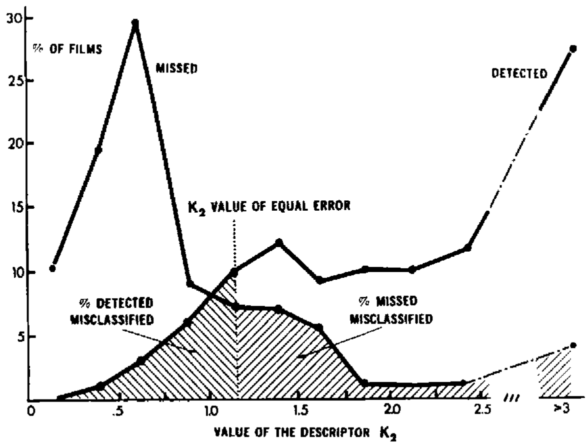

Fig. 1. Distribution of the two populations (missed and detected) as a function of the conspicuity, K_2. The K_2 value at which the two types of errors in classification are equal yields a measure of how much separation is obtained. In this case the error is 23% of each population, since each shaded area represents 23% of the total area under each curve.

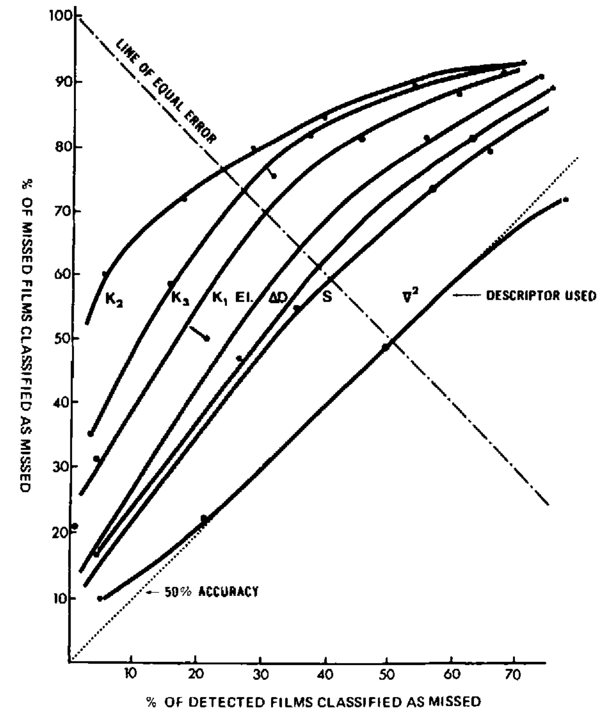

Fig. 2. The ability of various descriptors to separate correctly the two populations (missed and detected lesions). These curves were derived from experimental data similar to those in Figure 1, by plotting the area percentages of missed and detected lesions falling below various values of each descriptor.

sort the radiographs into two populations (missed and detected) rather than studying the probability of detection. For this purpose all of the data on each radiograph was recorded on punched cards and the sorting by descriptors carried out by computer.

Consider now a given descriptor. The distribution of the two populations, missed and detected lesions, can be plotted as a function of the value of this descriptor, resulting in the two distribution curves shown in Figure 1. Similar distributions can be plotted for any chosen descriptor. Since the two distribution curves overlap, there is no critical descriptor value which will completely separate them, and there will therefore be errors in classification. A measure of the "sorting ability" of the descriptor can be found by choosing that value for which the two types of classification errors are equal, as represented by the equality of the two shaded areas in Figure 1.

In order to compare the "sorting abilities" of different descriptors, a composite graph of performance curves (Fig. 2) can be derived. The curves are computed in the following manner. Each curve (in Figure 2) represents the performance of one descriptor and is derived from two distributions of the type shown in Figure 1. Each point on a curve in Figure 2 corresponds to one value of the descriptor, and its coordinates are the *areas* under the two distribution curves in Figure 1 up to that descriptor value. Increasing the descriptor value increases both areas, *i.e.*, the percentage of missed lesions classified as missed and the percentage of detected lesions classified as missed. In the best case, if the two populations were completely separate, the resulting performance curve would lie at the upper left hand corner of the coordinate system in Figure 2. In the worst case, if the two population distributions were completely overlapping, for each increase in descriptor value an identical increase in *both* area percentages would be obtained, and the performance curve would be a 45° straight line, as shown by the dotted line in Figure 2.

The advantage of using this graphic presentation instead of just computing a figure of merit is that the performance of various descriptors can be easily compared: the better the descriptor the closer the curve will fall to the upper left corner, and the worse it is the closer it will be to a 45° line.

The equal-error value for each descriptor can also be readily found: it is the intersection of the curve with the negative 45° line linking the two 100% points of the coordinate system, shown as a dashed line in Figure 2.

INTERPRETATION

The curves for each parameter taken singly and in combination were computed (Fig. 2). The parameters used, the parameter value at equal error, and the equal error rates are shown in TABLE I. The curves show that measuring the surround complexity ∇^2 alone does not separate the two populations at all. This is to be expected, since complexity is a property of the surround and is independent of the nodule. Some separation can be achieved using the nodule size, S, or the contrast ΔD, or the edge index EI.

Table I: Separability of Detected/Missed Lesions by Various Parameters

Parameter	Definition	Parameter Value at Equal Error	Equal-Error Rate (%)
Surround Complexity	∇^2	$\nabla e^2 = .12$	50
Lesion Size	$S = \sqrt{d_{max} \times d_{min}}$	$S_e = 17mm$	40
Lesion Contrast	ΔD	$D_e = .15$	39
Edge Index	$EI = \frac{value(2-10)}{6}$	$EI_e = 1$	36
Conspicuity K_1	$\Delta D/\nabla^2$	$K_{1e} = 1.2$	30
Conspicuity K_2	$\Delta D/\nabla^2 \cdot EI$	$K_{2e} = 1.2$	23
Conspicuity K_3	$S \cdot \Delta D/\nabla^2 \cdot EI$	$K_{3e} = 21mm$	27

These results are intuitively reasonable, since the radiograph in which the nodule was detected was obtained some time after the one on which it was missed. During this interval the nodules increased in size and, therefore, in contrast as well, so the detected nodules should, on the average, be larger and have more contrast than those which were missed.

Better separation is accomplished by using various combinations of the parameters expressed as conspicuity indices K_1, K_2, and K_3. K_1 is a first approximation of the conspicuity as defined before: $K_1 = \Delta D/\nabla^2$. This shows an improved separation, in agreement with previously published findings on simulated nodules (7). It also agrees very well with results based on preliminary data (1) based on about half the number of real lesions. Further improvements are achieved by broadening the definition of conspicuity to include the edge index, as in the case of K_2, defined as $K_2 = K_1/EI$. The improvement in separation is quite marked. No improvement was shown, however, by including the size of the nodule S, as defined in $K_3 = K_2.S$, even though nodule size taken alone was a discriminator.

DISCUSSION

A descriptor can be derived from measurements made on radiographs that will predict if a lesion can be detected by a skilled reader. It can do so with a probability of better than 75%. The importance of the complexity of the area immediately surrounding the nodule, which had earlier been demonstrated with the use of artificial nodules, has been verified using real cases. The measurements are still somewhat gross, but we are currently developing computer programs for use with an image digitizer that will increase the precision.

Once the exact functional relationship between the probability of detection p and the conspicuity K (as derived from physical measurements) is known, radiographic techniques can be compared in terms of physical measurements made with a scanning microdensitometer that will predict without observer tests the probable performance of readers. And reader performance is the ultimate test of any diagnostic imaging system.

Consider image processing. Increasing K will improve detection. K can be increased by increasing contrast or enhancing edges, provided that the surround is left unchanged. This is what is done in radiopaque contrast material studies. However, any attempt to devise an effective *general* image processing method either by computer, optical, or television means has been unsuccessful, because although a processor might enhance the contrast and edges of the lesion, it also affects the surrounding structures, and K is either unchanged or actually decreased (4, 6). Furthermore, our prior studies have also shown that image processing methods which increase the surround complexity also increase the false positive rate. The present study dealt with the retrospective analysis of abnormal radiographs only, and hence yielded data only on false negatives but not on false positives.

Another approach to improved detection is to reduce surround complexity without changing lesion contrast. One such procedure is image subtraction, where one image containing surround structures only is subtracted from another that contains the feature to be detected. This technique is often used in angiography. Recently efforts have been made to extend subtraction to other areas of radiography, such as in the subtraction of chest radiographs to detect changes (3, 5). Although it has been shown in theory that this can improve detection, it has yet to work in practice because of the technical difficulty of achieving good registration of images produced at different examinations.

Continuing studies of visual psychophysics as applied to radiological images will provide the necessary theoretical framework for image processing and image analysis, in addition to providing an insight into the causes of detection error in radiology.

Diagnostic Radiology Research Lab.
Temple University School of Medicine
Philadelphia, Penn. 19140

REFERENCES

1. Breckenridge JW, Bird GC: Errors of omission in pulmonary nodule detection. Appl Radiol (In press)
2. Guiss LW, Kuenstler P: A retrospective view of survey photofluorograms of persons with lung cancer. Cancer **13**:91–95, Jan 1960
3. Kinsey JH, Vannelli BD, Fontana RS, et al: Application of Digital Image Change Detection to Diagnosis and Follow-up of Cancer Involving the Lungs. [In]: Proc Meet Applicat Optical Instr Med, Atlanta, Geo, Sep 1975, vol 4, pp 99–112
4. Kundel HL, Revesz G, Stauffer HM: The electrooptical processing of radiographic images. Radiol Clin North Am **7**:447–460, Dec 1969
5. Kundel HL, Revesz G: Lesion conspicuity, structured noise and film reader error. Am J Roentgenol **126**:1,233–1,238, Jun 1976
6. Revesz G, Kundel HL, Haas C: Electronic techniques for radiological image processing. Med Bio Eng **7**:393–399, Oct 1969
7. Revesz G, Kundel HL, Graber MA: The influence of structured noise on the detection of radiologic abnormalities. Invest Radiol **9**: 479–486, Nov–Dec 1974