JCI The Journal of Clinical Investigation

Application of conditional probability analysis to the clinical diagnosis of coronary artery disease.

G A Diamond, … , D S Berman, H J Swan

J Clin Invest. 1980;65(5):1210-1221. https://doi.org/10.1172/JCI109776.

Research Article

Analysis of multiple noninvasive tests offers the promise of more accurate diagnosis of coronary artery disease, but discordant test responses can occur frequently and, when observed, result in diagnostic uncertainty. Accordingly, 43 patients undergoing diagnostic coronary angiography were evaluated by noninvasive testing and the results subjected to analysis using Bayes' theorem of conditional probability. The procedures used included electrocardiographic stress testing for detection of exercise-induced ST segment depression, cardiokymographic stress testing for detection of exercise-induced precordial dyskinesis, myocardial perfusion scintigraphy for detection of exercise-induced relative regional hypoperfusion, and cardiac fluoroscopy for detection of coronary artery calcification. The probability for coronary artery disease was estimated by Bayes' theorem from each patient's age, sex, and symptom classification, and from the observed test responses. This analysis revealed a significant linear correlation between the predicted probability for coronary artery disease and the observed prevalence of angiographic disease over the entire range of probability from 0 to 100% (P less than 0.001 by linear regression). The 12 patients without angiographic disease had a mean posttest likelihood of only 7.0 +/- 2.6% despite the fact that 13 of the 60 historical and test responses were falsely "positive." In contrast, the mean posttest likelihood was 94.1 +/- 2.8% in the 31 patients with angiographic coronary artery disease, although 45 of the 155 historical and test responses were […]

Find the latest version:

https://jci.me/109776/pdf

Application of Conditional Probability Analysis to the Clinical Diagnosis of Coronary Artery Disease

GEORGE A. DIAMOND, JAMES S. FORRESTER, MICHAEL HIRSCH, HOWARD M. STANILOFF, RAN VAS, DANIEL S. BERMAN, and H. J. C. SWAN,
Division of Cardiology, Department of Medicine, Cedars-Sinai Medical Center, and the Department of Medicine, University of California, Los Angeles School of Medicine, Los Angeles, California 90048

ABSTRACT Analysis of multiple noninvasive tests offers the promise of more accurate diagnosis of coronary artery disease, but discordant test responses can occur frequently and, when observed, result in diagnostic uncertainty. Accordingly, 43 patients undergoing diagnostic coronary angiography were evaluated by noninvasive testing and the results subjected to analysis using Bayes' theorem of conditional probability. The procedures used included electrocardiographic stress testing for detection of exercise-induced ST segment depression, cardiokymographic stress testing for detection of exercise-induced precordial dyskinesis, myocardial perfusion scintigraphy for detection of exercise-induced relative regional hypoperfusion, and cardiac fluoroscopy for detection of coronary artery calcification.

The probability for coronary artery disease was estimated by Bayes' theorem from each patient's age, sex, and symptom classification, and from the observed test responses. This analysis revealed a significant linear correlation between the predicted probability for coronary artery disease and the observed prevalence of angiographic disease over the entire range of probability from 0 to 100% ($P < 0.001$ by linear regression). The 12 patients without angiographic disease had a mean posttest likelihood of only 7.0±2.6% despite the fact that 13 of the 60 historical and test responses were falsely "positive." In contrast, the mean posttest likelihood was 94.1±2.8% in the 31 patients with angiographic coronary artery disease, although 45 of the 155 historical and test responses were falsely "negative." In 8 of the 12 normal patients, the final posttest likelihood was under 10% and in 26 of the 31 coronary artery disease patients, it was over 90%. These estimates also correlated well with the pooled clinical judgment of five experienced cardiologists ($P < 0.001$ by linear regression). The observed change in probability for disease for each of the 15 different test combinations correlated with their information content predicted according to Shannon's theorem ($P < 0.001$ by linear regression).

These results support the use of probability analysis in the clinical diagnosis of coronary artery disease and provide a formal basis for comparing the relative diagnostic effectiveness and cost-effectiveness of different test combinations.

Received for publication 13 August 1979 and in revised form 30 November 1979.

INTRODUCTION

The availability and clinical appeal of noninvasive techniques has led to their combined use in the diagnosis of coronary artery disease. These test combinations include electrocardiographic stress testing (1–3), cardiac fluoroscopy (4–6), cardiokymography (7–9), and thallium scintigraphy (10–12). The rationale for a multiple test approach resides in the fact that the predictive accuracy of any one of these tests is quite low when applied to a population with low disease prevalence, whereas the accuracy of two or more tests is very high when all results are in agreement (6, 9, 12–15). However, the frequent occurrence of discordant test responses and the increased cost of multiple testing limit the ultimate usefulness of these formats.

Some investigators have suggested the use of Bayes' theorem of conditional probability to assist in the interpretation of discordant test results (16–21). This method of analysis provides an explicit statement of the statistical probability that a given patient has disease, rather than an inflexible categorical diagnosis ("normal" vs. "abnormal"). The advantage of the probabilistic approach is that it realistically expresses the degree of diagnostic uncertainty in a manner that is relevant to the clinical decision-making process and is readily understood by both the patient and the physician. In addition, its quantitative nature makes it ideally

 J. Clin. Invest. © *The American Society for Clinical Investigation, Inc.* · *0021-9738/80/05/1210/12 $1.00*
Volume 65 May 1980 1210–1221

TABLE I
*Pretest Likelihood for CAD**

Age	Group TA	Group AA	Group NA	Group AS
Males				
60–69	0.943±0.004	0.671±0.013	0.281±0.019	0.123±0.005
50–59	0.920±0.006	0.589±0.015	0.215±0.017	0.097±0.004
40–49	0.873±0.010	0.461±0.018	0.141±0.013	0.055±0.003
30–39	0.697±0.032	0.218±0.024	0.052±0.008	0.019±0.003
Females				
60–69	0.906±0.010	0.544±0.024	0.186±0.019	0.075±0.006
50–59	0.794±0.024	0.324±0.030	0.084±0.012	0.032±0.004
40–49	0.552±0.065	0.133±0.029	0.028±0.007	0.010±0.002
30–39	0.258±0.066	0.042±0.013	0.008±0.003	0.003±0.001

* Each value represents the a priori probability for coronary artery disease (CAD) ±1 standard error. TA, typical angina; AA, atypical angina; NA, nonanginal pain; AS, asymptomatic.

suited to analysis of cost-effectiveness, a major consideration of health care planners, providers, and patients alike (22). A critical limitation of the approach, however, is that it has not been prospectively validated.

The purpose of this study, therefore, was twofold. The first goal was to determine the validity of probability analysis as applied to noninvasive testing for coronary artery disease by comparing the calculated probability of disease with subsequent coronary angiographic results. The second was to develop and apply methods for analyzing both the cost-effectiveness and the diagnostic effectiveness of different test combinations used in the noninvasive diagnosis of coronary artery disease.

METHODS

Clinical classification. 43 consecutive patients tested in the Cedars-Sinai Medical Center Diagnostic Stress Laboratory who subsequently underwent diagnostic coronary angiography were studied. The patients were all referred by their primary care physician for noninvasive diagnostic testing. The decision for subsequent coronary angiography was made by this same physician, presumably on the basis of his interpretation of the test responses and clinical history. The study population ranged in age from 37 to 69 yr, and 81% of the group were males. Patients currently taking digitalis preparations and those with coronary artery disease previously established by a documented history of myocardial infarction were excluded from the study. All studies were performed in the fasting state and all cardiac medications were discontinued at least 24 h before study. Before testing, each patient was assigned to one of four clinical groups on the basis of three characteristics of their presenting symptoms as determined by the examining physician, an experienced cardiologist. These characteristics related to the substernal location, exertional precipitation, and prompt relief of chest pain. The four groups were defined as follows: (*a*) typical angina (TA), patients with substernal chest discomfort precipitated by exertion, and relieved by rest or nitroglycerin within 10 min (17 patients); (*b*) atypical angina (AA), patients in whom only two of these three factors were present (6 patients); (*c*) nonanginal pain (NA), patients in whom only one of the characteristics was present (13 patients); and (*d*) asymptomatic (AS), patients without discomfort (7 patients).

The age, sex, and symptom groups were then used to classify each patient before diagnostic testing into 1 of 32 subsets associated with a specific likelihood for significant angiographic coronary artery disease. These "pretest likelihoods," which are based upon a previously published review of 28,948 patients in the medical literature (21), are listed in Table I.

Testing protocol. A resting 12-lead electrocardiogram was obtained in the upright and supine positions and during voluntary hyperventilation using a computerized Case Marquette ECG recording system. Cardiokymography (CKG)[1] was performed in the supine position over the V_3 precordial line between the fourth and fifth intercostal space, using a 5-cm capacitive transducer (9) at voluntary end-expiration. The recording position was marked on the chest with a felt pen and the transducer left in place during exercise. An indirect carotid and phonocardiographic tracing were obtained simultaneously for timing. The patient then underwent a standardized Bruce graded exercise test. A 12-lead electrocardiogram was obtained at the end of each minute of exercise. All patients achieved at least 85% of their age-predicted maximum heart rate, at which time 1.5 mCi 201-thallium chloride was injected through a previously placed intravenous line. The patient continued to walk for an additional 1 min before termination of exercise. A repeat 12-lead electrocardiogram was obtained during each minute of recovery and a repeat V_3 CKG was obtained in the 2nd, 3rd, and 4th min of recovery. 5 min into recovery, the patient was placed under a scintigraphic camera equipped with a low energy, high resolution collimater, and imaging was performed in the postero anterior, 40° and 70° left anterior oblique projections (200,000 counts per image). The patient then underwent cardiac fluoroscopy in the 30° right anterior oblique and 60° left anterior oblique projections, using a 6-inch high resolution image intensifier (70–100 kV, 1–3 mA). 4 h after exercise, scintigraphic imaging was repeated in the same three projections.

[1] *Abbreviations used in this paper*: Č/Ī, cost-effectiveness; CKG, cardiokymography; ROC, receiver-operating characteristic.

Data reduction. The exercise-induced magnitude of horizontal or downsloping ST segment depression was measured relative to the PR segment, in millimeters (=0.1 mV), for each lead in both the exercise and recovery tracings. The maximum magnitude of ST segment displacement at 0.08 s after the J point was expressed in 0.5 mm increments and the result assigned to one of six mutually exclusive ranges: <0.5, 0.5–0.9, 1.0–1.4, 1.5–1.9, 2.0–2.4, ≥2.5 mm.

The postexercise CKG tracings were reviewed for the development of exercise-induced outward motion referenced to the period of systolic ejection, defined from the delay-corrected upstroke of the simultaneously recorded indirect carotid pulse tracing to the first high frequency component of the phonocardiographic second heart sound. There were three mutually exclusive CKG interpretations, based upon movement of the tracing during the systolic ejection period. Outward systolic motion was termed "holosystolic" if not preceded by inward motion during ejection, and was termed "midsystolic" if its magnitude was >50% of the preceding inward motion. Systolic motion was otherwise termed "normal" (9).

Cardiac fluoroscopy was analyzed subjectively in real time for the presence of calcification in the left main/left anterior descending, left circumflex, and right coronary artery distributions.

The unenhanced exercise and redistribution scintigraphic images were analyzed visually by an experienced reader. One of three interpretations was given. Regional hypoperfusion was considered "fixed" if it remained visably unchanged over the 4-h period of redistribution, and was termed "reversible" if visible improvement in relative perfusion was observed. Uniform distribution of radioactivity after exercise and redistribution was considered normal.

The diagnosis of coronary artery disease was defined as the presence of at least one major coronary vessel with >50% Diam narrowing by coronary angiography.

Each test was analyzed by different observers without reference to the historical data or the results of the other tests.

Data analysis: calculation of coronary artery disease likelihood. The true positive rate, $p(T_j|D+)$, and the false positive rate, $p(T_j|D-)$, for each of the 16 possible test observations (six for ECG, four for fluoroscopy, and three each for CKG and thallium) were determined by a previous review of the medical literature emcompassing 6,599 patients who have undergone one or more of the four tests and have been studied by coronary angiography (21). These data are summarized in Table II. The posttest likelihood, $p(D+|T_j)$, and its associated standard error (see Appendix) were then calculated for each individual patient according to Bayes' theorem of conditional probability (16, 21, 23) using the patient's pretest likelihood, p(D+), obtained from Table I and the true and false positive rates of his specific test responses, from Table II:

$$p(D+|T_j) = p(D+)p(T_j|D+)/p(T_j),$$

where $p(T_j) = p(D+)\,p(T_j|D+) + p(D-)\,p(T_j|D-)$ and $p(D-) = 1 - p(D+)$. This equation was applied serially to each combination of the four test results. According to this assumption, the posttest likelihood for one (n) test is equivalent to the pretest likelihood for the next (n + 1) test (21):

$$[p(D+|T_j)]_n = [p(D+)]_{n+1}.$$

These posttest likelihood values, thus calculated, were then compared with the results of subsequent coronary angiography.

Correlation of likelihood and clinical judgment. The correlation between the statistical estimates of posttest likelihood calculated by Bayes' theorem and those by an experienced clinician's judgmental interpretation of the same test results was also determined. 30 cases exhibiting a uniform mix of both concordant and discordant test responses were selected and the data (consisting of age, sex, symptom class, and the results of testing) were presented to five experienced clinical cardiologists who had no knowledge of the patient and no formal experience with conditional probability analysis. Each was asked to state his opinion as to the probability for significant angiographic coronary artery disease as a percentage ranging from 0 to 100%. The 150 clinical estimates obtained (five physicians times 30 cases) were correlated directly with the 30 calculated according to Bayes' theorem.

TABLE II

True and False Positive Rates for Test Observations

Observation j	True positive rate $p(T_j\|D+)$*	False positive rate $p(T_j\|D-)$‡
ST Depression (mV)		
$0.00 \le j < 0.05$	0.143±0.033	0.625±0.057
$0.05 \le j < 0.10$	0.208±0.034	0.227±0.058
$0.10 \le j < 0.15$	0.233±0.025	0.110±0.014
$0.15 \le j < 0.20$	0.088±0.029	0.021±0.013
$0.20 \le j < 0.025$	0.133±0.023	0.012±0.008
$0.25 \le j < \infty$	0.195±0.016	0.005±0.005
Coronary calcification (vessels)		
0	0.420±0.025	0.961±0.017
1	0.235±0.022	0.023±0.013
2	0.201±0.021	0.016±0.011
3	0.145±0.018	0.000±0.000
Systolic outward motion		
None	0.257±0.052	0.942±0.034
Midsystolic	0.443±0.059	0.038±0.027
Holosystolic	0.300±0.055	0.019±0.019
Regional hypoperfusion		
None	0.147±0.017	0.836±0.031
Fixed	0.146±0.026	0.103±0.034
Reversible	0.707±0.019	0.061±0.013

* $p(T_j|D+)$ is the conditional probability for the test (T) observation (j) given coronary artery disease.

‡ $p(T_j|D-)$ is the conditional probability for the same test observation given no coronary artery disease.

Comparison of the diagnostic and cost-effectiveness of different test combinations. Because four tests were performed in each patient, there were 15 potentially different testing formats that could be evaluated: four individual tests, six two-test combinations, four three-test combinations, and one four-test combination. Each combination was analyzed as to "diagnostic effectiveness" and "cost-effectiveness."

Diagnostic effectiveness. Two independent methods were applied to assessment of diagnostic effectiveness. For each test combination, the mean change in calculated likelihood that occurred as a consequence of testing was determined, for each n of the N patients, as the arithmetic difference in likelihood before and after testing:

$$\Delta p_n = \left| p_n(D+|T_j) - p_n(D+) \right|.$$

This value was then multiplied by a sign constant (K_n) de-

pending on the results of coronary angiography. The sign constant was +1 if posttest likelihood was greater than pretest likelihood in the presence of angiographic disease or if posttest likelihood was less than pretest likelihood in the absence of disease. The sign constant was −1 when either of the opposite results were observed. The average of these 43 individual values of Δp for each of the 15 test combinations was termed the "mean correct change in likelihood," E_0:

$$E_0 = \frac{1}{N} \sum_n K_n \cdot \Delta p_n.$$

The mean correct change in likelihood for coronary angiography (E_c) was expressed similarly where $p(D+|T_j) = 1$ for the presence of disease and $p(D+|T_j) = 0$ in the absence of disease. The E_0 for each testing combination, was then expressed as a fraction of E_c, and termed "observed effectiveness:"

$$\bar{E}_0 = E_0/E_c.$$

The second method for assessing diagnostic test effectiveness was by determination of the information content (I) for each test combination from the literature pooled data in Table II. As defined by Shannon (24), information content provides a quantitative means of expressing the change in one's uncertainty concerning the probability of an event in respect to its a priori (pretest) likelihood and its a posteriori (posttest) likelihood (25, 26). This relationship may be stated in terms of the true and false rates of the tests used to assess the probability of the event. Thereby, this theorem expresses the reduction in uncertainty (i.e., addition of information) as a consequence of performing a test:

$$I = \sum_j p(T_j) \sum_i p(D_i|T_j) \log_2 p(D_i|T_j) - \sum_i p(D_i) \log_2 p(D_i).$$

The subscript i represents each of all possible disease states (in this case, D+ and D−). The derivation of this equation and its associated variance is outlined in the Appendix. A more complete discussion of information theory and formal mathematical proof of its theorems are available from several sources (24, 27).

Since information content is dependent upon prevalence, its integrated average ($\bar{I}$) relative to coronary angiography was used for comparisons between test combinations (see Apdendix). Because $\bar{I}$ was determined from the literature, and not the study data, it was termed "predicted effectiveness."

Cost-effectiveness. The cost ($\bar{C}$) of each test combination used in this study was determined as the prevailing monetary cost for the test divided by the prevailing monetary cost of coronary angiography (laboratory fee plus physician fee plus 2 d semi-private hospitalization). The ratio of cost to predicted effectiveness, $\bar{C}/\bar{I}$, was termed "cost-effectiveness." This definition of cost is distinguished from that used for "utility" analysis which considers additional nonmonetary factors such as the psychological "cost" of a therapeutic complication.

Statistical analysis. Statistical variances were calculated by methods described in the Appendix and differences in variance among the test combinations were assessed using Bartlett's test for homogeneity. Pooled variances were calculated by weighting the individual variances for differences in degrees of freedom. Two-sided 95% confidence intervals were determined from the t distribution ($\alpha = 0.025$).

RESULTS

There were 12 patients without significant angiographic coronary artery disease, 4 of whom had angina (group TA or AA), and 31 patients with angiographic coronary artery disease, 12 of whom did not have angina (group NA or AS). In only 10 of the 43 patients (23%) were the pretest symptom class and all four subsequent test results concordant.

Fig. 1 illustrates the average pretest likelihood (based on age, sex, and symptom class) compared with the average posttest likelihood (based on all four test procedures) for each of four groups defined by the number of vessels with >50% Diam narrowing at angiography. There was a significant relationship between the extent of angiographic disease, so defined, and the calculated likelihood of disease based on historical and noninvasive test data. In the 12 patients without disease, the average likelihood fell from a pretest level of 24.7±8.2% to 7.0±2.6% ($P < 0.01$, paired t test), even though 13 of the 60 (22%) historical and test results were falsely "positive." In contrast, likelihood increased from 56.1±6.5% to 94.1±2.8% ($P < 0.01$, paired t test) in the 31 coronary artery disease patients, even though 45 of the 155 (29%) historical and test results were falsely "negative."

Relationship between likelihood and angiographic prevalence. The 15 different combinations of the four tests provided a total of 645 estimates of posttest likelihood, 180 in the 12 patients without disease and 465 in the 31 patients with disease. The frequency distribution for these 645 estimates, grouped into deciles of likelihood, is illustrated in Fig. 2. These posttest likelihood estimates demonstrated a bimodal distribution ($P < 0.00001$, chi-square test). Of the 180 likelihood estimates in the nondiseased population,

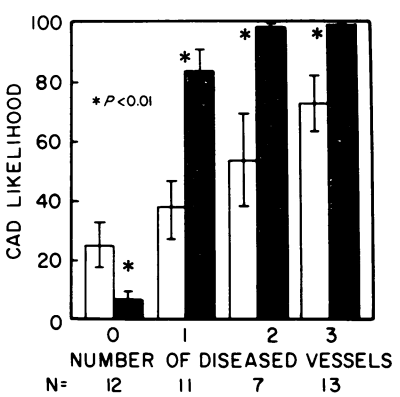

FIGURE 1 Mean pre- (□) and post- (▨) test likelihood for coronary artery disease (±1 SE) according to the number of coronary vessels with 50% Diam narrowing. A significant change in likelihood after noninvasive testing was observed in each group. N is the number of patients in each group. See text for further discussion. CAD, coronary artery disease.

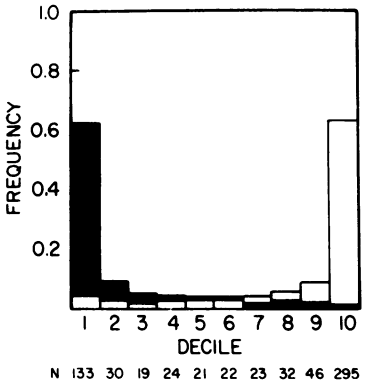

FIGURE 2 Frequency distribution of 645 estimates of posttest likelihood grouped into deciles. The majority of the 180 estimates in normal patients (■) were in the lowest decile and of the 465 estimates in coronary artery disease patients (□), in the highest decile. N is the number of estimates in each decile.

113 (63%) were in the first decile, from 0 to 9.9%, and only 2 (1%) in the tenth decile, from 90 to 100%. In contrast, of the 465 estimates in the diseased population, 293 (63%) were in the tenth decile, and only 20 (4%) were in the first decile. Thus, 99% (293 of 295) of the tenth decile values were associated with disease, whereas 85% (113 of 133) of the first decile values were associated with absence of disease.

The relationship between each decile of likelihood determined by noninvasive analysis and the subsequent prevalence of disease at coronary angiography, normalized for differences in size between the two groups, is illustrated in Fig. 3. A highly significant ($P < 0.001$, chi-square minimization and linear regression) relationship between calculated posttest likelihood and angiographic disease frequency was observed, ranging from a frequency of angiographic disease of 6.4±1.4% for patients in the first decile of likelihood to 98.9±0.9% for those in the tenth decile.

Information content of diagnostic testing. Although a probability statement has certain advantages as a method for expressing test results, clinical decisions based upon the test results are often binary rather than probabilistic (e.g., Should angiography now be performed? Yes or no—not 65% yes vs. 35% no.). The receiver-operating characteristic (ROC) curve (25, 28, 29) allows analysis of the magnitude of error associated with a binary judgment about the presence of disease as a function of the level of posttest likelihood. The ROC curve for this data base is illustrated in Fig. 4. The curve was constructed by considering that a given level of posttest likelihood indicates the presence of disease, and then determining the true positive and false positive rates at each likelihood level. This was done for 13 different levels of likelihood from >1 to >99%. As the likelihood value used for a positive diagnosis of coronary artery disease increased, both the true and false positive rates fell in a curvilinear fashion typical of all such curves.

This relationship is useful in defining the error levels associated with categorical interpretation of the cumulated test results, but cannot be used to assess the relative effectiveness of the 15 different test combinations nor to establish the operating criterion containing the maximum diagnostic information. Informational analysis was employed for this purpose. Fig. 5 illustrates the correlation between predicted information content of the 15 test combinations and the observed effectiveness relative to coronary angiography. The correlation between these variables was highly linear about the line of identity, indicating that information content is a valid quantitative index of test effectiveness, semantically equivalent to the average correct change in the probability of disease. Informational analysis of the ROC curve for our data revealed that if a single value for posttest likelihood were to be used to separate

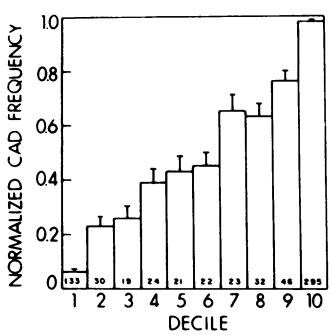

FIGURE 3 Normalized frequency of coronary artery disease (CAD) derived from the data in Fig. 2. A significant linear relationship between decile of posttest likelihood and coronary artery disease frequency (±1 SE) was observed. The number in each bar represents the number of likelihood estimates in each decile. See text for further discussion.

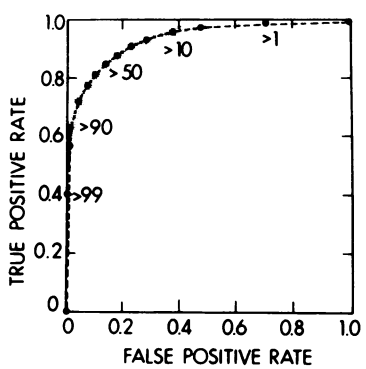

FIGURE 4 The ROC curve (– – –) derived from the data in Fig. 2 for the following threshold levels of posttest likelihood (in percent): 1, 5, 10, 20, . . . , 80, 90, 95, and 99. The strippled area represents the 95% confidence region calculated from the unpooled variances. See text for further discussion.

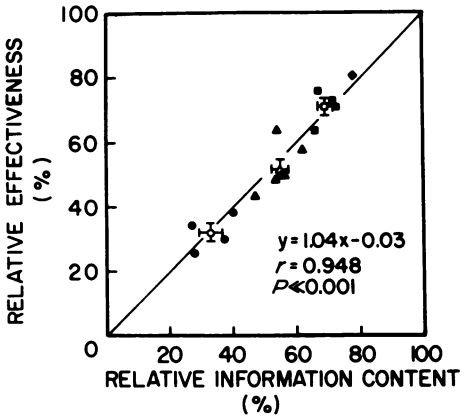

FIGURE 5 Correlation between observed effectiveness of the 15 different test combinations (y axis) and information content (x axis), both expressed as a percentage relative to coronary angiography. ● represent the single tests; ▲, the two-test combinations; ■, the three-test combinations; and ◆, the sole four-test combination. Open symbols represent the mean of each group of solid symbols (±1 SE). The diagonal represents the line of identity. See text for discussion.

patients into the two diagnostic categories of disease (D+) and no disease (D−), the 60% likelihood value would produce optimal separation. At the 60% level, where categorical information content was maximal, the true positive rate was 0.816±0.018 and the false positive rate was 0.094±0.022.

The ROC curve thereby defined the error levels associated with use of the single best likelihood value as the criterion for categorical diagnosis but such use does not discriminate between widely different values that both satisfy the threshold, 61 vs. 99% for example. Such a single threshold, therefore, is of little help in determining a rational endpoint to multiple testing. In this regard, Fig. 6 illustrates the distribution of information content for all 15 test combinations as a function of the achieved decile of posttest likelihood. Fully 95% of the total information content of testing was contained in the first and last deciles of likelihood. Thus, 441 (68%) of the 645 likelihood estimates were located in the region of maximal information content. Of these 441 likelihood estimates, 421 correctly predicted the presence or absence of angiographic coronary artery disease (predictive accuracy, 95.5±1.0%). These results indicate that, from an informational standpoint, when the level of posttest likelihood is <10% or >90%, an accurate endpoint criterion for diagnostic testing has been attained.

Cost vs. effectiveness of diagnostic testing. Although the cost and effectiveness of testing increased substantially as a function of the number of tests performed, the ratio of cost to information content (an index of cost-effectiveness) did not (Table III). Two distinctly different correlations between cost and information content were identified. As shown in Fig. 7, there was a group of eight test combinations with an average cost-effectiveness that was twice that of the other seven combinations (0.47±0.02 vs. 0.24±0.02, $P > 0.001$ unpaired t test). The single characteristic common to all of the less cost-effective test combinations was the inclusion of thallium scintigraphy, the most effective and the most costly of the individual test procedures.

Fig. 8 illustrates the relative effectiveness and cost-effectiveness of each of the 15 different test combinations compared with diagnostic coronary angiography. The figure shows that all 15 test combinations were cost-effective compared with diagnostic coronary angiography ($\bar{C}/\bar{I}$ ratios <1). As a generalization, although

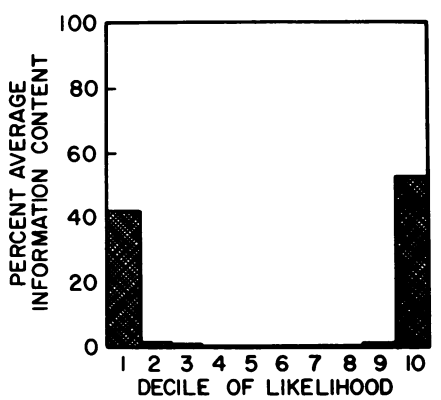

FIGURE 6 The distribution of information content for each decile of posttest likelihood obtained from the ROC data in Fig. 4. The majority of information was obtained when the posttest likelihood was under 10% (lst decile) or over 90% (10 decile). See text for discussion.

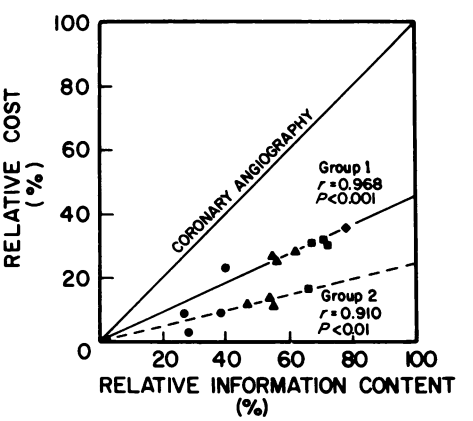

FIGURE 7 Relationship between cost and information content for the 15 test combinations relative to coronary angiography. The symbols are identical to those in Fig. 5. The diagonal represents maximum possible values relative to coronary angiography (relative cost = 100%, relative information content = 100%). See text for discussion.

TABLE III
Indices of Cost and Effectiveness of Testing

Test combination	Observed effectiveness	Predicted effectiveness	Relative cost	Relative cost-effectiveness
E	0.335±0.072	0.263±0.117	0.087	0.331±0.147
F	0.252±0.066	0.279±0.054	0.033	0.118±0.023
C	0.299±0.070	0.383±0.115	0.090	0.235±0.071
T	0.379±0.074	0.404±0.060	0.227	0.561±0.084
E:F	0.436±0.076	0.465±0.107	0.120	0.258±0.059
E:C	0.491±0.076	0.538±0.129	0.137	0.255±0.061
E:T	0.491±0.076	0.552±0.090	0.273	0.495±0.081
F:C	0.470±0.076	0.552±0.107	0.123	0.223±0.043
F:T	0.634±0.073	0.564±0.065	0.260	0.461±0.053
C:T	0.574±0.075	0.619±0.090	0.277	0.447±0.065
E:F:C	0.629±0.074	0.661±0.107	0.170	0.257±0.042
E:F:T	0.745±0.066	0.670±0.078	0.307	0.458±0.053
E:C:T	0.722±0.068	0.711±0.089	0.323	0.454±0.057
F:C:T	0.701±0.070	0.719±0.077	0.310	0.431±0.046
E:F:C:T	0.834±0.057	0.784±0.072	0.357	0.456±0.042

Each value represents the decimal equivalent of percent relative to coronary angiography±1 standard error. Relative costs are based on absolute figures and have no associated variance. E, electrocardiography; F, fluoroscopy; C, cardiokymography; T, thallium scintigraphy.

effectiveness increased as the number of tests employed increased, test combinations including fluoroscopy tended to be more cost-effective, whereas those including thallium scintigraphy were less cost-effective.

Correlation with clinical judgment. The 150 clinical estimates of post-test likelihood ranged from 1 to 100% (mean, 56.0±34.1% SD), whereas the 30 calculated estimates ranged from 0.1 to 99.9% (mean, 55.4±36.5%). The linear regression correlation coefficient between clinical and calculated posttest likelihood varied from $r = 0.820$ to $r = 0.949$ for the five cardiologists. Although a highly significant correlation between clinical judgment and calculated posttest likelihood was observed, an extraordinarily large amount of interobserver variability was noted in the range of 30–70% of calculated likelihood (Fig. 9). Accordingly, the average

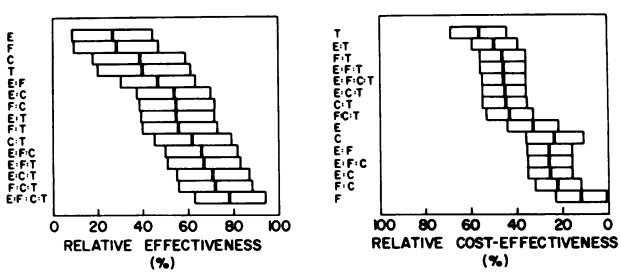

FIGURE 8 Relative effectiveness and cost-effectiveness for the 15 test combinations used. Each symbol represents the mean ±95% confidence interval. The variances were not distributed homogeneously ($P < 0.0005$), the difference between pooled and unpooled estimates averaging 63%. This difference produced <1% of artefactual narrowing of the confidence intervals calculated from the pooled data, and resulted in no change in the ranking of the tests relative to the lower confidence limit. These confidence intervals provide a very conservative decision threshold (in the range of $P < 0.005$) for comparison of test combinations, when the intervals do not overlap. This conservatism offsets, in part, the potential dependence between combinations that share a common test procedure, but simultaneously produces a larger β error that might mask existing significant differences. E, electrocardiography; F, fluoroscopy; C, cardiokymography; T, thallium scintigraphy.

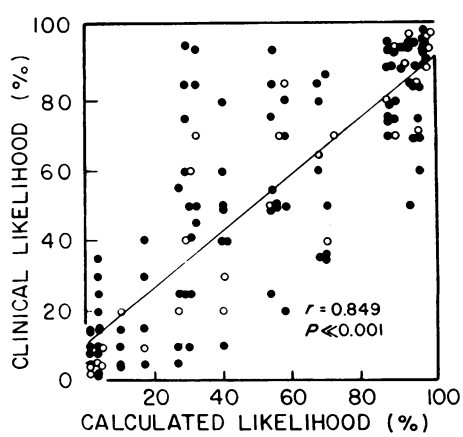

FIGURE 9 Relationship of the clinical estimate of likelihood for coronary heart disease, made by five cardiologists and the calculated posttest likelihood. There is wide variability especially in the 30–70% range of calculated likelihood. The overall correlation, however, is highly significant. Standard error of the estimate, 17.1.

estimate of the five cardiologists was substantially superior to that of any one individual (Fig. 10). In only 3 of the 30 average estimates was there wide disparity between clinical judgment and calculated likelihood. Each of these cases represented abnormal test responses in patients with low pretest likelihoods.

DISCUSSION

The goal of earlier and more accurate detection of coronary artery disease has stimulated the development of numerous noninvasive diagnostic techniques. Each of these techniques, however, is substantially limited in diagnostic accuracy: numerous studies have now demonstrated, both in theory and in practice, that a single positive test in a low prevalence population is of limited diagnostic value (13–21). Chaitman et al. (30), for example, recently demonstrated significant differences in the predictive accuracy of ECG stress testing in patients subgrouped according to symptoms before testing. Because the predictive accuracy of the ECG test alone was poor in those with minimal or equivocal symptoms, these authors advocated the use of additional test procedures such as cardiac fluoroscopy, CKG, and thallium scintigraphy in certain subgroups of patients. Others have made similar observations and recommendations (2, 6, 9, 31–33).

The relevance of probability analysis. In attempting to implement such recommendations, two problems immediately emerge. First, the accuracy of multiple testing is limited, especially when test responses are discordant, by the current custom of reporting the results categorically as "positive" or "negative" (21, 24). In our study, 46 (27%) of the 172 test responses were discordant in reference to coronary angiography and 30 of the 43 patients (70%) had at least one such discordant response. Our results indicate that in such circumstances, even cardiologists specializing in the field of stress testing, and with considerable knowledge of the relevant literature, experience difficulty in estimating the diagnostic probability of disease. When such estimates are inaccurate, subsequent clinical decisions may be inappropriately founded.

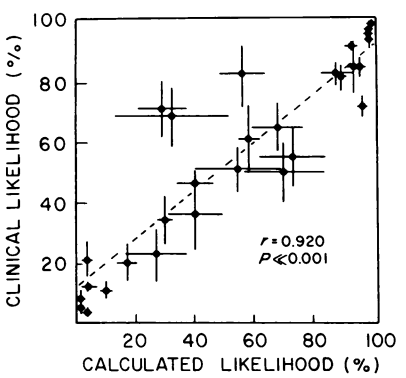

FIGURE 10 Comparison of the consensus of five clinical estimates to calculated likelihood. Each symbol represents the average (±1 SE). Although there was wide variation in individual predictions, the mean judgment was close to the calculated likelihood, and more precise than any one group of individual estimates. Standard error of the estimate, 4.7.

The second problem created by the emergence of multiple testing relates to cost-effectiveness. This point has been emphasized by Adelstein and McNeil (34) in reference to the diagnosis of pulmonary embolus, wherein they state that such an approach ". . . will surely increase the costs . . . regardless of how much it may improve accuracy." Although the goal of improving diagnostic effectiveness and simultaneously reducing cost has emerged as a major issue in medical practice, the lack of objective methods for cost-effectiveness comparisons has become increasingly apparent (22).

One method by which such a large amount of potentially conflicting information may be processed is through probability analysis. A number of investigators have advocated this approach as an aid to analysis of diagnostic test results on theoretical grounds, and its practicality relative to electrocardiographic stress testing has recently been demonstrated by Cohn et al. (35). Our study extends the probabilistic approach to the analysis of multiple procedures in diagnostic stress testing.

Advantages of probability analysis. The primary conclusion of this study is that the use of probability analysis contributes significantly toward solution of both problems associated with diagnostic testing, as alluded to above. First, these probability estimates correlated well with both angiographic prevalence and with experienced clinical judgment, despite frequently discordant test responses.

Second, probability analysis allows meaningful quantitation of the effectiveness of a given test or test combination relative to a defined standard such as coronary angiography by use of well established concepts embodied in the discipline of information theory.

The usefulness of this application may be best appreciated by an example. The effectiveness of a test is customarily expressed in relation to its "sensitivity" (true positive rate), its "specificity" (1 − false positive rate), and the "likelihood ratio" (true positive rate/false positive rate) (20, 28). When two tests differ only in one of the primary variables, then the test with the higher associated value is considered the more effective. Often, however, tests differ in ways more complex than this. If one test has a sensitivity of 51 and a specificity of 99%, for example, it is not apparent whether or not it is diagnostically superior to a second test with a sensitivity of 70 and a specificity of 90%. With multiple criteria for interpretation, such comparisons become even more difficult. Information theory greatly simplifies this analysis, and allows such comparative determinations

of diagnostic effectiveness. In the example above, although the likelihood ratio is 51 for the first test and only 7 for the second, their average information content is identical.

Limitations of probability analysis. Probability analysis has several potential and real limitations. The first relates to the assessment of likelihood before testing, and the use of specific true and false positive rates. In this study, factors of known importance (age, sex, and symptoms) were weighted according to published experience in the medical literature. Each of these estimates must be presumed to contain an element of unknown bias. We believe, however, that this limitation was minimized in the study design by use of estimates derived from the pooled experience of many investigators (36).

A second potential limitation to use of probability analysis is that there may be conditions other than coronary artery disease that cause more than one test result to be positive. This phenomenon of "dependence" should not materially influence serial application of Bayes' theorem, however, as long as the number of procedures employed is not large (37). Moreover, an earlier study from our laboratory supports the hypothesis that the procedures employed in this study are statistically independent.[2] Such a finding is not unexpected because the ST segment response is fundamentally electrophysiologic; coronary artery calcification, metabolic; cardiokymographic abnormalities, mechanical; and scintigraphic hypoperfusion, hydrodynamic.

A third limitation to the use of probability analysis involves the potential unawareness on the part of the physician that the calculated likelihood does not include consideration of other factors that may be of major significance in some cases. Awareness of what the analysis does and does not provide is, therefore, essential to the user. In this regard, probability estimates could be further improved by application of clinical judgement to those areas not encompassed by the formal analysis such as family history, exercise-induced hypotension, and depth of regional hypoperfusion.

Probability analysis and cost-effective decision making. The data in this study provide clinically relevant observations relative to effective and cost-effective decision making. For example, the widely used test combination of stress electrocardiography and thallium scintigraphy provided 27 concordant results in the 43 patients studied, all of which were correct relative to coronary angiography. Had the 43 patients instead been tested by stress electrocardiography and fluoroscopy, 26 would have attained a posttest likelihood of either <10 or >90%. Using these levels of likelihood as "establishing" the presence or absence of disease, all 26 patients would have been correctly identified relative to coronary angiography. If the remaining 17 patients with indeterminate posttest likelihoods had undergone CKG and thallium testing, 8 additional patients would have reached either the 10 or 90% likelihood level, all of whom would also have been correctly classified. This latter format, therefore, would have provided a 16% increase in total number of correct diagnoses at a simultaneous 16% reduction in the total cost of testing, compared with the electrocardiogram (ECG) scintigraphy combination.

Clearly, this approach represents a simplification of a more complex decision matrix because the physician often desires information beyond that related solely to the presence or absence of disease. The determination of functional exercise capacity, or the localization of disease to a specific anatomic region by fluoroscopy or perfusion scintigraphy, might be the primary reason for obtaining the test in certain individuals. Nevertheless, we believe it is reasonable to initiate such an analysis on the assumption that at least one goal of diagnostic testing is diagnosis itself. Although it is evident that the data base in this study is too limited to support the recommendation of any particular test combination at present, we believe that the concepts and analytic methods employed herein provide a basis for the future development of cost-effective formats for noninvasive testing, a growing concern of all involved in the health care system (22, 38, 39).

In summary, this study suggests that analysis of conditional probability provides an accurate, explicit means for the interpretation of multiple, even discordant, test responses relative to the diagnosis of angiographic coronary artery disease. The format is applicable to a wide spectrum of patients ranging from the asymptomatic to those with typical angina pectoris. The results of such analysis, which can be obtained in seconds using any of many available "personal computers" by technicians in the test facility, can be used by the physician as an aid to diagnosis and medical decision making. Because of the quantitative nature of the format, it may also aid in developing more cost-effective means for the diagnosis of coronary artery disease and may be of use to investigators conducting clinical trials in subsetting populations in whom coronary angiography cannot be used to establish a diagnosis with certainty.

APPENDIX

Calculation of information content. Information theory defines the average uncertainty for a set, i of mutually exclusive probabilistic events (such as the presence [D+] and absence [D−] of coronary disease) in terms of the probability associated with each event. If the events are represented

[2] Charuzi, Y., G. A. Diamond, M. Pichler, A. Waxman, R. Vas, R. Silverberg, D. Berman, and J. S. Forrester. Analysis of multiple noninvasive procedures for the diagnosis of coronary artery disease. Manuscript submitted for publication.

as D_i and the prior probability or frequency of occurrence by $p(D_i)$, then a priori uncertainty, expressed in binary digits or bits, is given by:

$$U\ pre = -\sum_i p(D_i) \log_2 p(D_i). \quad (1)$$

Similarly, the a posteriori uncertainty (resulting from performance of a test) is defined in relation to the set of posterior probabilities, $p(D_i/T_j)$, as the average uncertainty associated with all mutually exclusive test responses (T_j), weighted by their respective frequencies of occurrence, $p(T_j)$:

$$U\ post = -\sum_j p(T_j) \sum_i p(D_i|T_j) \log_2 p(D_i|T_j). \quad (2)$$

The information content of the test is defined as the average reduction in uncertainty, obtained by subtracting Eq. 2 from Eq. 1:

$$I = \sum_j p(T_j) \sum_i p(D_i|T_j) \log_2 p(D_i|T_j) - \sum_i p(D_i) \log_2 p(D_i). \quad (3)$$

The conditional probabilities in Eq. 3 are interrelated through Bayes' theorem:

$$p(D_i|T_j) = p(D_i)p(T_j|D_i)/p(T_j); \quad (4)$$

$$p(T_j) = p(D+)p(T_j|D+) + p(D-)p(T_j|D-). \quad (5)$$

The term $p(T_j|D+)$ is the conditional probability for the test observation j in a population of diseased patients (true positive rate), and $p(T_j|D-)$ is the conditional probability for j in the complementary population, D−, of nondiseased patients (false positive rate). By substitution of Eq. 4 and 5 into Eq. 3, therefore, one obtains the information content, I, expressed as a function of the three variables $p(D+)$, $p(T_j|D+)$, and $p(T_j|D-)$.

The average information content of the test is represented by the area under the information curve, given by integration of Eq. 3 with respect to $p(D+)$ over the inverval 0 to 1 (for simplicity of notation we will substitute P for p[D+]):

$$\bar{I} = \int_0^1 I dP. \quad (6)$$

If a test were perfect, then $U_{post} = 0$ and $I = U_{pre}$. "Maximum average information content" is thereby defined, according to Eq. 1 and Eq. 6:

$$\bar{I}_{max} = -\int_0^1 \sum_i p(D_i) \log_2 p(D_i) dP = \frac{2}{4 \ln 2} = 0.721. \quad (7)$$

"Relative information content" is then expressed as the ratio of Eq. 6 to Eq. 7:

$$\bar{I}/\bar{I}_{max} = 2 \ln 2 \int_0^1 I dP. \quad (8)$$

Multiplying Eq. 8 by 1/ln 2 converts all terms of this equation into natural logarithms:

$$\bar{I}/\bar{I}_{max} = 2 \int_0^1 \left(\sum_j p(T_j) \sum_i p(D_i|T_j) \ln p(D_i|T_j) - \sum_i p(D_i) \ln p(D_i)\right) dP. \quad (9)$$

Calculation of variances. The variances of the three probabilities $p(D+)$, $p(T_j|D+)$, and $p(T_j|D-)$ were determined from the equation:

$$\sigma^2 = \frac{pq}{N} = \frac{n(N-n)}{N^3}, \quad (10)$$

where $p = n/N$; $q = 1 - p = (N - n)/N$; N = total population; n = number of individuals in N with the characteristic, j. When these independent and uncorrelated errors are introduced into Bayes' equation for calculation of $p(D+|T_j)$, the variance (σ_L^2) is given by the partial differential equation:

$$\sigma_L^2 = \left(\frac{\partial L}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial L}{\partial B}\right)^2 \sigma_B^2 + \left(\frac{\partial L}{\partial P}\right)^2 \sigma_P^2; \quad (11)$$

where $L = p(D+|T_j)$; $A = p(T_j|D+)$; $B = p(T_j|D-)$; and $P = p(D+)$. The partials may be evaluated by substitution of Eq. 5 into Eq. 4 and expanding relative to $p(D_i)$:

$$\left(\frac{\partial L}{\partial A}\right) = \frac{\prod_i p(D_i|T_j)}{p(T_j|D+)}; \quad \left(\frac{\partial L}{\partial A}\right) = \frac{\prod_i p(D_i|T_j)}{p(T_j|D-)};$$

$$\left(\frac{\partial L}{\partial P}\right) = \frac{\prod_i p(D_i|T_j)}{\prod_i p(D_i)};$$

$$\sigma_L^2 = \left(\prod_i p(D_i|T_j)\right)^2 \left[\left(\frac{\sigma_A}{p(T_j|D+)}\right)^2 + \left(\frac{\sigma_B}{p(T_j|D-)}\right)^2 + \left(\frac{\sigma_P}{\prod_i p(D_i)}\right)^2\right]. \quad (12)$$

The variance associated with the calculated information content (σ_I^2) is given by the similar partial differential equation:

$$\sigma_I^2 = \left(\frac{\partial I}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial I}{\partial B}\right)^2 \sigma_B^2 + \left(\frac{\partial I}{\partial P}\right)^2 \sigma_P^2. \quad (13)$$

We may evaluate these partials by substituting Eq. 4 into Eq. 3 and expanding relative to $p(D_i)$:

$$\left(\frac{\partial I}{\partial A}\right) = \sum_j p(D+) \log_2 p(D+|T_j);$$

$$\left(\frac{\partial I}{\partial B}\right) = \sum_j p(D-) \log_2 p(D-|T_j);$$

$$\left(\frac{\partial I}{\partial P}\right) = \log_2 [p(D-)/p(D+)] + \sum_j p(T_j|D+) \log_2 p(D+|T_j) - \sum_j p(T_j|D-) \log_2 p(D-|T_j).$$

Similarly, the variance of $\bar{I}$ is given as the integrated average variance of I over the entire range of P:

$$\sigma_{\bar{I}}^2 = \int_0^1 \sigma_I^2 dP. \quad (14)$$

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Ms. Tina Orvis and Ms. Terry Collett for secretarial assistance, Mr. Lance Laforteza for graphics, and Ms. Patricia Edwards for photography.

This work was funded in part by National Institutes of Health Specialized Center of Research grant 17651.

REFERENCES

1. Bartel, A. G., V. S. Behar, R. H. Peter, E. S. Orgain, and Y. Kong. 1974. Graded exercise stress tests in angiographically documented coronary artery disease. *Circulation.* **49:** 348–356.
2. Detry, J. R., B. M. Kapita, J. Cosyn, B. Sottiaux, L. A. Brasseur, and M. F. Rousseau. 1977. Diagnostic value of history and maximal exercise electrocardiography in men and women suspected of coronary artery disease. *Circulation.* **56:** 756–769.
3. Goldschlager, N., A. Selzer, and K. Cohn. 1976. Treadmill stress tests as indicator of presence and severity of coronary artery disease. *Ann. Intern. Med.* **85:** 277–286.
4. Bartel, A. G., J. T. Chen, R. H. Peter, V. S. Behar, Y. Kong, and R. G. Lester. 1974. The significance of coronary calcification detected by fluoroscopy: a report of 360 patients. *Circulation.* **49:** 1247–1253.
5. Hamby, R. I., F. Tabrah, B. G. Wisoff, and M. L. Harstein. 1974. Coronary artery clacification: clinical implications and angiographic correlates. *Am. Heart J.* **87:** 565–570.
6. Aldrich, R. F., J. F. Brensike, J. W. Battaglini, J. M. Richardson, I. K. Loh, N. J. Stone, E. R. Passamani, H. Ackerstein, R. Seninger, J. S. Borer, R. I. Levy, and S. E. Epstein. 1979. Coronary calcifications in the detection of coronary artery disease. *Circulation.* **59:** 1113–1124.
7. Vas, R., G. A. Diamond, H. L. Wyatt, P. L. da Luz, H. J. C. Swan and J. S. Forrester. 1977. Noninvasive analysis of regional myocardial wall motion: cardiokymography. *Am. J. Physiol.* **233**(6): H700–H706.
8. Diamond, G. A., M. Chag, R. Vas, and J. S. Forrester. 1978. Cardiokymography: quantitative analysis of regional ischemic left ventricular dysfunction. *Am. J. Cardiol.* **41:** 1249–1257.
9. Silverberg, R. A., G. A. Diamond, R. Vas, D. Tzivoni, H. J. C. Swan, and J. S. Forrester. 1980. Noninvasive diagnosis of coronary artery disease: the cardiokymographic stress test. *Circulation.* **61:** 579–589.
10. Bailey, I. K., L. S. C. Griffith, J. Rouleau, H. W. Strauss, and B. Pitt. 1977. Thallium-201 myocardial perfusion imaging at rest and exercise: comparative sensitivity to electrocardiography in coronary artery disease. *Circulation.* **55:** 79–88.
11. Lenears, A., P. Block, P. E. Thiel, M. Lebedelle, P. Becquevort, F. Erbsmann, and A. M. Ermans. 1977. Segmental analysis of thallium-201 stress myocardial scintigraphy. *J. Nucl. Med.* **18:** 509–516.
12. Turner, D. A., W. E. Battle, H. Deshmukh, M. A. Colandrea, G. J. Snyder, E. W. Fordham, and J. V. Messer. 1978. The predictive value of myocardial perfusion scintigraphy after stress in patients without previous myocardial infarction. *J. Nucl. Med.* **19:** 249–255.
13. Froelicher, V. F., F. G. Yanowitz, and A. J. Thompson. 1973. The correlation of coronary angiography and the electrocardiographic response of maximal treadmill testing in 76 asymptomatic men. *Circulation.* **48:** 597–605.
14. Borer, J. S., J. F. Brensike, D. R. Redwood, S. B. Itschoitz, E. R. Passamani, N. J. Stone, J. N. Richardson, R. I. Levy, and S. E. Epstein. 1975. Limitations of the electrocardiographic response to exercise in predicting coronary artery disease. *N. Engl. J. Med.* **293:** 367–371.
15. Erikssen, J., I. Enge, K. Forfang, and O. Storstein. 1976. False-positive diagnostic tests and coronary angiographic findings in 105 presumably healthy males. *Circulation.* **54:** 371–376.
16. Galen, R. S., and S. R. Gambino. 1976. Beyond Normality: The Predictive Value and Efficiency of Medical Diagnosis. John Wiley & Sons, Inc. New York. 237.
17. Redwood, D. R., J. S. Borer, and S. E. Epstein. 1976. Whither the ST segment during exercise? *Circulation.* **54:** 703–706.
18. Epstein, S. E. 1978. Value and limitation of the electrocardiographic response to exercise in the assessment of patients with coronary artery disease. *Am. J. Cardiol.* **42:** 667–674.
19. Gorry, G. A., S. G. Pauker, and W. B. Schwartz. 1978. The diagnostic importance of the normal finding. *N. Engl. J. Med.* **298:** 486–489.
20. Rifkin, R. D., and W. B. Hood. 1977. Bayesian analysis of electrocardiographic exercise stress testing. *N. Engl. J. Med.* **297:** 681–686.
21. Diamond, G. A., and J. S. Forrester. 1979. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. *N. Engl. J. Med.* **300:** 1350–1358.
22. Enthoven, A. C. 1978. Cutting cost without cutting the quality of care. Shattuck Lecture. *N. Engl. J. Med.* **298:** 1229–1238.
23. Bayes, R. T. 1763. An essay toward solving a problem in the doctrine of chance. *Philos. Trans. R. Soc. London.* **53:** 370–418.
24. Shannon, E. C., and W. Weaver. 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana, Ill. 125.
25. Metz, C. E., D. J. Goodenough, and K. Rossman. 1973. Evaluation of receiver-operating characteristic curve data in terms of information theory, with applications in radiography. *Radiology.* **109:** 297–303.
26. Okada, M. 1978. A method for clinical data reduction based on "weighted entropy." *IEEE (Inst. Elect. Electron. Eng.) Trans. Biomed. Eng.* **25:** 462–467.
27. Khinchin, A. I. 1957. Mathematical Foundations of Information Theory. Dover Publications, Inc., New York.
28. McNeil, B. J., E. Keeler, and S. J. Adelstein. 1975. Primer on certain elements of medical decision making. *N. Engl. J. Med.* **293:** 211–215.
29. Metz, C. E. 1978. Basic principles of ROC analysis. *Semin. Nucl. Med.* **8:** 283–298.
30. Chaitman, B. R., D. D. Waters, M. G. Bourassa, J. F. Tubau, P. Wagniart, and R. J. Ferguson. 1979. The importance of clinical subsets in interpreting maximal treadmill exercise test results: the role of multiple-lead ECG systems. *Circulation.* **59:** 560–570.
31. Ransohoff, D. F., and A. R. Feinstein. 1976. Is decision analysis useful in clinical medicine? *Yale J. Biol. Med.* **49:** 165–168.
32. Hall, G. H. 1967. The clinical application of Bayes' theorem. Lancet. **II:** 555–557.
33. Hamilton, G. W., G. B. Trobaugh, J. L. Ritchie, K. L. Gould, T. A. De-Rouen, and D. L. Williams. 1978. Myocardial imaging with 201-thallium: an analysis of clinical use-

fulness based on Bayes' theorem. *Semi. Nucl. Med.* **8**: 358–364.
34. Adelstein, S. J., and B. J. McNeil. 1978. A new diagnostic test for pulmonary embolism. How good and how costly? *N. Engl. J. Med.* **299**: 305–307.
35. Cohn, K., B. Kamm, N. Feteih, R. Brand, and N. Goldschlager. 1979. Use of treadmill score to quantify ischemic response and predict extent of coronary disease. *Circulation.* **59**: 286–296.
36. Ransohoff, D. F., and A. R. Feinstein. 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N. Engl. J. Med.* **299**: 926–929.
37. Fryback, D. G. 1978. Bayes' theorem and conditional nonindependence of date in medical diagnosis. *Comput. Biomed. Res.* **11**: 423–434.
38. Schwartz, W. B., and P. L. Joskow. 1978. Medical efficacy versus economic efficiency: a conflict in values. *N. Engl. J. Med.* **299**: 1462–1464.
39. Kassirer, J. P., and S. G. Pauker. 1978. Should diagnostic testing be regulated? *N. Engl. J. Med.* **299**: 947–949.