Should Selection of Adjuvant Chemotherapy for Patients With Breast Cancer Be Based on erbB-2 Status?

Gary M. Clark

It has been more than 10 years since the initial report of an association between amplification of erbB-2 (also known as HER-2/neu and ERBB2) and poor clinical outcome of patients with primary breast cancer (1). Dozens of papers involving thousands of patients have now been published on this topic, and the general conclusion is that erbB-2 abnormalities, either gene amplification or protein overexpression, are associated with worse prognosis of patients with lymph node-positive breast cancer, but the relationship for patients with lymph node-negative disease is weak, at best. Since most of these correlative studies were retrospective and many of the patients were treated before adjuvant therapy was widely administered to patients with lymph node-negative disease, several clinical researchers speculated that these associations were due, in part, to specific adjuvant therapies received by lymph node-positive patients.

The first studies that examined the potential role of erbB-2 status for predicting response to adjuvant chemotherapy concentrated on regimens that contained CMF (cyclophosphamide, methotrexate, and 5-fluorouracil) (2,3). The general trend reported in these and subsequent studies is that patients whose tumors have little or no detectable levels of erbB-2 derive considerable benefit from CMF regimens, but patients whose tumors have amplified erbB-2 genes or overexpress erbB-2 protein do not benefit from these therapies. However, it must be stressed that these conclusions are based on retrospective analyses, and they have not been validated in prospective studies with sufficient statistical power to detect interactions between treatment and erbB-2 status.

In 1994, Muss et al. (4) published, to our knowledge, the first analysis of an interaction between expression of erbB-2 and adjuvant therapy with doxorubicin-containing regimens. The finding that tumors with high expression of erbB-2 responded better to dose-intensive treatment with CAF (cyclophosphamide, doxorubicin [Adriamycin], and 5-fluorouracil) was contrary to the previous results with the CMF regimens and spurred international efforts to validate or refute the published data and to understand the biologic mechanisms that might explain such results. But such validation studies are difficult to perform, and analyses of interactions between treatments and potential predictive factors can be accomplished only within a randomized clinical trial. Without the protection of randomization, selection biases will result in confounding effects between treatments administered and potential predictive factors that cannot be overcome with the most sophisticated multivariate statistical analyses.

A fact not often appreciated is that detection of an interaction effect requires many more patients and events (i.e., relapses or deaths) than detection of treatment effects when no interaction is present. It can take four times the number of failures to detect an interaction effect of the same magnitude as a treatment main effect when no interaction is present at the same level of type 1 and type 2 errors (significance level and power, respectively) (5). The net result is that randomized trials designed to detect treatment main effects have only limited statistical power to detect interaction effects. This problem is compounded if retrieval of archival tissue, required to perform assays of the biomarker under investigation, is incomplete.

Large, multicenter, cooperative group trials provide the best opportunities for addressing these questions. In this issue of the Journal, results from retrospective analyses of randomized clinical trials conducted by the CALGB (Cancer and Leukemia Group B) (6) and the NSABP (National Surgical Adjuvant Breast and Bowel Project) (7) are reported regarding relationships between erbB-2 expression and response to doxorubicin-based adjuvant therapy administered to patients with lymph node-positive breast cancer.

The CALGB study was designed as a validation of their previous published results (4). The original trial randomly assigned 1572 women to receive one of three regimens of adjuvant CAF to address questions about the efficacy of dose intensity (8). The initial retrospective study that generated the hypothesis about an association between erbB-2 expression and dose response to chemotherapy included 397 (25%) of these patients (4), and the validation study reported in this issue of the Journal (6) included an additional 595 patients (38%). It was disappointing that the hypothesized interaction between erbB-2 expression and dose of CAF was not statistically significant, although the trend was in the same direction as previously reported. Subsequent subset analyses that adjusted for an apparent failure of randomization did produce the hypothesized result. While this study suggests that patients whose tumors overexpress erbB-2 might benefit from dose-intensive CAF, the authors correctly concluded that the hypothesis should be further validated before clinical implementation.

The NSABP study was a retrospective analysis of patients enrolled on protocol B-11, a trial originally designed to compare PF (L-phenylalanine mustard plus 5-fluorouracil) with PAF (L-
phenylalanine mustard plus doxorubicin plus 5-fluorouracil). Unlike the CALGB study, this trial provided a comparison of a treatment regimen plus or minus doxorubicin and, therefore, a direct test of an interaction between doxorubicin and erbB-2 expression. It is remarkable that archival tissue was obtained for 638 (93.5%) of 682 eligible patients. The results clearly indicate that patients with erbB-2-negative tumors have the same clinical outcomes with or without doxorubicin. Patients with erbB-2-positive tumors who did not receive doxorubicin had a significantly worse prognosis, but the addition of doxorubicin improved clinical outcomes such that they were equivalent to those experienced by patients with erbB-2-negative tumors. Formal tests of interactions between doxorubicin and erbB-2 expression were statistically significant for disease-free survival ($P = .02$) and distant disease-free survival ($P = .02$) and not for overall survival ($P = .15$) or recurrence-free survival ($P = .06$).

These two studies, together with other studies that have recently been presented at national meetings (9) and preclinical data (10), strongly suggest that the hypothesized interaction between erbB-2 expression and adjuvant doxorubicin therapy is real. If indeed this is the case, then perhaps the time has come to standardize our assays and scoring systems for determining erbB-2 status.

Many years ago when estrogen receptor (ER) assays were first introduced, the scientific community agreed that a quality-control procedure needed to be established to assure clinicians and patients that ER values would be comparable between laboratories. Today, we need to determine which technique is best for assessing erbB-2 status. Should we report gene amplification, protein expression, or messenger RNA levels? Should gene amplification be determined by Southern blot analysis, differential polymerase chain reaction (PCR), fluorescence in situ hybridization, or some other technique? Should erbB-2 protein expression be evaluated by immunohistochemistry, western blot analysis, enzyme-linked immunosorbent assay, or other techniques? Immunohistochemistry has been the mostly widely used technique, but do all antibodies give the same results? What scoring systems should be used?

The primary erbB-2 assays used in each of the studies reported in this issue of the Journal were immunohistochemical, but the antibodies and the scoring systems were different. The CALGB used the CB11 monoclonal antibody and analyzed the results as a continuous variable. To display survival curves for patients with erbB-2-positive and erbB-2-negative tumors, erbB-2 positivity was defined as 50% or more cells showing positive expression. With this definition, 27% of the tumors were erbB-2 positive. In contrast, only 17% of the cases exhibited gene amplification as determined by differential PCR. The correlation between erbB-2 expression and amplification was statistically significant ($P<.001$) but only moderately strong ($r = .51$). In head-to-head analyses, erbB-2 expression was a better predictor of clinical outcome than was gene amplification.

The NSABP used a cocktail of two antibodies, i.e., mAb-1 (a mouse monoclonal antibody) and pAb-1 (a polyclonal rabbit antiserum). Any staining was regarded as positive, yielding an erbB-2-positivity rate of 37.5%.

In a recently reported study (9) from the Southwest Oncology Group (SWOG) that compared tamoxifen plus or minus CAF, erbB-2 status was determined by using both CB11 and mAb-1, separately, and each antibody produced a 16% overexpression rate. However, mAb-1 was a better predictor of response to therapy. Thus, there remains uncertainty about the best definition of erbB-2 overexpression.

It appears that the proof of principle that response to doxorubicin-based adjuvant therapy depends on erbB-2 status has been achieved. But before we begin to routinely select adjuvant chemotherapy for patients with breast cancer on the basis of erbB-2 status, we need a consensus about the most reliable, most reproducible, and most predictive method to determine erbB-2 status.

**REFERENCES**


Is There a Downside to Elderly Women Undergoing Screening Mammography?

Rebecca Smith-Bindman, Karla Kerlikowske

Whether the potential benefits of screening mammography outweigh the harms for elderly women is unknown. Randomized controlled trials of screening mammography either did not include women of age 70 years or older or included too few to provide meaningful results (1,2). The findings reported by Welch and Fisher (3) in this issue of the Journal provide evidence of a potential harm in mammography screening in elderly women. These authors found that additional downstream testing following screening mammography occurs in a considerable number of women of age 65 years or older enrolled in Medicare, the majority of whom do not have breast cancer. For every 1000 women screened, there were 83 abnormal mammogram results, 98 additional tests were performed, and nine cancers were detected. For most women, additional tests included diagnostic mammography (61 women) or breast ultrasound (23 women). Breast biopsy rates were somewhat less common than diagnostic imaging tests with 25 excisional biopsies for every 1000 women screened. The rates of additional diagnostic testing and positive predictive value of mammography were similar to those reported for elderly women by others (4,5).

Does downstream testing matter? Each woman differs in her willingness to undergo the possible consequences of a false-positive mammogram (i.e., psychological stress, invasive follow-up procedures, or morbidity from tests), given the very small probability that she will be one of the women who averts a breast cancer death as a result of routine screening. While some women may easily tolerate the additional tests that are recommended following an abnormal screening result, others may feel considerable psychological stress. Elderly women who are bothered by medical tests, visits to doctors, or the discomfort of undergoing mammography or those who experience substantial anxiety waiting for test results and are willing to accept a small risk of breast cancer might rationally defer screening. Downstream testing would seem particularly important to those elderly women who have little chance of benefiting from screening mammography because existing comorbid conditions make their chance of dying of diseases other than breast cancer much higher. Elderly women with three or more comorbid conditions (i.e., hypertension, diabetes, arthritis, history of myocardial infarction, stroke, respiratory disease, or other types of cancer) are 20-fold more likely to die of a cause other than breast cancer within 3 years, regardless of the stage of the breast cancer at diagnosis (6). Downstream testing also adds to the costs of screening mammography. As much as a third of the costs of a screening program arises from the evaluation of screen-detected abnormalities in women without breast cancer (4,7).

Welch and Fisher (3) also evaluated the amount of time it takes for downstream testing to be initiated in elderly women. Additional imaging generally occurred within 20 days of an abnormal screening result, and breast biopsies generally occurred within 30 days. A short time period of 20–30 days to initiate a diagnostic evaluation is unlikely to have an impact on stage of disease at detection or choice of treatment. However, the waiting time may actually feel considerably longer for women wondering whether or not they have breast cancer and may provoke much anxiety and worry (8,9). Educating elderly women who request or are offered screening mammography that most elderly women (92% of women of ages 65–69 years and 86% of women of age 70 years or older) (3) who have an abnormal screening result do not have cancer may alleviate some of the anxiety and worry.

For 40% of women, diagnostic testing occurred at approximately 6 months after the initial screening examination. This likely corresponds to a 6-month follow-up mammographic examination, an examination frequently used to further evaluate probably benign abnormalities (10). Recommending a 6-month follow-up examination to further evaluate probably benign abnormalities leaves a large number of women with diagnostic uncertainty for an extended period of time. Additional research is needed to determine whether the very low yield of small invasive cancers and ductal carcinoma in situ (DCIS) (10) for 6-month follow-up examinations outweighs the costs and anxiety that these examinations may provoke.

One potential harm of screening elderly women not addressed in the study by Welch and Fisher (3) is the identification of large numbers of cases of clinically insignificant lesions. Screening mammography tends to discover early cancers that may never have produced symptoms. The best example of this is DCIS. The incidence of DCIS increases with age (11), with 25% of mammographically detected cancers being DCIS among elderly women (5). The natural history of DCIS is unknown (in particular, the natural history of small mammographically detected lesions). Given that the natural history of DCIS is unknown but that some lesions will progress to invasive cancer, the vast majority of DCIS lesions are treated by some form of surgery (11). Identifying and treating DCIS lesions in elderly women are unlikely to have an impact on life expectancy, since, if DCIS progresses, it does so slowly, and the risk of death from DCIS progressing to invasive breast cancer is very low (12). Thus, it was concerning to see the prevalence of mammography use in the eldest women studied (3). Up to 26% of women...
older than age 80 years obtained screening mammography (3), despite its unproven benefit and likely marginal impact on life expectancy (13). An 80-year-old woman has an average life expectancy of 9 years, and an 85-year-old woman has an average life expectancy of only 6 years (14). For women aged 80 years and older who undergo screening mammography, early detection will increase the rate of surgical treatment of clinically insignificant lesions, with little hope of having an impact on overall mortality, given their short life expectancy and high risk of death from cardiovascular disease.

Decreasing the false-positive rate would help reduce the psychological and economic costs of screening mammography. Physicians need to inform elderly women that, if they choose to undergo screening mammography, about 8% will have an abnormal result that will require additional evaluation but that the vast majority (86%–92%) of these abnormal results do not represent cancer (3). Furthermore, about 25% of elderly women will have at least one abnormal result if they are screened regularly over a 10-year period (4,15). Elderly women who request or are offered screening mammography should be informed of the likely consequences of undergoing screening and that the benefits for women of age 70 years or older are unproven.

REFERENCES


NOTE

Supported in part by Public Health Service Breast Cancer Surveillance Consortium cooperative agreement U01CA63740 and Breast Cancer SPORE grant P50CA8207 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.