Estrogen Receptor Status by Immunohistochemistry Is Superior to the Ligand-Binding Assay for Predicting Response to Adjuvant Endocrine Therapy in Breast Cancer

Jennet M. Harvey, Gary M. Clark, C. Kent Osborne, and D. Craig Allred

Purpose: Immunohistochemistry (IHC) is a newer technique for assessing the estrogen receptor (ER) status of breast cancers, with the potential to overcome many of the shortcomings associated with the traditional ligand-binding assay (LBA). The purpose of this study was to evaluate the ability of ER status determination by IHC, compared with LBA, to predict clinical outcome—especially response to adjuvant endocrine therapy—in a large number of patients with long-term clinical follow-up.

Patients and Methods: ER status was evaluated in 1,982 primary breast cancers by IHC on formalin-fixed paraffin-embedded tissue sections, using antibody 6F11 and standard methodology. Slides were scored on a scale representing the estimated proportion and intensity of positive-staining tumor cells (range, 0 to 8). Results were compared with ER values obtained by the LBA in the same tumors and to clinical outcome.

Results: An IHC score of greater than 2 (corresponding to as few as 1% to 10% weakly positive cells) was used to define ER positivity on the basis of a univariate cut-point analysis of all possible scores and disease-free survival (DFS) in patients receiving any adjuvant endocrine therapy. Using this definition, 71% of all tumors were determined to be ER-positive by IHC, and the level of agreement with the LBA was 86%. In multivariate analyses of patients receiving adjuvant endocrine therapy alone, ER status determined by IHC was better than that determined by the LBA at predicting improved DFS (hazard ratios/P = 0.474/.0008 and 0.707/.3214, respectively) and equivalent at predicting overall survival (0.379/.0001 and 0.381/.0003, respectively).

Conclusion: IHC is superior to the LBA for assessing ER status in primary breast cancer because it is easier, safer, and less expensive, and has an equivalent or better ability to predict response to adjuvant endocrine therapy.

J Clin Oncol 17:1474-1481.

THE ESTROGEN RECEPTOR (ER) content of breast carcinomas is important as a prognostic and predictive biomarker, according to recently published guidelines,[1-3] and evaluation of ER status is part of the routine assessment of these neoplasms. Most of the data on the clinical utility of ER content have been generated using biochemical ligand-binding assays (LBAs), such as the dextran-coated charcoal assay (DCCA). Since the first report of its independent prognostic significance almost two decades ago,[4] the assessment of ER status by DCCA has been validated repeatedly and is generally regarded as the standard by which other methods are assessed. There are, however, problems associated with LBAs for ERs. They are technically challenging and expensive; require radioactive reagents and relatively large amounts of fresh-frozen tissue; and are insensitive and nonspecific in accounting for differences in the cellular composition of samples, such as those with a low tumor cellularity or contaminating benign cells that might be ER-positive.

The development of highly specific monoclonal antibodies[5] and immunohistochemistry (IHC) techniques to localize ERs[6] provided the potential to overcome most of the difficulties inherent to LBAs. Compared with LBAs, IHC is easier to perform, less expensive, safer, applicable to a wider variety of samples (eg, cytologic preparations, frozen tissue sections, fixed archival tissue sections, etc), and more sensitive and specific in the identification of rare ER-positive tumor cells or contaminating ER-positive benign epithelium under direct microscopic visualization.

The ultimate usefulness of ER status assessment by IHC, however, resides in its ability to predict clinical outcome, especially response to hormone therapy. Many studies have evaluated the clinical relevance of measuring ER status by IHC, and the large majority reported statistically significant relationships with clinical outcome.[7] Nonetheless, there were limitations associated with these generally positive studies. For example, the majority evaluated patient populations of mixed clinical stage and treatment status, making it nearly impossible to separate the prognostic from the predictive implications of their results. In addition, these

From the Department of Pathology, University of Western Australia, Nedlands, Western Australia, Australia; and Division of Medical Oncology and Department of Pathology, University of Texas Health Science Center at San Antonio, San Antonio, TX.

Submitted August 6, 1998; accepted January 21, 1999.

Supported by grant nos. P01-CA30195, P30-CA54174, and P50-CA58183 from the National Institutes of Health.

Address reprint requests to D. Craig Allred, MD, University of Texas Health Science Center at San Antonio, Department of Pathology, San Antonio, TX 78284-7750; email allred@uthscsa.edu.

0732-183X/99/1705-1474

studies used many different antibodies and detection systems with unequal sensitivities and specificities on tissue samples prepared in diverse ways. The most problematic aspect was the use of a wide variety of techniques for scoring and interpreting results with arbitrary rather than clinically calibrated definitions of ER positivity. Despite these largely unresolved issues, most laboratories today have already converted to assessing ER status almost exclusively by IHC on routine archival (ie, formalin-fixed paraffin-embedded) tissue samples.

The purpose of this study was to resolve some of these issues by developing an IHC assay for archival tissue, using inexpensive commercially available reagents and an easy, reproducible scoring system calibrated to clinical outcome. The prognostic and predictive usefulness of this IHC assay was evaluated and compared with a standard LBA in a large group of patients with primary breast cancer and long-term clinical follow-up.

PATIENTS AND METHODS

Patient Population

Tumor specimens from patients with primary breast cancer in the San Antonio Tumor Bank were included in this study. Patients were diagnosed between 1973 and 1993 and had their ER statuses evaluated by LBA at the time of diagnosis in our laboratory. Selection criteria included presentation with primary breast cancer, sufficient tumor tissue remaining after LBA for additional IHC assays, and long-term follow-up for disease recurrence and death. A total of 1,982 patients who satisfied these criteria were chosen: 997 with negative axillary lymph nodes and 985 with positive nodes. Surgical treatment included modified radical mastectomy in 91% of the patients and lumpectomy in 9%. Postoperative radiation was used in 21%. After surgery, 35% received no additional therapy. The remainder received systemic adjuvant therapy in a routine clinical setting; this therapy consisted of chemotherapy alone in 13%, endocrine therapy alone in 26%, and combined chemotherapy and endocrine therapy in 13%. The status of adjuvant therapy was unknown in 5%. Patients were observed for disease recurrence and death as previously described.[8] A total of 620 patients (31%) had experienced disease recurrence, and 734 (37%) had died by the time of analysis. The median follow-up period for patients still alive at the time of analysis was 65 months (range, 1 to 214 months).

LBA for ER

Breast tumor specimens were frozen in liquid nitrogen immediately after excision and then sent to the Steroid Receptor Laboratory at the University of Texas Health Science Center at San Antonio. The tumor tissues were pulverized in liquid nitrogen, and cytosols were prepared for the LBA as previously described.[9] From 1973 to 1984, ^{3}H estradiol was used as the labeled ligand. Since 1985, the standard multipoint DCCA had been modified to incorporate ^{125}I-labeled estradiol and ^{3}H-R5020 in a single assay, allowing for the simultaneous determination of both ER and progesterone receptor statuses. Tumors with an ER content of $\geq$ 3 fmol/mg protein (the limit of detection in this assay) were considered to be ER-positive, based on studies calibrated to clinical outcome.[10-12] The pulverized tissue that remained after the corticosteroid receptor assay was performed was stored at −70°C for future use.

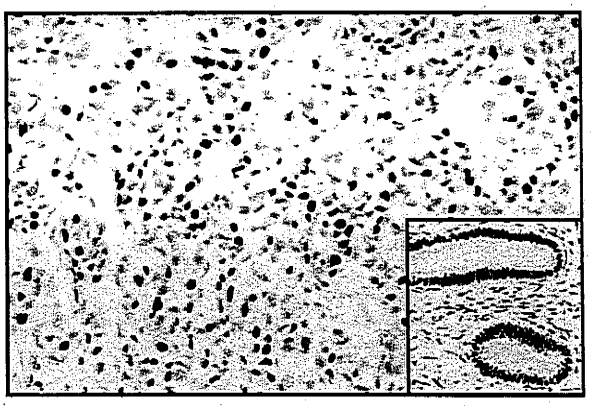

Fig 1. Photomicrograph of a representative invasive breast cancer tissue sample immunostained by the method used in this study (magnification, ×200). ER-positive cells showed a dark brown or black nuclear signal. Using this field, this tumor would get a total immunohistochemical score of 6 (proportion score [= 4] + intensity score [= 2]). The inset shows human endocervix tissue, which was used as a positive control because of its easy availability and relatively stable reactivity.

IHC for ER

Tissue sections for ER status determination by IHC were prepared from the pulverized frozen tumor specimens left over from the LBA as previously described.[13] Because of the ultracold temperature used during pulverization, the tissue was fractured into histologically intact fragments ranging from approximately 0.1 to 1.0 mm in size. Individual samples consisted of 100-mg pellets of this particulate tissue, which was fixed for 6 to 8 hours in 10% neutral buffered formalin and routinely processed to paraffin blocks. Histologic sections from these samples resemble the large-core needle biopsies in routine clinical use today.[13] The IHC assay was performed on 4 μm sections cut from the blocks and float-mounted on adhesive (silanized) glass slides. The essential techniques of the IHC assay included retrieving the antigen in 0.1 M boiling citrate buffer (pH 6.0) in a pressure cooker; blocking endogenous peroxidase with 0.1% sodium azide and 0.3% hydrogen peroxide; blocking nonspecific protein binding with 10% ovalbumin; binding with primary mouse monoclonal antibody 6F11 against the ER (Vector Laboratories, Burlingame, CA) at a dilution of 1:40 for 2 hours; linking with biotinylated rabbit antibody against mouse immunoglobulin G (Dako Corp, Carpenteria, CA) at a dilution of 1:100 for 30 minutes; enzyme labeling with streptavidin–horseradish peroxidase (Dako) at a dilution of 1:100 for 30 minutes; developing chromogen with 0.03% hydrogen peroxide and 1 mg/mL diaminobenzidine; enhancing the signal with 0.2% osmium tetroxide; and counterstaining with methyl green. Human endocervix tissue was used as a positive control because of its easy availability and relatively stable reactivity. The negative control consisted of nonimmune mouse immunoglobulin G substituted for the primary ER antibody. Controls were run with each batch of slides, at an average of approximately 50 slides per batch. The method produced a distinct nuclear signal in ER-positive tumor cells (Fig 1).

Immunostained slides were scored as previously described.[7,14] In brief, each entire slide was evaluated by light microscopy. First, a proportion score was assigned, which represented the estimated propor-

tion of positive-staining tumor cells (0, none; 1, < 1/100; 2, 1/100 to 1/10; 3, 1/10 to 1/3; 4, 1/3 to 2/3; and 5, > 2/3). Next, an intensity score was assigned, which represented the average intensity of positive tumor cells (0, none; 1, weak, 2, intermediate; and 3, strong). The proportion and intensity scores were then added to obtain a total score, which ranged from 0 to 8. Slides were scored by pathologists who did not have knowledge of ligand-binding results or patient outcome.

Two pathologists (J.M.H. and D.C.A.) were trained and calibrated to use of the IHC scoring system by simultaneously evaluating a panel of 200 breast cancer tissue samples that were immunostained for ER and which were not part of the study presented here. They then independently scored another 220 cases that were part of this study. Their results (total scores) on the second panel of tumors were in complete agreement in 71% of the cases and within one IHC score in the remaining 29% of the cases. The weighted kappa statistic for concordance was 0.87 ($P < .0001$). Taken together, these results indicated that the scoring method was easy to learn and highly reproducible. Because the concordance between the pathologists was so high during the training, all further scoring of cases in this study was carried out by one pathologist (J.M.H.).

Statistical Methods

Associations between continuous variables were analyzed using nonparametric Spearman rank correlation coefficients. Associations between categorical variables were assessed by χ^2 tests. Kappa statistics were used as measures of agreement between the different pathologists and between the two methods for determining ER status. An optimal cut point for defining ER positivity was determined by computing log-rank statistics for each of the seven possible cut points of the total IHC score. Adjustments were made to the resulting P values, as suggested by Hilsenbeck and Clark.[15] Univariate disease-free survival (DFS) and overall survival (OS) curves were estimated by the method of Kaplan and Meier and compared, using log-rank statistics. Cox proportional hazards regression models were created to assess the prognostic and predictive value of ER status in multivariate analyses. To adjust estimates of hazard ratios and their corresponding P values from Cox models for the multiple significance testing used to define the ER cut point, the following approach was used. The P value obtained from the Cox model was multiplied by seven (the number of possible cut points of the total IHC score). The Z statistic corresponding to this P value was obtained by inverting the cumulative normal distribution function. An adjusted parameter estimate for ER status was computed as the product of the Z statistic and the reported SE of the parameter estimate, based on the assumption that the bias associated with multiple significance testing primarily affects the magnitude of the parameter estimate rather than its SE. The adjusted hazard ratio and 95% confidence limits were obtained by exponentiation of the adjusted parameter estimate and its 95% confidence limits. Because all potential cut points are not biologically plausible and because this Bonferroni-type adjustment is known to be conservative, this technique probably overadjusts for the multiple significance testing used to define the IHC ER cut point. All analyses were performed using SAS (Version 6.11; SAS Institute, Cary, NC) on a Sun SparcServer 1000 system (Sun Microsystems, Inc, Mountain View, CA).

RESULTS

Agreement Between IHC and LBA for ER

A comparison of the distribution of IHC scores and ligand-binding values for ER status in the 1,982 tumors in this study is listed in Table 1. A nuclear signal for ERs, as assessed by IHC, was observed in 74% of the tumors, with positive scores ranging from 2 to 8. The mean and median ligand-binding values for the same group of tumors increased monotonically as the IHC score increased, although there was considerable variability among tumors with the same IHC score. The Spearman rank correlation coefficient between the two techniques was 0.68 ($P < .0001$).

Table 1. Comparison of ER Status Results, as Determined by IHC and LBA in 1,982 Primary Breast Cancer Cases

IHC Score	Patients		Ligand Binding Results (fmol/mg protein)				
	No.	%	Mean	SD	Median	Minimum	Maximum
0	517	26	10	49	1	0	758
2	67	3	50	100	8	0	548
3	117	6	59	95	23	0	623
4	190	10	67	73	39	0	428
5	320	16	104	139	56	0	1549
6	370	19	141	158	89	0	1181
7	318	16	193	215	142	0	1798
8	83	4	282	312	185	0	1439

Defining ER Positivity by IHC

To identify a clinically meaningful cut point for defining ER-positive tumors, we examined DFS curves for all possible IHC scores within the different treatment groups. For patients receiving no systemic adjuvant therapy (n = 701), ER status was only a weak prognostic factor, as expected. The log-rank P value for the best cut point (IHC score > 4) in untreated patients was only marginally significant ($P = .024$) and became nonsignificant ($P = .20$) after adjustment for multiple significance testing. For patients who received adjuvant chemotherapy alone (n = 407), no significant cut points were identified (all $P > .40$). However, for patients who received adjuvant endocrine therapy, either alone (n = 517) or in combination with chemotherapy (n = 260), ER status was a highly significant predictor of DFS. For these latter two groups combined (n = 777), the best cut point (IHC score > 2) was highly significant ($P < .0001$) and remained so ($P < .01$) after adjustment for multiple significance testing. On the basis of these results, tumors were defined as ER-positive if their total IHC score was greater than 2 and ER-negative if their score was 0 or 2. Note that a total score of 3, the lowest possible positive score, corresponds to as few as 1% to 10% weakly staining tumor cells. When this definition was applied to the 220 training cases that were independently scored by both study pathologists, only two cases (1%) showed a discrepancy (ie, positive versus negative) in ER status, and in both cases, the tumors received a score of 2 by one pathologist and a score of 3 by the other (Fig 2).

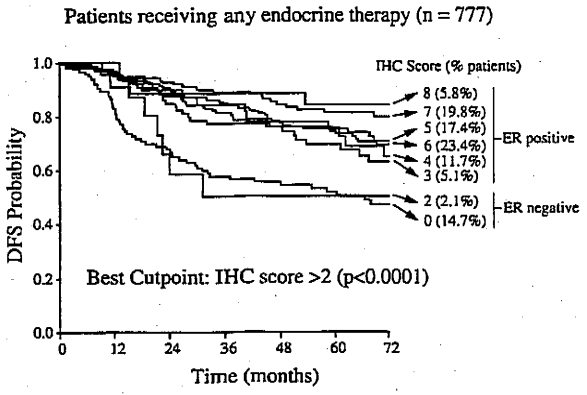

Fig 2. Univariate DFS curves for all possible total IHC scores in patients receiving any adjuvant endocrine therapy (almost always tamoxifen). An IHC score > 2 was the optimal cut point for predicting significantly improved outcome ($P < .0001$), and this value was used to define ER positivity throughout the study.

Using this definition of ER positivity, 70.5% of the 1,982 tumors in this study were ER-positive by IHC (ie, total score > 2), compared with 78.9% by LBA. Overall concordance between the tests was 85.5%. The kappa statistic for concordance was 0.62 ($P < .0001$). The remaining 14.5% of tumors had discordant results that fell into two groups. One group, comprising 11.4% of the tumors, was positive by LBA and negative by IHC. The LBA values were low (3 to 9 fmol/mg) in the majority of these tumors, but there was no overriding histologic explanation, such as the presence of ER-positive benign epithelium, to account for this discordant phenotype. The other group, comprising 3.1% of the tumors, was negative by LBA and positive by IHC. Again, there was no pervasive histologic feature, such as rare ER-positive tumor cells, that explained this discordant phenotype. When a cut point of 10 fmol/mg was used to define ER positivity by LBA, a standard used in many reference laboratories worldwide, the concordance between LBA and IHC assays increased slightly, to 87.7%.

Associations of ER by IHC With Other Standard Prognostic Factors

Table 2 shows the associations between ER status by IHC and other standard prognostic factors. Patients with positive axillary lymph nodes or with tumors larger than 2 cm in diameter had reduced frequencies of ER positivity ($P = .005$ and $P < .001$, respectively). ER positivity increased with advancing age, from 46% in patients younger than 35 years of age at diagnosis, to 65% in patients 35 to 65 years of age, to 82% in patients older than 65 ($P < .001$). Because this study was based on patients who had not been randomized to treatment, the rates of ER positivity differed with treatment status, as expected. For example, only 43% of patients who received chemotherapy alone had ER-positive tumors, compared with 88% of patients who received endocrine therapy alone.

Clinical Utility of Assessing ER by IHC Versus LBA

The associations between ER status and clinical outcome were independently evaluated for IHC with unadjusted cut points; for IHC with adjusted cut points; for LBA using a cut point of 3 fmol/mg protein (LBA3, our clinically validated laboratory standard for 15 years); and for LBA using a cut point of 10 fmol/mg protein (LBA10, a common international laboratory standard) (Table 3). All analyses were adjusted for the contributions of standard prognostic factors (including axillary lymph node status, tumor size, and patient age at diagnosis) by Cox modeling for proportional hazards regression.

In the subset of patients receiving no adjuvant therapy (n = 688), ER positivity by LBA10 showed a marginally significant association with improved DFS, whereas IHC, adjusted IHC, and LBA3 were not significantly associated with DFS. Positivity results determined by IHC, LBA3, and LBA10 all showed significant associations with prolonged OS, whereas the association with adjusted IHC was marginal. Overall, the fractional hazard ratios for all statistically significant associations observed in this nonrandomized, initially untreated group of patients were relatively large, emphasizing that ER status is a weak prognostic factor regardless of how it is measured, and were probably due in large part to responses to endocrine therapy given after first relapse in our study population.

In the subset of patients receiving adjuvant cytotoxic chemotherapy alone (n = 404), ER status by IHC, adjusted IHC, and LBA3 were not significantly related to DFS or OS.

Table 2. Relationships between ER Status Determined by IHC and Other Prognostic Factors

Factor	Patients		P
	No.	% ER-Positive	
Nodal status			.005
Node-negative	997	73	
Node-positive	985	68	
Tumor size, cm			< .001
≤ 2	667	78	
> 2	1294	67	
Patient age, years			< .001
< 35	81	46	
35-65	1181	65	
> 65	720	82	
Adjuvant therapy			< .001
None	701	72	
Chemotherapy alone	407	43	
Endocrine therapy alone	519	88	
Chemotherapy and endocrine therapy	261	73	

Table 3. Clinical Significance of ER Status Assessed by IHC; IHC Adjusted for Multiple Cut Points (IHC adj); LBA Using a Cut Point of 3 fmol/mg Protein (LBA3); and LBA Using a Cut Point of 10 fmol/mg Protein (LBA10)

	No Adjuvant Therapy (688 patients)					
	Disease-Free Survival (220 relapses)			Overall Survival (263 deaths)		
Factor	Hazard Ratio	95% CI	P	Hazard Ratio	95% CI	P
IHC	0.762	0.568-1.021	.069	0.685	0.518-0.906	.0079
IHC adj	0.900	0.672-1.207	.480	0.761	0.575-1.006	.056
LBA3	0.742	0.542-1.016	.062	0.793	0.586-1.075	.0001
LBA10	0.701	0.529-0.928	.013	0.679	0.519-0.887	.0046
	Chemotherapy only (404 patients)					
	Disease-Free Survival (149 relapses)			Overall Survival (154 deaths)		
Factor	Hazard Ratio	95% CI	P	Hazard Ratio	95% CI	P
IHC	1.008	0.734-1.383	.96	0.776	0.563-1.071	.12
IHC adj	1.008	0.734-1.383	.96	0.971	0.704-1.339	.86
LBA3	0.973	0.710-1.334	.86	0.823	0.597-1.134	.23
LBA10	0.748	0.536-1.043	.087	0.712	0.510-0.995	.047
	Endocrine Therapy Only (517 patients)					
	Disease-Free Survival (130 relapses)			Overall Survival (159 deaths)		
Factor	Hazard Ratio	95% CI	P	Hazard Ratio	95% CI	P
IHC	0.423	0.274-0.655	.0001	0.352	0.239-0.519	.0001
IHC adj	0.474	0.306-0.733	.0008	0.379	0.257-0.558	.0001
LBA3	0.707	0.356-1.404	.32	0.381	0.228-0.639	.0003
LBA10	0.699	0.426-1.145	.15	0.433	0.287-0.654	.0001
	Chemotherapy and Endocrine Therapy (260 patients)					
	Disease-Free Survival (98 relapses)			Overall Survival (91 deaths)		
Factor	Hazard Ratio	95% CI	P	Hazard Ratio	95% CI	P
IHC	0.491	0.320-0.753	.0011	0.502	0.315-0.801	.0038
IHC adj	0.559	0.365-0.858	.0078	0.590	0.370-0.944	.027
LBA3	0.513	0.307-0.856	.011	0.582	0.336-1.009	.050
LBA10	0.486	0.318-0.744	.0009	0.613	0.389-0.968	.036

NOTE. All analyses were adjusted for axillary lymph node status, tumor size, and age at diagnosis, by Cox modeling for proportional hazards regression.

LBA10 showed a marginally significant association with OS but was unrelated to DFS. Overall, the results in this nonrandomized group of high-risk patients initially treated with adjuvant chemotherapy also emphasize that ER status is a weak prognostic factor.

In clinical practice, ER status is used primarily as a predictive factor for response to adjuvant hormone therapy, rather than as a prognostic factor. In the subset of patients receiving adjuvant endocrine therapy alone (almost always tamoxifen; n = 517), ER positivity by IHC and adjusted IHC were both strongly associated with improved DFS (hazard ratios/P = 0.423/.0001 and 0.474/.0008, respectively) and OS (hazard ratios/P = 0.352/.0001 and 0.379/.0001, respectively). There were no significant associations between LBA3 or LBA10 and DFS, although ER positivity by both assays was associated with improved OS (hazard ratios/P = 0.381/.0003 and 0.433/.0001, respectively). Overall, the results in this group of nonrandomized but similar patients emphasize that ER status is a strong factor for predicting response to adjuvant endocrine therapy and that IHC is somewhat better than LBA in this setting.

In the subset of patients receiving combined adjuvant chemotherapy and endocrine therapy (n = 260), ER positivity results determined by IHC, adjusted IHC, LBA3, and LBA10 all showed strong and essentially equivalent correlations with improved DFS and OS, showing again that ER is a strong predictive factor for response to endocrine therapy.

DISCUSSION

ER and progesterone receptor statuses measured by LBAs were the only prognostic and predictive biomarkers recommended for routine clinical use in breast cancer by the Tumor Marker Panel of the American Society of Clinical Oncology.[3] In practice, their primary use is as predictive markers to distinguish patients who have little or no chance of benefiting from endocrine therapy from those who have some reasonable chance of benefiting. The justification for this endorsement was based on many studies conducted over the past two decades, involving patients in randomized clinical trials which showed that these tests were sufficiently sensitive, specific, and reproducible to reliably identify subsets of patients with significantly different risks for recurrence, survival, or treatment response.[2,7,16] Nonetheless, many problems associated with LBAs have become increasingly urgent over the years, including high cost, technical difficulty, reliance on radioactive reagents, and, especially, a need for relatively large amounts of fresh-frozen tumor tissue. In addition, because they are based on whole-tissue homogenates, they are somewhat insensitive and nonspecific in accounting for differences in the cellular composition of samples, such as rare tumor cells or contaminating benign cells that might be ER-positive. These problems motivated research into alternative methods of assessing ER status, including IHC. IHC has several potential advantages over LBA, including lower cost, easier technology, greater safety, the ability to evaluate a wide variety of samples (eg, fine-needle aspirates, frozen tissue, fixed archival tissue, etc), and higher sensitivity and specificity in the identification of rare tumor cells or contaminating benign cells under direct microscopic visualization.

Since appropriate antibodies became available over 10 years ago, many studies have used IHC to evaluate ER status in breast cancers. Several studies compared ER status measured in the same tumors, using both IHC and LBA, and found 80% to 90% agreement between these tests.[2,17] Many more studies, involving over 5,000 cumulative patients,

evaluated the relationship between ER status by IHC and clinical outcome in patients with breast cancer.[18-39] Nearly all of these studies showed a significant clinical benefit associated with the ER-positive phenotype, at least in univariate analyses and a few in multivariate analyses.[30,34,35] However, most of these studies involved patient populations of mixed clinical stage and treatment status, making it difficult to separate the prognostic from the predictive implications of their results.

The few studies that specifically addressed subsets of patients not receiving any type of systemic adjuvant therapy[19,27,33,35] found, on average, only approximately a 10% benefit in terms of DFS and/or OS associated with ER positivity as assessed by IHC, which is similar to results from earlier LBA studies and emphasizes that ER status is a very weak prognostic factor, regardless of how it is measured. The results of this study confirmed that ER status as determined by any method is a weak prognostic factor.

Several small studies have evaluated the ability of ER status determined by IHC to predict response to endocrine therapy in patients with advanced/metastatic breast cancer.[21,28,40-57] In these studies, cumulatively involving over 1,000 patients treated with a variety of endocrine therapies, an average of approximately 70% with ER-positive tumors showed a significant clinical response, whereas approximately 85% with ER-negative tumors did not, which was a little better than results with ER statuses measured by LBAs in some of the same studies.[7]

Much less is known about the ability of ER status determined by IHC to predict clinical outcome in the far larger number of patients with less advanced disease who receive adjuvant endocrine therapy, which was the primary focus of this study. In our study, multivariate analysis of the subset of patients receiving adjuvant endocrine therapy alone (almost always tamoxifen; n = 517) revealed that ER positivity determined by IHC was superior to that determined by LBA at predicting prolonged DFS (hazard ratios/ $P = 0.423/.0001$ *v* 0.707/.03, respectively) and roughly equivalent at predicting prolonged OS (hazard ratios/ $P = 0.352/.0001$ *v* 0.381/.0003, respectively). Fernö et al,[36] in—to our knowledge—the only other similar study, also showed that ER positivity determined by IHC in archival tissue predicted significantly improved DFS in 98 patients receiving adjuvant tamoxifen therapy alone.

In the sense that nearly all studies to date have shown some clinical significance to assessing ER status by IHC, this methodology is approaching clinical validation, relative to published guidelines.[1-3] However, there are still persistent shortcomings in the technical validation of this test. For example, these studies used many different antibodies (eg, H222, H226, D547, D75, 1D5) and a variety of usually arbitrary methods for scoring and interpreting results. In addition, the majority utilized frozen-section IHC with antibody H222, which is very expensive and relatively insensitive in archival tissue (which has become the standard in most laboratories).

Our study developed and used an IHC assay for measuring ER status, based on inexpensive, highly specific, commercially available reagents that are sensitive in archival tissue. We also developed a method of scoring results that was easy to learn and highly reproducible. Most importantly, the definition of ER positivity was calibrated to clinical outcome, in that it was based on the IHC score identifying the largest number of patients with significantly improved DFS in response to adjuvant endocrine therapy, the primary reason in clinical practice for measuring ER status. With minimal training, pathologists in our laboratory were in agreement on discriminating positive from negative tumors in 99% of cases.

The optimal cut point in our study was a total IHC score of greater than 2, meaning that even patients whose tumors scored 3 (corresponding to as few as 1% to 10% weakly positive cells) had a significantly improved response, compared with those who had lower scores, and tumors with scores of 3 comprised 6% of our total study population. Our low cut point by IHC essentially agrees with previous studies using LBA, in which ER levels as low as 4 to 10 fmol/mg protein were associated with significant rates of response to endocrine therapy.[10-12] There may be several explanations as to why such low IHC scores predict favorable clinical outcome, including the possibility that the sensitivity of the test underestimates the proportion of ER-expressing cells or that low scores correspond to an ER-positive stem-cell population. Our IHC cut point also provided clinically significant results in various subsets of our study population (eg, OS in untreated patients, DFS and OS in patients receiving endocrine therapy alone, and DFS and OS in patients receiving combined endocrine therapy and chemotherapy), which partially satisfies the recommendation that the utility of prognostic/predictive factor assays be demonstrated in "test" and "validation" sets of patients.[1]

Many hospital and commercial laboratories have converted to assessing ER status exclusively by IHC on archival tissue. They use diverse methodologies, and most have arbitrarily chosen 10% or even 20% positive tumor cells as their cutoff for defining ER positivity, potentially denying a substantial number of patients the benefits of adjuvant hormone therapy. Prudent laboratories offering ER status determination by IHC should perform rigorous validation studies themselves or follow the procedures of other laboratories that have.

REFERENCES

1. McGuire WL: Breast cancer prognostic factors: Evaluation guidelines. J Natl Cancer Inst 83:154-155, 1991 (editorial)
2. Clark GM: Prognostic and predictive factors, in Harris J, Lippman ME, Morrow M, et al (eds): Diseases of the Breast. Philadelphia, PA, Lippincott-Raven, 1996, pp 461-485
3. ASCO Tumor Marker Expert Panel: Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. J Clin Oncol 14:2843-2877, 1996
4. Knight WAI, Livingston RB, Gregory EJ, et al: Estrogen receptor as an independent prognostic factor for early recurrence in breast cancer. Cancer Res 37:4669-4671, 1977
5. Greene GL, Nolan C, Engler P, et al: Monoclonal antibodies to human estrogen receptor. Proc Natl Acad Sci U S A 77:5115-5119, 1980
6. King WJ, Greene GL: Monoclonal antibodies localize oestrogen receptor in the nuclei of target cells. Nature 307:745-747, 1984
7. Allred DC, Harvey JM, Berardo MD, et al: Prognostic and predictive factors in breast cancer by immunohistochemical analysis. Mod Pathol 11:155-168, 1998
8. Clark GM, Dressler LG, Owens MA, et al: Prediction of relapse or survival in patients with node-negative breast cancer by DNA flow cytometry. N Engl J Med 320:627-633, 1989
9. McGuire WL, De La Garza M, Chamness GC: Evaluation of estrogen receptor assays in human breast cancer tissue. Cancer Res 37:637-639, 1977
10. Clark GM, Osborne CK, McGuire WL: Correlations between estrogen receptor, progesterone receptor, and patient characteristics in human breast cancer. J Clin Oncol 2:1102-1109, 1984
11. Clark GM: Do we really need prognostic factors in breast cancer? Breast Cancer Res Treat 30:117-126, 1993
12. Clark GM, Wenger CR, Beardslee S, et al: How to integrate steroid hormone receptor, flow cytometric, and other prognostic information in regard to primary breast cancer. Cancer 71:2157-2162, 1993
13. Allred DC, Clark GM, Tandon AK, et al: Immunohistochemistry on histological sections from small (50 mg) samples of pulverized breast cancer. J Histotechnol 16:117-120, 1993
14. Allred DC, Clark GM, Elledge R, et al: Association of p53 protein expression with tumor cell proliferation rate and clinical outcome in node-negative breast cancer. J Natl Cancer Inst 85:200-206, 1993
15. Hilsenbeck SG, Clark GM: Practical p-value adjustment for optimally selected cutpoints. Stat Med 15:103-112, 1996
16. Osborne CK: Receptors, in Harris JR, Hellman S, Henderson IC, et al (eds): Breast Diseases. Philadelphia, PA, Lippincott, 1991, pp 301-325
17. Allred DC, Bustamante MA, Daniel CO, et al: Immunocytochemical analysis of estrogen receptors in human breast carcinomas: Evaluation of 130 cases and review of the literature regarding concordance with biochemical assay and clinical relevance. Arch Surg 125:107-113, 1990
18. DeSombre ER, Thorpe SM, Rose C, et al: Prognostic usefulness of estrogen receptor immunocytochemical assays for human breast cancer. Cancer Res 46:4256S-4264S, 1986 (suppl)
19. Walker KJ, Bouzubar N, Robertson J, et al: Immunocytochemical localization of estrogen receptor in human breast tissue. Cancer Res 48:6517-6522, 1988
20. Kinsel LB, Szabo E, Greene GL, et al: Immunocytochemical analysis of estrogen receptors as a predictor of prognosis in breast cancer patients: Comparison with quantitative biochemical methods. Cancer Res 49:1052-1056, 1989
21. Pertschuk LP, Kim DS, Nayer K, et al: Immunocytochemical estrogen and progestin receptor assays in breast cancer with monoclonal antibodies: Histopathologic, demographic, and biochemical correlations and relationship to endocrine response and survival. Cancer 66:1663-1670, 1990
22. Reiner A, Neumeister B, Spona J, et al: Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer. Cancer Res 50:7057-7061, 1990
23. Andersen J, Thorpe SM, King WJ, et al: The prognostic value of immunohistochemical estrogen receptor analysis in paraffin-embedded and frozen sections versus that of steroid-binding assays. Eur J Cancer 26:442-449, 1990
24. Cowen PN, Teasdale J, Jackson P, et al: Oestrogen receptor in breast cancer: Prognostic studies using a new immunohistochemical assay. Histopathology 17:319-325, 1990
25. Seymour L, Meyer K, Esser J, et al: Estimation of PR and ER by immunocytochemistry in breast cancer: Comparison with radioligand binding methods. Am J Clin Pathol 94:S35-S40, 1990 (suppl 1)
26. Andersen J, Thorpe SM, Rose C, et al: Estrogen receptor in primary breast cancer estimated in paraffin-embedded tissue. Acta Oncol 30:685-690, 1991
27. Querzoli P, Ferretti S, Marzola A, et al: Clinical usefulness of estrogen receptor immunocytochemistry in human breast cancer. Tumori 78:287-290, 1992
28. Robertson JFR, Bates K, Pearson D, et al: Comparison of two oestrogen receptor assays in the prediction of the clinical course of patients with advanced breast cancer. Br J Cancer 65:727-730, 1992
29. Hurlimann J, Gebhard S, Gomez F: Oestrogen receptor, progesterone receptor, pS2, ERD5, HSP27 and cathepsin D in invasive ductal breast carcinomas. Histopathology 23:239-248, 1993
30. Hanna W, McReady DR, Chapman JW, et al: The predictive value of ERICA in breast cancer recurrence: A univariate and multivariate analysis. Mod Pathol 6:748-754, 1993
31. Battifora H, Mehta P, Esteban JM: Estrogen receptor immunohistochemical assay in paraffin-embedded tissue: A better gold standard? Appl Immunohistochem 1:39-45, 1993
32. Esteban JM, Ahn C, Mehta P, et al: Biologic significance of quantitative estrogen receptor immunohistochemical assay by image analysis in breast cancer. Am J Clin Pathol 102:158-162, 1994
33. Kommoss F, Pfisterer J, Idris T, et al: Steroid receptors in carcinoma of the breast: Results of immunocytochemical and biochemical determination and their effects on short-term prognosis. Anal Quant Cytol Histol 16:203-210, 1994
34. Beck T, Weikel W, Brumm C, et al: Immunohistochemical detection of hormone receptors in breast carcinomas (ER-ICA, PgR-ICA): Prognostic usefulness and comparison with the biochemical radioactive ligand-binding assay (DCC). Gynecol Oncol 53:220-227, 1994
35. Stierer M, Rosen H, Weber R, et al: A prospective analysis of immunohistochemically determined hormone receptors and nuclear features as predictors of early recurrence in primary breast cancer. Breast Cancer Res Treat 36:11-21, 1995
36. Ferno M, Andersson C, Fallenius G, et al: Oestrogen receptor analysis of paraffin sections and cytosol samples of primary breast cancer in relation to outcome after adjuvant tamoxifen treatment. Acta Oncol 35:17-22, 1996

37. Layfield LJ, Saria EA, Conlon DH, et al: Estrogen and progesterone receptor status determined by the Ventana ES 320 automated immunohistochemical stainer and the CAS 200 image analyzer in 236 early-stage breast carcinomas: Prognostic significance. J Surg Oncol 61:177-184, 1996

38. Ferrer-Roca OF, Ramos A, Diaz-Cardama A: Immunohistochemical correlation of steroid receptors and disease-free interval in 206 consecutive cases of breast cancer: Validation of telequantification based on global scene segmentation. Anal Cell Pathol 9:151-163, 1995

39. Molino A, Micciolo R, Turazza M, et al: Prognostic significance of estrogen receptors in 405 primary breast cancers: A comparison of immunohistochemical and biochemical methods. Breast Cancer Res Treat 43:221-228, 1997

40. McCarty KSJ, Miller LS, Cox EB, et al: Estrogen receptor analyses: Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. Arch Pathol Lab Med 109:716-721, 1985

41. Pertschuk LP, Eisenberg KB, Carter AC, et al: Immunohistologic localization of estrogen receptors in breast cancer with monoclonal antibodies: Correlation with biochemistry and clinical endocrine response. Cancer 55:1513-1518, 1985

42. Ozzello L, DeRosa C, Habif DV, et al: Immunostaining of estrogen receptors in paraffin sections of breast cancers using anti-estrophilin monoclonal antibodies, in Ceriani RL (ed): Monoclonal Antibodies in Breast Cancer. Boston, MA, Martinus Nijhoff, 1985, pp 3-12

43. Jonat W, Maass H, Stegner HE: Immunohistochemical measurement of estrogen receptors in breast cancer tissue samples. Cancer Res 46:4296S-4298S, 1986 (suppl)

44. McClelland RA, Berger U, Miller LS, et al: Immunocytochemical assay for estrogen receptor: Relationship to outcome of therapy in patients with advanced breast cancer. Cancer Res 46:4241S-4243S, 1986 (Suppl)

45. Berger U, Mansi JL, Wilson P, et al: Detection of estrogen receptor in bone marrow from patients with metastatic breast cancer. J Clin Oncol 5:1779-1782, 1987

46. Burton GV, Flowers JL, Cox EB, et al: Estrogen receptor determination by monoclonal antibody in fine needle aspiration breast cancer cytologies: A marker of hormone response. Breast Cancer Res Treat 10:287-291, 1987

47. Coombes RC, Powles TJ, Berger U, et al: Prediction of endocrine response in breast cancer by immunocytochemical detection of oestrogen receptor in fine-needle aspirates. Lancet 2:701-703, 1987

48. De Lena M, Marzullo F, Simnone G, et al: Correlation between ERICA and DCC assay in hormone receptor assessment of human breast cancer. Oncology 45:308-312, 1988

49. Andersen J, Poulsen HS: Immunohistochemical analysis of estrogen receptors (ER) using formalin-fixed paraffin-embedded breast cancer tissue: Correlation with clinical endocrine response. J Steroid Biochem 30:337-339, 1988

50. Hawkins RA, Sangster K, Tesdale A, et al: The cytochemical detection of oestrogen receptors in fine needle aspirates of breast cancer: Correlation with biochemical assay and prediction of response to endocrine therapy. Br J Cancer 58:77-80, 1988

51. Gaskell DJ, Hawkins RA, Sangster K, et al: Relation between immunocytochemical estimation of oestrogen receptor in elderly patients with primary breast cancer and response to tamoxifen. Lancet 1:1044-1046, 1989

52. Andersen J, Poulsen HS: Immunohistochemical estrogen receptor determination in paraffin-embedded tissue: Prediction of response to hormonal treatment in advanced breast cancer. Cancer 64:1901-1908, 1989

53. Sklarew RJ, Bodmer SC, Pertschuk LP: Quantitative imaging of immunocytochemical (PAP) estrogen receptor staining patterns in breast cancer sections. Cytometry 11:359-378, 1990

54. McClelland RA, Finlay P, Walker KJ, et al: Automated quantitation of immunocytochemically localized estrogen receptors in human breast cancer. Cancer Res 50:3545-3550, 1990

55. Goulding H, Pinder S, Cannon P, et al: A new immunohistochemical antibody for the assessment of estrogen receptor status on routine formalin-fixed tissue samples. Hum Pathol 26:291-294, 1995

56. Pertschuk LP, Feldman JG, Kim YD, et al: Estrogen receptor immunocytochemistry in paraffin embedded tissues with ER1D5 predicts breast cancer endocrine response more accurately than H222Sp gamma in frozen sections or cytosol-based ligand-binding assays. Cancer 77:2514-2519, 1996

57. Barnes DM, Harris WH, Smith P, et al: Immunohistochemical determination of oestrogen receptor: Comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients. Br J Cancer 74:1445-1451, 1996

BRIEF COMMUNICATIONS

Real-World Performance of HER2 Testing—National Surgical Adjuvant Breast and Bowel Project Experience

Soonmyung Paik, John Bryant, Elizabeth Tan-Chiu, Edward Romond, William Hiller, Kyeongmee Park, Ann Brown, Greg Yothers, Steve Anderson, Roy Smith, D. Lawrence Wickerham, Norman Wolmark

Trastuzumab (Herceptin) provides clinical benefits for patients diagnosed with advanced breast cancers that have overexpressed the HER2 protein or have amplified the HER2 gene. The National Surgical Adjuvant Breast and Bowel Project (NSABP) Protocol B-31 is designed to test the advantage of adding Herceptin to the adjuvant chemotherapeutic regimen of doxorubicin and cyclophosphamide followed by paclitaxel (Taxol) in the treatment of stage II breast cancer with HER2 overexpression or gene amplification. Eligibility is based on HER2 assay results submitted by the accruing institutions. We conducted a central review of the first 104 cases entered in this trial on the basis of immunohistochemistry (IHC) results. We found that 18% of the community-based assays, which were used to establish the eligibility of patients to participate in the B-31 study, could not be confirmed by HercepTest™ IHC or fluorescence *in situ* hybridization (FISH) by a central testing facility. This report provides a snapshot of the quality of HER2 assays performed in laboratories nationwide. [J Natl Cancer Inst 2002;94:852–4]

Trastuzumab (Herceptin) is a humanized murine monoclonal antibody directed against the HER2 growth factor receptor, which provides clinical benefits for patients with metastatic breast cancer that overexpresses HER2 *(1,2)*. Several clinical trials are currently testing this therapy in combination with polychemotherapy in the adjuvant breast cancer setting. National Surgical Adjuvant Breast and Bowel Project (NSABP) Protocol B-31 compares four cycles of doxorubicin and cyclophosphamide followed by four cycles of paclitaxel (Taxol) to the same therapy combined with weekly Herceptin for a period of 1 year, beginning with the first cycle of paclitaxel (http://www.nsabp.pitt.edu/).

Eligibility for NSABP B-31 is based on HER2 assay results submitted by the accruing institutions. Until recently, assays from any accredited laboratory were accepted. Eligibility required a score of 3+ if the HercepTest™ (Dako HercepTest™; Carpinteria, CA) immunohistochemistry (IHC) assay was used, strong membrane staining of more than 33% of the tumor cells if other IHC assays were used, or gene amplification if fluorescence *in situ* hybridization (FISH) assays were used.

We tested the first 104 submitted cases for which eligibility was determined by using either HercepTest™ (n = 80) or other antibodies (n = 24) in IHC as part of the B-31 quality assurance program. Five-micrometer sections, cut from paraffin-embedded tumor blocks submitted by the accruing institutions, were centrally assayed by both the HercepTest™ and the PathVysion™ FISH assay (PathVysion™; Vysis, Inc., Downers Grove, IL) at Laboratory Corporation of America, Inc. (Research Triangle Park, NC). FISH results from the reference laboratory were validated by the NSABP Pathology Laboratory using a tissue array generated from a subset of cases (n = 81).

Assays submitted by the accruing institutions were confirmed to be strongly positive (3+) by central HercepTest™ in only 82 of 104 cases (79%; 95% confidence interval [CI] = 70% to 86%) (Table 1). They were confirmed positive for gene amplification by central FISH in 82 of 104 cases (79%; 95% CI = 70% to 86%). In 19 of 104 cases (18%; 95% CI = 11% to 27%), they were neither strongly positive by the HercepTest™ nor positive for gene amplification by central review. Among these 19 cases, 10 were scored 0 or 1+ and nine were scored 2+ by central HercepTest™.

To explain the lack of reproducibility between the accredited laboratory and the central testing facility, we examined the laboratories that performed the original assays according to the average volume of assays they perform (we used a cut point of 100 cases per month). There was less discrepancy with central HercepTest™ results in the large-volume laboratories (Table 1). Eighteen of 75 cases (24%) assayed as positive by the small-volume laboratories were found negative by both central assays, whereas only 1 of 29 cases (3%) assayed as positive by larger volume laboratories was found negative by the central assays. For small-volume laboratories, IHC assays other than the HercepTest™ could not be confirmed as positive more frequently (8 of 23 or 35% negative) than the HercepTest™ (10 of 52 or 19% negative). Large-volume laboratories used the HercepTest™ for 28 of 29 cases.

Altogether, 58 small-volume laboratories contributed 75 cases: 45 laboratories each contributed one, 10 laboratories each contributed two, two laboratories each contributed three, and one laboratory contributed four. The 18 negative assays came from 17 different laboratories (one laboratory contributed two cases). Nine large-volume laboratories contributed 29 cases: three laboratories each contributed one, three laboratories each contributed two, one laboratory contributed four, one laboratory contributed seven, and one laboratory contributed nine.

The concordance between central testing for FISH and HercepTest™ was good (98 of 104 cases in agreement; Table 2, A). To validate the central test-

Affiliations of authors: S. Paik, E. Tan-Chiu, W. Hiller, K. Park, R. Smith, D. L. Wickerham, N. Wolmark, National Surgical Adjuvant Breast and Bowel Project (NSABP) Operation Center, Pittsburgh, PA; J. Bryant, NSABP Biostatistical Center, Pittsburgh, Departments of Statistics and Biostatistics, University of Pittsburgh, PA; E. Romond, Division of Hematology/Oncology, Markey Cancer Center, University of Kentucky, Lexington; A. Brown, NSABP Biostatistical Center; G. Yothers, NSABP Biostatistical Center, and Department of Statistics, University of Pittsburgh; S. Anderson, Center for Molecular Biology and Pathology, Laboratory Corporation of America Holdings, Research Triangle Park, NC.

Correspondence to: Soonmyung Paik, M.D., Division of Pathology, National Surgical Adjuvant Breast and Bowel Project, Four Allegheny Center 5th Floor, East Commons Professional Bldg., Pittsburgh, PA 15212 (e-mail: soon.paik@nsabp.org).

See "Notes" following "References."

Table 1. Results from a central testing facility confirming original IHC assay results submitted by NSABP B-31 accruing institutions*

Test used for eligibility	Type of laboratory used	No. of negative cases detected by the central testing facility: HercepTest™†	PathVysion FISH‡
HercepTest™ 3+ (n = 80)	Small-volume§	10 of 52	12 of 52
	Large-volume‖	1 of 28	1 of 28
Other IHC assays (n = 24)¶	Small-volume	11 of 23	9 of 23
	Large-volume	0 of 1	0 of 1

*IHC = immunohistochemistry; NSABP = National Surgical Adjuvant Breast and Bowel Project; FISH = fluorescence *in situ* hybridization.

†HercepTest™ immunohistochemistry is scored on a three-point scale. For eligibility in NSABP B-31, a positive score of 3+ was required. A negative score was 0–2+.

‡PathVysion FISH is scored as either positive or negative for HER2 gene amplification.

§Small-volume laboratories were arbitrarily determined to perform no more than 99 tests per month.

‖Large-volume laboratories were arbitrarily determined to perform at least 100 tests per month.

¶Other IHC assays refers to any immunohistochemistry test that did not use the HercepTest™.

Table 2, A. Concordance between assays performed by the central testing facility (Lab Corp.)*

		PathVysion FISH†	
		Not amplified	Amplified
HercepTest™‡	3+	3	79
	0–2+	19 (18%)	3

B. Concordance between assays performed by the central testing facility (Lab Corp.) and those performed by the NSABP pathology laboratory

	PathVysion FISH by central testing facility	
	Not amplified	Amplified
FISH assay by NSABP pathology laboratory		
Not amplified	15	3
Amplified	1	62

*FISH = fluorescence *in situ* hybridization; NSABP = National Surgical Adjuvant Breast and Bowel Project.

†PathVysion FISH is scored as either positive or negative for HER2 gene amplification.

‡HercepTest™ immunohistochemistry is scored on a three-point scale. For eligibility in NSABP B-31, a positive score of 3+ was required. A negative score was 0–2+.

ing results, the NSABP Pathology laboratory also performed FISH on 81 of the cases (Table 2, B). The concordance between the two FISH assays was 77 of 81 (95%).

This brief communication provides a snapshot of the quality of HER2 assays nationwide. We found that an appreciable percentage of community-based assay results, which were used to establish the eligibility of patients to participate in B-31, could not be confirmed when tested in a central facility. These results may be surprising considering the studies *(3–12)* citing a high concordance between scores of 3+ in IHC and FISH. However, those studies were generally based on data obtained from laboratories with special expertise in HER2 research or from large-volume laboratories and, therefore, are consistent with our results showing good agreement between large-volume laboratories and central testing.

The reason for the trend favoring larger volume laboratories cannot be addressed directly because we have not performed a formal survey of laboratories. IHC results can vary substantially because of multiple factors, including time to fixation, duration of fixation, processing, antigen retrieval, staining procedure, and staining interpretation *(13)*. Because strongly positive (3+) cases represent only 15%–20% of newly diagnosed breast cancer cases, pathologists in small-volume laboratories may over-anticipate positive cases, leading to an interpretation bias. Such bias would be less likely to occur in a large-volume setting. Some U.S. laboratories have also recently introduced image analysis systems, which may improve the reproducibility of scoring.

The poor reproducibility of non-HercepTest™ IHC could be explained, in part, by the eligibility criteria that were used in the B-31 study. Some of the cases were enrolled on the basis of strong membrane staining of more than 33% of cells, which could have been 2+ intensity staining. Other antibodies can produce excellent results when used by qualified laboratories *(4,8)*.

FISH is generally accepted to be more reproducible than IHC for assessing HER2 status. Although studies demonstrate excellent portability when tested in multiple laboratories *(14,15)*, they used sections from a small number of cases or cell lines, which may not fully address potential problems associated with the variations in fixation and processing of tissue. In a real-world situation, where a limited number of cases are processed in small-volume laboratories, the reproducibility of FISH may require additional confirmation. Because only four cases were enrolled in B-31 on the basis of FISH assays that were performed before the analyses reported in this communication, it is not possible to comment on its reliability.

Our data suggest a need to improve quality control measures in laboratories that use IHC assays, including periodic testing for concordance with FISH. Given the cost and potential cardiotoxicity of Herceptin, it is reasonable to recommend that HER2 testing be done at large-volume reference laboratories. Since these data became available, we have implemented a laboratory approval process that considers both the laboratory volume and the quality of the assay. To date, 22 laboratories, all of which are experienced in both IHC and FISH, have been approved through this process. By performing both assays, quality can be cross-validated. We believe that such cross-validation may be the key to quality assurance of HER-2 assays performed in the community. In addition, all NSABP-approved laboratories use automated assay systems, probably reducing interassay variation. Accordingly, the NSABP has amended eligibility criteria for B-31: only patients whose tumors score 3+ by IHC performed by NSABP-approved reference laboratories or whose tumors demonstrate gene am-

plification by FISH from any laboratory would be allowed entry.

It is our position that the question of whether FISH or IHC is the better predictor of the response to Herceptin is still unanswered. Although the analysis of Mass et al. *(16)* suggested the superiority of FISH, the IHC used in that study was the Clinical Trials Assay. According to the package insert for Herceptin™ (http://www.gene.com/gene/products/information/oncology/herceptin/insert.jsp), concordance between the two assays is relatively poor, especially when the immunostaining is scored as 2+. Furthermore, the response of micrometastatic tumor cells in the adjuvant setting may be different from that of cancer cells in advanced disease, especially when given in combination with chemotherapy.

References

(1) Vogel C, Cobleigh MA, Tripathy D, Gutheil JC, Harris LN, Fehrenbacher L, et al. First-line, single-agent Herceptin(R) (trastuzumab) in metastatic breast cancer. a preliminary report. Eur J Cancer 2001;37 Suppl 1:25–9.

(2) Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 2001;344:783–92.

(3) Field AS, Chamberlain NL, Tran D, Morey AL. Suggestions for HER-2/neu testing in breast carcinoma, based on a comparison of immunohistochemistry and fluorescence *in situ* hybridisation. Pathology 2001;33:278–82.

(4) Thor A. HER2–a discussion of testing approaches in the USA. Ann Oncol 2001;12 Suppl 1:S101–7.

(5) Birner P, Oberhuber G, Stani J, Reithofer C, Samonigg H, Hausmaninger H, et al. Evaluation of the United States Food and Drug Administration-approved scoring and test system of HER-2 protein expression in breast cancer. Clin Cancer Res 2001;7:1669–75.

(6) Tubbs RR, Pettay JD, Roche PC, Stoler MH, Jenkins RB, Grogan TM. Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: apparent immunohistochemical false-positives do not get the message. J Clin Oncol 2001;19:2714–21.

(7) Lebeau A, Deimling D, Kaltz C, Sendelhofert A, Iff A, Luthardt B, et al. Her-2/neu analysis in archival tissue samples of human breast cancer: comparison of immunohistochemistry and fluorescence *in situ* hybridization. J Clin Oncol 2001;19:354–63.

(8) Bankfalvi A, Simon R, Brandt B, Burger H, Vollmer I, Dockhorn-Dworniczak B, et al. Comparative methodological analysis of erbB-2/HER-2 gene dosage, chromosomal copy number and protein overexpression in breast carcinoma tissues for diagnostic use. Histopathology 2000;37:411–9.

(9) Pauletti G, Dandekar S, Rong H, Ramos L, Peng H, Seshadri R, et al. Assessment of methods for tissue-based detection of the HER-2/neu alteration in human breast cancer: a direct comparison of fluorescence *in situ* hybridization and immunohistochemistry. J Clin Oncol 2000;18:3651–64.

(10) Hoang MP, Sahin AA, Ordonez NG, Sneige N. HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. Am J Clin Pathol 2000;113:852–9.

(11) Couturier J, Vincent-Salomon A, Nicolas A, Beuzeboc P, Mouret E, Zafrani B, et al. Strong correlation between results of fluorescent *in situ* hybridization and immunohistochemistry for the assessment of the ERBB2 (HER-2/neu) gene status in breast carcinoma. Mod Pathol 2000;13:1238–43.

(12) Ridolfi RL, Jamehdor MR, Arber JM. HER-2/neu testing in breast carcinoma: a combined immunohistochemical and fluorescence *in situ* hybridization approach. Mod Pathol 2000;13:866–73.

(13) O'Leary TJ. Standardization in immunohistochemistry. Appl Immunohistochem Mol Morphol 2001;9:3–8.

(14) Persons DL, Bui MM, Lowery MC, Mark HF, Yung JF, Birkmeier JM, et al. Fluorescence *in situ* hybridization (FISH) for detection of HER-2/neu amplification in breast cancer: a multicenter portability study. Ann Clin Lab Sci 2000;30:41–8.

(15) Masood S, Bui MM, Yung JF, Mark HF, Wong EY, Birkmeier JM, et al. Reproducibility of LSI HER-2/neu SpectrumOrange and CEP 17 SpectrumGreen Dual Color deoxyribonucleic acid probe kit. For enumeration of gene amplification in paraffin-embedded specimens: a multicenter clinical validation study. Ann Clin Lab Sci 1998;28:215–23.

(16) Mass RD, Press M, Anderson S. Improved survival benefit from Herceptin (trastuzumab) in patients selected by fluorescence *in situ* hybridization (FISH) [abstract]. Proc ASCO 2001;20:22a.

Notes

Supported by Public Health Service grants U10CA12027, U10CA69651, U10CA37377, and U10CA69974 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, and a grant from Genentech, Inc.

We thank Heather Theoret, Carole Donnelly, and Judy Jones at the NSABP Biostatistics Center for data management, Melanie Finnigan at the NSABP Division of Pathology for technical and data management support, Dr. Jeff Abrams at the National Cancer Institute, Drs. Robert Mass and Pamela Klein at Genentech for helpful comments, and Vysis, Inc., for providing the NSABP Division of Pathology with the fluorescence microscope used for the FISH analyses.

S. Paik is a member of Genentech's Pathology Advisory Panel, which provides recommendations for HER-2 testing. D. L. Wickerham is a member of AstraZeneca Speaker's Bureau.

Manuscript received November 29, 2001; revised April 1, 2002; accepted April 23, 2002.