

HER-2/*neu* in Breast Cancer: Interobserver Variability and Performance of Immunohistochemistry with 4 Antibodies Compared with Fluorescent *In Situ* Hybridization

Thomas A. Thomson, M.D., Malcolm M. Hayes, M.B.Ch.B., John J. Spinelli, Ph.D., Ernie Hilland, B.Sc., Christina Sawrenko, B.Sc., R.T., Don Phillips, R.T., Beverley Dupuis, R.T., Robin L Parker, M.D.

BC Cancer Agency (TT, MH, JS, EH, CS, DP), Vancouver Hospital (BD), and University of British Columbia (RP), Vancouver, British Columbia, Canada

The immunohistochemistry (IHC) performance of 4 anti-HER-2/*neu* antibodies was compared with fluorescent *in situ* hybridization (FISH) analysis of HER-2/*neu* gene expression in breast cancer patients considered for Herceptin (Trastuzumab) therapy. Interobserver variability in IHC interpretation was measured. Formalin-fixed tissue was received from 24 provincial hospital laboratories. The following anti-Her-2 antibodies were used: DAKO A0485 (polyclonal), Novacastra CB11 (monoclonal), Zymed TAB250 (monoclonal), and DAKO HercepTest (polyclonal). Additional sections were analyzed by FISH (Vysis). Three pathologists blinded to FISH results independently interpreted invasive tumor cell membranous staining on a scale of 0 to +3. The HER-2/*neu* gene was considered amplified when the FISH signal ratio of HER-2/CEP-17 was ≥ 2.0 . Blocks from all hospitals and of all ages were suitable for IHC and FISH analysis. No interlaboratory analysis variability was noted. The interobserver agreement (κ) for stain intensity for each antibody was good for 0 and +3 but poor for +1 and +2. Reasonable concordance between IHC and FISH was found with three of the four antibodies. TAB250 was the most sensitive antibody. For the three pathologists, the IHC sensitivities and specificities compared with FISH using 0/+1 as negative and +2/+3 as positive were as follows: A0485, 63–84/95–98; CB11, 63–66/97–98; TAB-250, 82–100/94–95; HercepTest,

59–77/91–93. The positive and negative predictive values varied by stain intensity. Stain scores of 0 and +3 were highly predictive of gene status. Stain scores of +1 and +2 were not sufficiently predictive to classify cases as amplified *versus* nonamplified. IHC is a reasonable first test to assess HER-2/*neu* status in patients with breast cancer. For most cases, DAKO A0485, TAB250, and HercepTest adequately predicted gene status. In cases with stain intensity of +1 or +2, the interobserver agreement is poor, and the predictive value is unsatisfactory for clinical use. Additional testing, preferably with FISH, is recommended.

KEY WORDS: Breast carcinoma, c-erbB2, Fluorescent *in situ* hybridization (FISH), HER-2/*neu*, Immunohistochemistry, Interobserver variability.

Mod Pathol 2001;14(11):1079–1086

HER-2/*neu* gene amplification or protein overexpression is seen in 20–30% of invasive breast cancer (1). Many studies have shown that this represents an adverse prognostic factor (2). Recent clinical studies have also suggested that overexpression may serve as a predictive marker of therapeutic response (3–5) to the monoclonal antibody, Trastuzumab (Herceptin, Genetech, Inc. CA), that directly targets the HER-2/*neu* protein. This new therapy highlighted the need for valid testing of HER-2/*neu* in our laboratory.

Immunohistochemistry (IHC) for HER-2/*neu* protein expression is the most attractive routine test based on issues of cost, convenience, and biological relevance. However, problems with variability in IHC staining have been reported. Specifically, differences in sensitivities and specificities among the various commercially available antibodies (6), variability in IHC interpretation (7), and technical artifact (8) have proven problematic. Although a true gold standard for testing does not exist, HER-

Copyright © 2001 by The United States and Canadian Academy of Pathology, Inc.

VOL. 14, NO. 11, P. 1079, 2001 Printed in the U.S.A.

Date of acceptance: July 12, 2001.

This work was supported by a grant from the Canadian Breast Cancer Foundation, B.C and Yukon Division. HercepTest kits were supplied without cost by DAKO Diagnostics Canada Inc. TAB250 was supplied by INTERMEDICO laboratory, Canada.

Presented in abstract form at the Canadian Association of Pathologists Annual Meeting, Ottawa, Ontario, Canada, June 17–21, 2000.

Address reprint requests to: Thomas A. Thomson, M.D., Department of Laboratory Medicine, BC Cancer Agency, Vancouver, British Columbia, Canada V5Z 4E6; e-mail: tthomson@bccancer.bc.ca.

2/*neu* gene amplification by fluorescent *in situ* hybridization (FISH) has been shown to correlate with protein overexpression by IHC and be similarly predictive and prognostic (9–14). The goal of this study was to assess the performance of IHC compared with that of FISH on fixed tissue from invasive breast carcinomas submitted to our laboratory from hospital laboratories in British Columbia, Canada, using several commercially available antibodies. Also, we sought to examine the effect of interobserver variability on IHC interpretation.

MATERIALS AND METHODS

Study Population

Study material was derived from patients with advanced breast cancer considered candidates for Herceptin therapy. Biopsy material was submitted from laboratories throughout the province of British Columbia to the BC Cancer Agency (BCCA) laboratory for HER-2/*neu* testing between January and December of 1999. Cases were initially tested by IHC with a polyclonal antibody without antigen retrieval (DAKO, c-erbB2, clone A0485). A proportion of these cases was difficult to classify, and these were considered by the original pathologist to be equivocal between negative and positive (IHC stain intensity, +1 *versus* +2). For this study, all of these equivocal cases plus a mix of negative (stain intensity 0 and +1) and positive cases (stain intensity +2 and +3) were chosen. Only cases with sufficient invasive carcinoma for multiple slides were included.

Paraffin blocks were obtained from breast biopsies ($n = 102$) and metastases to lymph nodes ($n = 11$), skin ($n = 3$), ovary ($n = 1$), and liver ($n = 1$). Blocks were received from 24 laboratories. All laboratories used 10% buffered formalin for primary tissue fixation. Duration of exposure to the fixative varied considerably, as typically encountered in surgical pathology practice. Two laboratories used postfixation treatment, one with alcohol and one with microwave. Two laboratories used a combination of alcohol and xylene in the tissue processor.

For each case, 4- μ m-thick tissue sections were cut from paraffin blocks. FISH was performed in the BCCA laboratory using the Vysis PathVysion HER-2/*neu* DNA probe Kit (Vysis, Inc. Downers Grove, IL). IHC was performed at the Vancouver Hospital with three anti-HER-2 antibodies (DAKO, c-erbB2, clone A0485; Zymed, HER2, clone TAB250; and Novocastra, HER2, clone CB11) and at the BCCA with HercepTest Kit (DAKO).

FISH for HER-2/*neu* Gene Amplification

FISH for HER-2/*neu* gene amplification was performed using the Vysis PathVysion HER-2/*neu* DNA

Probe Kit according to the manufacturer instructions. In brief, tissue sections were baked overnight at 56° C, deparaffinized in xylene, dehydrated in 100% ethanol, and air-dried. Pretreatment consisted of slide immersion in 0.2 N hydrochloric acid, a purified water wash, a wash in Vysis Pretreatment Wash Buffer and immersion in Vysis Pretreatment Reagent at 80° C for 30 minutes. Slides were then washed in purified water, washed in Vysis Wash Buffer, and immersed in Vysis Protease solution at 37° C for 20 minutes. Protease was removed with two changes of wash buffer and slides allowed to air dry. Ten microliters of PathVysion HER-2 Probe was added to the sample area of each section, and the section was coverslipped and the edges sealed with rubber cement. Slides were placed on a HyBrite (Vysis, Inc., Downers Grove, IL) for codenaturation using lengths of 5 minutes at 73° C to denature the DNA and 18 hours at 37° C to hybridize the probe and DNA. After hybridization, the rubber cement was removed, posthybridization solution was used to remove the coverslips, and slides were washed in posthybridization wash buffer and air-dried in the dark. Nuclei were counterstained with 10 μ l of DAPI/antifade and coverslipped. Slides were stored at -20° C for up to 7 days. Control slides, either negative or positive, were included in each Coplin jar of every run. Controls consisted of sections of cases of known HER-2 FISH status.

FISH Interpretation

Sections were examined with a fluorescent microscope (Zeiss Axioskop) using filter sets recommended by Vysis (DAPI/SpectrumOrange dual bandpass, DAPI/SpectrumGreen dual bandpass). A comparable hematoxylin and eosin (H&E) section was examined to assist with location of invasive tumor within each section. Forty malignant cell nuclei from invasive foci were scored for CEP-17 signal and HER-2 signal. Only nuclei with both signals interpretable were scored. A ratio of HER-2/CEP-17 was calculated for each specimen. A technologist and a pathologist independently enumerated each case.

IHC for HER-2/*neu* Protein Expression

Sections from each case were deparaffinized and rehydrated in graded alcohol solutions. Processing for each antibody was as follows: A0485: no antigen retrieval, dilution 1:25; TAB250: pretreatment with protease Type 1 (Ventana) for 4 minutes, dilution 1:50; CB11: no antigen retrieval, dilution 1:25; HercepTest: heat-induced antigen retrieval with temperature-controlled water bath. (Note: We now use antigen retrieval with A0485 as currently recommended by DAKO. At the time of this study,

DAKO was not recommending antigen retrieval for the A0485 antibody.)

At the Vancouver Hospital laboratory, all staining was performed on a Ventana ES Immunostainer (Ventana Medical Systems, Tucson, AZ), according to the manufacturer's instructions. The primary antibody was applied for 32 minutes. All cases were run with known positive- and negative-biopsy control material. The detection system used was AEC Detection Kit (Sigma), which included endogenous peroxidase inhibitor, biotinylated goat anti-mouse IgG/IgM, anti-rabbit IgG, peroxidase-labeled streptavidin, and AEC chromogen and substrate (hydrogen peroxide). Sections were counterstained with hematoxylin.

At the BCCA laboratory, staining for HER-2/*neu* protein expression was performed manually using the Dako HercepTest Kit according to the protocol in the manufacturer's guide. Paraffin sections were either freshly cut or were cut no longer than four weeks before the assay was performed. In brief, tissue sections were deparaffinized in xylene and rehydrated through a series of alcohols to distilled water. Epitope retrieval was in preheated (95° C) DAKO Epitope Retrieval Solution (0.01 mol/L citrate buffer, pH 6.0) in a water bath for 40 minutes. Peroxidase-blocking reagent was applied for 5 minutes. The primary antibody (prediluted rabbit anti-human antibody to HER-2/*neu* protein) was applied to test and positive-control sections for 30 minutes. The negative-control reagent (normal rabbit serum) was similarly applied to negative-control sections. The detection system consisted of DAKO visualization reagent (dextran polymer conjugated with horseradish peroxidase and goat anti-rabbit immunoglobulins) with 3,3'-diaminobenzidine chromogen solution. Sections were counterstained with hematoxylin. Controls for each run consisted of a negative and positive tissue control, a negative control for each patient, and the pellet control supplied with the kit.

IHC Interpretation

Slides were independently interpreted in a blinded fashion by an experienced breast pathologist (Observer 1), a general surgical pathologist (Observer 2), and a surgical pathology resident (Observer 3). Only invasive cancer was scored. Staining was interpreted on the maximum area of staining intensity as follows: 0 = no staining; +1 = weak, incomplete membranous staining; +2 = moderate, complete membranous staining of at least 10% of invasive tumor cells; and +3 = strong membranous staining of at least 10% of invasive tumor cells. Cases interpreted as 0 or +1 were considered negative, and cases interpreted as +2 or +3 were considered positive.

Statistical Methods

The distribution of the original A0485 antibody staining and FISH amplification for the study population were compared with the patient population using Pearson's χ^2 test. Patients with complete data (FISH and all four antibodies reviewed by each of the three observers) were compared with those with incomplete data using the *t* test for continuous measurements and the Pearson's χ^2 test for categorical measurements.

Positive predictive values (PPV) and negative predicted values (NPV) were calculated for each of the four antibodies and for each of the three observers. Sensitivity, specificity, false-negative proportion (1-sensitivity), and false-positive proportion (1-specificity) were also calculated using all possible cutpoints. Because the sensitivity and specificity were biased because of the enriched sampling scheme, adjusted sensitivity and specificity were computed using the following formulas. Let the T+, T- be positive and negative antibody stain for a particular cutoff, and F+, F- be amplified and nonamplified FISH assays. Then, the sensitivity, the probability that the antibody stain is positive given that FISH is amplified, can be written $\Pr(T+ | F+)$. Using Bayes' theorem, and with an estimate of the prevalence of a positive test, $\Pr(T+)$, the sensitivity can be expressed as a function of the PPV, $\Pr(F+ | T+)$ and the NPV, $\Pr(F- | T-)$, using that cutpoint:

$$\Pr(T+ | F+) = \frac{\text{PPV } \Pr(T+)}{\text{PPV } \Pr(T+) + (1 - \text{NPV}) \Pr(T-)}$$

The distribution of A0485 in the population (mixed, single reviewers) was used as the population prevalence for all antibodies. Receiver-operator characteristic curves (ROC) were also computed. For each observer and antibody stain, adjusted *c* statistics (the area under the ROC curve [AUC]) were computed.

Inter observer variability was assessed by multi observer kappa statistics(15).

RESULTS

FISH for HER-2/*neu* Gene Amplification

In total, 127 cases were assessed by FISH for HER-2/*neu* gene amplification and by IHC for protein expression. FISH analysis failed in 9 cases; 5 cases failed because of inadequate digestion, 3 cases failed because of loss of tumor from the section, and 1 case failed because of unknown reasons.

Of 118 cases with informative FISH results, 52 (44%) were gene amplified and 66 (56%) were not amplified. The FISH score (HER-2/CEP-17 ratio) for

amplified cases ranged from 2.0–11.0 (mean = 5.36).

Comparison of Study Group with Population

From June 1998 to February 2000, 484 cases were tested with the A0485 antibody. The results were negative (0) for 333 (68.8%), negative (+1) for 28 (5.8%), positive (+2) for 26 (5.4%), and positive (+3) for 97 (20%). By comparison, for the 114 patients with informative FISH analysis, the distribution of A0485 staining (Observer 1) was negative (0) for 46 (39.0%), negative (+1) for 20 (17.5%), positive (+2) for 14 (11.9%), and positive (+3) for 34 (28.8%). This was significantly different from the population distribution (Pearson $\chi^2 = 38.3$, $P < .001$). In addition, compared with the 26% prevalence of positive HER-2 by IHC in our laboratory, 44% of the study cases were gene amplified by FISH. Therefore, this study set was enriched with positive cases and cases equivocal between +1 and +2 over that encountered in our clinical practice.

Comparison of Complete and Incomplete Data

Complete data (FISH and 4 antibody results for 3 observers) was available in 95 cases. Statistical analysis was done on this data set. In 23 cases, a complete set of IHC stains was not available, as the original blocks could not be recovered for analysis. The profile of the complete (95 cases) and incomplete (118 cases) data sets were similar (Table 1). For the 95 cases with complete results, 43% (41 cases) were gene amplified.

Comparison of FISH and IHC Assays

Table 2 shows the complete data set (95 cases) for each observer for FISH score and antibody stain results.

The overall interobserver agreement (Kappa) for each antibody was as follows: A0485, 64.2; CB11, 74.4; TAB250, 79.2; and HerceptTest, 69.5 (Table 3). Although interobserver agreement was high (77–

TABLE 1. Comparison of Complete and Incomplete Cases

Parameter	Complete (n = 95)	Incomplete (n = 23)	P Value
Gene amplified by FISH	43%	48%	0.69
Mean Her-2 amplification	3.00	3.11	0.86
Mean FISH score (Her-2/CEP-17)	2.16	2.27	0.72
Mean FISH score >2	5.38	5.28	0.88
Specimen site—breast	70%	81%	0.14
FISH Score			
≤1	18%	13%	0.14
1.01–1.99	39%	39%	—
2–3.99	16%	17%	—
4+	27%	30%	0.95

FISH, fluorescent in situ hybridization.

TABLE 2. Cross-Tabulation of FISH Score by Antibody Staining for Each Observer

Fish Score	≤1	1–2	2–4	>4	Total
A485					
Observer 1					
0	12	25	1	0	38
1	4	9	1	1	15
2	0	2	3	7	12
3	1	1	10	18	30
Observer 2					
0	13	28	1	2	44
1	3	5	2	5	15
2	0	4	5	6	15
3	1	0	7	13	21
Observer 3					
0	14	29	1	2	46
1	3	6	1	5	15
2	0	2	3	3	8
3	0	0	10	16	26
Total	17	37	15	26	95
T250					
Observer 1					
0	15	19	0	0	34
1	2	10	3	0	15
2	0	6	2	2	10
3	0	2	10	24	36
Observer 2					
0	15	19	0	0	34
1	2	11	0	0	13
2	0	4	4	1	9
3	0	3	11	25	39
Observer 3					
0	13	16	0	0	29
1	4	13	0	0	17
2	0	6	4	1	11
3	0	2	11	25	38
Total	17	37	15	26	95
CB11					
Observer 1					
0	16	35	6	2	59
1	1	0	0	2	3
2	0	1	4	7	12
3	0	1	5	15	21
Observer 2					
0	15	35	5	3	58
1	2	0	2	2	6
2	0	1	3	5	9
3	0	1	5	16	22
Observer 3					
0	16	33	2	2	53
1	1	2	4	3	10
2	0	1	4	4	9
3	0	1	5	17	23
Total	17	37	15	26	95
HERC					
Observer 1					
0	13	25	0	1	39
1	3	4	3	0	10
2	0	5	0	5	10
3	1	3	12	20	36
Observer 2					
0	13	27	1	1	42
1	2	4	2	3	11
2	1	3	2	4	10
3	1	3	10	18	32
Observer 3					
0	13	25	2	1	41
1	2	5	1	6	14
2	2	4	1	6	13
3	0	3	11	13	27
Total	17	37	15	26	95

95.6) for 0 and +3 staining, it was generally poor (32.8–59.1) for +1 and +2 staining (Table 3). When combined scores of 0 and +1 were interpreted as negative and combined scores of +2 and +3 inter-

TABLE 3. Reliability (Kappa) with 95% Bootstrap Confidence Interval for Each of the Four Antibodies

Stain Intensity	A0485	TAB250	CB11	HercepTest
0	0.788 (0.678, 0.872)	0.875 (0.783, 0.949)	0.854 (0.746, 0.928)	0.871 (0.768, 0.942)
+1	0.393 (0.233, 0.581)	0.552 (0.375, 0.721)	0.328 (0.127, 0.625)	0.350 (0.171, 0.575)
+2	0.383 (0.170, 0.620)	0.516 (0.284, 0.726)	0.591 (0.363, 0.801)	0.384 (0.194, 0.600)
+3	0.770 (0.642, 0.875)	0.956 (0.883, 0.986)	0.822 (0.689, 0.915)	0.811 (0.700, 0.895)
All	0.642 (0.555, 0.737)	0.792 (0.709, 0.865)	0.744 (0.655, 0.830)	0.695 (0.606, 0.781)

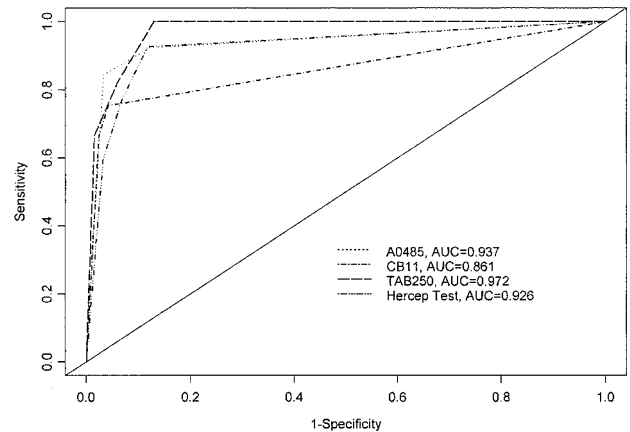
interpreted as positive, the kappa values were as follows: A0485, 79.5; CB11, 96.8; TAB250, 86.0; and HercepTest, 81.6.

The antibody sensitivity, specificity, and *c* statistic (AUC) for each observer are shown in Table 4. The results suggest similar overall test performance for A0485, HercepTest, and TAB250. The CB11 antibody had the lowest sensitivity for all observers (range, 63–66) but good specificity (range, 97–98). TAB250 was the most sensitive antibody (range, 82–100%) and showed very good specificity (range, 94–95%). These results are shown graphically as a sample ROC (Fig. 1) for the 4 antibodies for one observer.

The variation in individual-observer positive predictive value of a positive test (PPV) and negative predictive value of a negative test (NPV) for each level of membrane staining with each antibody is shown in Table 5. The NPVs and PPVs, respectively, were high for all antibodies when staining was absent (0) and when staining was strongly positive (+3). However, for TAB250 and HercepTest, a stain intensity of +2 was poorly predictive of gene amplification (TAB250, 40–56%; HercepTest, 50–54%), indicating a high false-positive rate for these antibodies at this stain level. Also, for CB-11 and HercepTest, +1 stain intensity was poorly predictive of no gene amplification (30–33% and 50–70%, respectively), indicating an increased false-negative rate at this stain level. Although the numbers are small, the data for CB11 suggest that any membrane stain at all favors gene amplification.

Cases Nonconcordant between IHC and FISH

For the FISH amplified cases, only 1 case (Her-2/Cep17 = 3.3) had no evidence of protein overexpression with all 4 antibodies. For the FISH nonamplified cases, 13 cases showed protein expression by at least one antibody. One case was IHC positive

**FIGURE 1.** Receiver operator curve (ROC) for Observer 1 with the four antibodies.

with all 4 antibodies (Her-2/Cep 17 = 1.3), one case was IHC positive with 3 of the 4 antibodies (Her-2/Cep-17 = 1.5), and two case were positive with 2 of 4 antibodies (HER-2/CEP = 1.2 & 1.3). In the remaining 9 cases, only one antibody was positive.

Cases without IHC Consensus

In 10 cases, each observer interpreted the stain intensity differently. These cases were reviewed by all observers at a multiheaded microscope. Heterogeneity of staining was noted in all these biopsies. Specific problems noted (Table 6) included enhanced edge staining on small biopsies, excess cytoplasmic staining mimicking membrane staining, and nonspecific staining of stroma around cell groups but not around individual cell membranes. Observers varied in the interpretation of equivocal staining (+1vs+2?) with some “scoring up” to positive (+2) and others “scoring down” to negative (+1).

TABLE 4. Range of Antibody Sensitivities and Specificities and *c*-Statistics (AUC) for the Three Observers*

Antibody	Obs 1			Obs 2			Obs 3		
	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
A0485	84	97	94	63	95	88	68	98	90
TAB250	82	94	97	100	95	98	100	95	99
CB11	66	98	86	63	97	85	65	98	90
HT	77	93	93	68	93	89	59	91	86

AUC, area under curve; HT, HercepTest.

* Values are adjusted to compensate for bias in the study group (see Methods and Discussion).

TABLE 5. Antibody Performance (Positive Predictive Value = PPV; Negative Predictive Value = NPV) for Individual Observers

Antibody	Stain Intensity	Observer 1		Observer 2		Observer 3	
		PPV	NPV	PPV	NPV	PPV	NPV
A0485	0		97		93		94
	1		87		53		60
	2	83		73		75	
	3	93		95		100	
TAB250	0		100		100		100
	1		80		80		100
	2	40		56		46	
	3	94		92		95	
CB11	0		86		86		93
	1		33		33		30
	2	92		89		89	
	3	95		96		96	
HercepTest	0		97		95		93
	1		70		55		50
	2	50		54		54	
	3	89		89		89	

TABLE 6. Factors Contributing to Interobserver Variability in IHC Interpretation with All Antibodies*

Apparent Causes of Variation in IHC Interpretation	No. of Cases
Heterogeneity of staining	6
Edge artifact	5
Stroma retraction artifact	4
Nonspecific staining of stroma and epithelium	3
Cytoplasmic staining	1
Pseudomembranous staining	1
Observer error	1

* In several cases, more than one factor was identified.

DISCUSSION

Clinical demand for HER-2/*neu* assessment is increasing, and there is a need for standardization of HER-2/*neu* assessment and reporting so that results are comparable between laboratories (16). As a reference hospital for the treatment of breast cancer, the BC Cancer Agency was particularly interested in defining a reliable approach to assess HER-2/*neu* status for the province of British Columbia, Canada.

The study goal was to examine the ability of IHC using four antibodies to predict gene amplification as defined by FISH analysis. An ideal antibody test would have near-perfect interobserver agreement on stain interpretation, would be sufficiently sensitive to capture all possible treatment candidates, and would have a high PPV (low false-positive rate) to minimize over treatment. None of the antibodies approached this ideal.

It should be emphasized that although a sampling bias was introduced in enriching the study with equivocal cases, this bias was corrected using statistical methods allowing comparison to our normal population prevalence of HER-2 positive cases.

Three antibodies (A0485, TAB250 and HercepTest) were found to have similar performance. TAB250 had the highest AUC for all three antibodies

(Table 4). Although TAB250 appeared to perform slightly better, in practical terms, this was not statistically significant. With all 3 antibodies, absent staining (score of 0) was highly predictive of a nonamplified case (NPV, 93–100%), and strong complete membrane staining (+3) was highly predictive of an amplified case (PPV, 89–100%). Stain intensity scores of +1 or +2 were variably predictive of gene amplification. Others have recently reported similar findings (16), and for this reason, we recommend FISH analysis for cases scored as +1 or +2.

Interobserver agreement varied with stain intensity. For stain intensities interpreted as 0 or +3, interobserver agreement was generally high. In contrast, interobserver agreement was generally poor for stain intensities interpreted as +1 or +2. This is significant, as distinction between these staining intensities is clinically relevant. At present, patients with breast cancers showing +1 staining for Her-2 protein are usually not offered Herceptin therapy, whereas those with +2 staining often are. Interobserver variability at these staining intensities further supports a role for FISH in evaluating cases scored as +1 and +2.

In 10 cases, there was no agreement on stain interpretation reflecting the number of equivocal cases in the study set. A particular problem was noted with small and/or fragmented biopsies in which significant edge enhancement of stain could occur. It would be prudent in these cases, if not clearly negative (0) or positive (+3), to repeat the test on another block and/or consider FISH analysis.

Another possible source of interobserver variation was suggested after the study was complete. The initial observer, who always had a positive control slide for comparison, noted that occasionally, the stain intensity of the positive control slide var-

ied. Because control comparison in the study was not rigidly adhered to by all observers, this may have accounted for some interpretation variation.

In this study, sections from tissues up to 18 years old were received from 24 different laboratories. Despite concern that variables in tissue fixation, processing, and embedding would be significant, tissue from all 24 laboratories was suitable for testing with both IHC and FISH.

Variability in sensitivity and specificity of commercially available HER-2/*neu* antibodies used with IHC has been previously noted (6, 17). In the study by Press *et al.* (6), the sensitivities and specificities of 28 antibodies were compared and showed variation in sensitivities ranging from 6–82% and specificities of 92–100%. In the present study, the sensitivities were found to vary between observers with the same antibody and between antibodies. Overall, the sensitivity ranged from 59 to 100% and the specificity ranged from 91–98%. CB11 showed the highest specificity (97–98%) but a low sensitivity (63–66%), and thus was unsuitable as a primary test. TAB250 had the highest sensitivity (82–100%) and very good specificity (94–94%). The negative predictive value of TAB250 was consistently high between observers (100% for absent stain and 80–100% for +1 stain). This suggests that if another antibody was used as the first test, TAB250 may have use as a second IHC test in cases with equivocal results (+1 *versus* +2?). HercepTest and A0485 showed a similar range of sensitivities between observers (A0485, 76–93%; HercepTest, 76–90%), but HercepTest was slightly less specific (A0485, 91–96%; HercepTest, 83–85%) and had a lower PPV and NPV than A0485.

With all antibodies, cases were interpreted as IHC negative that were FISH amplified. However, in only one FISH-amplified case did all 4 antibodies indicate that no protein overexpression was present. This one case may reflect a technical problem with the specimen or a true failure of protein production. For the other cases, the therapeutic significance of this discrepancy is not yet clear.

The performance of HercepTest, the only approved IHC method of HER-2 testing, has recently been questioned (18). The study reported a high level of staining of normal epithelium and a high number of apparent false-positive tests compared with FISH and two other antibodies. This may have been a result of the use of alcoholic formalin as a primary fixative. DAKO specifically advises the use of neutral-buffered formalin in the HercepTest package insert. All our cases were initially fixed with buffered formalin, and staining of normal epithelium was not a problem, emphasizing the importance of the initial fixative. We consider membranous staining of normal epithelium unacceptable for stain interpretation. In contrast, others have

found HercepTest to perform well (19). In the present study, HercepTest sensitivity and specificity were comparable to A0485 and TAB250. However, with TAB250 and HercepTest, the +2 stained cases had a low PPV in predicting gene amplification, and if +2 stain were interpreted as positive, more patients with FISH nonamplified status would be treated. In addition, the NPV of +1 with HercepTest and A0485 was poor.

Antigen retrieval may increase the sensitivity of IHC. However, using the same primary polyclonal antibody, this study found comparable results with antigen retrieval (HercepTest) and without antigen retrieval (A0485). We now use A0485 with a microwave antigen retrieval method as specified by the manufacturer. A higher antibody dilution is required to avoid excess cytoplasmic and background staining. This again highlights the need for controls that include nonstaining benign epithelium.

An unknown in this study is the number of false-negative FISH cases. It is possible that some of the IHC false-positive cases were actually FISH false-negative cases. Problems we noted in the interpretation of FISH staining included weak signal interpretation, variability in signal intensity over the section, separation of *in situ* from invasive cancer cells, and loss of tissue architecture with digestion. The biological significance of the IHC positive–FISH nonamplified cases remains to be seen, although a recent report (13) suggests that the FISH status may be relevant to therapy response.

There is a need to correlate clinical response to Herceptin therapy with Her-2 protein and gene expression levels. As this remains a critical issue in predictive testing, until we have examined our own outcome data in terms of test results and response to Herceptin therapy, we propose to continue to collect both IHC and FISH data on Herceptin-treated cases.

CONCLUSIONS

The results of this study can be summarized as follows.

1. Provided that initial fixation of tissue is with 10% buffered formalin, blocks from all hospitals and of all ages were suitable for IHC and FISH analysis.

2. For DAKO A0485, TAB250, and HercepTest there was generally good concordance between IHC and FISH.

3. Interobserver agreement for stain interpretation was high for the 4 antibodies when there was negative (0) or strong (+3) staining but was poor for weak (+1) and moderate (+2) staining.

4. The positive and negative predictive values varied by stain intensity. For most antibodies, stain

intensities of +1 and +2 were not sufficiently predictive, suggesting that FISH should be added as a second test. Intensity should always be compared with a positive (+3) control slide that includes non-staining normal epithelium.

5. Sections with scant invasive carcinoma, marked tissue fragmentation, excess cytoplasmic staining, or staining of normal epithelium should be rejected. Repeat IHC staining on a more suitable section, with a lower antibody concentration, or FISH analysis should be performed as needed.

Acknowledgment: *The authors thank Dr. Karen Gelmon, clinical professor, Head of the Breast Tumor Group, BC Cancer Agency for her encouragement and effort in obtaining funding for this study.*

REFERENCES

1. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, *et al.* Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 1989;244:707–12.
2. Ross JS, Fletcher JA. HER-2/neu (c-erb-B2) gene and protein in breast cancer. *Am J Clin Pathol* 1999;112(1 Suppl 1):S53–67.
3. Paik S, Bryant J, Park C, Fisher B, Tan-Chiu E, Hyams D, *et al.* erbB-2 and response to doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breast cancer [see comments]. *J Natl Cancer Inst* 1998;90:1361–70.
4. Thor AD, Berry DA, Budman DR, Muss HB, Kute T, Henderson IC, *et al.* erbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer [see comments]. *J Natl Cancer Inst* 1998;90:1346–60.
5. Ross JS, Fletcher JA. The HER-2/neu oncogene in breast cancer: prognostic factor, predictive factor, and target for therapy. *Stem Cells* 1998;16:413–28.
6. Press MF, Hung G, Godolphin W, Slamon DJ. Sensitivity of HER-2/neu antibodies in archival tissue samples: potential source of error in immunohistochemical studies of oncogene expression. *Cancer Res* 1994;54:2771–7.
7. Kay EW, Walsh CJ, Cassidy M, Curran B, Leader M. C-erbB-2 immunostaining: problems with interpretation. *J Clin Pathol* 1994;47:816–22.
8. Penault-Llorca F, Adelaide J, Houvenaeghel G, Hassoun J, Birnbaum D, Jacquemier J. Optimization of immunohistochemical detection of ERBB2 in human breast cancer: impact of fixation. *J Pathol* 1994;173:65–75.
9. Press MF, Bernstein L, Thomas PA, Meisner LF, Zhou JY, Ma Y, *et al.* HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *J Clin Oncol* 1997;15:2894–904.
10. Jacobs TW, Yaziji H, Gown AM, Barnes MJ, Schnitt SJ. Comparison of fluorescence in situ hybridization and immunohistochemistry for the evaluation of HER-2/neu in breast cancer. *J Clin Oncol* 1999;17:1974–82.
11. Pauletti G, Godolphin W, Press MF, Slamon DJ. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* 1996;13:63–72.
12. Jimenez RE, Wallis T, Tabasczka P, Visscher DW. Determination of Her-2/neu status in breast carcinoma: comparative analysis of immunohistochemistry and fluorescent in situ hybridization [see comments]. *Mod Pathol* 2000;13:37–45.
13. Slamon. 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.
14. Sastre-Garau. 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.
15. Shoukri MM, Edge VL. *Statistical methods for health sciences.* Boca Raton, FL: CRC Press; 1996.
16. Clark GM. Should selection of adjuvant chemotherapy for patients with breast cancer be based on erbB-2 status? [editorial; comment]. *J Natl Cancer Inst* 1998;90:1320–1.
17. Hanna W, Kahn HJ, Trudeau M. Evaluation of HER-2/neu (erbB-2) status in breast cancer: from bench to bedside. *Mod Pathol* 1999;12:827–34.
18. Jacobs TW, Gown AM, Yaziji H, Barnes MJ, Schnitt SJ. 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* 1983;17:1983–7.
19. Espinoza F. The HercepTest assay. Another perspective. *J Clin Oncol* 1999;17:2293–4.