Accurate Evaluation of Palpable Breast Masses by the Triple Test Score

Arden Morris, MD; Rodney F. Pommier, MD; Waldemar A. Schmidt, MD; Richard L. Shih; Priscilla W. Alexander, MD; John T. Vetto, MD

Background: We previously reported that the triple test (physical examination, mammography, and fine needle aspiration) for palpable breast masses yields 100% diagnostic accuracy when all 3 components are concordant (all benign or all malignant). However, 40% of cases are nonconcordant and require open biopsy.

Objective: To evaluate our experience with the triple test to develop a method to further limit the need for surgical biopsy.

Design: Diagnostic test study.

Setting: University hospital multidisciplinary breast clinic.


Intervention: The triple test was prospectively applied to each breast mass. Each component of the triple test was assigned 1, 2, or 3 points for a benign, suspicious, or malignant result, respectively, yielding a total triple test score (TTS).

Main Outcome Measures: The TTS was correlated with subsequent histopathologic examination results.

Results: Eighty-eight masses had a TTS of more than 6 points; all had malignant histopathologic characteristics. One hundred fifty-two masses had a TTS of 4 points or lower; all were benign. In both groups, diagnostic accuracy and predictive value were 100%, with \( P < .001 \). Twenty-one masses had a TTS of 5 points; of these, 13 (62%) were benign and 8 (38%) were malignant.

Conclusions: The TTS reliably guides evaluation and treatment of palpable breast masses. Masses that score 6 points or higher are malignant and should undergo definitive therapy; masses that score 4 points or lower are benign and may be clinically followed up. Only those masses that score 5 points (8% of our database) require open biopsy.
PATIENTS AND METHODS

Between October 1, 1991, and June 30, 1997, patients referred to our multidisciplinary breast clinic for evaluation of palpable breast masses were prospectively recruited for this diagnostic test study. Patients with palpable breast masses underwent temporally congruent physical examination, mammography, and fine needle aspiration prior to open biopsy for histopathologic confirmation or clinical follow-up. Patients with lesions for which all 3 elements of the TT were not performed or were not clearly recorded were excluded. Patients who underwent ultrasonography instead of mammography were also excluded.

Each element of the TT was classified as “benign,” “suspicious,” or “malignant” for each evaluated lesion, prospectively and independently. Physical examination results were determined by 2 skilled clinicians (R.F.P. and J.T.V.), mammography films were interpreted by 2 experienced mammographers (P.W.A.), and fine needle aspiration biopsy specimens were initially read by a single cytopathologist (W.A.S.), and later reviewed by both this same physician and by a second cytopathologist. After determining that the diagnostic accuracy of a concordant TT at our institution was 100%, lesions with a concordant benign TT were no longer subjected to open biopsy in a routine fashion but were followed up clinically.7

To better evaluate nonconcordant lesions, the TT was modified by assigning a score of 1, 2, or 3 points for a benign, suspicious, or malignant result, respectively. Individual element scores were then added together to yield a total TT score (TTS) for each lesion. This system results in a minimum score of 3 for a concordant benign test result and a maximum score of 9 for a concordant malignant test result. The TTS was correlated with subsequent histopathologic examination results, except in 112 lesions with benign concordant TT results (TTS = 3), for which clinicians and patients chose close follow-up based on multiple earlier studies.1-7

Accuracy rates for the TTs were then determined from standard formulas, as in previous reports: sensitivity = TP/(TP + FN); specificity = TN/(TN + FP); positive predictive value = TP/(TP + FP); negative predictive value = TN/(TN + FN); and accuracy = (TP + TN)/(TP + TN + FP + FN), wherein TP indicates true positive; TN, true negative; FP, false positive; and FN, false negative.

RESULTS

Two hundred fifty-nine patients with 261 palpable breast mass lesions were entered into the study. Two patients had 2 synchronous lesions and each lesion was subjected to an independent TT. Mean patient age was 49 years. Seventy-nine percent of patients presented with a chief complaint of a mass, 16% with an abnormal mammography result, 3% with pain, 1% with skin dimpling, and 1% with nipple discharge.

One hundred fifty-two masses had a TTS of 4 points or less. All masses with TTS of 4 had benign histopathologic characteristics on subsequent open biopsy (n = 17). Twenty-three masses with a TTS of 3 also had benign histopathologic characteristics, while the remaining 112 masses with a TTS of 3 were clinically followed up for a mean duration of 18 months (n = 112), via biannual physical examination and mammography. Thus, all 152 masses with a TTS of 4 or less were benign on open biopsy or clinical follow-up.

Eighty-eight masses had a TTS of 6 points or higher; all had malignant histopathologic characteristics on subsequent open biopsy.

Eighty-one masses (8% of total studied) had a TTS of 5. Of these, 13 (62%) were found to have benign histopathologic characteristics, while 8 (38%) had malignant characteristics on subsequent open biopsy.

Sensitivity was defined as the percentage of lesions in which histologically proven cancer was correctly predicted by the TTs, and specificity as the percentage of lesions in which benign lesions were correctly predicted by the TTs. Positive predictive value was defined as the percentage of lesions with a TTS of 6 or higher and subsequently ascertained to have malignant histopathologic characteristics; negative predictive value reflected the percentage of lesions that had a TTS of 4 or lower and were subsequently determined to be benign.

χ2 Analysis for masses with a TTS of 6 points or higher and a TTS of 4 points or lower revealed that diagnostic accuracy and positive and negative predictive value of histology were 100%, with P<.001 (Table).

COMMENT

The goals of the TT are to avoid expensive, potentially morbid negative open biopsies when no malignancy is present, and to allow the patient and clinician to proceed directly to definitive therapy without an intervening open biopsy if malignancy is present. Safely reducing the number of surgical breast biopsies can save patients the discomfort of undergoing open biopsy and eliminate unnecessary expenditures.

©1998 American Medical Association. All rights reserved.
Multiple previous studies have shown that a mass with a concordant benign TT result can be safely followed up without open biopsy.1-7 This is especially helpful in evaluating low-risk masses or multiple masses in a young patient with, for example, fibrocystic breast changes. Similarly, we have found that a patient with a concordant malignant TT result can safely proceed to definitive therapy, without undergoing an open biopsy.2 Thus, a concordant TT result can guide treatment with 100% accuracy.

However, patients with masses that have nonconcordant TT results (and would therefore still require open surgical biopsy based on these considerations) comprise more than 40% of our clinical experience. In an effort to better interpret the nonconcordant TT result and further limit the need for open surgical biopsy, we have developed a TT scoring system composed of the sum of the 3 components of the TT individually assigned 1, 2, or 3 points for a benign, suspicious, or malignant result, respectively. Mindful of the tendency among some practitioners to mitigate the unpleasant term “malignant” by substituting “suspicious,” our investigators rigorously labeled only indeterminate tests “suspicious” and otherwise used the categories “benign” or “malignant.”

In examining the TT elements individually, we note that fine needle aspiration is typically more accurate than physical examination or mammography (Table),7-12 yet each is given the same weight in the TTS. We decided against calculating a weighted scoring system, as we wished to derive a test that would be clinically easy to apply.

Eighty-eight masses (representing 34% of the patients in this study) had a TTS of 6 points or higher, which predicted subsequent malignant histopathologic characteristics with 100% accuracy (P<.001). Patients with these masses may undergo definitive therapy, as intervening open biopsy becomes superfluous. Confirmatory frozen section biopsy, prior to surgical resection, can be performed in masses with a TTS lower than 9. Our previous charge analyses7,8 have indicated that this practice does not negate the reduction in charges achieved by using the TT.

One hundred fifty-two masses (representing 59% of patients enrolled) had a TTS of 6 points or lower. Of these masses, 40 underwent confirmatory open biopsy, while 112 with a TTS of 3 were followed up clinically, for a mean duration of 18 months. All were benign on biopsy or on clinical follow-up (P=.001). All 17 masses with a TTS of 4 (by definition, 2 benign TT elements and 1 suspicious TT element) underwent confirmatory open biopsy and were benign. Based on these results, we recommend that patients whose masses have a TTS of 4 or lower can be safely observed, avoiding open biopsy, with the caveat that clinical follow-up of these lesions is mandatory.

The above recommendation implies that a mass with only 1 suspicious finding, including the possibility of a suspicious fine needle aspiration result associated with benign physical examination and mammography results, would remain unbiopsied. A falsely suspicious fine needle aspiration result can occur in the presence of a fibroadenoma, ductal papilloma, fat necrosis, mastitis, gynecomastia, postradiation changes, and epithelial proliferation, leading some authors to recommend avoiding fine needle aspiration altogether in the event of pregnancy or lactation.9,10 Our study contained 8 masses (3%) with benign physical examination and mammography results and suspicious fine needle aspiration results. However, a TTS of 4 for these masses uniformly correlated with benign histopathologic characteristics. Sixty-five patients had a TTS of 6, 7, or 8 points. All had malignant characteristics on open biopsy. Thus, the TTS can enable clinicians to discriminate between nonconcordant masses that are benign and those that may be malignant.

The TT is only unable to differentiate benign from malignant results for masses with a TTS of 5. Twenty-one masses (8% of patients) had a TTS of 5. Of these, 13 (62%) were benign and 8 (38%) were malignant by histopathologic examination. Therefore, we believe that the small group of masses with a TTS of 5 still require surgical biopsy for histopathologic diagnosis, although we continue to look for other ways to better characterize these lesions.

In the current health care climate, economic considerations remain compelling in every hospital setting. In the initial TT study, we reported a charge reduction for each breast mass of $825 using the TT to avoid outpatient surgical biopsy and $2001 for avoiding an inpatient open biopsy (average, $1412 per mass).3,7 Based on current costs, eliminating the charge for outpatient open biopsy of masses with concordant TT results yields a potential charge reduction of $49 500 per 100 patients evaluated. The TTS modification we describe here (no open biopsy for a TTS other than 5) would increase this figure to $75 900 per 100 outpatients evaluated, without missing any malignant neoplasms.

In summary, the TTS reliably guides evaluation and treatment of palpable breast masses, enhancing the value of the original TT. Masses that score 6 points or higher are malignant and should undergo definitive therapy; masses that score 4 points or lower are benign and may be clinically followed up. Only those masses that score 5 points require open biopsy. Modification of the standard TT by use of the TTS reduces the need for open biopsy from 40% to only 8% of our patients, with resultant greater potential decrease in patient charges and higher diagnostic accuracy.

Presented at the 69th Annual Session of the Pacific Coast Surgical Association, Maui, Hawaii, February 16, 1998. We thank Gary Sexton, PhD, for verification of statistical analysis.

Corresponding author: Rodney F. Pommier, MD, L223, 3181 SW Sam Jackson Park Rd, Portland, OR 97201-3098 (e-mail: pommierr@ohsu.edu).

REFERENCES

4. Butler JA, Vargas HI, Worthen N, et al. Accuracy of combined clinical-

**DISCUSSION**

**John A. Butler, MD, Orange, Calif:** Over the past 20 years, fine needle aspiration (FNA) cytology has been increasingly utilized in the diagnostic workup of women who present with palpable breast masses. The combination of physical examination, mammography, and FNA cytology has yielded multiple reports documenting the efficacy of using these tests to either safely watch a benign lesion or proceed to definitive treatment for a malignant mass. The authors of the current paper have contributed to that literature.

In today’s presentation, Drs Morris, Pommier, and Vetto attempt to extend the usefulness of this “triple test” in an effort to further reduce the number of patients requiring open biopsy for definitive diagnosis. In specifically addressing those cases in which the results of the 3 tests are nonconcordant, the authors’ results suggest the application of a triple test scoring system can safely obviate the need for open biopsy in 80% of those patients with discordant results. In their hands, this reduced the need for open biopsy to only 8% of the 260 patients prospectively evaluated.

Although their conclusions accurately reflect the results, there are several concerns I would have in terms of the broad application of this algorithm for managing patients with palpable breast masses. It is universally accepted that FNA cytology is clearly superior in overall accuracy when compared with mammography and physical examination. The Oregon data support this thesis, but the test score, in the interest of simplicity, has not been weighted to reflect this discrepancy. On the basis of this study, a test score of 4, with a suspicious FNA result, can be safely observed. This recommendation is made on the basis of a sample of 8 patients, which represents only 3% of the entire group.

In a similar study that we performed 10 years ago at Harbor UCLA, we had 18 cases with a single suspicious cytology result, and 1 of those 18 turned out to be a cancer. This problem is particularly relevant in young women with small masses in which both mammography and physical examination may be limited from the standpoint of diagnostic accuracy. I would ask the authors to comment on the strength of their conviction in this situation.

The other area where I have a major concern is the recommendation to proceed with definitive surgery in those patients with a test score of 6. Patients with only suspicious results in all 3 categories and patients whose masses are malignant on mammographic and physical examination but benign on FNA are potential candidates for this group. Lesions such as radial scars and fat necrosis may mimic malignancy on both physical examination and mammography, and I would strongly argue that definitive therapy be pursued only with a malignant cytologic examination.

I would also ask the authors to address the question of follow-up in those patients who are being observed. Are patients with a test score of 4 relegated to routine yearly follow-up, or should they have an interval examination to document the stability of the mass in question?

Finally, is there an age below which you would forgo mammography and rely solely on physical examination and FNA to guide further therapy?

I congratulate the authors on a well-done, large prospective study that will serve as an impetus for all of us to review our management of these patients. I would caution the audience, however, that these results reflect the work of an outstanding, dedicated, multidisciplinary group of surgeons, mammographers, and cytopathologists whose accuracy in each of the individual tests is among the best reported in the literature. It is imperative that individual institutions assess their own accuracy before embarking on a similar program.

**Lawrence A. Danto, MD, Davis, Calif:** Would the authors please say a few words about compliance? How do your patients with a dominant mass tolerate follow-up without excision?

**Howard Silberman, MD, Los Angeles, Calif:** I wonder if you could define what a lump in the breast is. Are women with lumpy, bumpy breasts considered to have a breast mass? Are they all subjected to FNAs? Are we talking about dominant lumps, and what are the features that allow you to categorize a mass of the breast as benign in your point system?

**Theodore X. O’Connell, MD, Los Angeles:** How exportable is this? Because a lot of the tests are subjective. Obviously the physical examination is subjective, and it may work in Dr Butler’s hands or with the people in Oregon. But can that be done by people with varying amounts of experience? Even mammograms with false positives and false negatives can vary by who is doing it, and certainly FNAs if you have an experienced cytopathologist can be very accurate, but if you have one who doesn’t have the experience, this could be detrimental. So how well can this be exported from Oregon back to California or elsewhere?

**F. Don Parsa, MD, Honolulu, Hawaii:** How do the authors follow patients who have breast implants? Do they proceed with FNA, or do they have alternative ways of following them?

**Dr Vetto:** As you have already heard, Dr Butler was a pioneer in this area, presenting the first American data in 1990 at the Western Surgical Association. At that time he made the startling recommendation that we reduce the use of the criterion standard, open surgical biopsy, from 100% to 85% by not doing biopsies for the 15% of the patients in his database who had concordant negative triple tests. As he and Dr Wilson have already alluded to, that was not met with great enthusiasm. In 1995 we recommended that we reduce that number further to 40%. Our task then was a little easier and we did not meet with quite the same resistance because 3 things have happened between 1990 and 1995. First, we see more younger patients in the breast clinics now, which increases the likelihood that the open biopsy will be benign (up to 97% of open biopsies are benign in younger women). Secondly, we have a new paradigm of breast cancer that emphasizes systemic over local control. And finally we were introduced between 1990 and 1995 to nonsurgical biopsies, such as stereotactic, so surgeons had more time to get used to this concept.

Nonetheless, today we have the daunting task of convincing the audience that we can reduce the use of the criterion standard to 8% and, therefore, we anticipated the general theme of
Dr Morris showed, of the 88 patients with a score of 6 or greater, all would have received a triple test score of 5 and would have gone on to definitive therapy, and the cost of frozen section is already built in to the charge reductions that you saw on our slides. Thus, we are not particularly worried about scores of 6 because they come to definitive therapy, and the cost of frozen section is already built in to the charge reductions that you saw on our slides. Therefore, if anything, the triple test score errs in the direction of safety, ie, it errs in the direction of falsely positive FNAs.

Dr Butler's third question was about young patients; can they forgo mammography? We apply ultrasound to women younger than 40 and have published a separate paper demonstrating that one can modify the triple test using ultrasound instead of mammography. Dr Danto asked about compliance on follow-up, which brings us to the general issue of follow-up. If the test is a 3 or 4 and the patient has a regular provider, we send them back to the provider for routine screening and this works very well in most managed care settings. If the patient has a 5, they will go on to biopsy, and we will send them either to ourselves or to the provider, depending on the result. If they have a 6 or greater, they become the “property” of the surgical oncologist. So follow-up and compliance are generally not a problem with this test.

Another question asked involved how we define “lump.” Every patient sent to the clinic with the diagnosis of “lump” is taken seriously. I agree that “lump” is hard to define, so we ask the provider and the patient where is the dominant area of concern, and we focus our attention on that area, because that was the purpose of the referral to our clinic.

Dr Parsa asked about following patients with implants. We use either ultrasound or MRI to determine whether the implant is ruptured; that is always the first question in a patient who has an implant and a lump. If there is no evidence of rupture, then we do the FNA under ultrasound to protect the implant.

Regarding the question on stereotactic breast biopsy (STBB), we reserve this for nonpalpable lesions, and therefore it does not “fit” into the clinical setting in which we use the triple test score (palpable breast masses). I am aware that some radiologists use STBB for palpable lesions, but we consider this practice an overuse of the technology.

Finally, I want to close with Dr O’Connell’s question about how we eliminate subjectivity by telling you exactly how we do this scoring system because, yes, I think it can and should be extrapolated to other breast clinics. We begin with the physical examination, and we try to do this in a blinded fashion looking directly at the lump, sometimes not even reviewing the mammograms first. Then we go to the imaging; in this area I think that the scoring has gotten simpler with the advent of the BIRADS system, which some of you are already familiar with. The BIRADS system translates directly to the triple test score; ie, mammograms scored with BIRADS of 1 or 2 get a triple test score of 1; mammograms with a BIRADS score of 3 or 4 get a triple test score of 2; and mammograms with a BIRADS score of 5 get a triple test score of 3. Thus, the mammogram interpretation has become less subjective in the last few years with the introduction of BIRADS. Further, I am happy to report that with new national legislation, BIRADS will be mandated for any mammogram purchased with federal monies. That is a good start in eliminating the English language from the interpretation of mammograms, which is something that I think all surgeons would welcome. And finally, the cytopathologist is asked to simply state whether his or her interpretation is clearly benign, clearly malignant, or suspicious.

We ask each of the 3 examiners (surgeons, cytopathologists, and radiologists) to not use the term “suspicious” when they mean “malignant”—to truly reserve the term “suspicious” for something in-between. This is where the multidisciplinary setting comes into play, because when you get back to your institutions and you want to try this, I encourage you to keep after your cytopathologists and radiologists to get in the habit of making this commitment. You will find that with encouragement they actually enjoy being asked to categorize their impressions into these 3 categories. The result will be, I think, a truly satisfying application of the multidisciplinary approach to the care of patients with breast lesions.

---

Surgical Anatomy

The left gastric artery is the smallest of the three branches of the celiac artery and the largest of the five arteries to the stomach. The splenic artery is the largest branch and passes through the lienorenal ligament and enters the hilum of the spleen as several branches. The hepatic artery takes a more sover course to the right along the upper border of the pancreas and, like the left gastric artery, creates a little mesentery which conducts it in front of the portal vein to the lesser omentum.