

Uses and Abuses of Tumor Markers in the Diagnosis, Monitoring, and Treatment of Primary and Metastatic Breast Cancer

N. LYNN HENRY, DANIEL F. HAYES

Department of Internal Medicine, Breast Oncology Program,
University of Michigan Comprehensive Cancer Center, Ann Arbor, Michigan, USA

Key Words. Tumor marker • Her-2/neu • Breast cancer • Estrogen receptor • Prognostic factor • Predictive factor

LEARNING OBJECTIVES

After completing this course, the reader will be able to:

1. Identify the difference between prognostic and predictive factors.
2. Describe the evaluation of tumor markers, including their utility, magnitude, and reliability.
3. Discuss issues surrounding tumor markers currently in routine use in breast cancer, including estrogen receptor, Her-2/neu, and Oncotype DX™.

CME

Access and take the CME test online and receive 1 AMA PRA category 1 credit at CME.TheOncologist.com

ABSTRACT

Although breast cancer incidence continues to increase, mortality has been decreasing, principally as a result of earlier detection and improvements in adjuvant systemic therapy. Nonetheless, because antineoplastic agents are associated with substantial morbidity and occasional mortality, efforts to individualize treatment strategies are desirable. In addition to classic histopathologic diagnosis, molecular and cellular tumor markers may help in establishing prognosis or prediction of benefit.

Recommendations for routine use of tumor markers in breast cancer have been conservative. Although several studies have been reported, few are of sufficiently high level of evidence to permit solid conclusions. Three key issues in tumor marker evaluation are utility, magnitude, and reliability. Poorly conceived

study designs cloud the issue of how the marker might be used. Reliance on *p*-values rather than the size of the differences in outcome between patients who are positive and those who are negative for the factor obscures the importance. Technical issues result in poor reproducibility and interpretability of assays. Analytical issues lead to poorly defined cutoff values for marker levels. Poor patient selection leads to difficulty interpreting results because of confounders such as differences in treatment regimens. This review focuses on these issues, with an emphasis on currently accepted tumor markers. Finally, new tumor marker reporting recommendations are discussed, the adoption of which may lead to improved design and publication of tumor marker studies in the future. *The Oncologist* 2006;11:541–552

Correspondence: Daniel F. Hayes, M.D., University of Michigan Comprehensive Cancer Center, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109, USA. Telephone: 734-615-6725; Fax: 734-615-3947; e-mail: hayesdf@umich.edu Received February 22, 2006; accepted for publication April 11, 2006. ©AlphaMed Press 1083-7159/2006/\$20.00/0

INTRODUCTION

Breast cancer is the most common cancer in women in the U.S., with an estimated 213,000 new cases diagnosed in the U.S. in 2005 [1]. Despite these increasing numbers, mortality from breast cancer continues to decline. This decline is felt to be a result of a combination of earlier detection of disease as a result of screening and improved treatment of disease with adjuvant systemic therapies [2, 3]. A majority of patients are cured with surgery and radiation therapy alone, and these patients will gain no additional benefit from adjuvant systemic therapies. In addition, having a high risk for recurrence does not imply that systemic therapy will prevent it. Even for those who recur, overall survival and palliation of symptoms for patients with metastatic breast cancer (MBC) has improved with the advent of new therapies. However, currently available methods are inadequate to help the clinician precisely predict a priori which patients will benefit from many of the available therapies.

For patients with early-stage breast cancer, it would be helpful to identify which patients will relapse without adjuvant systemic therapy, so that only patients who receive benefit are exposed to the inherent toxicities. Approach to treatment of metastatic disease is generally with palliative intent rather than for cure. In this setting, identification of those patients with rapidly progressive disease permits selection of more rapidly acting but perhaps more toxic therapy.

During the past few decades, with the explosion of molecular technology and understanding of the biology of breast cancer, numerous studies have been performed to identify prognostic and predictive factors in breast cancer, with mixed success. Multiple expert panels have convened to analyze available data in order to establish guidelines for the use of tumor markers, but their recommendations have been very conservative [4, 5]. In this review, we address the pitfalls that have led to difficulties establishing tumor markers for routine clinical use, with a specific focus on tumor markers in breast cancer.

AMERICAN SOCIETY OF CLINICAL ONCOLOGY GUIDELINES

The American Society of Clinical Oncology (ASCO) convened a panel of experts that first published recommendations regarding the use of circulating and tissue-based tumor markers in breast cancer in 1996 [6] and most recently updated these recommendations in 2001 (Table 1) [4]. The ASCO panel evaluated multiple serum markers for breast cancer, including assays for MUC1 protein (cancer antigen [CA] 15-3 and CA 27.29), carcinoembryonic antigen (CEA), and the circulating extracellular domain of Her-2/neu. The panel did not recommend monitoring of any of these markers for screening, diagnosis, staging, or routine

surveillance of patients free of detectable disease. Measurement of CA 15-3 or CA 27.29 and/or CEA was recommended, however, to monitor selected patients with MBC undergoing palliative therapy [7].

Routine measurement of multiple tissue markers was also discussed in the guidelines. The panel recommended routine measurement of estrogen and progesterone receptors (ER and PgR, respectively) to identify patients most likely to benefit from endocrine therapy in either the early or metastatic disease settings. In addition, measurement of Her-2/neu overexpression and possibly amplification was recommended for all patients at the time of initial diagnosis or recurrence, as it is predictive of response to trastuzumab (Herceptin®; Genentech, South San Francisco, CA), a monoclonal antibody directed against the Her-2/neu receptor [8–10]. The panel felt that data to support assessment of other tissue-based markers, including p53, cathepsin D, and flow cytometry-derived estimates of DNA content or S phase, were insufficient to recommend usage in routine clinical practice.

Therefore, despite the large number of research studies evaluating the prognostic and predictive ability of numerous tumor markers in breast cancer, the ASCO panel recommended few for routine use in clinical practice. Why were these recommendations so conservative? In the succeeding sections of this paper, we outline the multiple factors that underlie this conservative approach.

WHEN IS A TUMOR MARKER USEFUL (USE)?

When evaluating tumor markers for use in clinical practice, clinicians should consider their utility, the magnitude of their effects, and their reliability (Table 2). Tumor markers can be useful at multiple stages of cancer diagnosis and treatment (Table 3) [11, 12]. For example, for individuals who do not have cancer, a marker may be helpful in determining the risk for developing the disease and/or it may be beneficial for screening for disease. Once an abnormality is found, a tumor marker may be helpful for distinguishing between benign and malignant processes or between different malignant processes. After confirmation of a cancer diagnosis, tumor markers can help monitor disease status during and after therapy.

Tumor markers can also help determine prognosis independent of therapy and predict response to therapy. Prognostic factors reflect the metastatic potential and/or growth rate of the tumor and are used to select patient outcomes without consideration of treatment given [13]. Predictive factors, on the other hand, reflect the sensitivity or resistance of a tumor to a therapeutic agent and therefore are used to predict which patients are likely to respond to a specific treatment [14]. Pure prognostic and predictive factors are depicted schematically in Figures 1A and 1B, respectively.

Table 1. Summary of American Society of Clinical Oncology guidelines for use of tumor markers in breast cancer [6, 7]

	Marker	Assess in adjuvant setting	Assess in metastatic setting	Prognostic factor	Predictive factor	Use for monitoring only	
						NED	Met
Tissue-based	ER	Yes	Yes	Weak	Strong ^a	N/A	N/A
	PgR	Yes	Yes	Weak	Strong ^a	N/A	N/A
	Her-2/neu	Yes	Yes	Weak	Strong ^b	N/A	N/A
Circulating	CA 15-3	No	Yes	No	No	No	Yes
	CA 27.29	No	Yes	No	No	No	Yes
	CEA	No	Yes	No	No	No	Yes
	Her-2 ECD	No	No	No	No	No	No

^aFor benefit from endocrine treatment.

^bFor benefit from trastuzumab therapy.

Abbreviations: CA, cancer antigen; CEA, carcinoembryonic antigen; ECD, circulating extracellular domain; ER, estrogen receptor; NED, no evidence of disease; Met: metastatic; N/A: not applicable; PgR, progesterone receptor.

Table 2. Requirements essential for acceptance of a tumor marker

Determine utility of marker (Table 3)
Evaluate magnitude of effect (none, weak, moderate, strong)
Analyze reliability of marker
Technical issues (assay)
Analytical issues (cutoff points, test/validation sets, multivariate analysis)
Trial design issues (appropriate patient population)

Table 3. Clinical uses of tumor markers

Determine risk of developing disease
Screening for disease
Establish diagnosis
Differentiate benign versus malignant disease
Determine type of malignancy
Determine prognosis
For primary disease, predict relapse
For metastatic disease, predict progression
Predict survival
Predict response to therapy
Hormone therapy
Chemotherapy
Novel therapies (e.g., trastuzumab)
Monitor disease
For primary disease, predict relapse
For metastatic disease, follow detectable disease

Adapted from Stearns V, Yamauchi H, Hayes DF. Circulating tumor markers in breast cancer: accepted utilities and novel prospects. *Breast Cancer Res Treat* 1998;52:239–259, with kind permission of Springer Science and Business Media.

Few tumor markers are purely prognostic or predictive. In fact, most tumor markers have mixed prognostic and predictive features, and the utility typically depends on the therapeutic agent in question. For example, ER expression is weakly favorably prognostic but strongly predictive of response to treatment with endocrine therapy, as illustrated in Figure 1C. Her-2/neu overexpression, on the other hand, is an unfavorable prognostic factor and is strongly predictive of response to therapy with trastuzumab, as shown in Figure 1D. Until appropriate studies have been performed both in vitro and in vivo, it can be difficult to know how to use a tumor marker appropriately in the clinical setting. In breast cancer, tumor markers are currently used in only a few of these categories.

HOW USEFUL IS THE TUMOR MARKER (MAGNITUDE)?

Once a tumor marker use has been identified, it is important to determine the magnitude of the difference in outcomes for that particular use between those who are marker positive and those who are not. By evaluating the difference in outcome, regardless of treatment, between a patient positive for a given prognostic factor and one who is negative for the factor, the relative strength of a prognostic factor can be determined [15]. This assessment requires the selection of an appropriate outcome of interest, such as improvement in symptoms or survival, or surrogates of these end points, such as response rates or progression-free survival.

For example, a breast cancer patient with disease in the lymph nodes at the time of diagnosis is two to three times more likely to have a breast cancer event (local recurrence or distant metastasis) than a patient without lymph node involvement, regardless of treatment. Since lymph node status has classically been used to make clinical decisions,

we have arbitrarily designated it as a “strong” prognostic factor, using it as the gold standard to set the criteria for consideration of other, putative markers [16]. A strong prognostic factor is depicted by Factor 1 in Figure 1A. Alternatively, untreated patients with ER-positive breast cancer have only slightly better outcomes than those with ER-negative disease, and therefore we designated ER status as a weak prognostic factor, as portrayed by Factor 2 in Figure 1A [17, 18]. In a previous publication, we have suggested hazard ratios of <1.5 , $1.5\text{--}2$, and >2 , to distinguish weak, moderate, and strong prognostic factors, respectively, for breast cancer [16]. Such arbitrary designations would need to be established for other uses, as appropriate.

Predictive factors can also be classified as weak (Factor 1), moderate, or strong (Factor 2), depending on their ability to predict response to, and therefore benefit from, a given therapy, as illustrated in Figure 1B. One measure to permit comparison of the relative strengths has been designated

the “relative predictive value” (RPV), the ratio of the likelihood that a factor-positive patient will respond to treatment to the likelihood that a factor-negative patient will respond to treatment. As with prognostic factors, we have proposed arbitrary classes of prediction factors for breast cancer therapies based on what has been accepted by consensus, in this case ER [19]. Adjuvant tamoxifen therapy has been shown to decrease recurrence rates for ER-positive patients by 40%–50%, whereas ER-negative patients obtain minimal, if any, benefit from hormonal therapy [20]. Therefore, the RPV is >8 . Similarly, the majority of patients with ER-positive MBC have a clinical response to hormonal therapy, whereas patients with ER-negative disease do not respond [21, 22]. ER status is therefore a strong predictive factor for response to hormonal therapy. In this framework, we have arbitrarily proposed that, in breast cancer, weak, moderate, and strong predictive factors correspond to RPVs of 1–2, 2–4, and >4 , respectively [15]. It is important to understand

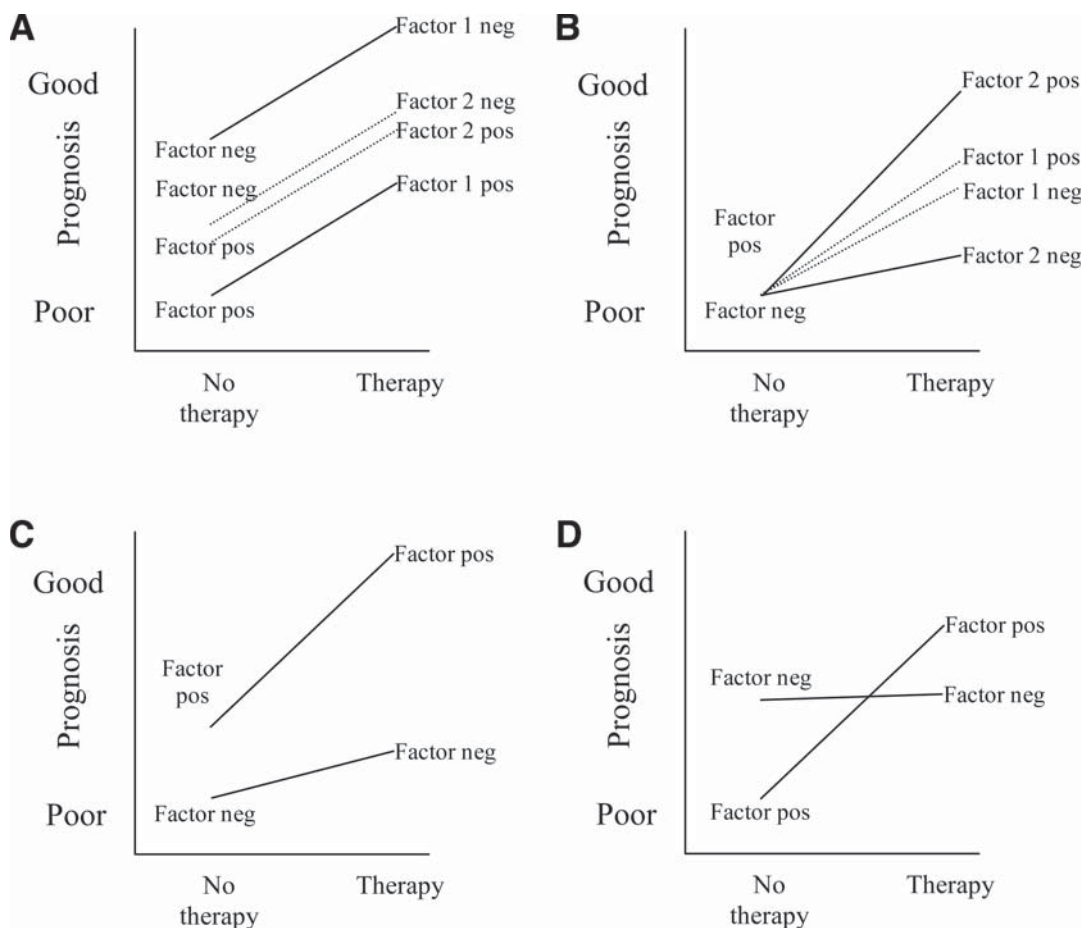


Figure 1. Schematic representation of prognostic and predictive factors. (A): Pure prognostic factor. (B): Pure predictive factor. (C): Mixed factor associated with weakly favorable prognosis and strong response to specific therapy. (D): Mixed factor associated with unfavorable prognosis and strong response to specific therapy. Modified from Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is “statistically significant” clinically useful? *Breast Cancer Res Treat* 1998;52:305–319, with kind permission of Springer Science and Business Media.

the clinical implications of the relative strength of a prognostic and/or predictive factor when integrating this information into routine practice, and to determine if the data support its use in a specific clinical situation.

HOW RELIABLE IS THE TUMOR MARKER (PRECISION AND ACCURACY)?

The preceding discussion illustrates the importance of estimating the magnitude of the relative tumor marker effect for a selected use. However, the marker is only useful if the estimate of its magnitude is reliable and reproducible. In this regard, many investigators conclude that their marker of interest has clinical utility if in their study the difference in outcomes between marker “positive” and marker “negative” patients is less than conventional measures of statistical significance ($p < .05$). This conclusion may be mistaken. Statistical significance only suggests that in the population chosen for that study, the differences observed are likely not to be a result of chance alone. It does not imply clinical utility, nor does a p -value < 0.05 document the validity of the tumor marker. Although it is important to determine that the differences in outcome achieve statistical significance, statistical significance alone does not determine clinical utility.

In addition to determining when to use a tumor marker and the magnitude of its effect, it is important to ensure that the technical aspects of the marker are reliable and reproducible and that the study design and conduct are appropriate to test the marker for a clinical use of interest. Several problems with tumor marker studies, including technical, analytical, and trial design issues, have limited the introduction of new prognostic and predictive factors into routine clinical practice [11].

What Technical Factors Influence Measurement of Markers?

From a technical standpoint, difficulties arise because of poor sensitivity and/or specificity of the assay for the analyte, poorly reproducible assays, and differences between assays that use different reagents for measurement of the same marker [11]. Even for the two most commonly used and accepted tumor markers, ER expression and Her-2/neu overexpression, standard methodologies have not yet been established [4, 5]. Two primary technical considerations are critical when measuring a tumor marker. The first is which type of assay should be used. The second is the reproducibility of the chosen assay, from both a technical and an analytical perspective.

For example, Her-2/neu status can be determined by measures of protein expression (by immunohistochemistry [IHC], Western blotting, or enzyme-linked immunosorbent assay), measures of RNA expression (by Northern

blotting or reverse transcriptase-polymerase chain reaction [RT-PCR]), and/or measures of DNA amplification (by fluorescence or chromogenic in situ hybridization [FISH and CISH, respectively]). Furthermore, even within these categories, different reagents (e.g., different antibodies in IHC assays) are used in different tests. The results are not interchangeable, either within or between classes of assays, and therefore researchers must decide which methodology they will employ. Once that decision is made, researchers must then decide how to perform the assay. For example, when assessing Her-2/neu overexpression by IHC, technical issues such as antibody concentration and antigen retrieval methods may cause unacceptably high false-positive or false-negative rates.

In one study, IHC and FISH resulted in only a 65% agreement for Her-2/neu status [23]. In a different study, results obtained from local laboratories were compared with those from a central laboratory for two Her-2/neu assays, the HercepTest™ IHC assay (Dako North America, Inc., Carpinteria, CA) and the FISH assay, with 79% concordance for HercepTest™ and 85% concordance for FISH [24]. Therefore, for the same test at multiple laboratories, and for different tests for the same marker, there is a significant degree of discordance for two commonly used tests for the evaluation of Her-2/neu status.

The stakes are high. Recently reported data suggest that adjuvant trastuzumab decreases recurrence rates by 50%. However, up to 5% of patients who receive trastuzumab develop cardiac dysfunction, and the cost of 1 year of therapy may exceed \$100,000. Therefore, it is essential that Her-2/neu, the target for trastuzumab, be assayed accurately and precisely for every tissue sample. Expert panels are now being convened to establish guidelines for the conduct and interpretation of common tumor marker assays, including ER and Her-2/neu. These guidelines should lead to standardization of the assays, which should allow for more reliable results both for routine clinical practice and for use of these assays in clinical trials.

What Analytical Issues Are Important to Consider?

Assay Interpretation

Determination of assay results can also vary, even for a single type of assay. For example, with visual assays such as IHC for ER and Her-2/neu, intra- and interobserver variability leads to differences in interpretation [25, 26]. Some attempts have been made to standardize interpretation, such as development of the so-called “Allred score” for semiquantitation of ER expression [27], but these have not been universally adopted. Automated and semiautomated systems appear to

be highly accurate and are likely to be more reproducible. Examples of automated systems include the ChromaVision ACIS® system (ChromaVision Medical Systems, Inc., San Juan Capistrano, CA) for ER expression measurement, the CellSearch™ assay (Veridex, LLC, Warren, NJ) for detection of circulating tumor cells [28], and the Oncotype DX™ assay (Genomic Health, Inc., Redwood City, CA), a new prognostic tool for patients with hormone receptor (HR)-positive, lymph node-negative breast cancer [29].

Cutoff Point Determination

Regardless of the assay, one has to select some value or level that distinguishes positive from negative results. However, there is no consensus regarding correct methods to establish cutoff points, and different studies of the same prognostic or predictive factor can have widely varying “optimal” cutoff points [30].

Cutoff points may be defined using either arbitrary or data-derived methods (Table 4). One approach is to consider any value greater than two standard deviations above the mean for normal subjects to be positive. Cutoff points can also represent arbitrary values within affected patients; for example, one might decide that 10%, 50%, or 90% of affected patients will be classified as “positive.” Others have defined cutoff points based on technical factors, such as the limit of detection of the assay [21]. Finally, the cutoff point for a new assay can be defined by comparing it with an older assay [27].

Deriving cutoff points based on patient outcome data may provide more accurate values. For example, the cutoff point for ER expression was first defined by limits of the assay and later by determining the optimal level that distinguished those patients who respond to hormonal therapy from those who do not. In another example, the cutoff point for the CellSearch™ assay for circulating tumor cells was initially determined based on differences in time to progression of a test set of patients with metastatic disease, and this cutoff was then validated with an independent but simi-

lar patient cohort from the same study [31]. Another common method to generate a data-derived cutoff point is to construct a receiver operating characteristic curve, which demonstrates the tradeoff between the sensitivity and specificity of an assay at different cutoff points.

Recently, a novel data-derived method to select cutoff points, designated subpopulation treatment effect pattern plot (STEPP) analysis, has been proposed [32]. STEPP analysis evaluates outcomes to specific treatments in subpopulations of patients within randomized clinical trials or meta-analyses [32]. For example, it has been proposed that recurrence rates after treatment with chemotherapy should be evaluated in the context of the endocrine responsiveness of tumors, since HR-positive and -negative tumors appear to behave differently. Rather than arbitrarily defining cutoff points for ER positive and negative, the authors performed a STEPP analysis of data from a previously conducted randomized clinical trial and were able to demonstrate a benefit from chemotherapy only in the subset of patients with very low ER values.

Cutoff Point Validation

Regardless of whether cutoff points are chosen arbitrarily or are data-derived, the selected cutoffs require subsequent validation. The initial evaluation should be performed using a “test set” of patients. In the second part of the study, the utility of the cutoff point should then be confirmed using a separate “validation set” composed of a similar, but completely independent, patient population.

For example, the Oncotype DX™ assay is based on the principle of evaluating expression of multiple candidate genes using quantitative RT-PCR [29]. The investigators initially screened more than 200 candidate genes with the aim of developing a test that would predict the likelihood of recurrence of cancer in patients with HR-positive, lymph node-negative breast cancer. Breast cancer tissues from 447 patients with HR-positive, lymph node-negative tumors were used retrospectively to generate an algorithm using 16 of these genes that permitted division of patients into subgroups with very low, intermediate, or very high risk for recurrence. Patients are assigned to these groups based on a “recurrence score” derived from the algorithm. The majority of the test samples were obtained from patients treated with tamoxifen alone in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-20 trial [33], and the data-derived algorithm was then validated using a separate retrospective cohort of patients from the NSABP B-14 trial [34], who had similar clinical characteristics and were also treated with tamoxifen alone. If the cohorts had different clinical characteristics or had been treated differently, the validation would not

Table 4. Methods used for selection of cutoff points

Arbitrary
Based on limit of detection of the assay
Defined as two standard deviations above the mean of normal patients
Defined based on mean value in affected versus normal patients
“Appropriate” percentage of positive cells
Data-derived
Plot <i>p</i> -value versus outcome
Plot magnitude of effect of marker versus outcome
Receiver operating characteristic curve
Subpopulation treatment effect pattern plot analysis

have been legitimate. Validation of the results in a separate patient population strongly suggests that the test is reliable and that the results are likely to be meaningful in a larger population, as long as the patients tested are similar to the cohorts included in the original studies.

Statistical Analysis

Of course, statistical analysis is necessary to determine that the observations are not a result of chance alone. However, once a tumor marker has been identified and validated, it is important to determine the relative value of the marker in the context of previously identified prognostic and/or predictive factors, such as lymph node status and tumor size, using some type of a multivariate analysis. If such an evaluation is not performed, clinicians will be unable to determine the usefulness of incorporating the new marker into routine clinical practice.

TRIAL DESIGN AND FRAMEWORKS FOR GENERATING, REPORTING, AND EVALUATING TUMOR MARKER RESULTS

One of the key steps in identifying and confirming the benefit of a new tumor marker is appropriate study design. The Tumor Marker Utility Grading System (TMUGS) was initially developed by members of the ASCO panel to provide a framework within which the utility of tumor markers can be graded based on published information [11]. Markers are assigned a grade based on the level of evidence (LOE) available. These LOE reflect the relative quality of the studies used to generate an estimate of the effect of the marker (Table 5). LOE I and II studies are the most beneficial for evaluating the utility of a tumor marker. According to the TMUGS framework, the ideal

clinical trial is a properly powered, prospective, randomized trial designed specifically to evaluate the clinical utility of a tumor marker in question for a specific and pre-designated use. In such a study, diagnostic and/or therapeutic decisions for the study arm are based on the tumor marker in question, whereas the decisions for the control arm are made independently [11].

Systematic overviews and/or pooled analyses of well-conducted LOE II studies are equivalent to LOE I studies, especially if the correlative studies address a specific use but are underpowered in a single study. It has been estimated that a clinical trial that has been appropriately powered to determine a clinical end point, such as progression-free survival, is underpowered for analysis of tumor marker-designated subgroups by one fourth, even if tissue from 100% of enrollees is available. Moreover, interpretation still requires judgment regarding the clinical importance of the finding.

For example, from the prospective randomized clinical trials of adjuvant trastuzumab versus placebo, we anticipate combined analyses of the multiple underpowered LOE II studies evaluating novel markers for benefit from this drug. These pooled results should help focus trastuzumab therapy in the subgroups of Her-2/neu-positive patients most likely to benefit. The combined analyses of small LOE III studies that contain patients with variable clinical characteristics and treatments, however, are more likely useful to generate new hypotheses than to provide clinically useful and validated results.

For all studies, whether prospective or retrospective, it is important to identify the appropriate patient population to be investigated. All patients should have a similar profile based on known prognostic factors. Importantly, the effects

Table 5. Levels of evidence for determining utility of tumor marker

Level	Type of evidence
I	Evidence from a single, high-powered, <i>prospective</i> controlled study that is specifically designed to test marker, or evidence from well-done meta-analysis and/or overview of level I studies. Ideally, the study is a prospective, randomized controlled trial in which diagnostic and/or therapeutic clinical decisions in one arm are determined at least in part on the basis of marker results, and diagnostic and/or therapeutic clinical decisions in the control arm are made independently of marker results, or Evidence from overview of level of evidence II studies addressing specific use.
II	Evidence from study in which marker data are determined in relationship to <i>prospective</i> therapeutic trial that is performed to test therapeutic hypothesis but not specifically designed to test marker utility. Specimen collection for marker study and statistical analysis are <i>prospectively determined</i> in protocol as secondary objectives.
III	Evidence from large but <i>retrospective</i> studies from which variable numbers of samples are available or selected. Statistical analysis for tumor marker was not dictated prospectively at time of therapeutic trial design.
IV	Evidence from small <i>retrospective</i> studies that do not have prospectively dictated therapy, follow-up, specimen selection, or statistical analysis.
V	Evidence from <i>small pilot</i> studies designed to determine or estimate distribution of marker levels in sample populations.

Reprinted from Hayes DF, Bast RC, Desch CE et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88(20):1456–1466, permission of Oxford University Press.

of systemic treatment are critical and must be considered. If the study is addressing the *prognostic* value of a marker, all patients should have been treated uniformly and without whatever treatment might be considered if the patients have a “poor prognosis.” If the question is addressing the utility of adding any treatment, all patients should be untreated. If the question is whether more treatment should be given, then all patients should have received the same treatment.

If the study is addressing *predictive* factors, a control group that has been treated identically to the study group, with the exception that they did not receive the treatment in question, is essential. Although the control group might be from a selected historical control, predictive factors are ideally studied in the context of prospective, randomized, controlled trials comparing the patients who received the treatment in question with those who did not receive that treatment.

Tumor marker studies should be carefully designed, using the above criteria, to obtain clinically useful information. Researchers frequently have practical difficulties designing such studies, however, because of the need for significant numbers of patients with particular clinical characteristics in order to address a specific clinical question. As discussed above, this can sometimes be overcome by pooling the results of several well-done but underpowered studies. Another significant drawback is obtaining funding for tumor marker studies, as pharmaceutical companies and third-party payers derive relatively smaller financial benefit from the results compared to the enormous payoffs for a “blockbuster” therapeutic agent. Regardless, given the consequences, one has to question why it is acceptable for tumor marker studies to be performed with less scientific rigor than studies of new pharmaceutical agents.

Reporting of tumor marker studies has also been historically haphazard. Recently, in order to standardize reporting of tumor marker study results, the National Cancer Institute-European Organization for Research and Treatment of Cancer (NCI-EORTC) Working Group on Cancer Diagnostics developed REporting recommendations for tumor MARKer prognostic studies (REMARK) [35]. The guidelines outline items that should be addressed by researchers when reporting the results of tumor marker studies, including prospectively defining the question the study is trying to address, identifying the appropriate patient population and controls, determining the end point, and identifying potential confounding factors. Explicit recommendations are given regarding which information must be contained in publications of tumor marker studies, including patient and treatment information, specimen characteristics, assay methods, study design, and statistical analysis methods.

REAL-WORLD CLINICAL EXAMPLES

In the preceding sections we identified the essential elements for establishing the usefulness, strength, and reliability of tumor markers. Let us now discuss the data supporting two currently used tumor markers, Her-2/neu and Oncotype DX™.

Her-2/neu

The first report of Her-2/neu as a prognostic factor in breast cancer was published in 1987 [36]. Since then, more than 200 papers addressing this topic have been published, with widely mixed and disparate results [37]. Different authors have concluded that Her-2/neu is associated with poor outcomes, no difference in outcome, or even favorable outcomes. Indeed, a great deal of this confusion could have been avoided if the investigators would have addressed the components described above: (a) What is the intended use, (b) What is the magnitude of difference between positive and negative for that use, and (c) How reliable is the estimate of the magnitude?

The potential uses for Her-2/neu are for prognosis and prediction, as outlined in Table 6. Issues related to the reliability of the assays have been described in detail above. Overall, studies support that Her-2/neu overexpression is a poor prognostic factor, although its magnitude appears weak. For example, in Adjuvant! Online it has only a relative predictive value of 1.5 [38]. Its role as a prognostic factor thus remains unclear.

More data exist to support the role of Her-2/neu status for prediction of response to standard therapies and trastuzumab. For selective estrogen receptor modulators (SERMs), such as tamoxifen, preclinical and clinical studies suggest that Her-2/neu positivity confers a relative resistance, with moderate magnitude, although the data are LOE III at best [39]. Data for aromatase inhibitors (AIs) are mixed, although in one pilot study of neoadjuvant endocrine therapy, Her-2/neu overexpression correlated with lower response to tamoxifen than to AIs [39]. At present, Her-2/neu status is not used to determine which endocrine therapy to use, because of the poor level of available evidence and conflicting data. Confirmation of these results may lead to preferential use of AIs in patients with Her-2/neu-positive disease.

Patients with tumors that overexpress Her-2/neu appear to have relative resistance to some chemotherapy regimens, such as cyclophosphamide, methotrexate, and 5-fluorouracil (CMF), but not to others, such as anthracycline-containing regimens [37, 40, 41]. Her-2/neu status is not generally a consideration when choosing a chemotherapy regimen, however, because, as is the case with endocrine therapy, the level of available evidence does not support using Her-2/neu status to predict response to chemotherapy.

Table 6. Theoretical uses for tissue-based Her-2/neu assessment

Use	Effect ^a	Magnitude	LOE
Prognosis	Negative	Moderate	III
Prediction			
Hormone therapy			
SERM	Negative	Weak-moderate	III
AI	Neutral	N/A	III
Chemotherapy			
Nonanthracycline	Negative	Weak-moderate	III
Anthracycline	Positive or Neutral	Moderate or N/A	III
Trastuzumab	Positive	Strong	II

^aNegative indicates worse outcome if Her-2/neu-positive; positive indicates better outcome if Her-2/neu-positive; neutral indicates outcome same regardless of Her-2/neu result.

Abbreviations: AI, aromatase inhibitor; LOE, level of evidence; N/A, not applicable; SERM, selective estrogen receptor modulator.

In contrast, Her-2/neu status appears to be strongly predictive of response to trastuzumab. A patient with a tumor that overexpresses Her-2/neu is usually treated with trastuzumab in either the adjuvant or metastatic settings [8, 9, 42, 43] because the benefits outweigh the risks in the majority of cases. A patient with a tumor that fails to overexpress Her-2/neu appears not to respond to treatment with trastuzumab [10, 44] and therefore would not be treated with trastuzumab to avoid both unnecessary toxicity and cost. Thus, use of Her-2/neu to select trastuzumab is recommended based on “use” and “magnitude.” However, as discussed above, there are substantial apparent difficulties with the technical reliability of all available assays for the marker. Nonetheless, despite the shortcomings of Her-2/neu studies outlined above, at present there is sufficient LOE II evidence to support the routine clinical use of Her-2/neu overexpression for selection of trastuzumab therapy, as indicated in the most recent ASCO tumor marker guidelines [4].

Multigene Expression: Oncotype DX™

A more recent example of development of a new tumor marker is provided by the case of Oncotype DX™. The investigators specifically developed the test to determine prognosis in ER-positive, lymph node-negative patients who were treated with tamoxifen [29]. The available results suggest that, in this group of patients, Oncotype DX™ is a strong prognostic factor because the ratio of the hazard ratios of the high and low recurrence score cohorts is >2 in both the test and validation cohorts [29]. The Oncotype DX™ test is also reliable because it fulfills the criteria outlined above for technical, analytical, and trial design issues (Table 2). The assay is reproducible and was validated in independent test and validation cohorts of patients, as described above.

Oncotype DX™ has also been evaluated as a predictive factor, although less rigorously. In the NSABP B-14 trial, the patient cohorts were treated with tamoxifen versus observation, and comparison of the cohorts suggested that the Oncotype DX™ assay is predictive for tamoxifen [45]. Similarly, analysis of NSABP B-20, in which patients were randomized to CMF and tamoxifen versus tamoxifen alone, permitted the investigators to determine that the assay is predictive for response to chemotherapy [45].

Are the available data sufficient to conclude that Oncotype DX™ has been validated to the extent that patient treatment decisions should be based on the results? Perhaps, but because these studies were all proposed using available samples from trials performed many years ago and represented only subsets of the overall population entered into the trials, concerns have been raised about wholesale clinical adoption of this assay. In that regard, the North American Breast Cancer Intergroup is developing the TailorRx clinical trial to further validate and extend the Oncotype DX™ results. The trial design assumes that the assay is prognostic, and will confirm the ability of the assay to predict response to chemotherapy.

In the TailorRX trial, tumors of patients with ER-positive and lymph node-negative breast cancer will be tested using the Oncotype DX™ assay. Patients with low recurrence scores, who have good prognoses without chemotherapy, will receive hormonal therapy alone. At the other end of the spectrum, patients with high recurrence scores will receive chemotherapy in addition to hormonal therapy. Those patients whose scores fall in the intermediate range will all receive hormonal therapy and be randomly assigned to chemotherapy or not. This trial design will permit validation of the Oncotype DX™ results in a similar patient population in a large prospective clinical trial,

and will allow for generation of new data on which to base treatment recommendations for patients whose recurrence scores are intermediate.

Given the substantial technical, analytical, and trial design problems with previously performed tumor marker studies, it is imperative to address these issues. It is especially important to standardize the assays for commonly used tumor markers. Otherwise, patients with false-positive test results for predictive factors will receive treatments that are not beneficial but which may cause significant toxicity, and those with false-negative results will not be offered potentially life-saving therapies. In addition, a better understanding of the potential pitfalls in tumor marker study design will allow for the development of new, potentially more useful assays.

CONCLUSIONS

Tumor markers, when well defined, can play a significant role in prediction and prognosis for breast cancer patients. Because of the abundance of poorly designed tumor marker

studies to date, however, very few markers have been accepted for routine use by groups such as ASCO. When designing studies to establish a new tumor marker, or new use for an old marker, it is important to address the utility, magnitude, and reliability of the marker (Table 2).

Frameworks such as TMUGS can be useful when designing and conducting these studies to ensure that appropriate components are included, thereby leading to the establishment of new tumor markers for routine clinical use [11]. By progressively generating and refining a hypothesis, based on data derived from increasingly well-developed studies, tumor markers with clinical utility can be identified (Fig. 2). In addition, the new REMARK guidelines should promote better design and conduct of studies specifically focused on tumor marker validation [35]. Implementation of these recommendations when designing tumor marker studies will result in the generation and publication of appropriate and complete clinical data, leading to the adoption of new, well-validated tumor markers for routine clinical use.

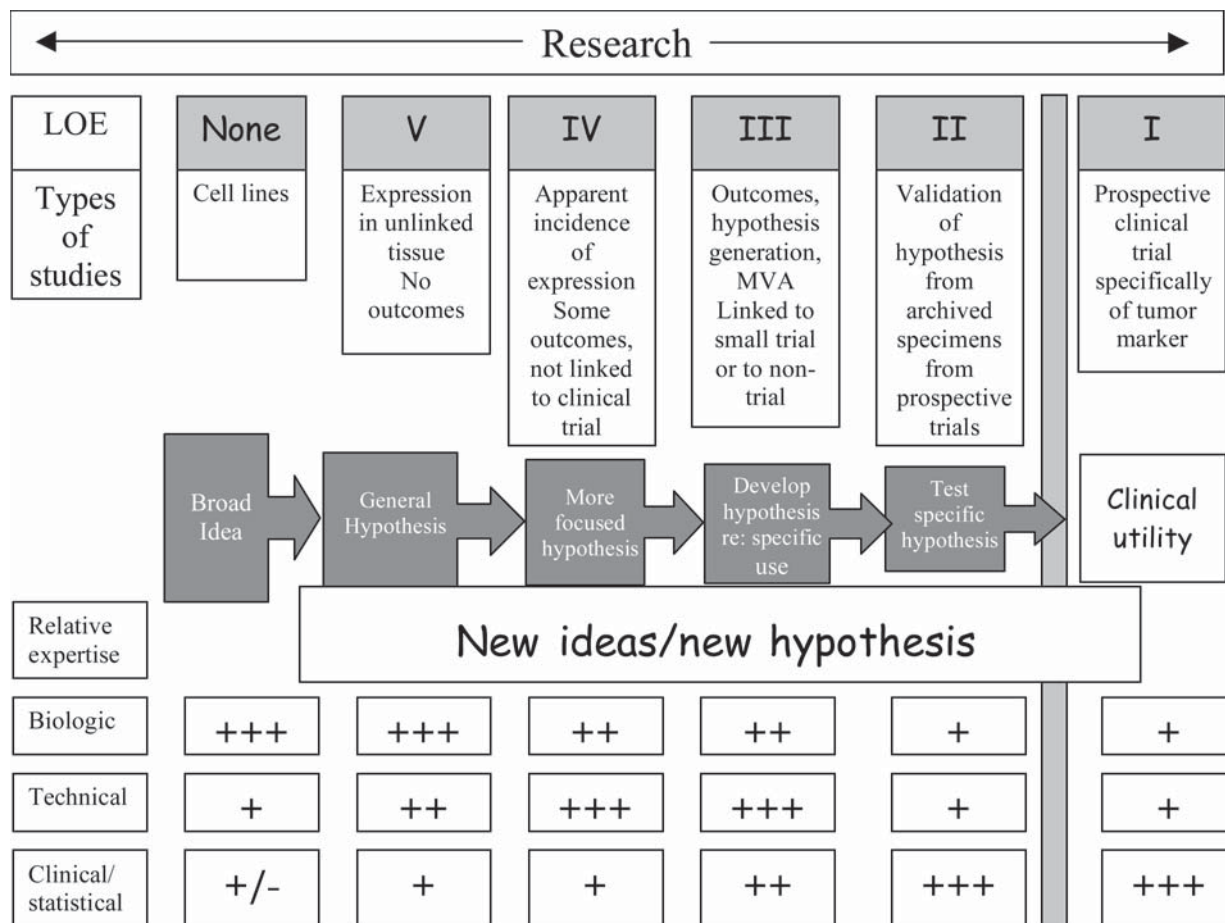


Figure 2. Tumor marker development flow chart. LOE, level of evidence; MVA, multivariate analysis.

AUTHORS' NOTE

Supported by National Institutes of Health grant 5R01CA092461-03 and Fashion Footwear Charitable Foundation of New York/QVC Presents Shoes-on-Sale™.

DISCLOSURES

D.F.H. has served as an unpaid consultant for Genomic Health, Inc. in the past year and has received research

funding from Immunicon. D.F.H. has been a consultant/advisory panel participant or held lecture/honorarium position during past year for Dendreon, Immunicon, Novartis, Pfizer, Precision Therapeutics, Inc., Oncotech, and Veridex. D.F.H. was the principle or coinvestigator during the past year on studies funded by Immunicon, Wyeth Ayerst-Genetics Institute, Pfizer, and Novartis.

REFERENCES

- American Cancer Society. Cancer Facts and Figures, 2005. Available at <http://www.cancer.org/downloads/STT/CAFF2005f4PWSecured.pdf>. Accessed March 10, 2005.
- Peto R, Boreham J, Clarke M et al. UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. *Lancet* 2000;355:1822.
- Berry DA, Cronin KA, Plevritis SK et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;353:1784-1792.
- Bast RC Jr, Ravdin P, Hayes DF et al. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001;19:1865-1878.
- Hammond ME, Fitzgibbons PL, Compton CC et al. College of American Pathologists Conference XXXV: solid tumor prognostic factors-which, how and so what? Summary document and recommendations for implementation. *Arch Pathol Lab Med* 2000;124:958-965.
- Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. *J Clin Oncol* 1996;14:2843-2877.
- Bast RC Jr, Ravdin P, Hayes DF et al. Errata: 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001;19:4185-4188.
- Cobleigh MA, Vogel CL, Tripathy D et al. Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J Clin Oncol* 1999;17:2639-2648.
- Slamon DJ, Leyland-Jones B, Shak S et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783-792.
- Mass RD, Press MF, Anderson S et al. Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer* 2005;6:240-246.
- Hayes DF, Bast RC, Desch CE et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456-1466.
- Stearns V, Yamauchi H, Hayes DF. Circulating tumor markers in breast cancer: accepted utilities and novel prospects. *Breast Cancer Res Treat* 1998;52:239-259.
- McGuire WL, Clark GM. Prognostic factors and treatment decisions in axillary-node-negative breast cancer. *N Engl J Med* 1992;326:1756-1761.
- Gasparini G, Pozza F, Harris AL. Evaluating the potential usefulness of new prognostic and predictive indicators in node-negative breast cancer patients. *J Natl Cancer Inst* 1993;85:1206-1219.
- Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? *Breast Cancer Res Treat* 1998;52:305-319.
- Isaacs C, Stearns V, Hayes DF. New prognostic factors for breast cancer recurrence. *Semin Oncol* 2001;28:53-67.
- Fisher B, Costantino JP, Wickerham DL et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 1998;90:1371-1388.
- Knight WA, Livingston RB, Gregory EJ et al. Estrogen receptor as an independent prognostic factor for early recurrence in breast cancer. *Cancer Res* 1977;37:4669-4671.
- Hayes DF. Do we need prognostic factors in nodal-negative breast cancer? *Arbiter. Eur J Cancer* 2000;36:302-306.
- Early Breast Cancer Trialists' Collaborative Group (EBCTG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;365:1687-1717.
- McGuire WL, Chamness GC, Costlow ME et al. Steroids and human breast cancer. *J Steroid Biochem* 1975;6:723-727.
- Allegra JC, Lippman ME, Thompson EB et al. Estrogen receptor status: an important variable in predicting response to endocrine therapy in metastatic breast cancer. *Eur J Cancer* 1980;16:323-331.
- Dressler LG, Berry DA, Broadwater G et al. Comparison of HER2 status by fluorescence in situ hybridization and immunohistochemistry to predict benefit from dose escalation of adjuvant doxorubicin-based therapy in node-positive breast cancer patients. *J Clin Oncol* 2005;23:4287-4297.
- Perez EA, Suman VJ, Davidson NE et al. HER2 testing by local, central, and reference laboratories in the NCCTG N9831 Intergroup Adjuvant Trial. *J Clin Oncol* 2004;22:567a.
- Press MF, Hung G, Godolphin W et al. Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. *Cancer Res* 1994;54:2771-2777.
- Jacobs TW, Gown AM, Yaziji H et al. Specificity of HercepTest in determining HER-2/neu status of breast cancers using the United States Food and Drug Administration-approved scoring system. *J Clin Oncol* 1999;17:1983-1987.
- Harvey JM, Clark GM, Osborne CK et al. Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. *J Clin Oncol* 1999;17:1474-1481.

- 28 Allard WJ, Matera J, Miller MC et al. Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin Cancer Res* 2004;10:6897–6904.
- 29 Paik S, Shak S, Tang G et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817–2826.
- 30 Altman DG, Lausen B, Sauerbrei W et al. Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* 1994;86:829–835.
- 31 Cristofanilli M, Budd GT, Ellis MJ et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N Engl J Med* 2004;351:781–791.
- 32 Gelber RD, Bonetti M, Castiglione-Gertsch M et al. Tailoring adjuvant treatments for the individual breast cancer patient. *Breast* 2003;12:558–568.
- 33 Fisher B, Dignam J, Wolmark N et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. *J Natl Cancer Inst* 1997;89:1673–1682.
- 34 Fisher B, Costantino J, Redmond C et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. *N Engl J Med* 1989;320:479–484.
- 35 McShane LM, Altman DG, Sauerbrei W et al. Reporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93:387–391.
- 36 Slamon DJ, Clark GM, Wong SG et al. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177–182.
- 37 Yamauchi H, Stearns V, Hayes DF. The role of c-erbB-2 as a predictive factor in breast cancer. *Breast Cancer* 2001;8:171–183.
- 38 Adjuvant! Online breast cancer help files: prognostic estimates. 2006. <https://www.adjuvantonline.com/breathelp0306/breastindex.html>. Accessed February 17, 2006.
- 39 Ellis MJ, Coop A, Singh B et al. Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial. *J Clin Oncol* 2001;19:3808–3816.
- 40 Muss HB, Thor AD, Berry DA et al. c-erbB-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *N Engl J Med* 1994;330:1260–1266.
- 41 Thor AD, Berry DA, Budman DR et al. erbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. *J Natl Cancer Inst* 1998;90:1346–1360.
- 42 Piccart-Gebhart MJ, Procter M, Leyland-Jones B et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005;353:1659–1672.
- 43 Romond EH, Perez EA, Bryant J et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;353:1673–1684.
- 44 Seidman AD, Berry D, Cirincione C et al. CALGB 9840: phase III study of weekly (W) paclitaxel (P) via 1-hour(h) infusion versus standard (S) 3h infusion every third week in the treatment of metastatic breast cancer (MBC), with trastuzumab (T) for HER2 positive MBC and randomized for T in HER2 normal MBC. *J Clin Oncol* 2004;22:512a.
- 45 Paik S, Shak S, Tang G et al. Expression of the 21 genes in the Recurrence Score assay and prediction of clinical benefit from tamoxifen in NSABP study B-14 and chemotherapy in NSABP study B-20. *Breast Cancer Res Treat* 2004;88:24a.

ADDITIONAL READING

- Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. *J Clin Oncol* 1996;14:2843–2877.
- Hayes DF, Bast RC, Desch CE et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456–1466.
- McShane LM, Altman DG, Sauerbrei W et al. Reporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93:387–391.