Alternative diagnostic strategies for coronary artery disease in women: demonstration of the usefulness and efficiency of probability analysis

JACQUES A. MELIN, M.D., WILLIAM WIENS, M.D., ROLAND J. VANBUTSELE, B.S., ANNIE ROBERT, M.S., PATRICK DE COSTER, M.D., LUCIEN A. BRASSEUR, M.D., CHRISTIAN BECKERS, M.D., and JEAN-MARIE R. DETRY, M.D.

ABSTRACT  Alternative strategies using conditional probability analysis for the diagnosis of coronary artery disease (CAD) were examined in 93 infarct-free women presenting with chest pain. Another group of 42 consecutive female patients was prospectively analyzed. For this latter group, the physician had access to the pretest and posttest probability of CAD before coronary angiography. These 135 women all underwent stress electrocardiographic, thallium scintigraphic, and coronary angiographic examination. The pretest and posttest probabilities of coronary disease were derived from a computerized Bayesian algorithm. Probability estimates were calculated by the four following hypothetical strategies: S0, in which history, including risk factors, was considered; S1, in which history and stress electrocardiographic results were considered; S2, in which history and stress electrocardiographic and stress thallium scintigraphic results were considered; and S3, in which history and stress electrocardiographic and scintigraphic results were used, but in which stress scintigraphic results were considered only if the poststress probability of CAD was between 10% and 90%, i.e., if a sufficient level of diagnostic certainty could not be obtained with the electrocardiographic results alone. The strategies were compared with respect to accuracy with the coronary angiogram as the standard. For both groups of women, S2 and S3 were found to be the most accurate in predicting the presence or absence of coronary disease (p < .05). However, it was found with use of S3 that more than one-third of the thallium scintigrams could have been avoided without loss of accuracy. It was also found that diagnostic catheterization performed to exclude CAD as a diagnosis could have been avoided in half of the patients without loss of accuracy. Studies in the prospective group of 42 women confirmed that S2 and S3 had the best diagnostic accuracy. We also observed a higher prevalence of angiographically documented coronary disease in this group (48% vs 26% in the first group of 93 women). In this prospectively studied group, there was a smaller proportion of patients with a low probability estimate of CAD (29% of the patients with a probability of ≤10% in the prospective group vs 58% of the patients in the first group). Our results suggest that the treating physician's prior knowledge of probability estimates reduced the number of unnecessary diagnostic coronary angiographic procedures performed.


THE APPLICATION of probability methods to the diagnosis of coronary artery disease (CAD) is an important conceptual advance.1–4 The validity of probability analysis for diagnosing CAD has already been demonstrated by comparison of the calculated probability of disease with coronary angiographic results.5–10 Another potential application of probability analysis is to guide the choice of diagnostic procedures. A variety of diagnostic strategies are conceivable and all involve either a systematic approach or the sequential use of the various tests in diagnosing CAD (i.e., exercise electrocardiography, thallium scintigraphy, coronary angiography). The serial probability estimation obtained by the probability analysis might provide the physician with a sufficient level of diagnostic certainty to make a therapeutic decision or, alternatively, indicate the need to perform other diagnostic tests.

In this study we compared, in terms of diagnostic accuracy, strategies using tests on a probabilistic mode with strategies based on blanket-testing decision rules. We performed this analysis on data from symptomatic women without previous myocardial infarction, be-
cause the value of noninvasive testing is still controversial in this group in which there is low-to-intermediate prevalence of CAD.\textsuperscript{11-14} Moreover, if unnecessary scintigraphy and diagnostic coronary angiography in this female population could be avoided it would improve the quality and decrease the cost of care for these patients.

Two consecutive groups of women who all underwent stress electrocardiographic, thallium scintigraphic, and coronary angiographic examination were studied. Data from the second group were prospectively analyzed to assess the influence of availability of posttest probability estimates of CAD on the test-ordering pattern of the physician.

Methods

Patient selection. The study population consisted of two consecutive groups of female patients (n = 135) who all underwent stress electrocardiographic and thallium scintigraphic examination within the 2 weeks preceding coronary angiography. The first group (n = 93) was evaluated by noninvasive testing and coronary angiography for suspected CAD between January 1979 and December 1982; the second group (n = 42) was evaluated between January 1983 and April 1984. The two consecutive groups of female patients were formed in January 1983 because, on this date, the pretest and posttest probability estimates of coronary disease (see below) were sent to the referring and attending physicians, along with the routine exercise electrocardiographic and thallium scintigraphic reports.

Patients were excluded from the study if they had evidence of myocardial infarction. Other exclusion criteria were valvular heart disease, cardiomyopathy, previous coronary bypass surgery, electrocardiographic signs of bundle branch block or left ventricular hypertrophy, and the use of digitalis. Therapy with \(B\)-blocking drugs and calcium blockers was discontinued at least 3 days before the exercise test.

Clinical classification. Before the exercise test, patients were carefully questioned and classified into three groups on the basis of three criteria relative to location, precipitation, and relief of chest pain.\textsuperscript{5, 9} Patients were considered to have typical angina if they complained of substernal discomfort that was precipitated by physical exertion and that was relieved within 10 min by rest or nitroglycerin. Patients were considered to have atypical angina if their discomfort was not substernal, not precipitated by exertion, or not relieved by rest or nitroglycerin. Patients were considered to have nonanginal chest pain if more than one of the three characteristics listed above were absent. The coronary risk factors were calculated in each patient as in Framingham study.\textsuperscript{15} The clinical classification was assessed before the exercise test and was not subsequently modified according to the results of the exercise test or of angiography.

Exercise electrocardiography. The exercise tests were performed on a bicycle ergometer with an initial workload of 20 W for 1 min; the exercise intensity was subsequently increased by 20 W every minute until each patient experienced typical anginal pain or until subjective exhaustion. Isolated depression of the ST segment was never a reason for discontinuing the exercise test. Three orthogonal leads (X, Y, and Z of the Frank system) were constantly monitored and recorded on paper every minute during exercise and for the first 5 min of recovery. A 20 sec electrocardiographic sample was also recorded on a digital magnetic tape at the end of every minute, at rest, during exercise, and during the 5 min of the recovery. These digitized electrocardiographic samples were later averaged and analyzed by computer.\textsuperscript{16} The amount of ST segment depression during maximal exercise was measured 80 msec after the end of QRS, with the PR interval (30 to 10 msec before QRS onset) as reference level. A positive exercise electrocardiogram was defined by 1 mm ST segment depression 80 msec after the J point.

Myocardial perfusion imaging. During the same testing session in which the exercise electrocardiogram was obtained thallium-201 (1.5 to 2 mCi iv) was given 1 min before maximal exercise. Each patient underwent imaging, 5 to 10 min after injection of thallium, under a Searle Pho-Camera V or an Apex 215 M Elscint camera. Images (250,000 counts) were obtained in the anterior and 45 and 65 degree left anterior oblique positions. Redistribution scintigrams were obtained in the majority of the patients (85%) and in all the patients in whom the stress scintigram was initially interpreted as abnormal. The scintigrams were interpreted without computer enhancement by consensus of two observers who had no knowledge of the clinical, electrocardiographic, or angiographic findings.\textsuperscript{17} In each of the three views, the thallium image was divided into five equal circumferential segments and activity in each segment was graded as normal or slightly, moderately, or markedly reduced. A slight reduction in thallium activity localized to the apex was not considered abnormal. Breast shadows were identified on scintigrams as a fixed perfusion defect in the anterosapetal or anterolateral segment with concurrent irregularities in the background activity; such a pattern was considered normal.\textsuperscript{18, 19} Cardiac catheterization. Selective coronary arteriography including multiple projections was performed by a percutaneous transfemoral approach. Coronary arteriograms were interpreted by two angiographers not directly involved in the study and were considered abnormal in the presence of a reduction of 50% or more in diameter of at least one major coronary vessel.

Probability of coronary artery disease. For each patient, the probability of coronary disease was calculated with a microcomputer and a software program called CADENZA (Cardiokinetics, Seattle).\textsuperscript{20} This program is based on previously published data\textsuperscript{2-21} analyzed by Bayes’ theorem.\textsuperscript{22} The following variables are considered in this algorithm: (1) history and coronary risk factors (age, sex, type of chest pain, blood pressure, cholesterol, smoking, resting electrocardiogram), (2) exercise test results (heart rate, blood pressure, ammonium excretion, arrhythmias, R wave amplitude, degree and slope of the ST segment response), and (3) results of thallium scintigraphy (regional hypoperfusion always reversible) and magnitude [mild, moderate, severe] of hypoperfusion.

For each patient, probability of coronary disease was calculated by the four following hypothetical diagnostic strategies: S0, no test was performed and the probability of coronary disease was derived only from the history (including the coronary risk factors); S1, the probability was calculated based on the history and the results of stress electrocardiography; S2, the probability was calculated based on the history and results of stress electrocardiography and thallium scintigraphy; and S3, the probability was calculated based on the history and results of stress electrocardiography and/or thallium scintigraphy (all patients underwent exercise electrocardiography and thallium scintigraphy was performed only in patients with a postelectrocardiography probability of between 10% and 90%).

Statistical analysis. Accuracy of individual probability estimates by the four diagnostic strategies was measured by an accuracy score with the use of the coronary angiogram as the diagnostic standard.\textsuperscript{23, 24} The accuracy score was calculated as follows:

\[
Si = \left(\frac{1}{1 + x_i} \log (1 + pi) + (2 - x_i) \log (2 - pi) - \log (2!) / \log (2)\right) \times 100
\]
where \( p_i \) is the estimated probability of CAD in patient "i," \( S_i \) is the accuracy of this estimate, \( x_i = 1 \) if patient "i" has CAD, and \( x_i = 0 \) if patient "i" does not have CAD. This accuracy score ranges from a minimum of 0 to a maximum of 100, and is highest when the actual outcome is accurately predicted. The score is also related to the amount of information contained in the probability estimate. The difference in individual accuracy scores between the four diagnostic strategies was tested statistically by analysis of variance followed by the contrast method of Scheffé.

The overall accuracy of the probability given by each strategy was assessed with the accuracy coefficient of Zagoria and Reggia. This coefficient is given by

\[
Q = \frac{(2/N) \sum_{K=1}^{n} (P^*_K - 0.5)}
\]

where \( P^*_K \) is the poststrategy probability of CAD in patients with CAD and is 1 minus the poststrategy probability of CAD in patients without CAD; \( n \) is the number of cases. The difference \((P^*_K - 0.5)\) represents how much \( P^*_K \) differs from 0.5, the noninformative prediction. \( Q \) takes the value 1 for a perfect prediction, the value 0 for a perfectly incorrect prediction, and a value of 0 indicates no predictive skill. This overall accuracy coefficient is complementary to the individual accuracy scores previously described because it provides a quantification of the accuracy of each individual strategy.

After calculations by each diagnostic strategy, patients were classified into probability subgroups as follows: low (probability \( \leq 10\% \)), intermediate (probability \( >10\% \) and \( <90\% \)), and high (probability \( \geq 90\% \)). According to the presence or absence of angiographically evident coronary disease, the patients were placed into correctly classified (poststrategy probability \( \geq 90\% \) if CAD and \( \leq 10\% \) if no CAD), misclassified (probability \( \leq 10\% \) if CAD and \( \geq 90\% \) if no CAD), and intermediate groups (probability \( >10\% \) and \( <90\% \)). McNemar's test of symmetry was used to compare the strategies with respect to the division of the patients into the three groups: patients correctly classified, misclassified and with an intermediate probability.

**Results**

**Characteristics of the population.** The clinical and angiographic data on the patients are listed in table 1. In group 1 the prevalence of CAD was lower (26%, 24/93) than in group 2 (48%, 20/42). The majority of the patients with CAD (29/44, 66%) had a history of typical angina pectoris. In 75% of these patients, double- or triple-vessel disease was present. Conversely, of patients with normal coronary angiograms \((n = 91), 85\% (77/91)\) presented with a history of atypical angina pectoris or nonanginal chest pain. The prevalence of CAD was 67% (29/43) in women with typical angina and 25% (15/61) in women with atypical angina.

**Exercise test results.** Classic parameters, sensitivity (true positives/all patients with CAD), and specificity (true negatives/patients without CAD) were calculated in the study sample. In group 1 sensitivity (58%, 14/24) and specificity (80%, 55/69) of exercise electrocardiography were lower than those of exercise thallium scintigraphy (sensitivity 71%, 17/24%; specificity 91%, 63/69). The figures for sensitivity (electrocardiography 65%; scintigraphy 70%) and specificity (electrocardiography 77%; scintigraphy 95%) were similar in group 2.

**Distribution of probability estimates.** The distribution of probability estimates of CAD in group 1 after the four diagnostic strategies were applied is illustrated in figure 1. After calculating by S2 and S3, the majority of the patients were in the first and tenth deciles (65 [70%] by S2, and 73 [78%] by S3). In contrast, with S0 and S1 there was a more uniform distribution of probabilities, indicating a greater level of uncertainty with these strategies (the level of uncertainty being maximal in the middle range of probabilities). With S0, only 35 (38%) of the patients were in the first and tenth deciles, while with S1 this number was 43 (46%).

**Distribution of accuracy scores.** The distribution of accuracy scores for probability estimates in group 1 after calculations by the four diagnostic strategies is illustrated in figure 2. The accuracy scores were better with S2 and S3: with S2, 84% of the patients had an accuracy score of between 90 and 100 and 87% of the patients had a score with S3. The differences between accuracy scores for the four diagnostic strategies were tested statistically by analysis of variance. In group 1 there was a difference \((p = .04)\) between S0 and S1 vs S2 and S3. No significant difference was noted either between S0 and S1 or between S2 and S3. In group 2 accuracy scores were also similar for S2 and S3. There was also a difference \((p = .02)\) between S0 and S1 vs S2 and S3. The accuracy coefficients (perfect prediction = 1; perfectly incorrect prediction = -1) were, for S0, S1, S2, and S3, respectively, 0.492, 0.504, 0.686, and 0.702 (group 1) and 0.253, 0.386, 0.610,
FIGURE 1. Histogram showing the distribution of Bayesian probabilities of coronary disease after four diagnostic strategies for group 1. The probabilities are grouped into deciles. The bars represent the number of patients classified in a given probability decile at the end of each strategy.

FIGURE 2. Histogram showing the distribution of accuracy scores of probability estimates after four diagnostic strategies for group 1. The accuracy scores are grouped into deciles. The bars represent the number of patients classified in a given accuracy score decile at the end of each strategy.
and 0.591 (group 2). Thus, S2 and S3 were the most accurate in predicting the presence or absence of CAD and their accuracy scores were similar.

Classification into high-, low-, and intermediate-probability subgroups. This classification is illustrated in figures 3 and 4. The percentage of patients correctly classified was similar when S2 and S3 were used (group 1, 68% for S2 and 74% for S3; group 2, 55% for S2 and 62% for S3). Among the misclassified patients, only one (in group 1) had CAD and was classified in the low-probability subgroup; this patient had one-vessel disease (50% stenosis in the middle portion of the left anterior descending artery). According to results with S3, only in the patients in the intermediate-probability subgroup would it have been necessary to perform thallium scintigraphy after the electrocardiographic examination. This amounted to 56% of the patients in group 1 and 67% of those in group 2. The major difference between group 1 and group 2 was in the proportion of patients in each in the low-probability subgroup: more than half in group 1 (54% for S2 and 58% for S3) and less than one-third in group 2 (24% for S2 and 29% for S3).

**Choice of the optimal strategy.** The choice of the optimal diagnostic strategy can be accomplished by use of the criteria of accuracy and risk and cost to the patient. The accuracy of each strategy refers to its ability to accurately diagnose or exclude the diagnosis of CAD: S2 and S3 were found to be the most accurate. The strategy (S3) based on electrocardiographic results and in which a threshold probability was used in guiding the use of thallium scintigraphy was the most risk and cost effective: without loss of accuracy, 44% of the patients in group 1 and 33% of those in group 2 could have been spared scintigraphic examination; 58% and 29%, respectively, could have been spared diagnostic catheterization.

**Discussion**

This study demonstrates that a diagnostic strategy using certain tests on a probabilistic mode is as accurate as a strategy using blanket testing in every patient. In a group of 135 symptomatic women, this sequential strategy would have eliminated the need for a significant number of thallium scintigrams and for cardiac catheterizations without loss of diagnostic accuracy. In a prospectively studied group of 42 patients, the percentage of patients undergoing coronary angiography

---

**FIGURE 3.** Distribution of probability subgroups at the end of strategy 2 (S2; history, exercise ECG and thallium scintigram in all patients) and of strategy 3 (S3; history and exercise ECG in all patients; thallium scintigram only in patients with a probability between 10 and 90% after history and exercise ECG) in group 1. There are three probability subgroups: p ≤10%, p between 10 and 90% and p ≥90%. The results of coronary angiography are indicated for the subgroups p ≤10% and p ≥90% as CAD− (absence of coronary disease) and CAD+ (presence of coronary disease).

**FIGURE 4.** Distribution of probability subgroups at the end of S2 and of S3 in group 2.
who had a low posttest probability of CAD was markedly reduced (29% vs 58%); this reduction likely results from the fact that in this prospective study the physician was aware of the posttest probability before coronary angiography was performed.

A patient population of women without myocardial infarction was selected for the following reasons: (1) the diagnosis of CAD in women who have a chest pain syndrome remains a difficult clinical problem, (2) the specificity of exercise testing for CAD in women is reportedly relatively poor, and (3) there is a need to diminish diagnostic coronary angiography in this group of patients in which there is a low-to-intermediate prevalence of CAD. The angiographically determined prevalence of CAD in the two groups of women selected for this study is similar to that reported in the Coronary Artery Surgery Study of 2810 women who underwent coronary angiography for chest pain and who had no previous history of myocardial infarction.21 The sensitivity and specificity of the two diagnostic noninvasive tests for CAD in this study are within the range of values for women reported by others.11-14, 27-31 Thallium scintigraphy is more sensitive and more specific for diagnosing CAD than exercise electrocardiography. Quantitative methods for interpreting thallium uptake and clearance and single-photon emission tomographic imaging would further improve the sensitivity of thallium scintigraphy. Nevertheless, the combination of exercise electrocardiography and planar thallium scintigraphy has been reported to be a better predictor of the presence or absence of disease than a single noninvasive test.14, 17, 32, 33 However, it is not feasible, from a cost and risk point of view, to obtain a stress thallium scintigram from every woman complaining of chest pain.

In this study, we used conditional probability analysis as an aid to diagnosis and medical decision making. The advantages of probability analysis are as follows: first, based on Bayes’ theorem, the prevalence of CAD in the studied population and the sensitivity and specificity of the tests are all taken into account; second, in a sequential use of multiple tests, potentially conflicting information may be processed; and third, this method yields the probability that a given patient has disease, rather than a categorical diagnosis. Besides their value for the diagnosis of CAD,5-20 posttest probabilities given by the Bayesian method can be used for medical decision making. If one accepts that the diagnostic certainty is sufficient when posttest probability for CAD is above or below a given threshold, then only the patients with posttest probabilities between these thresholds would need the next test in a sequential diagnostic tree.

In this study using conditional probability analysis, we first showed that the strategies combining exercise electrocardiography and thallium scintigraphy were the more accurate in two consecutive populations of women. We found no statistical difference between results with S0, based on history and risk factors only, and S1, in which stress electrocardiograms were also considered. This confirms the findings of Goldman et al.,34 who showed that the incremental value of the stress electrocardiogram for the diagnosis of CAD is limited after other easily obtainable clinical data are taken into account. The better diagnostic accuracy obtained with strategies including thallium scintigraphic results was similarly assessed by Hlatky et al.36 These authors documented that perfusion scintigraphy significantly improved cardiologists’ diagnostic accuracy beyond that achieved with a detailed history, physical examination, and data from the stress electrocardiogram. Moreover, the extent of improvement depended on the patient’s clinical presentation, being greatest in the patients with atypical angina and a positive electrocardiographic result, a population in which prevalence of CAD is similar to that in the present group of patients.

A further step in the search for optimal strategies for diagnosing CAD in women was to show that a decision rule based on a threshold probability and used to guide the choice of whether or not to perform thallium scintigraphy is as accurate as the strategy in which both stress electrocardiography and thallium scintigraphy are considered for all patients. This was true in two consecutive groups of patients. Patterson et al.36 showed theoretically in a quantitative model of cost-effectiveness analysis that a clinical policy based on the sequential use of exercise electrocardiography before thallium imaging was the most cost effective. Cutoff points of 10% and 90% were reported to contain the majority of diagnostic information.3 With 10% and 90% as threshold probabilities, thallium scintigraphy could have been avoided in more than one-third of the patients, diminishing considerably the risk and the cost of the strategy, without loss of accuracy.
angiography in the low-probability group was markedly reduced. In group 1, using S3 we found that 58% of the patients were classified in the low-probability group, and this percentage fell to 29% in group 2. While the staff members of the Division of Cardiology were the same, we may assume that the knowledge of the posttest probabilities by the physician before the decision to perform coronary angiography was made is an important causal factor. The published reports on the utility of combining exercise electrocardiography and thallium scintigraphy for excluding the diagnosis of CAD have probably also influenced the confidence of the physician in the results of these tests. Despite this potential influence of posttest probabilities on the test-ordering pattern, it is clear that probability analysis will never replace clinical judgment in the decision to submit a patient to coronary angiography, but it could provide useful complementary information by integrating complex or discordant test results.

To better understand the advantages of S3, the risk and cost issues raised with respect to the thallium scan and angiography need to be discussed. For thallium scintigraphy, the potential risks are reviewed by Stason and Fortress, and include the risk of exposure to radiation (total body = 360 mrad; ovaries = 850 mrad), which is a concern in this group of patients. These risks and the high average cost of thallium scintigraphy could have been drastically reduced by use of S3, by which more than one-third of the patients would not have undergone the procedure. For coronary angiography, the risks are principally those of radiation in this group of women with a low prevalence of CAD. The reported morbidity and mortality figures for the procedure are indeed less applicable for a group with a low prevalence of coronary and multivessel disease. The cost of catheterization must be weighed against the fact that the knowledge of a normal coronary anatomy could significantly contribute to the rehabilitation of the patient and reduce their use of medical care facilities in the future. However, verification that a normal coronary angiogram is “therapeutic” has been lacking to date. For this reason and because of the high accuracy of S3, it seems justified to avoid performing coronary angiography in the low-probability subgroup.

We conclude from this study that in a group of women referred for suspicion of CAD, a diagnostic strategy that uses probability analysis to guide the use of diagnostic tests offers the best balance of diagnostic accuracy, patient comfort, and minimal financial cost. Also, the knowledge of the posttest likelihood of disease before angiography is performed would likely reduce the need for the procedure in patients with a low posttest likelihood of CAD.

References
Alternative diagnostic strategies for coronary artery disease in women: demonstration of the usefulness and efficiency of probability analysis.
J A Melin, W Wijns, R J Vanbutsele, A Robert, P De Coster, L A Brasseur, C Beckers and J M Detry

Circulation. 1985;71:535-542
doi: 10.1161/01.CIR.71.3.535

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/71/3/535