BREASTAID

Clinical Results From Early Development of a Clinical Decision Rule for Palpable Solid Breast Masses

Janet Rose Osuch, MD, MS, Mathew J. Reeves, BVSc, PhD,† Dorothy R. Pathak PhD, MS,†‡ and Tosca Kinchelow, MD§*

Objective: To develop a clinical decision rule (entitled BREASTAID) that will predict the probability of malignancy in women with palpable solid breast masses.

Summary Background Data: Currently, 80% of open breast biopsies are benign, resulting in excessive economic, psychologic, and physical morbidity.

Methods: A total of 452 solid breast masses were evaluated in a surgical breast clinic between November 1994 and February 1998. Breast cancer status was defined histologically as ductal carcinoma in situ or invasive cancer. Noncancer status included benign histology, mass resolution, or stability at 12-month follow-up. Data were collected on risk factors, clinical breast examination, mammography, and cytology results. Three multiple logistic regression models were used to generate the probability of cancer at 3 logical steps in the workup; Bayes' theorem was applied in a stepwise fashion to generate a final probability of cancer.

Results: A model incorporating only clinical breast examination and mammography resulted in an excessive number of either missed cases or biopsies compared with one that included cytology. Using a cut-point of 4%, this latter BREASTAID model had 97.6% sensitivity and 85.1% specificity. Compared with triple diagnosis, BREASTAID would have reduced the open biopsy rate from 39.8% (180 of 452) to 22.3% (101 of 452), improving the diagnostic yield from 22.7% to 40.6%.

Conclusions: This study convincingly demonstrates that at minimum, clinical, radiologic, and cytologic evaluations are required to accurately evaluate a solid breast mass. BREASTAID has the potential to minimize the number of open biopsies performed while allowing safe triage to follow-up. Before widespread application, further validation studies are required.

(*Ann Surg* 2003;238: 728–737)

From the Departments of *Surgery, †Epidemiology, and ‡Family Practice, Michigan State University College of Human Medicine, East Lansing, MI; and the §Department of Surgery, UMDNJ New Jersey Medical School, Newark, NJ.

Supported in part by U.S. Army Grant No. DAMD17-97-1-7182, and by the Institute for Managed Care and the Office of the Dean for Research, College of Human Medicine, Michigan State University.

This work is an expansion of a thesis written by the first author of this article to meet the requirements for a Master of Science degree in Epidemiology from Michigan State University.

Reprints: Janet Rose Osuch, MD, Department of Epidemiology, Michigan State University, 4660 South Hagadorn, Suite 600, East Lansing, MI 48823. E-mail:janet.osuch@ht.msu.edu.

Copyright © 2003 by Lippincott Williams & Wilkins

0003-4932/03/23805-0728
DOI: 10.1097/01.sla.0000094446.78844.ae

In 1970, breast cancer presented as a palpable mass in 90% of patients.[1] Recent reports demonstrate that the majority (55%-68%) of breast cancer cases continue to present with palpable masses, despite the widespread use of screening mammography.[2,3] While the approach to a solid palpable breast mass is often open surgical biopsy, the malignant/benign breast biopsy ratio in the United States averages only about 1:4, or 20%.[2,4] Use of this figure and the annual incidence of breast cancer predicts that more than 1 million breast biopsies will be performed in 2002. Using an average reimbursement for open breast biopsy of $2400,[5] the financial cost to the healthcare system in 2002 will be more than $2.4 billion. While many physicians consider open biopsy a minor procedure with few consequences, women pay a significant emotional toll, not only because of the anxiety of the procedure itself but also because of the delay between the clinical evaluation, the operation, and the pathology result. Some women also experience disfigurement as a consequence of the procedure, and subsequent breast cancer screening interpretation can be more difficult because of tissue scarring. Competing with these issues is the acknowledged difficulty of ruling out breast cancer short of open biopsy. In addition, delay in the diagnosis of breast cancer is the most common reason for malpractice litigation in the United States.[6] This reality provides strong incentives for surgeons to continue to perform the procedure, despite its low malignant/benign ratio and associated morbidities.

Triple diagnosis has been used to decrease the open biopsy rate for solid breast masses, a technique that combines

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

the results of clinical impression, mammography, and fine needle aspiration biopsy (FNAB). When all three results are concordant for benign disease, follow-up rather than biopsy is advised.[7,8] However, triple diagnosis has several limitations. First, the ill-defined interpretive criterion associated with the clinical impression limits its application to experienced examiners. Triple diagnosis often ignores hypocellular FNAB results, rendering the test useless in up to 20% to 30% of cases.[9] Reporting only positive or negative outcomes for the test results in loss of information, since all three components use multiple categories in their interpretation. Triple diagnosis assumes independence of the tests comprising it, but this has not been tested. Finally, the pretest likelihood of disease (as evaluated by risk factor data), is not formally assessed.

Development of a clinical decision rule (CDR) that could overcome the limitations of triple diagnosis has the potential to significantly impact the healthcare of women. A CDR quantifies the contributions that various components of the history, physical examination, and basic laboratory results make toward the diagnosis.[10–12] The purposes of this study are threefold: 1) to develop a CDR that will accurately predict the probability of malignancy in women with palpable solid breast masses; 2) to develop a CDR that is practical enough to be used by a general surgeon or primary care clinician; and 3) to compare the CDR to our current method of practice, which employs triple diagnosis. Termed BREASTAID (*B*reast *R*isk *E*valuation *A*nd *S*coring System *T*o *A*id *I*n the *D*iagnosis of Mammary Masses), this article reports on these goals and the clinical results of the first stages of CDR development.

METHODS

Setting. The study was conducted at the Comprehensive Breast Health Clinic (CBHC) at Michigan State University, a single-site referral center for both physician and self-referred patients attended by 5 surgeons and 1 nurse practitioner. Of the new patients, approximately 85% of the patients are physician-referred and 15% are self-referred. Approximately 35% of the clinic population consists of follow-up high-risk patients and patients with a previous history of breast cancer, and 65% of newly referred women with active problems. The majority of the clinic population is drawn from a 30-mile radius around Lansing, Michigan. The study was performed with the informed consent of the patients and followed the ethical standards of the Institutional Review Board at Michigan State University. It was approved by the Institutional Review Board prior to its onset and renewed at appropriate intervals.

Selection of study population. Women with palpable solid breast masses who had a fine needle aspiration biopsy performed between November 1994 and February 1998 and a mammographic evaluation of the mass within 6 months of initial presentation were eligible for the study. Eligibility also required open biopsy, follow-up for a minimum of 12 months, or disappearance of the mass during the follow-up period.

Routine management of study population. The breast masses used in the development of BREASTAID were managed in the CBHC using the routine clinical protocol of triple diagnosis. Open biopsy was recommended for any triple diagnosis result that was other than benign.

Predictor variables. We collected two different sets of predictor variables, epidemiological and clinical. These were collected prospectively and in a blinded fashion, prior to determination of cancer status. The epidemiological predictor variables were collected using a standardized questionnaire completed by each patient at the time of the initial evaluation and updated at each subsequent visit. Clinical predictor variables were collected using a standardized form that was completed at every visit by the examiner.

Epidemiologic predictor variables included age, height, weight, personal history of breast cancer, reproductive factors (age at menarche, parity, age at first delivery, number of deliveries, menopausal status, age at menopause); exogenous hormone use (oral contraceptives, estrogen alone or estrogen and progesterone replacement therapy, including duration of use); family history of breast cancer (first- and second-degree relatives, including age and unilateral or bilateral involvement); history of early-life radiation exposure; and tobacco and alcohol exposure, including duration and frequency. Menopausal status was assigned according to the previously developed criteria of Colditz and Frazier.[13] The 11 patients who did not know their family history were assigned to the negative family history category. Clinical predictor variables included characteristics of the breast mass on clinical breast examination (CBE), mammography results, and FNAB results (consistency and cytology). Standard clinical definitions were used to describe mass characteristics on CBE according to size, shape, external texture, and mobility.[14,15] Twelve masses were described as subtle thickenings on CBE. These lesions represented two-dimensional thickenings without easily definable borders and did not have a mass size recorded. To include them in the analysis, they were assigned the average mass size for the total study population (1.6 cm). Mammography results in this study refer to the radiographic results at the same location as the palpable lesion in question. Results were recorded by the surgeons at the same time as mass evaluation using the BI-RADS lexicon system of the American College of Radiology.[16] The timing of mammograms in relation to FNAB followed published guidelines.[17,18] FNAB was performed using a modification of the technique originally introduced by Martin and Ellis in 1930.[19] Internal consistency during FNAB was classified according to the degree of hardness (soft, rubbery, or hard) and the presence or absence of a gritty texture during the procedure

© 2003 Lippincott Williams & Wilkins

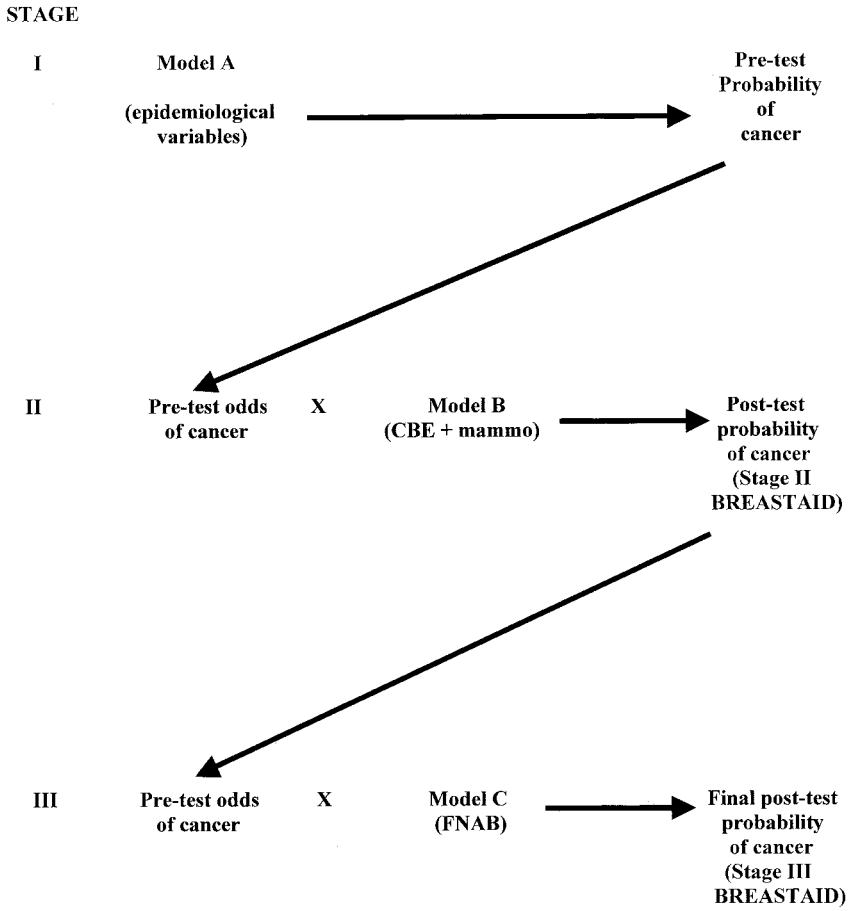
Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

using descriptions previously published.[20,21] Fine needle aspiration cytology was classified according to the 1996 consensus conference guidelines from the National Institutes of Health,[22] with the exception that an additional category of mild atypia was added. Thus, cytology results were classified as hypocellular, benign epithelial cells, mild atypia, moderate-severe atypia, suspicious for malignancy, and overtly malignant.

Outcome variable. The outcome variable was defined as histologic confirmation of breast cancer (defined as ductal carcinoma in situ, invasive ductal carcinoma, or invasive lobular carcinoma) at open biopsy. Noncancer status was defined in 1 of 3 ways: 1) benign disease histologically, 2) resolution of the mass within 12 months of presentation, or 3) stability of the mass on CBE after a minimum of 12-month follow-up. All patients who did not undergo immediate biopsy were seen in follow-up at 3-month intervals unless they were initially evaluated during a suboptimal hormonal environment. These patients had a 6-week evaluation following the initial visit and then were followed at 3-month intervals. Masses that persisted despite a decreased hormonal milieu, that increased in size, or whose consistency became more pronounced on follow-up underwent open biopsy at that time.

Data Analysis

Model development. Details of model development are provided in a separate publication (Reeves MJ, Osuch, JR, Pathak DR. Development of a clinical decision rule for triage of women with palpable breast masses. J Clin Epidemiol. 2003;56:635–645). Briefly, a three-step approach was used to model the probability of breast cancer in a way as similar as possible to the typical workup that a clinician uses in evaluating a breast mass. Three independent multivariable logistic regression models were generated, with the outcome of interest being the probability of breast cancer from each model. Model A assessed the epidemiological variables, model B assessed the combination of CBE and mammography results, and model C assessed the results of FNAB. Variables that



CBE = Clinical Breast Examination; mammo = mammogram; FNAB = fine needle aspiration biopsy

FIGURE 1. Schema for the development of BREASTAID: the sequential application of Bayes' theorem.

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

achieved statistical significance at the $P < 0.05$ level were retained in the final models. Odds ratios and 95% confidence intervals were calculated to assess the magnitude of the association between each predictor variable and the outcome (breast cancer). Bayes' theorem was then applied in a sequential fashion to the three models to arrive at a final probability of disease, as has been described previously.[23,24]

For purposes of clarity, the generation of the three models, followed by the application of Bayes' theorem to revise disease probabilities, was referred to as a "stage" in CDR development. Thus, stage I was used to establish the pretest likelihood of cancer based on the epidemiological variables alone. Stage II (typical of most primary care and some surgical practices) assessed the contributions of CBE and mammography results to the history. Stage III (typical of some primary care and many surgical practices) assessed the additional contribution of FNAB results to the history and the clinical data collected as part of stages I and II. A summary of the steps used in BREASTAID development is shown in Figure 1. Specifically, model A assessed the epidemiologic predictor variables gathered at the initial interview. The output of this model was used to set an initial pretest probability of breast cancer (stage I), which was then converted to pretest odds in stage II. Model B assessed the contributions of the CBE and mammography results, and the results of this model were then combined with the pretest odds of breast cancer (from model A), using Bayes' theorem, to generate a post-test probability of breast cancer (stage II BREASTAID; Fig. 1). Model C assessed the FNAB contributions to the diagnosis of breast cancer. The output from model C was combined with the pretest odds of cancer derived from stage II to generate a final post-test probability of breast cancer (stage III BREASTAID; Fig. 1).

Performance Evaluation

The test performance of stages II and III was evaluated separately and compared by measurement of test characteristics (ie, sensitivity and specificity and receiver operating characteristic [ROC] curve analysis). First, 2 × 2 tables were constructed for multiple cut-points for the posttest probabilities of both stage II and stage III BREASTAID. Sensitivity, specificity, positive predictive value, and negative predictive value were calculated for each cut-point. A cut-point defines a range of "positive" test results and a 2 × 2 table evaluates the number of breast cancer cases diagnosed at the chosen cut-point and compares it to the number of benign cases diagnosed under the same conditions. Formulas and the definitions of each test characteristic along with an example of a 2 × 2 table are shown in Figure 2. Keeping in mind that sensitivity and specificity are inversely related, an optimal cut-point for a disease such as breast cancer, in which one would choose a very high rate of true positive results, would maximize sensitivity at the expense of specificity (the true negative rate). The best cut-point for BREASTAID was defined as the one that would achieve a clinically acceptable sensitivity of at least 95%, while optimizing specificity to avoid large numbers of biopsies among women with benign lesions.

TERM	DEFINITION	CHARACTER FORMULA	SYMBOL FORMULA
Sensitivity	Proportion of patients with breast cancer who have a positive test	$\frac{TP}{TP+FN}$	$\frac{a}{a+c}$
Specificity	Proportion of patients without breast cancer who have a negative test	$\frac{TN}{TN+FP}$	$\frac{d}{b+d}$
Positive predictive value (PPV)	Proportion of patients with a positive test who have breast cancer	$\frac{TP}{TP+FP}$	$\frac{a}{a+b}$
Negative predictive value (NPV)	Proportion of patients with a negative test who do not have breast cancer	$\frac{TN}{TN+FN}$	$\frac{d}{c+d}$

	CANCER	NO CANCER	
TEST + (≥ cut-point)	True Positive (TP) a	False Positive (FP) b	a + b
TEST – (< cut-point)	False Negative (FN) c	True Negative (TN) d	c + d
	a + c	b + d	N

FIGURE 2. Definition of test characteristics used in BREASTAID and an example of a 2 × 2 table.

Second, ROC curve analysis was performed using available software.[25] A ROC curve is a graphic representation of test results, with sensitivity (true positive rate) plotted on the "y" axis and 1–specificity (false positive rate) plotted on the "x" axis. A ROC curve depicts the inherent tradeoff between sensitivity and specificity as cut-points for a "positive" test are changed. The area under the curve (AUROC) represents the overall accuracy of the test and ranges between 0 and 1.0, with an area of 0.5 representing test accuracy no better than chance alone.[26] A ROC plot can also be used visually and quantitatively to assess the various cut-points that best predict outcome. Curves whose plateau is located in the farthest "northwest" corner of the graph have the greatest discrimination ability. In this study, AUROC was used to compare the performance of stage II and stage III BREASTAID.

RESULTS

Population Studied. Of 443 women identified as potentially eligible for the study, 36 (8%) were excluded because a

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

mammogram was not performed within 6 months of presentation and 27 (6%) were excluded because the medical record could not be located after three attempts. This resulted in 380 eligible women for the study, in whom 452 individual lesions were evaluated. The average age of the total study population was 47.7 years, and 64.2% (N = 244) were younger than 50 years. Forty-one women (9.1%) were diagnosed with breast cancer; their average age was 54.3 years compared with 46.9 for the noncancer cases. Eighteen (43.9%) of the 41 cancer patients diagnosed with malignancy were younger than 50 years, and 3 (7.3%) were younger than 40 years.

Routine management and biopsy results on the studied population. Results of the clinical disposition for the 452 lesions are summarized in Figure 3. Using triple diagnosis to triage patients, open biopsy was performed in 180 (39.8%) of the 452 lesions and follow-up without subsequent biopsy in the remaining 272 (61.2%). Of the 180 masses that were biopsied, 41 or 22.8% were malignant, resulting in a malignant/benign biopsy ratio of 1:3.4. Of the 272 masses that were followed, 263 disappeared on follow-up. Only 9 masses persisted without change after 12 months of follow-up; these continued to be followed according to routine protocol. A surgical decision for an open procedure was made on the basis of triple diagnosis in 108 cases at initial presentation and included 38 of the total 41 malignancies (92.7%). Open biopsy was done in 53 additional cases within 3 months of initial presentation, and 3 of these masses were malignant. A surgical decision recommending biopsy was made at the 6-week follow-up visit for each of these 3 malignancies, all of whom had subtle CBE results. One of the patients was premenopausal and at a suboptimal phase of her menstrual cycle at initial evaluation, and the other 2 were taking hormone replacement therapy, which was discontinued at the initial visit. Six-week reevaluation of these 3 cases noted persistence of subtle CBE findings and biopsy was recommended. Nineteen additional lesions had recommendations for biopsy between 3 and 6 months of initial presentation. All of these were benign. The histopathological results for the 180 open biopsies were as follows: invasive ductal carcinoma, n = 30 (16.7%); invasive lobular carcinoma, n = 6, (3.3%); ductal carcinoma in situ, n = 5; (2.8%); lobular carcinoma in situ, n = 2 (1.1%); atypical epithelial hyperplasia, n = 9 (5.0%); fibrocystic change, n = 90 (50.0%); fibroadenoma, n = 25 (13.9%); fat necrosis, n = 4 (2.2%); lipoma, n = 3 (1.7%); and miscellaneous benign lesions, n = 6; (3.3%).

Test Performance Characteristics of BREASTAID

Stage II. Table 1 represents a classification table demonstrating the test characteristics for various cut-points for stage II BREASTAID. The best posttest probability cut-point for maximizing sensitivity and specificity, while minimizing the number of biopsies needed, was 2%. Using a 2% cut-point, stage II BREASTAID had a sensitivity of 97.6% and a specificity of 49.9%. The positive predictive value and negative predictive value for the population were 16.3% and 99.5%, respectively. Application of stage II BREASTAID using a cut-point of 2% would have required 246 biopsies, whereas the usual method of diagnosis in the CBHC resulted in fewer biopsies (180) in the same population. Using stage II BREASTAID would have resulted in a malignant/benign biopsy ratio of 1:5.2. The error rate (ie, 1–negative predictive value) among negative results (the 206 masses triaged into follow-up) for stage II BREASTAID would have been 0.49% (1 of 206). The false negative rate (ie, 1–sensitivity) among cancer cases would have been 2.4% (1 of 41 missed malignancies at the first visit). Use of a higher cut-point to decrease the number of breast biopsies would have resulted in an unacceptable number of missed cancers (Table 1). Figure 4 illustrates the ROC curves for stages II and III of BREASTAID. The AUROC for stage II in Figure 4 was 0.93 (95% confidence interval = 0.87–0.97).

Stage III. Table 2 represents a classification table demonstrating the test characteristics for various cut-points for stage III BREASTAID (that includes FNAB data). The best posttest probability cut-point for maximizing sensitivity and specificity, while minimizing the number of biopsies needed, was 4%. Using a 4% cut-point, stage III BREASTAID had a similar sensitivity of 97.6% compared with stage II, but a much higher specificity of 85.2%. The positive predictive value for the population using stage III BREASTAID was 39.6% and the negative predictive value was 99.7%. Appli-

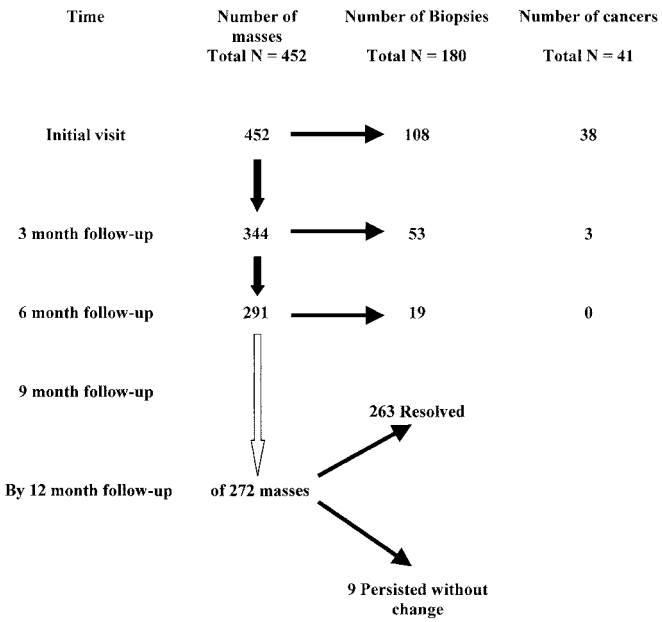


FIGURE 3. Clinical disposition of 452 solid breast masses.

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

TABLE 1. Test Characteristics for Stage II BREASTAID for Various Cut-points by Final Probability of Cancer (Based on 452 Masses)

≥% Final Probability	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	No. of Biopsies Needed	% Masses Biopsied	Malignant/ Benign Ratio	No. of Cancers (total 41) Missed	% Missed Cancers
95	39.0	100	100	94.3	16	3.5	1:0	25	61.0
90	41.5	100	100	94.5	17	3.8	1:0	24	58.5
80	41.5	100	100	94.5	17	3.8	1:0	24	58.5
70	41.5	99.8	94.4	94.5	18	4.0	1:0.06	24	58.5
60	46.3	99.5	90.5	94.9	21	4.6	1:0.1	22	53.7
50	46.3	99.5	90.5	94.9	21	4.6	1:0.1	22	53.7
40	61.0	98.5	80.6	96.2	31	6.9	1:0.2	16	39.0
30	63.4	98.3	78.8	96.4	33	7.3	1:0.3	15	36.6
20	80.5	92.0	50.0	97.9	66	14.6	1:1.0	8	19.5
10	85.4	86.1	38.0	98.3	92	20.4	1:1.6	6	14.6
5	87.8	77.1	27.7	98.4	130	28.8	1:2.6	5	12.2
4	90.2	69.8	23.0	98.6	161	35.6	1:3.4	4	9.8
3	92.7	62.8	19.9	98.8	191	42.3	1:4.0	3	7.3
2*	97.6	49.9	16.3	99.5	246	54.4	1:5.2	1	2.4
1	97.6	41.6	14.3	99.4	280	62.0	1:6.0	1	2.4
0.5	100	20.9	11.2	100	366	81.0	1:7.9	0	0
0.4	100	19.2	11.0	100	373	82.5	1:8.1	0	0
0.3	100	13.1	10.3	100	398	88.0	1:8.7	0	0
0.2	100	5.8	9.6	100	428	94.7	1:9.4	0	0
0.1	100	5.1	9.5	100	431	95.4	1:9.5	0	0

PPV, positive predictive value; NPV, negative predictive value.
*Optimal cut-point for maximizing sensitivity followed by specificity.

cation of stage III BREASTAID to the study population using this cut-point would have required only 101 biopsies, as compared with 180 biopsies using triple diagnosis, and 246 using stage II. Using stage III BREASTAID would have improved the malignant/benign biopsy ratio from 1:3.4 (22.7%) using triple diagnosis to 1:1.5 (40%). The error rate (ie, 1–negative predictive value) among negative results (the 351 masses triaged into follow-up) for BREASTAID stage III would have been 0.28% (1 of 351). The false negative rate (ie, 1–sensitivity) among cancer cases would have been the same as for stage II, 2.4% (1 of 41 missed malignancies at the first visit). Figure 4 demonstrates that the ROC curve for stage III BREASTAID was better than that for stage II. It is closer to the northwest corner of the graph and has a higher AUROC of 0.97 (95% confidence interval = 0.92–0.99).

DISCUSSION

Stage III BREASTAID (herein referred to simply as "BREASTAID") is a unique prediction model for breast cancer in women with palpable solid breast masses that has the potential to reduce the number of diagnostic open biopsies. It must be cautioned, however, that BREASTAID is still in research stages and should not be applied clinically until all of the methodological standards for a CDR have been met.[12] Most notably, multi-institutional prospective validation will be necessary to test the generalizability of the method in other settings.

Data analysis from stage II BREASTAID, which does not use FNAB results, convincingly demonstrates that evaluation of a solid breast mass requires, at minimum, cytologic evaluation in addition to clinical and radiographic assessment. This has important implications to the practice of medicine. Although FNAB is not an especially difficult procedure to perform technically, it has a number of interpretive pitfalls both cytologically and clinically.[21] Clinicians who routinely perform FNAB in settings where cytologic expertise is available are those most likely to be comfortable with the application of BREASTAID. If future studies demonstrate BREASTAID to be accurate in different practice settings, those who do not routinely perform FNAB in their clinical practices but who wish to apply BREASTAID will not only need to learn the technique of FNAB and its pitfalls,

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

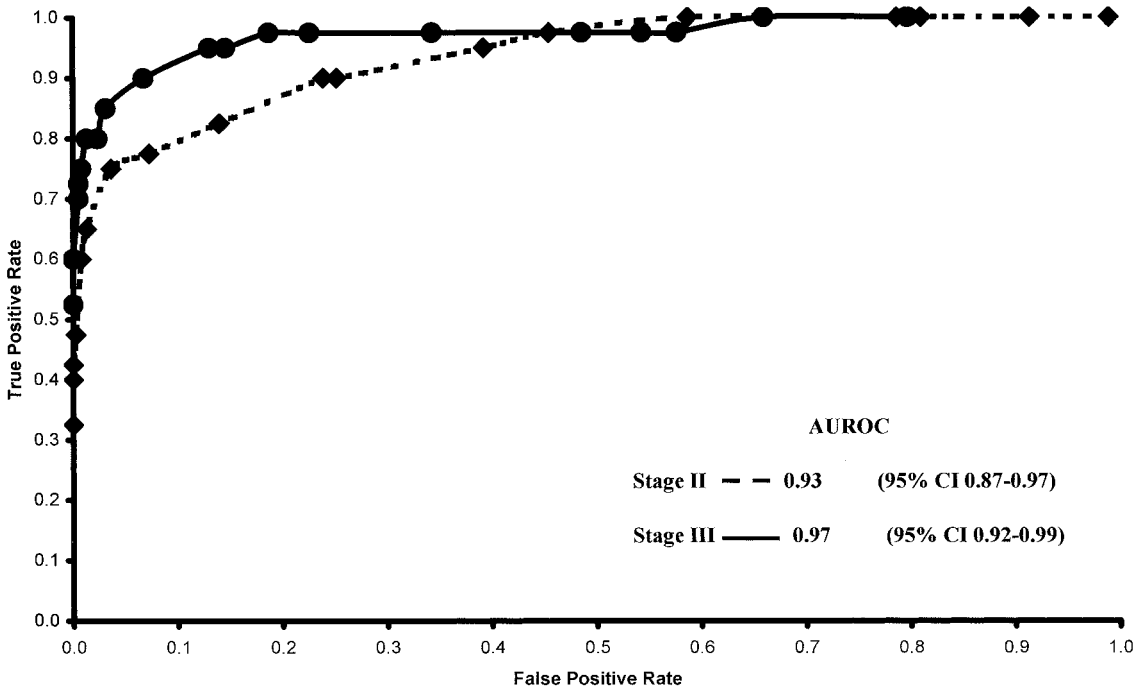


FIGURE 4. Receiver operator characteristic (ROC) curves for Stage II BREASTAID (dotted line) and Stage III BREASTAID (solid line).

TABLE 2. Test Characteristics for Stage III BREASTAID for Various Cut-points by Final Probability of Cancer (Based on 452 Masses)

≥% Final Probability	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	No. of Biopsies Needed	% Masses Biopsied	Malignant/ Benign Ratio	No. of Cancers (total 41) Missed	% Missed Cancers
95	51.2	99.8	95.4	95.4	22	4.9	1:0.05	20	48.8
90	53.7	99.8	95.7	95.6	23	5.1	1:0.05	19	46.3
80	58.5	99.8	96.0	96.0	25	5.5	1:0.04	17	41.5
70	65.8	99.5	93.1	96.7	29	6.4	1:0.1	14	34.2
60	68.3	99.0	87.5	96.9	32	7.1	1:0.1	13	31.7
50	68.3	99.0	87.5	96.9	32	7.1	1:0.1	13	31.7
40	73.2	98.3	81.1	97.4	37	8.2	1:0.2	11	26.8
30	82.9	97.6	77.3	98.3	44	9.7	1:0.3	7	17.1
20	85.4	96.8	72.9	98.5	48	10.6	1:0.4	6	14.6
10	92.7	93.2	57.6	99.2	66	14.6	1:0.7	3	7.3
5	95.1	87.4	42.9	99.5	91	20.1	1:1.3	2	4.9
4*	97.6	85.2	39.6	99.7	101	22.4	1:1.5	1	2.4
3	97.6	82.5	35.7	99.7	112	24.8	1:1.8	1	2.4
2	97.6	76.9	29.6	99.7	135	29.9	1:2.4	1	2.4
1	97.6	67.4	23.0	99.6	174	38.5	1:3.4	1	2.4
0.5	97.6	49.2	16.1	99.5	249	55.1	1:5.2	1	2.4
0.4	97.6	43.1	14.6	99.4	274	60.6	1:5.8	1	2.4
0.3	97.6	36.5	13.3	99.3	301	66.6	1:6.5	1	2.4
0.2	97.5	29.7	12.2	99.2	329	72.8	1:7.2	1	2.4
0.1	100	16.1	10.6	100	386	85.4	1:8.4	0	0

PPV, positive predictive value; NPV, negative predictive value.
*Optimal cut-point for maximizing sensitivity followed by specificity.

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

but will also need to ensure that their patient volume can support sustained expertise. In addition, variability in cytologic expertise in the general pathology community[22] necessitates that those using the procedure have confidence in the cytopathologic results provided to them.

The derivation of BREASTAID uses a multistep approach employing multiple regression analysis in the application of Bayes' theorem. Quantification of the results of triple diagnosis using a less complicated scoring system triple test score (TTS) has also been done, where the total score is used to determine which patients need open biopsy. While TTS performs very well for the authors who originated it, concern has been raised about the applicability of this method in less specialized settings,[27,28] especially because the "clinical impression" component of triple diagnosis, which requires considerable expertise to be accurately applied, remains undefined. In contrast, BREASTAID has defined the clinical examination results and simultaneously incorporates all of the fundamental principles of triple diagnosis. Moreover, it adds to triple diagnosis by incorporating pertinent risk factor information and internal consistency characteristics on FNAB into the interpretation of the test results.

There are potentially negative clinical consequences in following a solid breast mass, rather than subjecting it to immediate biopsy. While most lesions will be benign using either BREASTAID or triple diagnosis, the false negative rate for lesions that are followed will never be zero, and some may argue that every solid breast mass requires open surgical biopsy to avoid diagnostic delay. However, many authors who use triple diagnosis suggest that not every breast mass needs an open biopsy, but that nonbiopsied lesions need follow-up.[9,27,29,30] Unfortunately, the interval between follow-up visits and the duration of follow-up necessary are not clearly defined in the surgical literature. Since medicolegal consequences of a diagnostic delay are dependent on the biology of the primary tumor as well as the length of delay,[31] a 3-month follow-up schedule as suggested in this paper would minimize the risk for both the patient and the physician. In this study, the false negative rate at initial presentation was 7.3% (3 of 41) for triple diagnosis, and 2.4% (1 of 41) for BREASTAID, which seems well within acceptable limits, especially since the cancers missed at initial presentation were all diagnosed at 6-week follow-up, based on persistence of CBE findings despite a decreased hormonal milieu. Using the data from this paper, the authors suggest that a false negative rate following a 6-month visit should be as close to zero as possible and certainly $<1\%$. However, the definition of an acceptable false negative rate is likely to vary among physicians as well as patients and needs to be studied using clinical decision analysis. We are planning such a study in the context of determining the risk/benefit ratio of BREASTAID.

The economic impact of follow-up on a 3-month schedule rather than open biopsy can be estimated from this study, in which 263 of 272 unbiopsied breast masses disappeared within 12 months of clinical follow-up (Fig. 3). A clinical examination of the breast is all that is required on follow-up examination. Reimbursement for three follow-up visits after

TABLE 3. Results of Indeterminate or Less Findings on Mammography and FNAB in 12 Cases of Breast Cancer: Comparison With Final Probability of Cancer Predicted by BREASTAID

Age (yr)	Menopausal Status	Maternal History of Breast Cancer	Lesion Size (cm)	Mammogram	Stage II BREASTAID Probability (%)	Internal Consistency on FNAB	Gritty on FNAB	FNAB Cytology	Stage III BREASTAID Probability (%)
42	Pre	No	1.2	Normal	0.88	Rubbery	No	Benign	0.17
45	Pre	No	1.0	Benign	3.18*	Rubbery	No	Many atypia	4.95†
59	Post	Yes	1.0	Benign	22.40*	Rubbery	No	Normal	5.24†
48	Pre	No	2.0	Indeterminate	26.57*	Rubbery	No	Benign	11.65†
54	Post	No	1.5	Indeterminate	49.06*	Rubbery	No	Hypocellular	12.89†
57	Post	Yes	6.0	Benign	48.04*	Rubbery	No	Normal	15.04†
45	Pre	No	1.2	Indeterminate	26.57*	Soft	Yes	Normal	33.81†
71	Post	Yes	1.0	Benign	21.96*	Rubbery	Yes	Few atypia	34.93†
45	Pre	No	1.5	Indeterminate	26.57*	Rubbery	No	Many atypia	36.46†
68	Post	No	3.0	Indeterminate	68.97*	Soft	No	Hypocellular	39.38†
71	Post	Yes	1.5	Benign	27.56*	Rubbery	Yes	Few atypia	42.06†
57	Post	Yes	1.5	Benign	28.08*	Hard	Yes	Normal	74.37†

*Probability of $>2\%$ an indication for biopsy.
†Probability of $>4\%$ an indication for biopsy.

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

the initial evaluation is estimated to be about $180, compared with $2400 for each open biopsy.

The question of whether patients with a breast mass referred to a specialty clinic are representative of all patients with breast masses is an important one to address. This study documents a prevalence of breast cancer of 9.1%. Most of the surgical literature reports much higher malignancy rates, between 20% and 50%.[9,27–30,32,33] However, most of these studies give rates only for those patients who underwent open biopsy. In our study, the open biopsy malignancy rate was 22.8%. Two studies exist that have reported on consecutive patients seen in specialty clinics like ours.[2,34] In these series, malignancy rates were between 8.2% and 13.2%, similar to our study. It therefore seems likely that the malignancy rate in our study is representative of the population of women with palpable solid breast masses who are referred.

One could convincingly argue that all cases of either suspicious or malignant mammograms or FNAB cytology results should result in open biopsy. Twelve of 41 (29.3%) malignancies in this study had indeterminate or lower findings on both mammography and FNAB. The challenge for the surgeon in such cases is to decide which lesions assessed to be equivocal or less by the diagnostic tests should be biopsied. These decisions often are based on the clinical impression component of triple diagnosis. Table 3 illustrates that stage II and stage III BREASTAID were capable of sorting out this dilemma in 11 of the 12 cases. As expected, the 1 cancer missed by BREASTAID was 1 of the same 3 patients initially missed by the triple diagnosis method. The other 2 cases missed by the triple diagnosis method were triaged to biopsy by both versions of BREASTAID.

BREASTAID requires much more study before it can be used in clinical practice. Goals of future research for BREASTAID first include an internal validation study using boot-strap analysis.[10] This initial evaluation of accuracy and misclassification will then be followed by external validation studies, first using a new group of patients treated in the same setting as the original patient population, and then several groups of patients treated at a wide variety of clinics. Assessing reliability of both individual clinical predictor variables and of BREASTAID as a whole will also be required. Finally, design of a user-friendly device for BREASTAID's calculation and performance of an impact analysis to measure its effectiveness will also be necessary. If the predictive ability of BREASTAID is confirmed and the people for whom it is designed find it useful, BREASTAID has the potential to conserve health care dollars while simultaneously offering both patients and physicians the assurance that clinical follow-up is a safe alternative to open biopsy.

REFERENCES

1. Leis HP Jr. Clinical diagnosis of breast cancer. *J Reprod Med*. 1975;14: 231–240.
2. Seltzer MH. The significance of breast complaints as correlated with age and breast cancer. *Am Surg*. 1992;58:413–417.
3. Reeves M, Newcomb P, Remington P, et al. Determinants of breast cancer detection among Wisconsin (United States) women, 1988–1990. *Cancer Causes Control*. 1995;6:103–111.
4. Seltzer MH. Preoperative prediction of open breast biopsy results. *Cancer*. 1997;79:1822–1827.
5. Burkhardt JH, Sunshine JH. Core-needle and surgical breast biopsy: comparison of three methods of assessing cost. *Radiology*. 1999;212: 181–188.
6. Physician's Insurance Association of America. *Breast Cancer Study*. Physician's Insurance Association of America; 1995, Washington, DC.
7. Vetto J, Pommier R, Schmidt W, et al. Use of the 'triple test' for palpable breast lesions yields high diagnostic accuracy and cost savings. *Am J Surg*. 1995;169:519–522.
8. Layfield LJ. Can fine-needle aspiration replace open biopsy in the diagnosis of palpable breast lesions? [editorial]. *Am J Clin Pathol*. 1992;98:145–147.
9. Ciatto S, Bonardi R, Cariaggi MP. Performance of fine-needle aspiration cytology of the breast: multicenter study of 23,063 aspirates in ten Italian laboratories. *Tumori*. 1995;81:13–17.
10. Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules: applications and methodological standards. *N Engl J Med*. 1985;313:793–799.
11. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. *JAMA*. 1997;277:488–494.
12. McGinn TG, Guyatt GH, Wyer PC, et al. Users' guides to the medical literature: XXII. How to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA*. 2000;284:79–84.
13. Colditz GA, Frazier AL. Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. *Cancer Epidemiol Biomarkers Prev*. 1995;4:567–571.
14. Holleb AI. The technique of breast examination. *CA Cancer J Clin*. 1966;16:7–9.
15. Haagenson C. Physician's role in the detection and diagnosis of breast disease. In: Haagenson C, ed. *Diseases of the Breast*, 3rd ed. Philadelphia: W.B. Saunders; 1986.
16. American College of Radiology (ACR). Illustrated breast imaging reporting and data system (BI-RADS). 3rd edition. Reston [VA]: American College of Radiology. 1998.
17. Horobin JM, Matthew BM, Preece PE, et al. Effects of fine needle aspiration on subsequent mammograms. *Br J Surg*. 1992;79:52–54.
18. Sickles EA, Klein DL, Goodson WH, et al. Mammography after needle aspiration of palpable breast masses. *Am J Surg*. 1983;145:395.
19. Martin H, Ellis E. Biopsy by needle puncture and aspiration. *Ann Surg*. 1930;92:169–181.
20. Roberts JC, Rainsbury RM. 'Tactile sensation': a new clinical sign during fine needle aspiration of breast lumps. *Ann R Coll Surg Engl*. 1994;76:136–138.
21. Kline TS, Kline IK. *Guides to Clinical Aspiration Biopsy: Breast*. New York: Igaku-Shoin; 1989.
22. The uniform approach to breast fine-needle aspiration biopsy. National Cancer Institute Fine-Needle Aspiration of Breast Workshop Subcommittees. *Diagn Cytopathol*. 1997;16:295–311.
23. Weinstein M, Feinberg H. *Clinical Decision Analysis*. Philadelphia: W.B. Saunders; 1980.
24. Sackett DL, Haynes RB, Guyatt GH, et al. *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Boston: Little, Brown; 1991.
25. Metz CE. *ROCKIT*, vol. 2002, ed. version 0. 9B. Chicago: University of Chicago, 1998.
26. Renshaw AA. Accuracy of thyroid fine-needle aspiration using receiver operator characteristic curves. *Am J Clin Pathol*. 2001;116:477–482.
27. Morris A, Pommier R, Schmidt W, et al. Accurate evaluation of palpable breast masses by the triple test score. *Arch Surg*. 1998;133:930–934.
28. Morris KT, Pommier RF, Morris A, et al. Usefulness of the triple test score for palpable breast masses; discussion 1012–1013. *Arch Surg*. 2001;136:1008–1012.
29. Di Pietro S, Fariselli G, Bandieramonte G, et al. Diagnostic efficacy of the clinical-radiological-cytological triad in solid breast lumps: results of a second prospective study on 631 patients. *Eur J Surg Oncol*. 1987;13: 335–340.

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.

30. Martelli G, Pilotti S, Coopmans de Yoldi G, et al. Diagnostic efficacy of physical examination, mammography, fine needle aspiration cytology (triple-test) in solid breast lumps: an analysis of 1708 consecutive cases. *Tumori*. 1990;76:476–479.
31. Osuch J, Bonham VL. Medicolegal pitfalls in breast cancer diagnosis and management. In: Jatoi I, ed. *Manual on Breast Diseases*. Philadelphia: Lippincott Williams & Wilkins, 2002;519–538.
32. Hermansen C, Skovgaard Poulsen H, Jensen J, et al. Palpable breast tumours: 'triple diagnosis' and operative strategy. Results of a prospective study. *Acta Chir Scand*. 1984;150:625–628.
33. Kreuzer G, Boquoi E. Aspiration biopsy cytology, mammography and clinical exploration: a modern setup in the diagnosis of tumors of the breast. *Acta Cytol*. 1976;20:319–323.
34. Horgan PG, Waldron D, Mooney E, et al. The role of aspiration cytologic examination in the diagnosis of carcinoma of the breast. *Surg Gynecol Obstet*. 1991;172:290–292.

© 2003 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.