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.