II-6

J Clin Epidemiol Vol. 46, No. 10, pp. 1151–1158, 1993
Printed in Great Britain

0895-4356/93 $6.00 + 0.00
Pergamon Press Ltd

A NEW SYSTEM OF VISUAL PRESENTATION OF ANALYSIS OF TEST PERFORMANCE: THE "DOUBLE-RING" DIAGRAM

MILTIADIS A. STEFADOUROS*
Department of Medicine, King Fahad Hospital, Riyadh, Saudi Arabia

(*Received in revised form 30 April 1993*)

Abstract—Substitution of graphic representation for extensive lists of numerical statistical data is highly desirable by both editors and readers of medical journals, faced with an exploding abundance of contemporary medical literature. A novel graphic tool, the "double-ring diagram", is described herein which permits visual representation of information regarding certain statistical variables used to describe the performance of a test or physical sign in the diagnosis of a disease. The diagram is relatively easy to construct on the basis of a number of primary data such as the prevalence and the true positive, true negative, false positive and false negative test results. These values are reflected in the diagram along with the values of other statistical variables derived from them, such as the sensitivity, specificity, predictive values for positive and negative test result, and accuracy. This diagram may be useful in visualizing a test's performance and facilitating visual comparison of performance of two or more tests.

Bayesian Sensitivity Specificity Predictive value Accuracy Statistics

INTRODUCTION

The recent immense proliferation of biomedical literature [1] has resulted in a large number of articles competing for limited space in medical journals and, once published, for the readers' attention. Editors recommend conciseness and brevity [2, 3], discourage lengthy tables [4], advise that tables be designed in ways that combine economy of space with readability [5], or urge the use of appropriate figures as substitutes for extensive lists of quantitative data [6]. The length of an article beyond certain stated limits may adversely affect the probability of its acceptance for publication [7] while in borderline cases the chances for publication may improve if the author is prepared to consent to omission of less vital parts of the article [8].

*All correspondence should be addressed to: Miltiadis A. Stefadouros, MD, Chairman, Department of Medicine, King Fahad Hospital, PO Box 22490, Riyadh 11426, Saudi Arabia.

This article introduces a new concept of a special figure, called a "double-ring diagram", that is the visual counterpart of a list of numbers describing the performance of a diagnostic test according to established principles of analysis (Appendix).

DESCRIPTION OF THE METHOD

The double-ring diagram consists of a system of two concentric rings each one typically (but not invariably) having a shaded and a blank segment (Fig. 1). The outer ring represents all subjects in the sample who either have (shaded) or do not have (blank) the disease or attribute in question, as established by means other than the test itself. The inner ring represents the results of the test under evaluation; positive test results are shown by the shaded segment and negative tests by the blank segment of the inner ring. The rules for the construction of the double-ring diagram are described below.

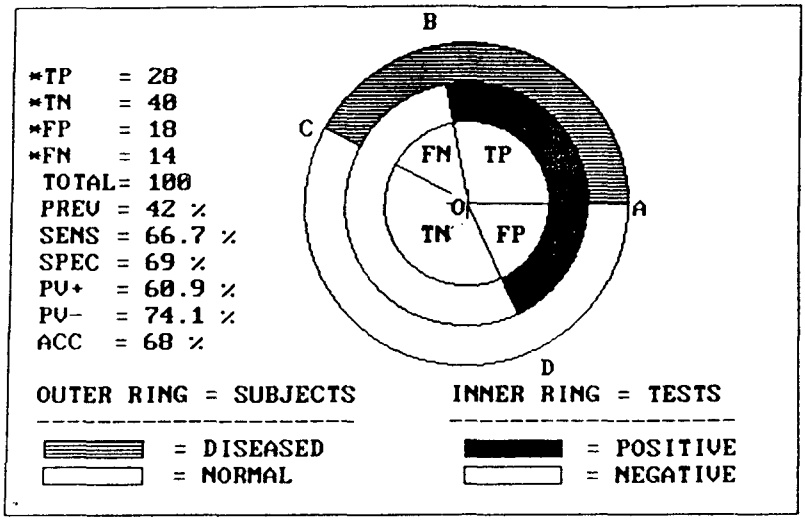

Fig. 1. Typical double-ring diagram depicting the performance of a hypothetical test including TP, TN, FP, and FN results. Arbitrary values for these variables were entered in the computer which constructed the diagram and calculated and displayed values for prevalence (PREV), sensitivity (SENS), specificity (SPEC), predictive value for positive (PV+) and negative result (PV−), and accuracy (ACC). For the principles of construction see text. *Values entered.

In a typical case, each of the 2 rings has a shaded and a blank segment but neither of the 2 segments of one ring corresponds exactly with the respective segment of the other ring (Fig. 1). In such a case 4 radii can be drawn from the common center to the 4 transition points (2 for each ring), defined as the boundaries between the shaded and the blank segment of the same ring. These 4 radii divide the circle into 4 sectors. Starting from the "3 o'clock" radius (OA) and proceeding counterclockwise, the first sector (AOB) comprises all subjects with the disease (shaded outer ring) whose test results are positive (shaded inner ring), i.e. the true positives (TP). The next sector (BOC) encompasses all subjects with the disease (shaded outer ring) who tested negative (blank inner ring), that is false negatives (FN). Proceeding in a similar fashion and direction, there follows the sector COD of true negative results (TN) where both rings have blank segments. The fourth sector (DOA) is the one where the test results were positive (shaded inner ring) in subjects devoid of the disease (blank outer ring); these are the false positives (FP).

Construction of the double-ring diagram

The double-ring diagram can be drawn manually using a compass for drawing circles, a ruler and a protractor. The steps are as follows (Fig. 1):

1. Draw 3 concentric circles of radii R1 (inner), R2 (middle) and R3 (outer). The absolute magnitude of these radii is not important but a 3:4:5 ratio for R1:R2:R3 is convenient because it defines 2 rings of equal thickness and provides sufficient room for labelling the sectors on the inner circle.
2. Define origin and direction. We have followed the standard convention used in trigonometry, setting the origin (0°) at the "3 o'clock" position and defining the counterclockwise direction as positive. It should be noted that any other convention would be acceptable if clearly stated.
3. Determine values for TP, TN, FP and FN by comparing individual test results with the known presence or absence of disease on each subject in the sample. For the purpose of plotting, translate the values of each of these four variables into degrees by multiplying by 360 and dividing by the total number of subjects in the sample ($N = TP + TN + FP + FN$):

   Variable° (in degrees)
   $= \text{Variable (in numbers)} \times (360/N)$

4. Shade the rings. Starting from the origin (A) in a counterclockwise direction, define and shade on the outer ring an arc AC whose measure in degrees is equal to the sum of TP° and FN°, representing all subjects with the disease. Repeat the same procedure to define and shade on the inner ring an arc AB equal to TP° representing the true positive test

Table 1. Visual comparisons of segments of the 'double ring' diagram and their correspondence to variables of analysis

Variable	Symbol/formula	Corresponding segment(s) or segment ratios in the 'double ring' diagram
	Primary	
True positive	TP	Shaded segment AB, inner ring
True negative	TN	Blank segment CD, inner ring
False positive	FP	Shaded segment AD, inner ring
False negative	FN	Blank segment BC, inner ring
	Derived	
Prevalence (in the sample)	$\frac{TP + FN}{TP + FN + FP + TN}$	Shaded segment AC, outer ring / Complete outer ring
Sensitivity	$\frac{TP}{TP + FN}$	Shaded segment AB, inner ring / Shaded segment AC, outer ring
Specificity	$\frac{TN}{TN + FP}$	Blank segment CD, inner ring / Blank segment CA, outer ring
Predictive value +	$\frac{TP}{TP + FP}$	Shaded segment AB, inner ring / Shaded segment BD, inner ring
Predictive value −	$\frac{TN}{TN + FN}$	Blank segment CD, inner ring / Blank segment BD, inner ring
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	(Shaded segment AB + Blank segment CD)* / (Shaded segment AD + Blank segment BC)*

*On the inner ring.

results. Finally, starting from the origin A and proceeding in a negative, clockwise direction, define and shade on the inner ring an arc AD equal to FP, representing the false positives.

5. Draw the radii to points A, B, C and D thereby defining sectors AOB, BOC, COD and DOA which correspond to variables TP, FN, TN and FP, respectively. On each sector write the numerical value of the appropriate variable.

In addition to depicting these 4 primary variables, the double-ring diagram permits visual assessment of a number of derived variables such as the prevalence, sensitivity, specificity, predictive values and accuracy of the test in question. For each of these variables, this is done by visually comparing the length of the appropriate segments, as listed in detail in the Table 1. For example, if the length of the shaded segment AB on the inner ring is equal to 2/3 of the length of the shaded segment AC on the outer ring—as is the case on the double-ring diagram of Figure 1—then the sensitivity of the test is 66.7%.

The above description and rules of construction of the double-ring diagram referred to a typical case, where all 4 possible outcomes—TP, TN, FP, and FN—are represented in the sample i.e. none of the 4 variables has zero value. This is not always the case since, occasionally, 1 or more of the 4 variables is zero. In such cases, adherence to the same rules for construction mentioned above results in a double-ring diagram that has no graphic representation of segments (and sectors) depicting variables with a value of zero.

Constructing double-ring diagrams with a computer

The process can be greatly facilitated by use of a computer with graphics capability, suitably programmed to perform the task according to an algorithm based on these steps. Figure 2 illustrates such a computer graphics output obtained by use of a simple program we wrote in GWBASIC. On the menu, a total of 11 variables are listed in 3 groups. The first group consists of the 4 primary variables (TP, TN, FP, FN); the second group includes the sample size (TOTAL), prevalence (PREV), sensitivity (SENS), and specificity (SPEC); and the last group provides predictive values for positive (PV+) and negative (PV−) results, and accuracy (ACC). The operator enters absolute values for either the first (option 1) or the second group

(option 2) and the computer calculates values for the other two groups, using the appropriate equations (see Appendix). From the entered or the calculated values, the computer then constructs, displays on the screen and prints on the attached printer the double-ring diagram together with values for all 3 groups.

Option 1 is used in the course of evaluating the performance of a test as applied on a sample. In this case, PREV represents the prevalence in the sample. To the extent that predictive values depend on prevalence as well as on the test's sensitivity and specificity, the computed values for PV+ and PV− pertain to the sample, or to a population whose prevalence of the disease is the same as that in the sample.

Option 2 is used to determine the performance of a test of known sensitivity and specificity, as applied on a population of known size and disease prevalence. Values for these 4 variables are entered and the expected values for TP, TN, FP, FN, as well as those for the PV+, PV−, and accuracy of the test, as they apply to that population, are calculated. An example of a diagram obtained by the option 2 is given in Fig. 2.

Application of the double-ring diagram

The usefulness of the double-ring diagram can be demonstrated by applying the method to the results of a published, multicenter study conducted by the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) [9]. In that study, the performance of lung ventilation/perfusion (V/Q) scan was tested on a randomly selected group of 933 patients for whom a V/Q scan had been requested because of clinical symptoms suggestive of acute pulmonary embolism. Pulmonary angiography was the "gold standard" on the basis of which the presence or absence of pulmonary embolism was established. After exclusion of patients in whom either of the two studies was not completed, there remained 755 patients with complete and interpretable V/Q scans and pulmonary angiograms.

The V/Q scans were classified in 4 categories: high probability, intermediate probability, low probability, and near normal/normal, according to a set of preestablished study criteria. Similarly, the angiograms were classified with regard to pulmonary embolism into 3 categories: present, absent, and uncertain. Table 2 is a modified, composite presentation of the data listed in Tables 4 and 5 of the cited article [9]. Only data related to the angiographic categories "present" and "absent", together totalling 731 subjects, are included in Table 2, since these were the 2 categories used to compute values for sensitivity and specificity.

As can be seen in Table 2, high probability V/Q scans were highly specific (97%) in excluding pulmonary embolism but not sufficiently sensitive (41%) in detecting its presence. When

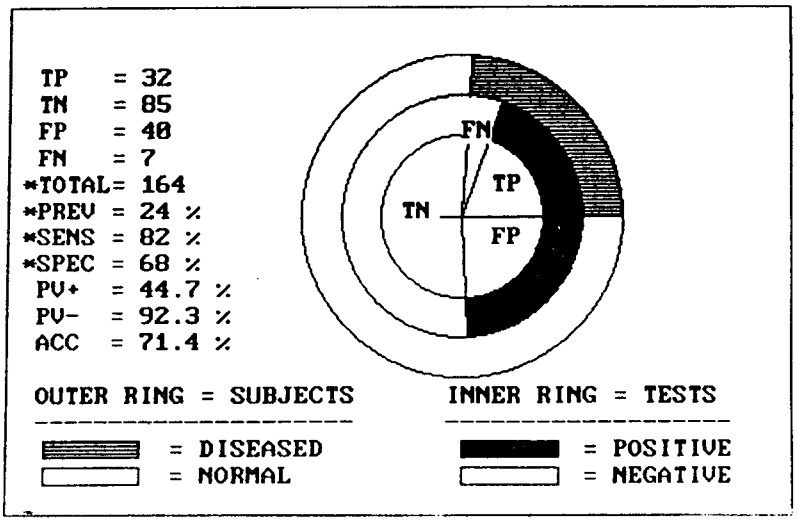

Fig. 2. Diagram obtained using option 2 (see text). The known values for sensitivity and specificity of a hypothetical test in diagnosing a given disease were entered, together with the known prevalence of that disease in a population of size = TOTAL. Values for the predictive values for positive and negative results and accuracy, as applied to that population, are calculated as shown. Also shown are values for TP, TN, FP, and FN expected from applying the test to this population. All abbreviations and symbols are as in Fig. 1.

Table 2. Comparison of lung V/Q scan category with angiographic findings*

V/Q scan category	Pulmonary embolism by pulmonary angiography: Present	Absent	Sensitivity %	Specificity %
High probability	102	14	↓41	↓97
			↑ 82	↑ 52
Intermediate probability	105	217	↓	↓
			↑ 98	↑ 10
Low probability	39	199	↓	↓
Near normal/Normal	5	50		
Total	251	480		

Abbreviations: V/Q = Ventilation/Perfusion.
*Modified composite table of Tables 4 and 5 from Ref. [9].

a V/Q scan was considered positive if it belonged to either high or intermediate probability, sensitivity rose to 82% but at a substantial cost to specificity, which fell to 52%. Finally, when the designation "positive" was broadened enough to include a V/Q scan of either high intermediate, or low probability, the test became highly sensitive (98%) but its specificity plummeted to 10%. Thus, none of the 3 designations of "positive" V/Q scan—high probability; high or intermediate probability; and high, intermediate, or low probability—reached clinically reassuring levels for both sensitivity and specificity. The beneficial effect of combining clinical judgement regarding the likelihood of pulmonary embolism with results of V/Q scans on the performance of the latter, was a most interesting finding of the study [9]. It should be mentioned that the study had a number of other, equally important aspects which are not considered here.

Figure 3 illustrates the statistics of Table 2 in the form of 3 double-ring diagrams, 1 for each of the 3 examined schemes for "positive" V/Q scan result; diagram 1 (left panel) for the high probability scan; diagram 2 (middle panel) for the high or intermediate probability scan; and diagram 3 (right panel) for the high, intermediate, or low probability scan. Each diagram is accompanied by a list of values of pertinent statistics for reference. In the following discus-

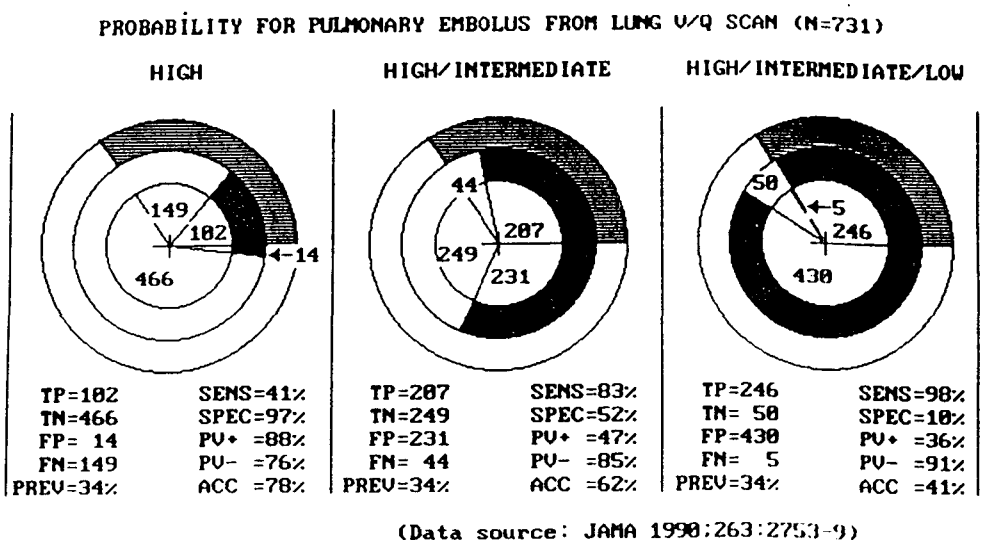

Fig. 3. Double-ring diagrams comparing performance of lung ventilation/perfusion (V/Q) scans in the diagnosis of pulmonary embolus, depending upon the definition used for "positive" V/Q scan. Note progressive increase in SENS and PV−, and a parallel decrease in SPEC, PV+, and ACC as the definition of positive scan expands from high probability (left panel), to high or intermediate probability (middle panel), and to high, intermediate or low probability (right panel). These changes are reflected on the size and degree of overlap of the shaded and blank segments of the inner and outer rings of each diagram. See text for details. All abbreviations and symbols are as in Fig. 1. (Primary data obtained from Ref. [9].)

sion these values, rounded to the nearest integer, are quoted in parentheses.

A visual impression of the *prevalence* of the disease (i.e. pulmonary embolism) in the sample, can be obtained by noting that on each diagram the shaded segment on the outer ring—denoting subjects with embolism present on the angiogram—occupies roughly one third (34%) of the total ring representing the 731 subjects with unambiguous angiograms.*

To glean information on *sensitivity*, one may observe what proportion of the shaded outer ring (positive angiograms) is overlapped by the shaded part of the inner ring (positive scans). On visual inspection of diagram 1, this proportion is a little less than $\frac{1}{2}$ (41%), and increases to about $\frac{4}{5}$ (82%) in diagram 2; and in diagram 3, almost all (98%) of the shaded outer ring is paralleled by shaded inner ring.

For *specificity*, attention is focused on the fact that in diagram 1 almost all (97%) of the blank segment on the outer ring (negative angiograms) is matched by blank inner ring (negative scans). This proportion is about $\frac{1}{2}$ (52%) in diagram 2, and falls to about $\frac{1}{10}$ (10%) in diagram 3.

In a similar fashion, one can obtain a rough impression of *predictive value positive* by noting that in diagram 1 almost $\frac{9}{10}$ (88%) of the shaded inner ring—denoting positive scans—are overlapped by the shaded segment of the outer ring, i.e. belongs to patients with positive angiograms. This proportion progressively falls to about $\frac{1}{2}$ (47%) and close to $\frac{1}{3}$ (36%) in diagrams 2 and 3, respectively.

In diagram 1, the *predictive value negative* (76%) of the test is appreciated by noting that almost $\frac{3}{4}$ (76%) of the blank segment of the inner ring (negative scans) corresponds to blank outer ring, i.e. to patients with negative angiograms. The corresponding proportion is about $\frac{4}{5}$ (85%) in diagram 2 whereas in diagram 3 almost all (91%) of the blank inner ring segment is matched by a blank outer ring corresponding to patients with negative angiograms.

Finally, a visual estimate of the test's overall *accuracy* can be made by noting that the "true" sectors labeled TP and TN taken together occupy roughly $\frac{3}{4}$ (78%) of the full circle in diagram 1, $\frac{2}{3}$ (62%) of the circle in diagram 2, and about $\frac{1}{3}$ (40%) of the circle in diagram 3.

*The authors [9] quote a 33% prevalence of pulmonary embolism among 755 patients with pulmonary angiograms, a figure that includes 731 in whom pulmonary embolism was "present" or "absent" plus 24 patients with "uncertain" angiograms. The double-ring method makes no provision for uncertain test results (see Discussion). Accordingly, it calculated prevalence of 34% among 731 patients with unambiguous angiograms.

DISCUSSION

The usefulness of a test (or a symptom) as a means for confirming the presence or establishing the absence of a particular disease depends on its sensitivity and specificity. Calculation of these performance indices requires that each positive or negative test result can be designated as true or false according to whether the outcome is concordant or discordant, respectively, with the designation of the subject as having (positive) or not having (negative) the disease as determined by a different, presumably perfect method that serves as the "gold standard" against which the test under consideration is compared. The number of true positives, true negatives, false positives and false negatives are then fed into simple algebraic formulas that permit computation of the test's sensitivity and specificity. When these performance indices are applied to a population with known prevalence of the disease in question, the predictive values and accuracy of the results can also be computed.

Presenting statistics in the form of graphics permits the reader to easily grasp and retain the essence of information hidden in a multitude of numerical data, and facilitates drawing the proper conclusions and inferences. This paper describes a new method for presenting data related to test performance in a concise fashion that is easy to grasp and retain without sacrificing its informational content. Its utility is analogous to that of a scatter plot which is an effective way of conveying the degree of association between two variables, an association intrinsically present in, albeit difficult to discern from, a list of the individual values.

The double-ring diagram fulfills the 2 major requirements of an effective statistical graphics tool, viz. visual simplicity and richness of informational content. Its visual simplicity is due to the fact that it involves 2 concentric rings, each with a shaded and a blank segment, plus values for 4 variables (TP, TN, FP, and FN). Once the principles of construction and the meaning of each ring segment are understood, simple inspection enables the viewer to grasp the essentials: whether the test performance is poor, average, good, or excellent, both, in a general sense and also with regard to specific statistical

variables, such as in tests with, say, high specificity but low sensitivity etc. With a certain degree of familiarity, comparison of performance of two or more tests is easier by inspecting their double-ring diagrams than by studying the tables listing their individual performance indices.

The degree to which a particular graphic diagram fulfills the other requirement mentioned, that of the richness of informational content, can be appreciated by considering the additional amount of information that can be extracted from the diagram itself, over and above that presented in the form of numerical data displayed in it. This requirement is met to a satisfactory degree by the double-ring diagram since it contains information regarding all those statistical variables derived from the 4 displayed primary variables (TP, TN, FP, FN), viz. prevalence (in the sample), sensitivity, specificity, predictive values (positive and negative), and accuracy. Since these derivative variables represent ratios of one primary variable (or, in the case of accuracy, the sum of 2 primary variables) and the sum of 2 or more primary variables, their visual appreciation on the double-ring diagram is based on comparisons between ring segments (or sums of ring segments) with other segments (or sums of ring segments), as shown on Table 1. Such visual comparisons give a rough impression of the magnitude of the derived variable rather than its accurate numerical value which, however, can still be computed from the displayed primary variables. Of course, the need for such computations could be avoided if a more comprehensive—but visually more "cluttered"—version of the double-ring diagram were adopted, that would display values for all variables, primary and derived alike.

Notwithstanding its merits, the double-ring diagram has two important limitations: first, it can accept only dichotomous values for the disease (present/absent) or the test results (positive/negative or normal/abnormal). Indeterminate, intermediate, and uninterpretable test results are not uncommon in most studies, and Simel *et al.* [10] have drawn attention to the need for their inclusion in the data assessment. Such ambiguous test results are not depicted in the double-ring diagram in the form proposed here. Although technically feasible, such depiction would render the diagram sufficiently complex to defeat the object of its very existence, which is facilitating the transmission of information. Also, although the double-ring diagram permits appreciation of such variables as prevalence ("pre-test probability") and predictive value ("post-test probability") by means of visual comparisons of different sectors, it does not visually represent the *relation* between these 2 variables—a relation that is central to the Bayesian analysis. These deficiencies subtract somewhat, but do not negate, the usefulness of the double-ring diagram.

The way the double-ring diagram presents certain special cases is worth mentioning. Thus, in the case where all test results are positive (TN = 0; FN = 0), the probability of a subject having the disease or attribute in question is:

$$\text{Prevalence} = \frac{TP + FN}{TP + FP + TN + FN} = \frac{TP}{TP + FP} = PV+.$$

Similarly, in the case where all test results are negative (TP = 0; FP = 0), the probability of a subject not having the disease is:

$$1 - \text{Prevalence} = 1 - \frac{FN}{FN + TN} = \frac{TN}{FN + TN} = PV-.$$

Accordingly, when all (or nearly all) test results are positive or negative, calculation of PV+ or PV− offers no advantage over using prevalence or 1 − prevalence, respectively, as a means of predicting the likelihood of presence or absence of the disease in a given subject. Such tests are of marginal value, and on the double-ring diagram are depicted in the form of an inner ring which is either totally shaded or totally blank. For example, in the diagram of Fig. 3(C) almost all (93%) of the inner ring is shaded, reflecting the percentage of positive lung scans (681/731); the very small difference between prevalence (34%) and PV+ (36%) renders calculation of the latter redundant.

In conclusion, this paper introduces a new concept for presenting results of analysis of test performance in a graphic form. This concept may be of value in summarizing the essential facts of the analysis, thereby facilitating their retention and permitting comparison of the performance of 2 or more tests.

REFERENCES

1. Bernier CL, Yerkey AN. **Cogent Communication: Overcoming Information Overload.** Westport, CT: Greenwood Press; 1979: 39.
2. Instructions to Authors. **Gastroenterology** 1989; 96.
3. Information for Authors. **Am J Clin Nutr** 1987; 45: 141.
4. Writing for the Lancet. **Lancet** 1991; 338: ii.
5. Information for Authors. **N Engl J Med** 1991; 325: 748.
6. Instructions to Authors. **Circulation** 1991; 84: A10.
7. Roberts WC. In search for the masterful figure [Editorial]. **Am J Cardiol** 1987; 60: 633–634.
8. Instructions to Authors. **Am J Cardiol** 1991; 68: 740.
9. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism: results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED). **JAMA** 1990; 263: 2753–2759.
10. Simel DL, Feussner Jr, DeLong ER, Matchar DB. Intermediate, indeterminate, and uninterpretable diagnostic test results. **Med Decis Making** 1987; 7: 107–114.
11. Rosner B. **Fundamentals of Biostatistics.** Boston: PWS Publishers; 1982: 56–57.

APPENDIX

The test's sensitivity (SENS), specificity (SPEC) and accuracy (ACC) are calculated from the values of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) results, as:

$$\mathrm{SENS} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{SPEC} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

The disease (or attribute) prevalence (PREV), predictive value positive (PV+) and predictive value negative (PV−) in the sample are computed as:

$$\mathrm{PREV} = \frac{\mathrm{TP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

$$\mathrm{PV+} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$\mathrm{PV-} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}$$

When applied to a population of known probability of disease (i.e. prevalence), the test's predictive values can be calculated as follows [11]:

$$\mathrm{PV+} = \frac{\mathrm{PREV} \times \mathrm{SENS}}{[\mathrm{PREV} \times \mathrm{SENS} + (1 - \mathrm{PREV}) \times (1 - \mathrm{SPEC})]}$$

$$\mathrm{PV-} = \frac{(1 - \mathrm{PREV}) \times \mathrm{SPEC}}{[(1 - \mathrm{PREV} \times \mathrm{SPEC}) + \mathrm{PREV} \times (1 - \mathrm{SENS})]}.$$

Values for all derived variables vary from 0 to 1. A popular expression as percentages (0–100%) can be obtained by multiplying their absolute values by 100.