

American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer

Antonio C. Wolff, M. Elizabeth H. Hammond, Jared N. Schwartz, Karen L. Hagerty, D. Craig Allred, Richard J. Cote, Mitchell Dowsett, Patrick L. Fitzgibbons, Wedad M. Hanna, Amy Langer, Lisa M. McShane, Soonmyung Paik, Mark D. Pegram, Edith A. Perez, Michael F. Press, Anthony Rhodes, Catharine Sturgeon, Sheila E. Taube, Raymond Tubbs, Gail H. Vance, Marc van de Vijver, Thomas M. Wheeler, and Daniel F. Hayes

From the American Society of Clinical Oncology, Alexandria, VA; and the College of American Pathologists, Northfield, IL.

Submitted September 20, 2006; accepted September 27, 2006; published online ahead of print at www.jco.org on December 11, 2006.

Antonio C. Wolff, M. Elizabeth H. Hammond, Jared N. Schwartz, and Daniel F. Hayes are American Society of Clinical Oncology/College of American Pathologists Expert Panel co-chairs.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Address reprint requests to American Society of Clinical Oncology, Cancer Policy and Clinical Affairs, 1900 Duke Street, Suite 200, Alexandria, VA 22314; e-mail: guidelines@asco.org.

This guideline was developed through a collaboration between American Society of Clinical Oncology and College of American Pathologists and has been published jointly by invitation and consent in both the *Journal of Clinical Oncology* and the *Archives of Pathology & Laboratory Medicine*. Copyright © 2006 American Society of Clinical Oncology and College of American Pathologists. All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission by the American Society of Clinical Oncology or College of American Pathologists.

0732-183X/07/2501-1/\$20.00

DOI: 10.1200/JCO.2006.09.2775

A B S T R A C T

Purpose

To develop a guideline to improve the accuracy of human epidermal growth factor receptor 2 (HER2) testing in invasive breast cancer and its utility as a predictive marker.

Methods

The American Society of Clinical Oncology and the College of American Pathologists convened an expert panel, which conducted a systematic review of the literature and developed recommendations for optimal HER2 testing performance. The guideline was reviewed by selected experts and approved by the board of directors for both organizations.

Results

Approximately 20% of current HER2 testing may be inaccurate. When carefully validated testing is performed, available data do not clearly demonstrate the superiority of either immunohistochemistry (IHC) or in situ hybridization (ISH) as a predictor of benefit from anti-HER2 therapy.

Recommendations

The panel recommends that HER2 status should be determined for all invasive breast cancer. A testing algorithm that relies on accurate, reproducible assay performance, including newly available types of brightfield ISH, is proposed. Elements to reliably reduce assay variation (for example, specimen handling, assay exclusion, and reporting criteria) are specified. An algorithm defining positive, equivocal, and negative values for both HER2 protein expression and gene amplification is recommended: a positive HER2 result is IHC staining of 3+ (uniform, intense membrane staining of > 30% of invasive tumor cells), a fluorescent in situ hybridization (FISH) result of more than six *HER2* gene copies per nucleus or a FISH ratio (*HER2* gene signals to chromosome 17 signals) of more than 2.2; a negative result is an IHC staining of 0 or 1+, a FISH result of less than 4.0 *HER2* gene copies per nucleus, or FISH ratio of less than 1.8. Equivocal results require additional action for final determination. It is recommended that to perform HER2 testing, laboratories show 95% concordance with another validated test for positive and negative assay values. The panel strongly recommends validation of laboratory assay or modifications, use of standardized operating procedures, and compliance with new testing criteria to be monitored with the use of stringent laboratory accreditation standards, proficiency testing, and competency assessment. The panel recommends that HER2 testing be done in a CAP-accredited laboratory or in a laboratory that meets the accreditation and proficiency testing requirements set out by this document.

J Clin Oncol 25. This guideline was developed through a collaboration between American Society of Clinical Oncology and College of American Pathologists and has been published jointly by invitation and consent in both the *Journal of Clinical Oncology* and the *Archives of Pathology & Laboratory Medicine*. Copyright © 2006 American Society of Clinical Oncology and College of American Pathologists. All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without written permission by the American Society of Clinical Oncology or the College of American Pathologists.

INTRODUCTION

The human epidermal growth factor receptor 2 gene *ERBB2* (commonly referred to as *HER2*) is amplified in approximately 18% to 20% of breast cancers.¹ *ERBB2* is the official name provided by the HUGO Gene Nomenclature Committee for the v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 gene that encodes a member of the epidermal growth factor receptor family of receptor tyrosine kinases. Several aliases have been used in the literature (for example, NEU, NGL, HER2, TKR1, HER-2, c-erb B2, HER-2/*neu*) and the panel opted to adopt the commonly used term HER2 throughout this article.² Amplification is the primary mechanism of HER2 overexpression and abnormally high levels of a 185-kd glycoprotein with tyrosine kinase activity are found in these tumors.³ HER2 overexpression is associated with clinical outcomes in patients with breast cancer.⁴⁻⁶ There are several possible uses of HER2 status. HER2 positivity is associated with worse prognosis (higher rate of recurrence and mortality) in patients with newly diagnosed breast cancer who do not receive any adjuvant systemic therapy. Thus, HER2 status might be incorporated into a clinical decision, along with other prognostic factors, regarding whether to give any adjuvant systemic therapy.

HER2 status is also predictive for several systemic therapies.⁶ In this regard, HER2 positivity appears to be associated with relative, but not absolute, resistance to endocrine therapies in general.⁷ Although controversial, preclinical and clinical studies have suggested that this effect may be specific to selective estrogen receptor modulator therapy, such as tamoxifen, and perhaps not to estrogen depletion therapies, such as with aromatase inhibitors.⁸ HER2 status also appears to be predictive for either resistance or sensitivity to different types of chemotherapeutic agents. HER2 may be associated with relative, but not absolute, lower benefit from nonanthracycline, nontaxane-containing chemotherapy regimens.⁹ In contrast, retrospectively obtained results from prospectively conducted randomized clinical trials appear more definitive in suggesting that HER2 positivity is associated with response to anthracycline therapy; although, this effect may be secondary to coamplification of HER2 with topoisomerase II, which is the direct target of these agents.¹⁰⁻¹³ Preliminary data also suggest that HER2 may predict for response and benefit from paclitaxel in either the metastatic or adjuvant settings.^{14,15}

Perhaps most importantly, several studies have now shown that agents that target HER2 are remarkably effective in both the metastatic and adjuvant settings. Trastuzumab (Herceptin; Genentech, South San Francisco, CA), a humanized monoclonal antibody, improves response rates, time to progression, and even survival when used alone or added to chemotherapy in metastatic breast cancer.¹⁶ Trastuzumab is also active as a single agent^{17,18} and was approved in 1998 by the US Food and Drug Administration for the treatment of metastatic disease.

Importantly, five international, prospective randomized clinical trials have demonstrated that adjuvant trastuzumab reduces the risk of recurrence and mortality by one half and one third, respectively, in patients with early-stage breast cancer.¹⁹⁻²³ Furthermore, recently reported results suggest that a small molecule dual HER1/HER2 tyrosine kinase inhibitor of HER2 tyrosine kinase activity, lapatinib (Tykerb, GlaxoSmithKline, Philadelphia, PA), improves clinical outcome in patients with advanced disease when added to capecitabine.²⁴ Taken together, these results imply that HER2 is a useful marker for

therapeutic decision making for patients with breast cancer, and they emphasize the importance of evaluating the assay accurately.

Gene amplification was initially detected by Southern hybridization in frozen tumor specimens, and was subsequently found to correlate with overexpression at the mRNA and protein levels.²⁵ The early trials of trastuzumab in metastatic breast cancer enrolled patients after central testing using an immunohistochemistry (IHC) assay with the anti-HER2 antibodies 4D5 (the parent antibody of trastuzumab) and CB11 on formalin-fixed, paraffin-embedded tissue, and this clinical trials assay identified staining patterns for HER2 as negative (0 and 1+) or positive (2+ and 3+). In these studies, only patients with 2+ or 3+ staining were eligible. Retrospective analyses have suggested that only patients with IHC 3+ staining and/or *HER2* gene amplification by fluorescent in situ hybridization (FISH) benefited¹⁶ (also see Appendix C). Concordance data subsequently showed that only 24% of the IHC 2+ tumors had gene amplification when tested by FISH.²⁶ Preliminary findings from the only randomized trial of trastuzumab in patients with 0 or 1+, nonamplified HER2 status have been reported.²⁷ In this study, there was no statistically significant benefit from the addition of trastuzumab to paclitaxel in women with HER-negative breast cancer, but the study was underpowered and limited by the lack of central testing for HER2.

Early studies suggested that as many as 30% of breast cancers have HER2 overexpression.^{1,25} However, it is likely that HER2 positivity was overestimated in these studies, and that its true frequency is lower in a general unselected population, since those data came largely from high-risk early-stage breast cancer cohorts and from patients with metastatic disease. The frequency of HER2-positive breast cancer appears to be lower when considering all patients with a new diagnosis of invasive breast cancer. Yaziji et al²⁸ recently reported their experience with large volume testing and observed that 18% of samples tested ($n = 2,913$) showed gene amplification by FISH (defined as a *HER2*:CEP17 ratio ≥ 2). A similar descriptive study by Owens et al²⁹ observed a frequency of HER2 overexpression of 20% among 116,736 specimens tested by IHC and 22.7% among 6,556 specimens tested by FISH.

The results of five randomized trials of adjuvant trastuzumab versus no trastuzumab have been reported since 2005, and various strategies to determine or confirm HER2 overexpression were used (Table 1). Adjuvant trastuzumab given during and/or after chemotherapy to women with early-stage breast cancer and evidence of HER2 overexpression results in a significant improvement in disease-free survival^{19,20} and overall survival.²¹⁻²³ HER2 overexpression is now accepted as a strong predictive marker for clinical benefit from trastuzumab in both the metastatic³⁹ and adjuvant settings. Indeed, the American Society of Clinical Oncology (ASCO) Tumor Marker Guidelines Panel has recommended routine testing of HER2 on newly diagnosed and metastatic breast cancer since 2001.³⁹

In summary, HER2 testing should be routinely performed in patients with a new diagnosis of invasive breast cancer. However, the best method to assess HER2 status, in regards both to the type of assay used and the optimal method to perform each assay, remains controversial. For most of the prospective randomized adjuvant trials of trastuzumab, testing algorithms for HER2 were somewhat arbitrarily developed, consisting of either IHC testing with reflex FISH if IHC 2+ or reliance on ISH testing alone to detect gene amplification ratios of 2.0 or higher.^{19-22,30,31,33,40} Those with evidence of amplification by FISH or overexpression by IHC (3+) were considered suitable

Guideline for HER2 Testing in Breast Cancer

Table 1. HER2 Testing Requirements of Five Randomized Trials of Adjuvant Trastuzumab

| Trials | No. Patients Enrolled | IHC | | FISH | | CISH | | Role of Central Laboratory |
|--------------------------|-----------------------------------|---------------------------------|---|---------------------------------|--|---------------|--|---|
| | | Assay | Positive | Assay | Positive | Assay | Positive | |
| NSABP B31 ²¹ | 2,119 | Institutional assay of any type | 3+ (2+ if FISH+); required 10% staining for 3+ | Institutional assay of any type | Amplified (HER2: CEP17 ratio \geq 2) | — | — | Institutional testing at first, but excessive initial false-positive HER2 results ³⁰ led to a change to mandatory confirmation by central (Dako IHC HercepTest using package insert definition of 10% or more staining) or approved labs (Dako IHC HercepTest or another IHC assay with more than 30% cells with strong staining); institutional IHC3+ required central confirmation with Dako HercepTest; central FISH testing was done using Vysis PathVysion, though the study continued to accept local FISH testing by any method |
| NCCTG 9831 ²¹ | 3,505 | Institutional assay of any type | 3+ (2+ if FISH+); required 10% staining for 3+ | Institutional assay of any type | Amplified (HER2: CEP17 ratio \geq 2) | — | — | Institutional testing at first, but changed to mandatory confirmation by central or approved labs due to excessive false-positive HER2 results; ³¹ institutional IHC3+ required central confirmation with Dako HercepTest; institutional FISH+ required central confirmation with Vysis PathVysion |
| HERA ^{22,23} | 5,081 | Institutional assay of any type | 3+ (2+ if FISH+) Definitions not described in article | Institutional assay of any type | Amplified (HER2: CEP17 ratio \geq 2) | — | — | Institutional IHC 3+ required central confirmation with Dako HercepTest; institutional FISH+ required central confirmation with Vysis PathVysion |
| BCIRG 006 ¹⁹ | 3,222 | — | — | Vysis PathVysion | Amplified (HER2: CEP17 ratio \geq 2) | — | — | Mandatory central testing |
| FinHer ²⁰ | 1,010 total (232 in HER2+ subset) | Institutional assay of any type | Local criteria | — | — | Central assay | Amplified (\geq 6 or more signals/nucleus in > 50% cells; add CEP17 hybridization if equivocal) ³² | Institutional IHC 2+ or 3+ required central confirmation by CISH |

Abbreviations: HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization; CISH, chromogenic in situ hybridization; NSABP B-31, National Adjuvant Breast and Bowel Project trial B31; NCCTG 9831, North Central Cancer Treatment Group trial 9831; HERA, Herceptin Adjuvant trial; BCIRG 006, Breast Cancer International Research Group trial 006; FinHer, Finland Herceptin trial.

candidates for participation in these trials. For the most part, these algorithms have been adopted into clinical practice.

The assays that are used to obtain the required data to populate these algorithms have not been standardized. Several assays have been used for HER2 determination in tissue (Table 2). US Food and Drug Administration regulations also allow pathology laboratories to develop and implement so called “home brew assays” using US Food and Drug Administration-approved analyte specific reagents.⁴¹ While some assays have been carefully validated, others, especially “home brew assays” have not. Prospective substudies from two of the adjuvant randomized trials of trastuzumab versus nil have demonstrated that approximately 20% of HER2 assays performed in the field (at the primary treatment site’s pathology department) were incorrect when the same specimen was re-evaluated in a high volume, central laboratory.^{30,40} Such a disorganized practice and high rate of inaccuracy, for such an important test that dictates a critically effective yet potentially life-threatening and expensive treatment, is not acceptable.

Trastuzumab therapy is not without its drawbacks. Although treatment duration in the metastatic setting varies widely, currently

adjuvant trastuzumab is recommended for 12 months. The drug cost of 52 weeks of trastuzumab in the community setting in the United States is approximately \$100,000 based on average sales price (www .acc-cancer.org). In addition, there is a requirement for 9 to 12 months of intravenous therapy after completion of adjuvant chemotherapy. Importantly, trastuzumab is associated with a small risk of serious cardiac toxicity.⁴² In the prospective randomized adjuvant trials, careful serial cardiac monitoring has demonstrated that at median follow-up times of 3 years or fewer, approximately 5% to 15% of patients develop cardiac dysfunction, and approximately 1% to 4% develop significant cardiac events (including symptomatic congestive heart failure) while taking trastuzumab.⁴³⁻⁴⁵

Taken together, the significant benefits coupled with the high cost and potential cardiotoxicity of trastuzumab demand accurate HER2 testing. If response to therapy were to be considered a gold standard, then the ideal test for HER2 would approach 100% sensitivity (ie, identify as HER2-positive all patients who will benefit from a specific therapy—the true-positives) and 100% specificity (ie, identify as HER2-negative all patients who would not benefit from a specific

Table 2. HER2 Assays Used in Tissue Specimens: Clinical Trials, Clinical Practice, and Under Development

| Parameter | HER2 Protein | | | HER2 Gene | | | |
|------------------|--|---|--|---|---|--|--|
| | IHC | | | FISH | CISH | SISH | |
| Assay | CTA | HercepTest | Pathway | PathVysion HER2 DNA Probe Kit | INFORM HER2/ <i>neu</i> Probe | SPoT-Light | EnzMet GenePro |
| Manufacturer | Home-brew assay | DAKO* | Ventana† | Abbott‡ | Ventana§ | Invitrogen/Zymed¶ | Ventana |
| Methodology | CB11 and 4D5 monoclonal antibodies | A085 polyclonal antibody | CB11 monoclonal antibody | Hybridization of fluorescent DNA probes to <i>HER2</i> gene (orange) and chromosome 17 centromere (green) | Hybridization of biotin-labeled DNA probe to <i>HER2</i> gene and fluorescently-labeled avidin | Hybridization of digoxigenin-labeled DNA probe to <i>HER2</i> gene; detection via mouse antidigoxigenin antibody followed by antimouse-peroxidase | Hybridization of dinitrophenol-labeled DNA probe to <i>HER2</i> gene; detection via peroxidase labeled multimer followed by enzyme metallography |
| Scoring criteria | 0 and 1+ negative, 2+ weakly positive, 3+ strongly positive | Weakly positive (2+): weak to moderate complete membrane staining in > 10% of tumor cells; strongly positive (3+): strong complete membrane staining in > 10% of tumor cells* | Positive (2+): weak complete staining of the membrane, > 10% of cancer cells; positive (3+): intense complete staining of the membrane, > 10% of cancer cells† | HER2 amplification: <i>HER2</i> /CEP17 ratio ≥ 2 on average for 60 cells; results at or near the cut off point (1.8-2.2) should be interpreted with caution ^{34,35} | HER2 amplification: average of > 6 <i>HER2</i> gene copies/nucleus; an average of > 4.0 < 6.0 gene copies/nucleus for 60 cells described as equivocal in one publication ^{35,36} | High HER2 amplification defined as > 10 dots, or large clusters, (low if > 5 dots to 10 dots, or small clusters) or mixture of multiple dots and large clusters of the <i>HER2</i> gene present per nucleus in > 50% tumor cells ³⁷ | Amplification defined as six or more dots, or large clusters of dots, in 30% or more of invasive tumor cells ³⁸ |
| Status | Research assay used in trials of trastuzumab in metastatic breast cancer | FDA approved | FDA approved | FDA approved | FDA approved | DNA probe kit not available in the United States | DNA probe kit not available in the United States |

Abbreviations: HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization; CISH, chromogenic in situ hybridization; SISH, silver enhanced in situ hybridization; CTA, clinical trials assay; HercepTest, Dako A/S, Glostrup, Denmark; FDA, US Food and Drug Administration.
*http://www.dakousa.com/prod_downloadpackageinsert.pdf?objectid=105073003
†<http://www.ventanamed.com/products/files/ScoringGuide.pdf>
‡http://www.vysis.com/PathVysionHER2DNAProbeKit_35793.asp
§http://www.ventanamed.com/catalog/search_detail.html?id=402&categories_id=4
¶<https://catalog.invitrogen.com/index.cfm?fuseaction=viewCatalog.viewProductDetails&productDescription=10,952&CMP=LEC-GCMSSEARCH&HQS=HER2>

therapy—the true-negatives). However, two points must not be forgotten. First, a precise definition of accuracy is how close the measured values are to a supposed true value, and it incorporates both variability (ie, precision) and bias (ie, a systematic difference between average measured value and true value). Implicit in this discussion is that a suitable gold standard has been established for purposes of determining true status for each specimen. Second, accurate determination of HER2 status must not be viewed exclusively in terms of benefit from anti-HER2 therapy, like trastuzumab. Patients with breast cancers that overexpress HER2 differ greatly in their response to trastuzumab. Available clinical data indicate the near certainty that there are patients who truly overexpress HER2 but have upstream or downstream anomalies that render the interaction with trastuzumab ineffective, and it would not be appropriate to consider these patients as having HER2-negative disease. Rather the challenge remains to define the additional defects that place this HER2 positivity in the appropriate therapeutic context.

Despite attempts within the international pathology community to improve the status of HER2 testing in routine practice,⁴⁶⁻⁵⁰ testing inaccuracy remains a major issue with both IHC and FISH.^{30,31,40} Various factors can explain the large variability observed in clinical

practice and in clinical trials. These factors are summarized in Table 3. The use of laboratory assays as the sole determinant for therapy selection poses a significant challenge to pathologists performing and interpreting the results and to oncologists who must rely on them for clinical decisions. Therefore, ASCO and the College of American Pathologists (CAP) established a clinical practice guideline expert panel charged with developing recommendations regarding HER2 testing in breast cancer. Information regarding the scope of the problem can be found in Appendices C (Evidence of HER2 Status and Trastuzumab Benefit) and D (Evidence on HER2 Testing Variation).

Guideline Questions

This guideline addresses two principal questions regarding HER2 testing. Table 4 summarizes the recommendations.

1. What is the optimal testing algorithm for the assessment of HER2 status?
2. What strategies can help ensure optimal performance, interpretation, and reporting of established assays?
 - a. What is the regulatory framework that permits enhanced testing scrutiny?

Table 3. Sources of HER2 Testing Variation

| | |
|---|--|
| Preanalytic | |
| Time to fixation | |
| Method of tissue processing | |
| Time of fixation | |
| Type of fixation | |
| Analytic | |
| Assay validation | |
| Equipment calibration | |
| Use of standardized laboratory procedures | |
| Training and competency assessment of staff | |
| Type of antigen retrieval | |
| Test reagents | |
| Use of standardized control materials | |
| Use of automated laboratory methods | |
| Postanalytic | |
| Interpretation criteria | |
| Use of image analysis | |
| Reporting elements | |
| Quality assurance procedures | |
| Laboratory accreditation | |
| Proficiency testing | |
| Pathologist competency assessment | |
| Abbreviation: HER2, human epidermal growth factor receptor 2. | |

b. What are the optimal external quality assurance methods to ensure ongoing accuracy in HER2 testing?

c. How can these efforts be implemented and the effects measured?

Practice Guidelines

Practice guidelines are systematically developed statements to assist practitioners and patients in making decisions about appropriate health care for specific clinical circumstances. Attributes of good guidelines include validity, reliability, reproducibility, clinical applicability, clinical flexibility, clarity, multidisciplinary process, review of evidence, and documentation. Guidelines may be useful in producing better care and decreasing cost. Specifically, utilization of clinical guidelines may provide the following:

1. Improvement in outcomes;
2. Improvement in medical practice;
3. Means for minimizing inappropriate practice variation;
4. Decision support tools for practitioners;
5. Points of reference for medical orientation and education;
6. Criteria for self-evaluation;
7. Indicators and criteria for external quality review;
8. Assistance with reimbursement and coverage decisions;
9. Criteria for use in credentialing decisions;
10. Identification of areas where further research is needed.

In formulating recommendations for HER2 testing in breast cancer, ASCO and CAP considered these tenets of guideline development, emphasizing review of data from appropriately conducted and analyzed clinical trials. However, it is important to note that guidelines cannot always account for individual variation among patients. Guidelines are not intended to supplant physician judgment with respect to particular patients or special clinical situations and cannot be considered inclusive of all proper meth-

ods of care or exclusive of other treatments reasonably directed at obtaining the same result. Accordingly, ASCO considers adherence to these guidelines to be voluntary, with the ultimate determination regarding their application to be made by the physician in light of each patient's individual circumstances. In addition, these guidelines describe the use of procedures and therapies in clinical practice; they cannot be assumed to apply to the use of these interventions performed in the context of clinical trials, given that clinical studies are designed to evaluate or validate innovative approaches in a disease for which improved staging and treatment is needed. In that guideline development involves a review and synthesis of the latest literature, a practice guideline also serves to identify important questions and settings for further research.

METHODS

Panel Composition

The ASCO Health Services Committee (HSC) and the CAP Council on Scientific Affairs (CSA) jointly convened an expert panel consisting of experts in clinical medicine and research relevant to HER2 testing, including medical oncology, pathology, epidemiology, statistics, and health services research. Academic and community practitioners and a patient representative were also part of the panel. Representatives from the US Food and Drug Administration, the Centers for Medicare and Medicaid Services, the National Cancer Institute, and the National Academy of Clinical Biochemistry served as ex-officio members. The opinions of panel members associated with official government agencies represent their individual views and not necessarily those of the agency with which they are affiliated. The panel members are listed in Appendix A Table A1. Representatives of commercial laboratories and assay/drug manufacturers (Appendix B Table A2) were invited as guests to attend the open portion of the panel meeting held at ASCO headquarters in March 2006.

Literature Review and Analysis

Literature search strategy. The following electronic databases were searched from January 1987 through February 2006: MEDLINE, PreMEDLINE, and the Cochrane Collaboration Library. In addition, abstracts presented at ASCO or CAP from 2000 to 2005 and at the San Antonio Breast Cancer Symposium from 2003 to 2005 were identified. Results were supplemented with hand searching of selected reviews and personal files. The following MeSH terms were used in a MEDLINE search: "immunohistochemistry," "in situ hybridization, fluorescence," "genes, erbB2," "receptor, erbB2," "receptor, epidermal growth factor," "breast neoplasms," and the substance name "epidermal growth factor receptor-*neu* receptor." The search was expanded by the addition of the following text words, in varying combinations: immunohistochemistry, immunocytochemistry, "IHC," fluorescence in situ hybridization, "FISH," chromogenic hybridization, "CISH," gold-facilitated hybridization, autometallographic, bright field, "GOLDFISH," HER2, erbB2, breast cancer, and breast tumor. All searches were limited to the English language.

Study design was not limited to randomized controlled trials, but was expanded to include any study type, including cohort designs, case series, evaluation studies, comparative studies, and prospective studies. Also included were testing guidelines and proficiency strategies of various United States and international organizations. Letters, commentaries, and editorials were reviewed for any new information. Case reports were excluded.

Articles were selected for inclusion in the systematic review of the evidence if they met the following criteria: (1) the study compared, prospectively or retrospectively, the negative predictive value (NPV) or positive predictive value (PPV) of FISH or IHC; the study described technical comparisons across various assay platforms; the study examined potential testing algorithms for HER2 testing; or the study examined the correlation of HER2 status in primary versus metastatic tumors from the same patients; and (2) the study population consisted of patients with a diagnosis of invasive breast cancer; and (3) the primary outcomes included the PPV and NPV of FISH and IHC to determine

Table 4. Summary of Guideline Recommendations

| | Recommendation |
|---|--|
| Optimal algorithm for HER2 testing | <p>Positive for HER2 is either IHC HER2 3+ (defined as uniform intense membrane staining of > 30% of invasive tumor cells) or FISH amplified (ratio of <i>HER2</i> to CEP17 of > 2.2 or average <i>HER2</i> gene copy number > six signals/nucleus for those test systems without an internal control probe)</p> <p>Equivocal for HER2 is defined as either IHC 2+ or FISH ratio of 1.8-2.2 or average <i>HER2</i> gene copy number four to six signals/nucleus for test systems without an internal control probe</p> <p>Negative for HER2 is defined as either IHC 0-1+ or FISH ratio of < 1.8 or average <i>HER2</i> gene copy number of < four signals/nucleus for test systems without an internal control probe</p> <p>These definitions depend on laboratory documentation of the following:</p> <ol style="list-style-type: none"> 1. Proof of initial testing validation in which positive and negative HER2 categories are 95% concordant with alternative validated method or same validated method for HER2 2. Ongoing internal QA procedures 3. Participation in external proficiency testing 4. Current accreditation by valid accrediting agency |
| Optimal FISH testing requirements | <p>Fixation for fewer than 6 hours or longer than 48 hours is not recommended</p> <p>Test is rejected and repeated if</p> <ul style="list-style-type: none"> • Controls are not as expected • Observer cannot find and count at least two areas of invasive tumor • > 25% of signals are unscorable due to weak signals • > 10% of signals occur over cytoplasm • Nuclear resolution is poor • Autofluorescence is strong <p>Interpretation done by counting at least 20 cells; a pathologist must confirm that counting involved invasive tumor</p> <p>Sample is subjected to increased counting and/or repeated if equivocal; report must include guideline-detailed elements (see Table 10)</p> |
| Optimal IHC testing requirements | <p>Fixation for fewer than 6 hours or longer than 48 hours is not recommended</p> <p>Test is rejected and repeated or tested by FISH if</p> <ul style="list-style-type: none"> • Controls are not as expected • Artifacts involve most of sample • Sample has strong membrane staining of normal breast ducts (internal controls) <p>Interpretation follows guideline recommendation</p> <ul style="list-style-type: none"> • Positive HER2 result requires homogeneous, dark circumferential (chicken wire) pattern in > 30% of invasive tumor • Interpreters have method to maintain consistency and competency <p>Sample is subjected to confirmatory FISH testing if equivocal based on initial results</p> <p>Report must include guideline-detailed elements (see Table 9)</p> |
| Optimal tissue handling requirements | <p>Time from tissue acquisition to fixation should be as short as possible; samples for HER2 testing are fixed in neutral buffered formalin for 6-48 hours; samples should be sliced at 5-10 mm intervals after appropriate gross inspection and margins designation and placed in sufficient volume of neutral buffered formalin</p> <p>Sections should ideally not be used for HER2 testing if cut > 6 weeks earlier; this may vary with primary fixation or storage conditions</p> <p>Time to fixation and duration of fixation if available should be recorded for each sample</p> |
| Optimal internal validation procedure | <p>Validation of test must be done before test is offered</p> <p>Initial test validation requires 25-100 samples tested by alternative validated method in the same laboratory or by validated method in another laboratory</p> <p>Proof of initial testing validation in which positive and negative HER2 categories are 95% concordant with alternative validated method or same validated method for HER2</p> <p>Ongoing validation should be done biannually</p> |
| Optimal internal QA procedures | <p>Initial test validation</p> <p>Ongoing quality control and equipment maintenance</p> <p>Initial and ongoing laboratory personnel training and competency assessment</p> <p>Use of standardized operating procedures including routine use of control materials</p> <p>Revalidation of procedure if changed</p> <p>Ongoing competency assessment and education of pathologists</p> |
| Optimal external proficiency assessment | <p>Participation in external proficiency testing program with at least two testing events (mailings)/year</p> <p>Satisfactory performance requires at least 90% correct responses on graded challenges for either test</p> <ul style="list-style-type: none"> • Unsatisfactory performance will require laboratory to respond according to accreditation agency program requirements |
| Optimal laboratory accreditation | <p>Onsite inspection every other year with annual requirement for self-inspection</p> <ul style="list-style-type: none"> • Reviews laboratory validation, procedures, QA results and processes, results and reports • Unsatisfactory performance results in suspension of laboratory testing for HER2 for that method |

Abbreviations: HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization; QA, quality assurance.

HER2 status, alone and in combination; concordance across platforms; accuracy in determining HER2 status and benefit from anti-HER2 therapy, sensitivity, and specificity of specific tests. Consideration was given to studies that directly compared results across assay platforms.

The panel reviewed the results of randomized controlled trials in breast cancer testing anti-HER2 therapies like trastuzumab and lapatinib. The panel also reviewed unblinded trials comparing various testing methods, describing test characteristics, and defining strategies for quality assurance of testing in the literature. Individuals representing regulatory agencies (CMS and US Food and Drug Administration) also provided information about the regulatory framework. Individuals involved with quality assurance in the United States (CAP), Great Britain, and Canada (Province of Ontario) also provided information about programs to measure and improve HER2 testing. Survey data from the maker of trastuzumab (Genentech) was also evaluated as well as testimony provided by testing manufacturers (Ventana, Dako, Abbott) and large clinical laboratories (Clariant, Mayo Medical Labs, Phenopath, Quest, and US Labs) to define the current status of training and testing for HER2. This information was used to help the panel define the best algorithm for testing, specify testing requirements and exclusions, and the necessary quality assurance monitoring that will make the testing less variable and more accurate.

ASCO/CAP expert panel literature review and analysis. An initial abstract screen was performed by ASCO staff. The ASCO/CAP panel reviewed all remaining potentially relevant abstracts identified in the original literature searches to select studies pertinent to its deliberations. Two panel members independently reviewed each abstract for its relevance to the clinical questions, and disagreements were resolved by third-party review. Full-text articles were then reviewed for all selected abstracts. Evidence tables were developed based on selected studies that met the criteria for inclusion.

Consensus Development Based on Evidence

The entire panel met in March 2006; additional work on the guideline was completed through electronic mail and teleconferences of the panel. The purposes of the panel meeting were to refine the questions addressed by the guideline and to make writing assignments for the respective sections. All members of the panel participated in the preparation of the draft guideline, which was then disseminated for review by the entire panel. Feedback from external reviewers was also solicited. The content of the guideline and the manuscript were reviewed and approved by the ASCO HSC and board of directors and by the CAP CSA and board of governors before dissemination.

Guideline and Conflict of Interest

All members of the expert panel complied with ASCO policy on conflict of interest, which requires disclosure of any financial or other interest that might be construed as constituting an actual, potential, or apparent conflict. Members of the expert panel completed ASCO's disclosure form and were asked to identify ties to companies developing products that might be affected by promulgation of the guideline. Information was requested regarding employment, consultancies, stock ownership, honoraria, research funding, expert testimony, and membership on company advisory committees. The panel made decisions on a case-by-case basis as to whether an individual's role should be limited as a result of a conflict. No limiting conflicts were identified.

Revision Dates

At annual intervals, the panel co-chairs and two panel members designated by the co-chairs will determine the need for revisions to the guideline based on an examination of current literature. If necessary, the entire panel will be reconvened to discuss potential changes. When appropriate, the panel will recommend revision of the guideline to the ASCO HSC, the CAP CSA, the ASCO board, and the CAP board for review and approval.

Summary of Outcomes Assessed

The primary outcome of interest was the correlation between HER2 status and benefit from anti-HER2 therapy. Other outcomes of interest included the PPV and NPV of FISH and IHC to determine HER2 status, alone

and in combination; concordance across platforms; and accuracy in determining HER2 status, sensitivity, and specificity of specific tests. Additional outcomes considered were the performance of newer testing techniques, including bright field in situ hybridization. However, data on these newer assays were considered supplementary as these assays are not US Food and Drug Administration approved at present. The primary focus of this guideline is on IHC and FISH assays.

RESULTS

Literature Search

Preliminary searches identified 1,802 MEDLINE abstracts. The initial abstract screen performed by ASCO staff eliminated 1,010 abstracts that failed to meet any of the inclusion criteria. The ASCO panel conducted dual independent review of all remaining 792 potentially relevant abstracts identified in the original systematic review. The panel eliminated 667 abstracts at this stage of the review; the remaining 125 articles were reviewed in full for the interventions and outcomes described herein. A meta-analysis was not performed because the studies were judged to be too heterogeneous for meaningful quantitative synthesis.

Previous Guidelines and Consensus Statements

ASCO and CAP have previously published separate guideline and position statements regarding HER2 testing.^{39,51} However, these were developed before newly published adjuvant data from adjuvant trastuzumab trials were available, and most have either simply stated that HER2 testing should be performed without recommending specific methodology or without addressing quality assurance measures.

Testing algorithms described in existing guidelines assume a high level of correlation between IHC and FISH assays. An example from the United Kingdom in 2004 recommends a testing algorithm that uses IHC as the primary test, with a score of 0 or 1+ interpreted as HER2-negative, a score of 3+ interpreted as HER2-positive, and a score of 2+ interpreted as equivocal (or inconclusive) and automatically sent for FISH testing.⁴⁸ The United Kingdom panel emphasizes several requirements for a laboratory to be approved for HER2 testing, such as a minimum annual case load (250 cases of IHC and 100 cases of FISH) below which laboratories should consider using a reference laboratory, use of standardized and validated assays, and adherence to ongoing quality assurance programs. Other recommendations include the use of tissue-based controls, limiting the reading to the invasive component of the tumor, and strict adherence to kit assay protocol and scoring methodology.⁴⁸ CAP issued a similar set of recommendations in the United States after a Strategic Science Symposium sponsored by CAP, Rosemont, IL, May 4-5, 2002.⁴⁷ It emphasized the need for individual laboratories to document their own concordance experience of FISH v IHC (90% for IHC 0 and 3+ and 95% for IHC 1+) before limiting reflex FISH testing only to IHC 2+ results, and also offered recommendations on the use of a standardized report format and defined terminology. Note that these concordance requirements were set based on a palliative role of trastuzumab. Efforts to improve HER2 testing accuracy have been reported by several groups.⁵²

RECOMMENDATIONS

What Is the Optimal Testing Algorithm for the Assessment of HER2 Status?

Summary and recommendations. The literature review and resultant panel discussion elucidated three categories of HER2 testing results leading to different clinical decisions for patients with breast cancer. The test, regardless of method used, can be found to be positive, equivocal, or negative. Each of these test results triggers defined patient management algorithms as shown in Figures 1 (IHC) and 2 (FISH).

In all cases, it is assumed that the test being used is accurate and reproducible based on good laboratory practices as defined later in this article. In order to classify a HER2 test as either positive or negative, the laboratory must have performed concordance testing with a validated FISH assay and confirmed that only 5% or less of samples classified as either positive or negative disagree with that validated assay on an ongoing basis. If the laboratory cannot satisfy this criterion, it should not perform HER2 testing and should send specimens to a reference laboratory. Equivocal cases are not expected to 95% concordant, rather they should be subjected to a confirmatory test. Concordance testing should be annually confirmed. A minority view expressed within the panel was that IHC is not a sufficiently accurate assay to determine HER2 status and that FISH should be preferentially used.

It is important to note that concordance of assays does not assure accuracy (ie, how close the measured values are to a supposed true value; Appendix F). Evaluating accuracy of a test requires comparison to a gold standard. There is no gold standard at present; no assay

currently available is perfectly accurate to identify all patients expected to benefit or not from anti-HER2 therapy.

The following definitions have been accepted for analysis of HER2. It is critical that these analyses be conducted on the invasive component of the breast cancer, because HER2 is, for unclear reasons, frequently increased (overexpressed and/or over amplified) in in situ breast cancer, and the clinical implications of this finding are uncertain.⁵³

Positive HER2 test. Based on a literature review of clinical trials, international studies and protocols, expert consensus, and US Food and Drug Administration Panel findings, a positive HER2 test is defined as either IHC result of 3+ cell surface protein expression (defined as uniform intense membrane staining of > 30% of invasive tumor cells) or FISH result of amplified *HER2* gene copy number (average of > six gene copies/nucleus for test systems without internal control probe) or *HER2/CEP 17* ratio of more than 2.2, where CEP 17 is a centromeric probe for chromosome 17 on which the *HER2* gene resides. The 30% criteria for a positive IHC is further discussed in Appendix G. The original FISH test results were defined as either positive or negative, but an intermediate range (from hereon referred to as equivocal range) has since been described and the clinical significance of this observation remains unclear.³⁴⁻³⁶ This strategy classifies patients as having HER2-positive disease based on positive results with either test. It is recognized that current data are insufficient to define whether these patients represent true- or false-positives. Although the large prospective randomized clinical trials of trastuzumab were not prospectively designed to answer these questions, we anticipate

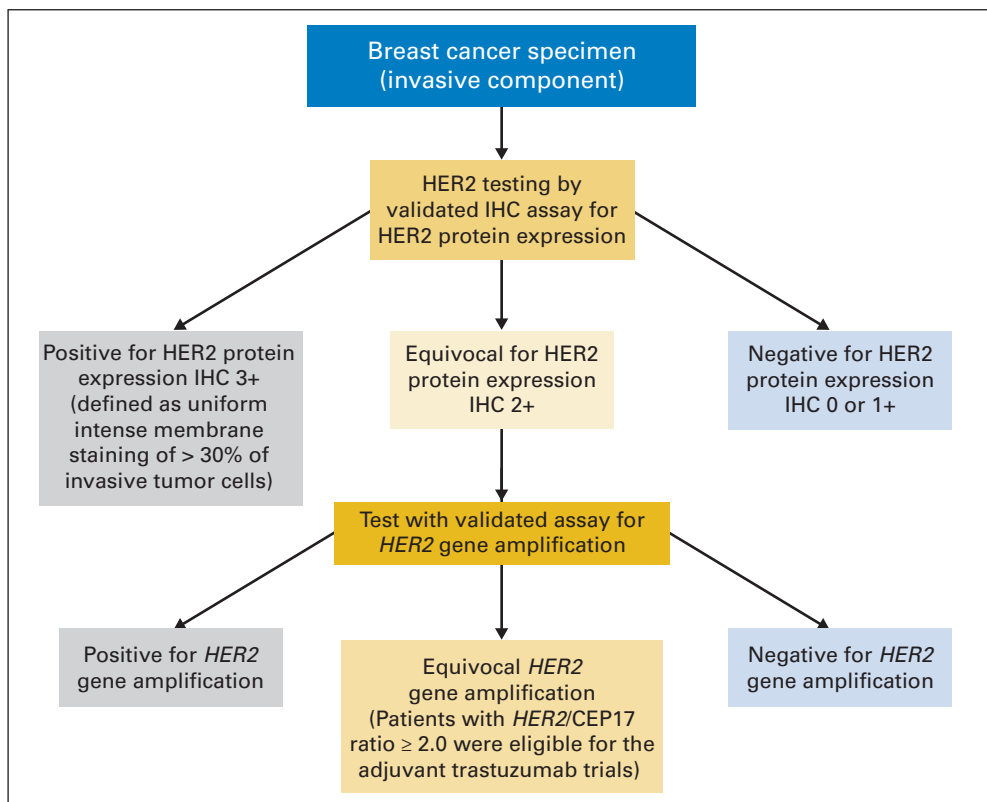


Fig 1. Algorithm for immunohistochemistry (IHC). For additional information regarding adjuvant trastuzumab trials see Slamon et al,¹⁹ Romond et al,²¹ Piccart-Gebhart et al,²² and HERA trial study.²³ HER2, human epidermal growth factor receptor 2.

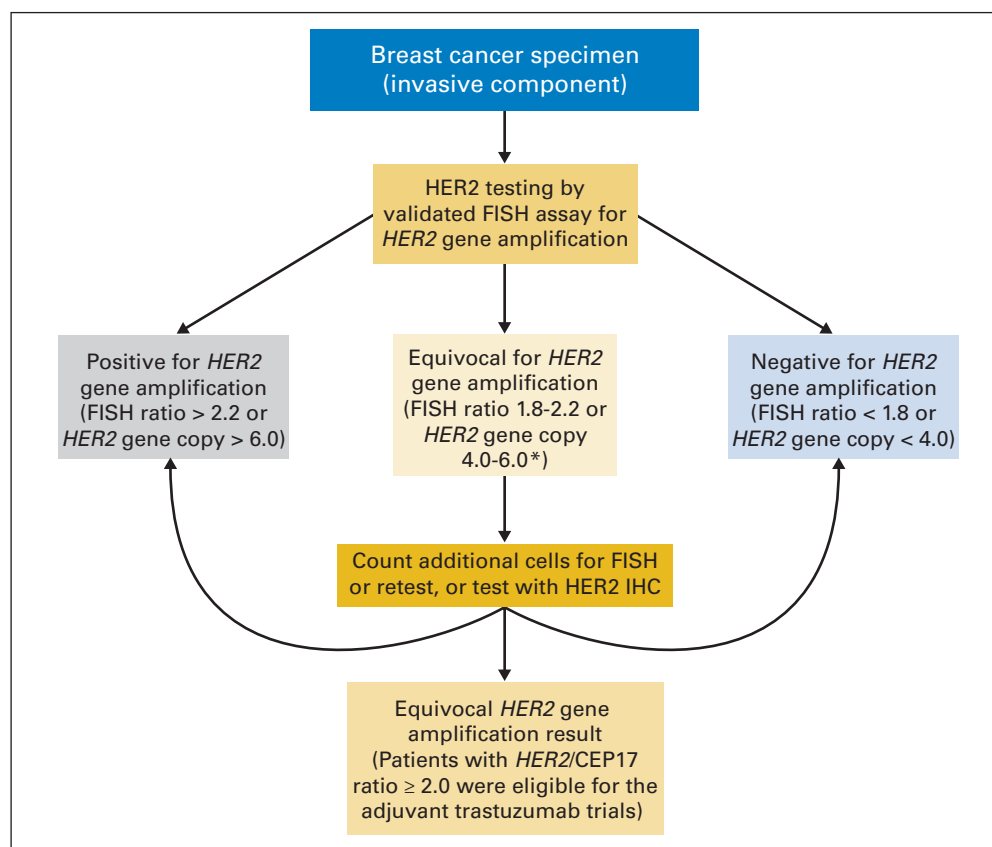


Fig 2. Algorithm for fluorescent in situ hybridization (FISH). For additional information regarding adjuvant trastuzumab trials, see Slamon et al,¹⁹ Romond et al,²¹ Piccart-Gebhart et al,²² and HERA trial study.²³ (*)³⁴⁻³⁶ HER2, human epidermal growth factor receptor 2.

and recommend that such analyses will be forthcoming as correlative studies.

Equivocal HER2 test. Much of the confusion about HER2 testing has resulted from the need to define trastuzumab treatment (yes or no) based on test results that represent a continuous rather than a categorical variable. Furthermore, there is significant variation in the intermediate (equivocal) ranges for both the IHC and FISH assays. The equivocal range for IHC consists of samples scored 2+, and this may include up to 15% of samples.²⁹ An equivocal result (2+) is complete membrane staining that is either nonuniform or weak in intensity but with obvious circumferential distribution in at least 10% of cells. Very rarely, in the experience of panel members, invasive tumors can show intense, complete membrane staining of 30% or fewer tumor cells. These are also considered to be equivocal in this guideline. Some but not all of these samples may have *HER2* gene amplification and require additional testing to define the true HER2 status (Figs 1 and 2).^{54,55}

The equivocal range for FISH assays is defined as *HER2/CEP17* ratios from 1.8 to 2.2 or average gene copy numbers between 4.0 and 6.0 for those systems without an internal control probe.³⁴⁻³⁶ Note, however, that patients with a *HER2/CEP17* FISH ratio between 2.0 and 2.2 were formerly considered HER2-positive and were eligible for treatment in the adjuvant trastuzumab trials. Therefore, available efficacy data do not support excluding them from therapy with trastuzumab. This group is much smaller, probably fewer than 3% of samples.⁵⁶ Polysomy 17 is observed in approximately 8% of all specimens, mostly among cases with four to six *HER2* gene copies (equivocal range).^{57,58} There is no accepted definition of what constitutes

polysomy and authors have used different criteria to define it. If polysomy 17 is defined as three or more copies of CEP17, most are not associated with protein or mRNA overexpression⁵⁷ and the same has been observed in tumors with a *HER2* gene copy number between 4 and 6.³⁵

Discordant results (IHC 3+/FISH-negative or IHC < 3+/FISH-positive) have also been described, and were observed in approximately 4% among 1,503 patients screened centrally (LabCorp, Burlington, NC) with both methods for eligibility for a clinical trial with trastuzumab.⁵⁹ However, clinical outcome data for these two groups are not yet available. We anticipate and recommend that such analyses will be forthcoming as correlative studies of the large prospective randomized clinical trials of trastuzumab.

It is also clear from the panel discussion and literature review that patients with equivocal HER2 test results constitute a poorly studied subgroup with uncertain association of test scores to benefit from HER2-directed therapy.⁶⁰ The panel suggested that further studies of this patient group would be promoted by defining these test results as equivocal or borderline. The panel elected to use the term equivocal to avoid confusion with borderline positive and borderline negative terminology which is sometimes used in the interpretation of FISH assays. Equivocal results of a single test require additional action which should be specified in the initial report. Equivocal IHC samples must be confirmed by FISH analysis of the sample. Equivocal FISH samples are confirmed by counting additional cells or repeating the FISH test. If FISH remains equivocal after additional cells counted or assay repeated, confirmatory IHC is recommended so that HER2 protein

expression is known for the sample with true equivocal gene amplification status.

Negative HER2 test. A negative HER2 test is defined as either an IHC result of 0 or 1+ for cellular membrane protein expression (no staining or weak, incomplete membrane staining in any proportion of tumor cells), or a FISH result showing *HER2/CEP17* ratio of less than 1.8 or an average of fewer than four copies of *HER2* gene per nucleus for systems without an internal control probe. The upper limit of 5% false-negatives should be considered high in view of the potential curative potential of trastuzumab treatment in the adjuvant setting, and laboratories should aim at bringing this percentage of false negative tests as close to 0% as possible.

HER2 assay exclusions. Each assay type has diagnostic pitfalls to be avoided. The panel agreed that there were situations where one assay type was preferred because of assay or sample considerations. Exclusion criteria to perform or interpret an IHC or FISH assay for HER2 are presented in Tables 5 and 6, respectively. The pathologist who reviews the histologic findings on the sample in question should determine the optimal assay type.

Review of Relevant Literature

The panel reviewed data from existing and completed clinical trials, published reports, and panel presentations by representatives of other national groups where stringent internal and external quality assessment measures have been implemented.

What Strategies Can Help Ensure Optimal Performance, Interpretation, and Reporting of Established Assays?

Summary and recommendations. The recommendations that follow are based on consensus conferences held in the United States,^{47,61} single institution studies,⁶² experience from reference laboratories,^{28,29,55} international reports,^{48-50,63,64} regulations currently in force in the United States (Clinical Laboratory Improvement Amendment [CLIA] 88 and US Food and Drug Administration regulations), and expert consensus at the panel meeting. The recommendations are summarized in Table 4. See Appendix E for tissue handling requirements and control materials.

Testing validation requirements. This section describes technical validation requirements of an assay. It is important that any new test be compared with a reference test for which there has been

Table 5. Sample Exclusion Criteria to Perform or Interpret a HER2 IHC Assay

| |
|--|
| Tissues fixed in fixatives other than neutral buffered formalin* |
| Needle biopsies fixed less than 1 hour in neutral buffered formalin |
| Excisional biopsies fixed in formalin for less than 6 hours or longer than 48 hours† |
| Core needle biopsies with |
| Edge or retraction artifact involving entire core |
| Crush artifact (thin gauge, vacuum extraction needle samples) |
| Tissues with strong membrane staining of internal normal ducts or lobules |
| Tissues where controls exhibit unexpected results |

Abbreviations: HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry.

*If a laboratory uses fixatives other than buffered formalin, it must validate the performance characteristics of the assay to show that they are concordant with results using buffered formalin in same samples.

†This is not an absolute exclusion criterion, but if known to be fixed shorter than 6 hours or longer than 48 hours, the report should qualify any negative result with this information.

Table 6. Sample Exclusion Criteria to Perform or Interpret a HER2 FISH Assay

| |
|---|
| Samples with only limited invasive cancer difficult to define under UV light |
| Tissue fixed in fixatives other than buffered formalin* |
| Tissue fixed for prolonged intervals in formalin (greater than 48 hours)† |
| Controls with unexpected results |
| FISH signals nonuniform (< 75% identifiable) |
| Background obscures signal (> 10% of signals over cytoplasm) |
| Nonoptimal enzymatic digestion (poor nuclear resolution, persistent autofluorescence) |

Abbreviations: HER2, human epidermal growth factor receptor 2; FISH, fluorescent in situ hybridization.

*If a laboratory uses fixatives other than buffered formalin, it must validate the performance characteristics of the assay to show that they are concordant with results using buffered formalin in same samples.

†This is not an absolute exclusion criterion, but if known to be fixed longer than 48 hours or unknown, the report should qualify any negative result with this information.

clinical validation, which means that the reference test predicts clinical outcome. The validation procedure for any new test offered by a laboratory involves several steps. The laboratory must select and acquire appropriate equipment, assure that personnel are trained in the use of the equipment, and develop a standard operating procedure for the test to be offered. Personnel must then be trained on the standard operating procedure with a standardized training plan. The new procedure must then be tested on a group of clinical cases representative of those on which the test will be offered. This testing must be done in parallel along with a validated clinical test for the same analyte (for example, HER2). If the new test (for example, HER2 by IHC) is to be compared with a previously validated complementary test (for example, HER2 by FISH) the samples are tested by both methods and the results compared. Alternatively, for laboratories that have not previously validated either test, the test can be validated by having it run in parallel by another laboratory in which a validated assay is already offered. The number of tests required for a reliable validation is not well defined, but ranges from 25 to 100 cases and depends on the variety of results possible and the amount of variation in results encountered in the test. A new test should show at least 95% concordance with the validated assay to which it is compared. Individuals interpreting the assay must also have their concordance compared with each other and this concordance should also be at least 95%. If laboratories choose to use alternative fixatives other than buffered formalin, the laboratory is obligated to validate that fixative's performance against the results of testing of the same samples fixed also in buffered formalin and tested with the identical HER2 assay, and concordance in this situation must also be 95%. Appendix F discusses statistical considerations for determining appropriate numbers of cases to include in test sets and for setting reasonable performance goals.

Ongoing competency assessment. As part of the laboratory's internal quality assurance program, the competency of laboratory professionals and pathologists interpreting assays must be continuously addressed. The laboratory director has responsibility to assure the competency of those performing the test, using established laboratory procedures available for review at the time of inspection. The review of competency for pathologists should include periodic or continuous peer comparisons among reviewing pathologists for the laboratory's HER2

specimens. If variation in interpretations is encountered, remediation must be done and documented.

The panel agreed that acceptable performance standards for such tests were as follows:

- The assay should only be evaluated in invasive breast cancer or the invasive component of the breast cancer.
- The tissue handling requirements must be standardized and reported on every specimen. All previous conferences on HER2 testing as well as the original premarket approval for trastuzumab required fixation of breast tissue samples in 10% neutral buffered formalin. Optimal fixations times are 6 to 48 hours and should be documented in the pathology report.
- There are specific sample exclusion criteria for HER2 immunohistochemistry and FISH assays which should be disseminated (Tables 5 and 6).
- Assay procedures must be validated by the laboratory before offering the test clinically. The new test should show 95% concordance with a validated reference assay.
- Assay procedures must be standardized. Any deviation from the standardized method must be recorded and justified by revalidation of the method. Such changes in procedure must be documented in the report. The panel agreed that optimal performance is more easily obtained using automated staining platforms rather than manual methods. Personnel performing assays must have their competency assessed at regular intervals.

Standardized control materials, either purchased products or conforming to defined manufacturing standards (for example, cell lines of the European Collection of Cell Cultures or those produced by the National Institute of Standards and Technology [NIST]) or defined by the laboratory director, must be consistently used by each laboratory with each run of tests. Adequate control materials include cell lines or tumor blocks with well defined negative, equivocal, and positive expression and gene amplification assay results. Faux tissue blocks or xenografts with variable HER2 expression levels may also be used if the results expected are well characterized. If controls do not show usual results, assay must be repeated rather than interpreted.

Image analysis can be an effective tool for achieving consistent interpretation. However, a pathologist must confirm the image analysis result. Image analysis equipment, just as other laboratory equipment, must be calibrated and subjected to regular maintenance and internal quality control evaluation. Image analysis procedures must be validated before implementation. One issue identified during the panel discussion was lack of calibration of the optical microscopes used by pathologists, something which certainly contributes to interpretive variation. If pathologists use several different microscopes to read assays, a system of calibration of these instruments should be implemented to ensure consistent interpretation.

Interpretation criteria for all types of HER2 tests must be standardized and refined, based on the interpretation criteria from recent clinical trials and international experience (Table 2). Criteria are delineated in Tables 7 and 8 and see also Appendix G.

Reporting elements for HER2 testing must be standardized. Lists of elements are provided in Tables 9 and 10.

While evidence was provided that specifying minimal testing volumes is an effective strategy to qualify laboratories for HER2 testing, there was no systematic evidence for a relationship of volume to test accuracy. Panel members agreed it is advisable to consider not offering HER2 testing if few tests will be performed annually. Rather,

Table 7. IHC Interpretation Criteria

| | |
|--|--|
| Review controls; if not as expected, test should be repeated | |
| More than 30% of tumor must show circumferential membrane staining for positive result* | |
| Membrane staining must be intense and uniform | |
| A homogeneous, dark circumferential (chicken wire) staining pattern should be seen | |
| Ignore incomplete or pale membrane staining | |
| Quantitative image analysis is encouraged for cases with weak membrane staining (1-2+) to improve consistency of interpretation | |
| If cytoplasmic staining obscures membrane staining, repeat assay or do FISH | |
| Reject sample if normal ducts and lobules show obvious staining | |
| Reject sample if there are obscuring artifacts | |
| Avoid scoring DCIS; score only infiltrating ductal carcinoma | |
| Abbreviations: IHC, immunohistochemistry; FISH, fluorescent in situ hybridization; DCIS, ductal carcinoma in situ. | |
| *This percentage is greater than that specified in US Food and Drug Administration-approved tests but value was based on expert panel consensus that this level is a more realistic threshold. | |

the specimen should be sent from the primary institution to a central laboratory with more experience and volume. In laboratories with multiple pathologists, it may also be advisable to consider limiting the number who interpret HER2 tests so that each pathologist will interpret a greater number of cases.

Review of Relevant Literature

Previous consensus conferences described the laboratory requirements for HER2 testing.^{47,61} Internal quality assurance requirements are mandated by CLIA 88 (legislation passed by United States Congress in 1988) as the basis for these recommendations. Reporting recommendations are also defined broadly in CLIA 88 requirements and have been specified in CAP consensus conference⁴⁷ and in expert opinion.⁶⁵ Literature substantiating the testing exclusion criteria and interpretation criteria were reviewed and used to establish the criteria.^{47,66}

What Is the Regulatory Framework that Permits Enhanced Testing Scrutiny?

Summary and recommendations. CLIA 88 provides stringent quality standards for highly complex tests, which include all predictive

Table 8. FISH Interpretation Criteria

| | |
|---|--|
| Review corresponding hematoxylin and eosin and/or IHC slide to localize the invasive cancer; carcinoma in situ should not be scored | |
| Review controls; if not as expected, test should be repeated | |
| Count at least 20 nonoverlapping cells in two separate areas of invasive cancer | |
| Reject if signals are nonuniform (> 25%) | |
| Reject if autofluorescence high or nuclear resolution poor | |
| Reject if background obscures signal resolution (> 10% over cytoplasm) | |
| If HER2/CEP17 ratio between 1.8 and 2.2, have additional person recount | |
| If heterogeneous expression, have additional person recount | |
| Counting can be done by a trained technologist, but pathologist must confirm that result (count) is correct and that invasive tumor was counted | |
| Note: Tumor heterogeneity, monosomy or polysomy of chromosome 17, and gene deletion may influence the interpretation of the absolute ratio value. | |
| Abbreviations: FISH, fluorescent in situ hybridization; IHC, immunohistochemistry; HER2, human epidermal growth factor receptor 2. | |

Table 9. Reporting Elements for IHC

| |
|--|
| Patient identification information |
| Physician identification |
| Date of service |
| Specimen identification (case and block number) |
| Specimen site and type |
| Specimen fixative type |
| Time to fixation (if available) |
| Duration of fixation (if available) |
| Antibody clone/vendor |
| Method used (test/vendor and if FDA approved) |
| Image analysis method (if used) |
| Controls (high protein expression, low-level protein expression, negative protein expression, internal) |
| Adequacy of sample for evaluation |
| Results |
| Percentage of invasive tumor cells exhibiting complete membrane staining |
| Uniformity of staining: present/absent |
| Homogeneous, dark circumferential pattern: present/absent |
| Interpretation |
| Positive (for HER2 protein expression); equivocal (FISH will be done and reported); negative (for HER2 protein expression); not interpretable |
| Comment |
| If an FDA-approved method is used, it should be stated; if the FDA-approved method has been modified, a statement in the report should be included indicating what modifications were made and that the changes have been validated; if the test is not FDA approved or an FDA-approved test has been modified, a clear statement must be made that the laboratory reporting results takes responsibility for test performance |
| Abbreviations: IHC, immunohistochemistry; FDA, US Food and Drug Administration; FISH, fluorescent in situ hybridization; HER2, human epidermal growth factor receptor 2. |

Table 10. Reporting Elements for FISH

| |
|---|
| Patient identification information |
| Physician identification |
| Date of service |
| Specimen identification (case and block number) |
| Specimen site and type |
| Specimen fixative type |
| Time to fixation (if available) |
| Duration of fixation (if available) |
| Probe(s) identification |
| Method used (specifics of test/vendor and if FDA approved) |
| Image analysis method |
| Controls (amplified, equivocal, and nonamplified, internal) |
| Adequacy of sample for evaluation (adequate number of invasive tumor cells present) |
| Results |
| Number of invasive tumor cells counted |
| Number of observers |
| Average number of <i>HER2</i> signals/nucleus or tile |
| Average number of CEP 17 chromosome probes/nucleus or tile |
| Ratio of average <i>HER2</i> signals/CEP 17 probe signals |
| Note: Tile is unit used for image system counting |
| Interpretation |
| Positive (amplified); equivocal; negative (not amplified); not interpretable; if IHC is being done because of problems with assay or results, this should also be indicated |
| Comment |
| If an FDA-approved method is used, it should be noted; if the FDA-approved method has been modified, a statement in the report should be included indicating what modifications were made and that the changes have been validated; if the test is not FDA approved or an FDA-approved test has been modified, a clear statement must be made that the laboratory reporting results takes responsibility for test performance |
| Abbreviations, FISH, fluorescent in situ hybridization; FDA, US Food and Drug Administration; HER2, human epidermal growth factor receptor 2. |

cancer factor assays. This legislation also requires biannual surveys of laboratories performing highly complex tests with defined criteria and actions required when performance is deficient. CAP and the Joint Commission on Accreditation of Healthcare Organizations have been given the deemed status to perform these inspections. US Food and Drug Administration regulates medical devices as a result of the 1976 Medical Devices Amendments. HER2 testing, which has potentially high impact on patient mortality and morbidity, is considered a high risk device (class 3) that requires intense premarket scrutiny. After review of the legislation and regulations that apply (Appendix H), the panel agreed that the current regulatory framework provided sufficient justification for the guideline recommendations without modification. We propose that this framework now be used to apply our proposed guidelines for HER2 testing.

What Are the Optimal External Quality Assurance Methods to Ensure Ongoing Accuracy in HER2 Testing?

Summary and recommendations. Currently there are no mandatory requirements for proficiency testing of HER2 analytes, such as protein expression by IHC and gene expression by FISH; although, this testing is offered as a voluntary educational program. The current guideline will make proficiency testing mandatory and require enhanced levels of scrutiny at the time of laboratory accreditation. The guideline is based on regulatory requirements of CLIA 88, published studies, previous CAP experience,^{34,67} experience of other groups,^{50,68,69} and expert panel consensus. The panel recommends

that HER2 testing be done in a CAP-accredited laboratory or in a laboratory that meets the accreditation and proficiency testing requirements set out within this article.

External quality assurance (laboratory accreditation). Beginning in 2007, the CAP Laboratory Accreditation Program will require that every CAP-accredited laboratory performing HER2 testing participate in a guideline concordant proficiency testing program for that testing. In the future, the panel recommends that all accrediting agencies require guideline concordant proficiency testing and laboratory accreditation requirements for HER2 testing. The Laboratory Accreditation Program will monitor performance in the required proficiency testing. Performance below 90% will be considered unsatisfactory and will require internal or external response consistent with accreditation program requirements. Responses must include identification of the cause of the poor performance, actions taken to correct the problem, and evidence that the problem has been corrected. The checklist of requirements for laboratories is presented in Table 11. International external quality assessment initiatives are described in Appendix I.

Proficiency testing requirements. All laboratories reporting HER2 results must participate in a guideline concordant proficiency testing (PT) program specific for each assay method used (ie, separate programs for IHC, FISH, brightfield ISH, image analysis). To be concordant with this guideline, PT programs must distribute specimens at least twice per year including a sufficient number of challenges (cases)

Table 11. CAP Laboratory Accreditation Elements Requiring Documentation at Inspection

| |
|---|
| Validation of test method |
| Use of standard operating procedures |
| Training of personnel involved in testing |
| Proper labeling of samples and reagents |
| Proper storage of samples and reagents |
| Equipment calibration and QC |
| Internal QA plan and evidence that it is followed |
| Quality of tests for interpretation |
| Ongoing competency assessment of technologists and pathologists |
| Report adequacy and quality |
| Accurate submission of results |

NOTE. Competency assessment is monitored by periodic or continuous review of performance of those doing tests against peers. When failure is documented, remediation is undertaken.
Abbreviations: CAP, College of American Pathologists; QC, quality control; QA, quality assurance.

to ensure adequate assessment of laboratory performance. For programs with 10 or more challenges per event, satisfactory performance requires correct identification of at least 90% of the graded challenges in each testing event. Laboratories with less than 90% correct responses on graded challenges in a given PT event are at risk for the next event. Laboratories that have unsatisfactory performance will be required to respond according to accreditation program requirements up to and including suspension of HER2 testing for the applicable method until performance issues are corrected.

Statistical Considerations for Proficiency Testing

Ideally, the test case set for proficiency testing should reflect as closely as possible the types of specimens that a laboratory will encounter in routine practice with respect to the range of inherent specimen characteristics and specimen handling and fixations conditions. There are likely to be practical limitations on the range of inherent specimen characteristics that can be represented in the test case set. For example, the specimens will be biased toward larger specimens, and the specimens must be ones for which there is sufficient biologic homogeneity and staining or hybridization consistency so that an unambiguous gold standard value can be determined. Test specimens may be presented in either tissue microarray (TMA) format or as conventional full sections.

Standard measures of performance for diagnostic tests having binary outcomes include sensitivity, specificity, and overall accuracy. Overall accuracy combines sensitivity and specificity into a single measure of the percentage of cases (positive and negative) for which the assay result is concordant with the true status (concordance rate). See Appendix F for a more detailed discussion of the statistical considerations involved in testing.

How Can These Efforts Be Implemented and the Effects Measured?

Summary and recommendations. To be effective, these recommendations must be widely communicated to the medical community and to patients both by educational efforts and by modifying the regulatory oversight of laboratories doing HER2 testing. We recommend coordinated educational efforts by both CAP and ASCO to provide such education and coordinate standardized review criteria among all agencies performing laboratory accreditation. In addition,

CAP will periodically publish the aggregate results of the proficiency testing results to make the oncology community aware of the improvements resulting from this strategy.

Educational requirements and communication strategies. For this guideline to be effectively implemented by laboratories anywhere in the world, there will need to be effective and widespread educational efforts of pathologists, oncologists, patients, and advocacy groups. CAP will offer online and live educational sessions about clinical necessity, testing requirements, test interpretation guidelines, and methods by which acceptable performance will be measured through laboratory accreditation and proficiency testing, and organizations in other parts of the world could play a similar role. ASCO will create education materials for oncologists and patients about how laboratory quality can be evaluated through review of reports and laboratory quality assurance activities. Pathologists must actively monitor the quality of their test procedures and oncologists on behalf of their patients must seek assurance that laboratories providing test results are appropriately accredited. These actions should improve the consistency of testing for HER2, although quantifying this improvement will be difficult. One of the important outcomes resulting from accurate HER2 testing is to ensure that every breast cancer patient who might benefit from anti-HER2 therapy be accurately and promptly identified, while those who would not benefit be spared a costly and potentially harmful placebo.

Review of Previous Educational Efforts of the College of American Pathologists

Since 1994 CAP has sponsored several consensus conferences about predictive and prognostic factors in cancer. The first conference on this subject in 1994 resulted in a publication detailing the potential value of testing for HER2 in breast cancer without specific recommendations for testing method.⁷⁰ The second conference, held in 1999, endorsed the evaluation of breast cancer for HER2 antigen or gene expression without specifying the method.⁵¹ In 2002, a specific Strategic Science symposium was convened by CAP to make recommendations to pathologists about how testing should be done for HER2, which led to the publication detailing very specific recommendations to pathologists (Table 12).⁴⁷

In 2003, CAP cosponsored (along with the National Cancer Institute [Bethesda, MD], NIST, and US Food and Drug Administration) a session to discuss the development of a standardized HER2 reference material for use in immunohistochemic HER2 assays. There was general consensus at the meeting that the availability of reference

Table 12. Summary Recommendations From the 2002 CAP HER2 Strategic Science Symposium

| |
|--|
| Use only formalin fixation for HER2 testing samples |
| Use standardized approved methods after laboratory validation |
| Train pathologists and technologists and measure performance |
| Use positive and negative controls with each batch and reference them |
| Use image analysis for interpretation |
| Use reporting template created with oncologist input |
| Engage in regular proficiency testing and do root cause analysis on deficiencies |
| Measure performance regularly, do correlations with other tests, and take remedial action when necessary |

Abbreviations: CAP, College of American Pathologists; HER2, human epidermal growth factor receptor 2.

material would help to standardize HER2 testing. Recommendations from that session are summarized in Table 13.⁶¹ NIST has since been funded by the National Cancer Institute to develop this reference material, consisting of cell lines with specific different levels of HER2 protein expression. Cell lines have been identified and standardized production is in process. CAP has also held numerous educational sessions on HER2 testing at each of its national meetings since 2002. There have also been sessions given at industry-sponsored workshops at various national pathology meetings, but the message provided by these sessions has not been uniform.

Modification of the regulatory environment. This guideline will be made available for review by organizations involved in laboratory accreditation and proficiency testing services in the USA. ASCO and CAP will jointly work to facilitate the dissemination of these guidelines. Efforts will be directed at enhancing the education of laboratories by requesting publication of guideline information in Morbidity and Mortality Weekly Report published by the Centers of Disease Control and Prevention. CAP will engage in significant live and online educational activities to help pathologists understand the significance of these changes in accreditation practice, beginning at the CAP annual meeting in September 2006. ASCO and CAP will provide educational opportunities (print, online, and society meetings) to educate health care professionals, patients, third party payers, and regulatory agencies. CAP will urge its members and participants in accreditation and proficiency testing programs to provide information in its reports specifying participation in laboratory accreditation. ASCO and CAP will work to coordinate these recommendations with those of other organizations, such as the National Comprehensive Cancer Network,⁷¹ the National Cancer Advisory Board, and patient advocacy organizations.

We are confident that these measures will improve performance of laboratories using these and future predictive testing methods. CAP will actively review results of proficiency testing and laboratory accreditation activities and periodically publish performance results. The organization will also work to include quality monitoring activities of HER2 testing in its programs designed for ongoing quality assessment, similar to CAP's Q-tracks and Q-probes.⁷²

Table 13. Summary Recommendations From 2003 NIST Consensus Workshop on Reference Material for HER2 Testing

| |
|--|
| NIST certifiable standard should be available breast cell lines (three) with variable receptor and gene expression to bracket the range of receptor quantity |
| Grown in optimal form |
| Characterized by NIST as to HER2 |
| Receptor content (scatchard plots) |
| Physical protein characteristics |
| Gene copy number |
| RNA expression levels |
| Second standard will be made commercially |
| Must be used in all HER2 laboratory testing by vendors/commercial labs/researchers |
| Fixative for HER2 testing and for reference standard must be required to be 10% buffered formalin |
| Must be included with each slide (or run) and used in all interlab and interassay comparisons |
| Abbreviations: NIST, National Institute of Standards and Technology; HER2, human epidermal growth factor receptor 2. |

Limitations of the Literature

Whether in the context of trastuzumab clinical trials or of studies comparing HER2 testing platforms, interpretation of the literature in the field of HER2 testing is complicated by a lack of standardization across trials in assay utilization and interpretation, presence or absence of confirmatory testing, and local versus central laboratory testing, among other considerations. Testing algorithms for HER2 were somewhat arbitrarily developed and assays used within algorithms have not always been standardized. While some assays have been carefully validated, others, especially the "home brew assays" have not, which complicates direct comparisons across trials and platforms, and we maintain this situation leads to either over- or undertreatment of a substantial percentage of patients with breast cancer.

In addition to published studies, the panel also considered previous guidelines and position statements from national and international professional organizations. Most of these earlier guidelines simply stated that HER2 testing should be performed, without addressing specific methodology, quality control, or associations with clinical outcomes. Guidelines have also emphasized the need for individual laboratories to document their own concordance experience of FISH versus IHC (90% for IHC 0 and 3+, and 95% for IHC 1+) before limiting reflex FISH testing only to IHC 2+ results. The guidelines considered were developed before the publication of the adjuvant trastuzumab data, and thus these concordance requirement parameters were set taking into account the palliative role of trastuzumab, and not the survival advantage shown in the adjuvant trials. Finally, other organizations have recommended algorithms based on best available data, which in fact have been quite sparse. It should be noted that testing algorithms described in existing guidelines assume a high level of correlation between IHC and FISH assays, which the existing literature shows may be unfounded.

An important gap in the literature identified by the panel concerns those patients with test results in the intermediate or equivocal range. The decision to treat with specific therapies like trastuzumab is by necessity dichotomous (yes or no). However, HER2 test results are derived from a continuous variable, which can be expected to lead to some results falling into a gray area. Adding to this confusion is the fact that there is significant variation in the intermediate ranges for both the IHC and FISH assays. The literature is lacking in this subgroup of patients with intermediate results, and there are also limited efficacy data in the subgroup tested with both high-quality IHC and FISH and found to have a discordant result. Patients with such results constitute poorly studied subgroups with less confidence in the scores and actual benefit from trastuzumab therapy. As these patient subgroups (and number of events) found within each of the individual adjuvant trastuzumab trials are relatively small, we urge those principal investigators to pool their data for a joint analysis to attempt to address some of these questions.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following authors or their immediate family members indicated a financial interest. No conflict exists for drugs or devices used in a study if they are not being

evaluated as part of the investigation. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment: N/A **Leadership:** N/A **Consultant:** Richard J. Cote, Clariant, Roche, Abbott; Mitchell Dowsett, Roche, Abbott Diagnostics; Soonmyung Paik, Genentech, GlaxoSmithKline; Mark D. Pegram, Genentech; Michael F. Press, Genentech; Anthony Rhodes, European Collection of Cell Cultures; Marc van de Vijver, Roche; Daniel F. Hayes, Immunicon, Pfizer, Novartis **Stock:** Richard J. Cote, Clariant **Honoraria:** M. Elizabeth H. Hammond, M2 Communications, Serinomics; Mitch Dowsett, Roche; Soonmyung Paik, Genentech, GlaxoSmithKline; Mark D. Pegram, Genentech; Michael F. Press, Genentech; Raymond Tubbs, Ventana, Abbott/Vysis; Marc van de Vijver, Roche **Research Funds:** Antonio C. Wolff, Roche; Richard J. Cote, Roche, Abbott; Raymond Tubbs, Ventana, Abbott/Vysis; Daniel F. Hayes, Pfizer, Novartis **Testimony:** Michael F. Press, Genentech **Other:** Raymond Tubbs, Nanoprobe

AUTHOR CONTRIBUTIONS

Conception and design: Antonio C. Wolff, M. Elizabeth H. Hammond, Jared N. Schwartz, Daniel F. Hayes

REFERENCES

1. Slamon DJ, Clark GM, Wong SG, et al: Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235:177-182, 1987
2. NCBI: PubMed. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=full_report&list_uids=2064
3. Akiyama T, Sudo C, Ogawara H, et al: The product of the human c-erbB-2 gene: A 185-kilodalton glycoprotein with tyrosine kinase activity. *Science* 232:1644-1646, 1986
4. Press M, Pike M, Chazin V, et al: Her-2/neu expression in node-negative breast cancer: Direct tissue quantitation by computerized image analysis and association of overexpression with increased risk of recurrent disease. *Cancer Res* 53:4960-4970, 1993
5. Press MF, Bernstein L, Thomas PA, et al: HER-2/neu gene amplification characterized by fluorescence in situ hybridization: Poor prognosis in node-negative breast carcinomas. *J Clin Oncol* 15:2894-2904, 1997
6. Yamauchi H, Stearns V, Hayes DF: When is a tumor marker ready for prime time? A case study of c-erbB-2 as a predictive factor in breast cancer. *J Clin Oncol* 19:2334-2356, 2001
7. Konecny G, Pauletti G, Pegram M, et al: Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. *J Natl Cancer Inst* 95:142-153, 2003
8. Ellis MJ, Coop A, Singh B, et al: Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: Evidence from a phase III randomized trial. *J Clin Oncol* 19:3808-3816, 2001
9. Menard S, Valagussa P, Pilotti S, et al: Response to cyclophosphamide, methotrexate, and fluorouracil in lymph node-positive breast cancer according to HER2 overexpression and other tumor biologic variables. *J Clin Oncol* 19:329-335, 2001

Administrative support: Karen L. Hagerty

Collection and assembly of data: Antonio C. Wolff, Karen L. Hagerty

Data analysis and interpretation: Antonio C. Wolff, M. Elizabeth H. Hammond, Jared N. Schwartz, Karen L. Hagerty, D. Craig Allred, Richard J. Cote, Mitchell Dowsett, Patrick L. Fitzgibbons, Wedad Hanna, Lisa M. McShane, Soonmyung Paik, Mark D. Pegram, Edith A. Perez, Michael F. Press, Anthony Rhodes, Catharine Sturgeon, Sheila Taube, Raymond Tubbs, Gail H. Vance, Marc van de Vijver, Thomas M. Wheeler, Daniel F. Hayes

Manuscript writing: Antonio C. Wolff, M. Elizabeth H. Hammond, Jared N. Schwartz, Karen L. Hagerty, D. Craig Allred, Richard J. Cote, Mitchell Dowsett, Patrick L. Fitzgibbons, Wedad Hanna, Amy Langer, Lisa M. McShane, Soonmyung Paik, Mark D. Pegram, Edith A. Perez, Michael F. Press, Anthony Rhodes, Catharine Sturgeon, Sheila Taube, Raymond Tubbs, Gail H. Vance, Marc van de Vijver, Thomas M. Wheeler, Daniel F. Hayes

Final approval of manuscript: Antonio C. Wolff, M. Elizabeth H. Hammond, Jared N. Schwartz, D. Craig Allred, Richard J. Cote, Mitchell Dowsett, Patrick L. Fitzgibbons, Wedad Hanna, Amy Langer, Lisa M. McShane, Soonmyung Paik, Mark D. Pegram, Edith A. Perez, Michael F. Press, Anthony Rhodes, Catharine Sturgeon, Sheila Taube, Raymond Tubbs, Gail H. Vance, Marc van de Vijver, Thomas M. Wheeler, Daniel F. Hayes

10. Thor AD, Berry DA, Budman DR, et al: ErbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. *J Natl Cancer Inst* 90:1346-1360, 1998

11. Pritchard KI, Shepherd LE, O'Malley FP, et al: HER2 and responsiveness of breast cancer to adjuvant chemotherapy. *N Engl J Med* 354:2103-2111, 2006

12. Cardoso F, Durbecq V, Larsimont D, et al: Correlation between complete response to anthracycline-based chemotherapy and topoisomerase II-alpha gene amplification and protein overexpression in locally advanced/metastatic breast cancer. *Int J Oncol* 24:201-209, 2004

13. Villman K, Sjostrom J, Heikkila R, et al: TOP2A and HER2 gene amplification as predictors of response to anthracycline treatment in breast cancer. *Acta Oncol* 45:590-596, 2006

14. Konecny GE, Thomssen C, Luck HJ, et al: Her-2/neu gene amplification and response to paclitaxel in patients with metastatic breast cancer. *J Natl Cancer Inst* 96:1141-1151, 2004

15. Hayes DF, Thor A, Dressler L, et al: HER2 predicts benefit from adjuvant paclitaxel after AC in node-positive breast cancer: CALGB 9344. *J Clin Oncol* 24:5S, 2006 (abstr 510)

16. Slamon DJ, Leyland-Jones B, Shak S, et al: Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 344:783-792, 2001

17. Cobleigh MA, Vogel CL, Tripathy D, et al: Multinational study of the efficacy and safety of humanized anti-HER2 monoclonal antibody in women who have HER2-overexpressing metastatic breast cancer that has progressed after chemotherapy for metastatic disease. *J Clin Oncol* 17:2639-2648, 1999

18. Vogel CL, Cobleigh MA, Tripathy D, et al: Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J Clin Oncol* 20:719-726, 2002

19. Slamon D, Eiermann W, Robert N, et al: Phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel (ACT) with doxorubicin and cyclophosphamide fol-

lowed by docetaxel and trastuzumab (ACTH) with docetaxel, carboplatin and trastuzumab (TCH) in HER2 positive early breast cancer patients: BCIRG 006 study. *Breast Cancer Res Treat* 94:S5, 2005 (supp 1; abstr 1)

20. Joensuu H, Kellokumpu-Lehtinen P-L, Bono P, et al: Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *N Engl J Med* 354:809-820, 2006

21. Romond EH, Perez EA, Bryant J, et al: Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 353:1673-1684, 2005

22. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al: Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 353:1659-1672, 2005

23. The Herceptin Adjuvant (HERA) Trial Study Team: Trastuzumab following adjuvant chemotherapy in HER2-positive early stage breast cancer (HERA trial): Disease-free and overall survival after 2 year follow-up. *J Clin Oncol* 24, 2006 presented as a late breaking abstract at the 42nd Annual Meeting of the American Society of Clinical Oncology, Atlanta, GA, June 2-6, 2006

24. Geyer CE, Cameron D, Lindquist D, et al: A phase III randomized, open-label, international study comparing lapatinib and capecitabine vs. capecitabine in women with refractory advanced or metastatic breast cancer (EGF100151). *J Clin Oncol*; 2006 presented as a late breaking abstract at the 42nd Annual Meeting of the American Society of Clinical Oncology, Atlanta, GA, June 2-6, 2006

25. Slamon DJ, Godolphin W, Jones LA, et al: Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* 244:707-712, 1989

26. Mass RD, Sanders C, Charlene K, et al: The concordance between the clinical trials assay (CTA) and fluorescence in situ hybridization (FISH) in the Herceptin pivotal trials. *Proc Am Soc Clin Oncol* 19:75a, 2000 (abstr 291)

27. Seidman AD, Berry D, Cirincione C, et al: CALGB 9840: Phase III study of weekly (W) paclitaxel (P) via 1-hour(h) infusion versus standard (S) 3h

infusion every third week in the treatment of metastatic breast cancer (MBC), with trastuzumab (T) for HER2 positive MBC and randomized for T in HER2 normal MBC. *J Clin Oncol* 22: 6s, 2004 (abstr 512)

28. Yaziji H, Goldstein LC, Barry TS, et al: HER-2 testing in breast cancer using parallel tissue-based methods. *JAMA* 291:1972-1977, 2004

29. Owens MA, Horten BC, Da Silva MM: HER2 amplification ratios by fluorescence in situ hybridization and correlation with immunohistochemistry in a cohort of 6556 breast cancer tissues. *Clin Breast Cancer* 5:63-69, 2004

30. Paik S, Bryant J, Tan-Chiu E, et al: Real-world performance of HER2 testing-National Surgical Adjuvant Breast and Bowel Project experience. *J Natl Cancer Inst* 94:852-854, 2002

31. Perez EA, Suman VJ, Davidson NE, et al: HER2 testing by local, central, and reference laboratories in specimens from the North Central Cancer Treatment Group N9831 Intergroup adjuvant trial. *J Clin Oncol* 24:3032-3038, 2006

32. Isola J, Tanner M, Forsyth A, et al: Interlaboratory comparison of HER-2 oncogene amplification as detected by chromogenic and fluorescence in situ hybridization. *Clin Cancer Res* 10:4793-4798, 2004

33. Press M, Bernstein L, Sauter G, et al: Topoisomerase II-alpha gene amplification as a predictor of responsiveness to anthracycline-containing chemotherapy in the Cancer International Research Group 006 clinical trial of trastuzumab (Herceptin) in the adjuvant setting. *Breast Cancer Res Treat* 94:S54, 2005 (suppl 1; abstr 1045)

34. Persons DL, Tubbs RR, Cooley LD, et al: HER-2 fluorescence in situ hybridization: Results from the survey program of the College of American Pathologists. *Arch Pathol Lab Med* 130:325-331, 2006

35. Dal Lago L, Durbecq V, Desmedt C, et al: Correction for chromosome 17 is critical for determination of true Her-2/neu gene amplification status in breast cancer (BC). *Mol Cancer Ther* 5:2572-2579, 2006

36. Vera-Roman JM, Rubio-Martinez LA: Comparative assays for the HER-2/neu oncogene status in breast cancer. *Arch Pathol Lab Med* 128:627-633, 2004

37. Hanna WM, Kwok K: Chromogenic in-situ hybridization: A viable alternative to fluorescence in-situ hybridization in the HER2 testing algorithm. *Mod Pathol* 19:481-487, 2006

38. Downs-Kelly E, Pettay J, Hicks D, et al: Analytical validation and interobserver reproducibility of EnzMet GenePro: A second-generation bright-field metallography assay for concomitant detection of HER2 gene status and protein expression in invasive carcinoma of the breast. *Am J Surg Pathol* 29:1505-1511, 2005

39. Bast RC Jr, Ravdin P, Hayes DF, et al: 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: Clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 19:1865-1878, 2001

40. Roche PC, Suman VJ, Jenkins RB, et al: Concordance between local and central laboratory HER2 testing in the breast intergroup trial N9831. *J Natl Cancer Inst* 94:855-857, 2002

41. Gutman S: Regulatory issues in tumor marker development. *Semin Oncol* 29:294-300, 2002

42. Hayes DF, Picard MH: Heart of darkness: The downside of trastuzumab. *J Clin Oncol* 24:4056-4058, 2006

43. Tan-Chiu E, Yothers G, Romond E, et al: Assessment of cardiac dysfunction in a randomized

trial comparing doxorubicin and cyclophosphamide followed by paclitaxel, with or without trastuzumab as adjuvant therapy in node-positive, human epidermal growth factor receptor 2-overexpressing breast cancer: NSABP B-31. *J Clin Oncol* 23:7811-7819, 2005

44. Perez E, Suman V, Davidson N, et al: Exploratory analysis from NCCTG N9831: Do clinical and laboratory characteristics predict cardiac toxicity of trastuzumab when administered as a component of adjuvant therapy? *Breast Cancer Res Treat* 94: 2005 (suppl 1; abstr 2038)

45. Geyer CE, Bryant JL, Romond EH, et al: Four year update of cardiac dysfunction on NSABP B-31: A randomized trial of AC->paclitaxel vs. AC-> paclitaxel with trastuzumab in HER2 positive, node positive operable breast cancer. *J Clin Oncol* 24: 23S, 2006 (abstr 581)

46. Cell Markers and Cytogenetics Committees College Of American Pathologists: Clinical laboratory assays for HER-2/neu amplification and overexpression: Quality assurance, standardization, and proficiency testing. *Arch Pathol Lab Med* 126:803-808, 2002

47. Zarbo RJ, Hammond ME: Her-2/neu testing of breast cancer patients in clinical practice. *Arch Pathol Lab Med* 127:549-553, 2003

48. Ellis IO, Bartlett J, Dowsett M, et al: Best practice No 176: Updated recommendations for HER2 testing in the UK. *J Clin Pathol* 57:233-237, 2004

49. Hanna W, O'Malley F: Updated recommendations from the HER2/neu consensus meeting. *Current Oncology* 9:S18-S19, 2002 (suppl 1)

50. Bilous M, Dowsett M, Hanna W, et al: Current perspectives on HER2 testing: A review of national testing guidelines. *Mod Pathol* 16:173-182, 2003

51. Fitzgibbons PL, Page DL, Weaver D, et al: Prognostic factors in breast cancer: College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* 124:966-978, 2000

52. Vincent-Salomon A, MacGrogan G, Couturier J, et al: Calibration of immunohistochemistry for assessment of HER2 in breast cancer: Results of the French multicentre GEFPICS study. *Histopathology* 42:337-347, 2003

53. Allred DC, Clark GM, Tandon AK, et al: HER-2/neu in node-negative breast cancer: Prognostic significance of overexpression influenced by the presence of in situ carcinoma. *J Clin Oncol* 10:599-605, 1992

54. Tubbs RR, Pettay JD, Roche PC, et al: Discrepancies in clinical laboratory testing of eligibility for trastuzumab therapy: Apparent immunohistochemical false-positives do not get the message. *J Clin Oncol* 19:2714-2721, 2001

55. Perez EA, Roche PC, Jenkins RB, et al: HER2 testing in patients with breast cancer: Poor correlation between weak positivity by immunohistochemistry and gene amplification by fluorescence in situ hybridization. *Mayo Clin Proc* 77:148-154, 2002

56. Press MF, Sauter G, Bernstein L, et al: Diagnostic evaluation of HER-2 as a molecular target: An assessment of accuracy and reproducibility of laboratory testing in large, prospective, randomized clinical trials. *Clin Cancer Res* 11:6598-6607, 2005

57. Downs-Kelly E, Yoder BJ, Stoler M, et al: The influence of polysomy 17 on HER2 gene and protein expression in adenocarcinoma of the breast: A fluorescent in situ hybridization, immunohistochemical, and isotopic mRNA in situ hybridization study. *Am J Surg Pathol* 29:1221-1227, 2005

58. Ma Y, Lespagnard L, Durbecq V, et al: Polysomy 17 in HER-2/neu status evaluation in breast

cancer: Effect on daily practice. *Clin Cancer Res* 11:4393-4399, 2005

59. Reddy JC, Reimann JD, Anderson SM, et al: Concordance between central and local laboratory HER2 testing from a community-based clinical study. *Clin Breast Cancer* 7:153-157, 2006

60. FDA/CBER: Clinical Review Briefing Document sBLA STN: 103792/5008/0 [Original BLA 98-0369] Trastuzumab. http://www.fda.gov/ohrms/dockets/ac/01/briefing/3815b1_08_HER2%20FISH.pdf

61. Hammond ME, Barker P, Taube S, et al: Standard reference material for Her2 testing: Report of a National Institute of Standards and Technology-sponsored consensus workshop. *Appl Immunohistochem Mol Morphol* 11:103-106, 2003

62. Press MF, Slamon DJ, Flom KJ, et al: Evaluation of HER-2/neu gene amplification and overexpression: Comparison of frequently used assay methods in a molecularly characterized cohort of breast cancer specimens. *J Clin Oncol* 20:3095-3105, 2002

63. Hanna W, O'Malley FO: Updated recommendations from the HER2/neu consensus meeting: Toronto, Ontario, 7 September 2001. *Curr Oncol* 9:S18-S20, 2002 (suppl 1)

64. Dowsett M, Bartlett J, Ellis IO, et al: Correlation between immunohistochemistry (HercepTest) and fluorescence in situ hybridization (FISH) for HER-2 in 426 breast carcinomas from 37 centres. *J Pathol* 199:418-423, 2003

65. Taylor CR: The total test approach to standardization of immunohistochemistry. *Arch Pathol Lab Med* 124:945-951, 2000

66. Hicks DG, Tubbs RR: Assessment of the HER2 status in breast cancer by fluorescence in situ hybridization: A technical review with interpretive guidelines. *Hum Pathol* 36:250-261, 2005

67. Fitzgibbons P, Murphy D, Dorfman D, et al: Interlaboratory comparison of immunohistochemical testing for HER2: Results of the 2004 and 2005 College of American Pathologists HER2 immunohistochemistry tissue microarray survey. *Arch Pathol Lab Med* 130:1440-1445, 2006

68. Rhodes A, Jasani B, Anderson E, et al: Evaluation of HER-2/neu immunohistochemical assay sensitivity and scoring on formalin-fixed and paraffin-processed cell lines and breast tumors: A comparative study involving results from laboratories in 21 countries. *Am J Clin Pathol* 118:408-417, 2002

69. O'Malley F, Thomson T, Julian J, et al: HER2 status: A Canadian experience of concordance between central and local testing laboratories. *Breast Cancer Res Treat* 82:305, 2003

70. Henson DE, Fielding LP, Grignon DJ, et al: College of American Pathologists Conference XXVI on clinical relevance of prognostic markers in solid tumors. *Arch Pathol Lab Med* 119:1109-1112, 1995

71. Carlson RW, Moench SJ, Hammond ME, et al: HER2 testing in breast cancer: NCCN Task Force report and recommendations. *J Natl Compr Canc Netw* 4:S1-S22, 2006 (suppl 3); quiz S23-S24, 2006

72. Zarbo RJ, Jones BA, Friedberg RC, et al: Q-tracks: A College of American Pathologists program of continuous laboratory monitoring and longitudinal tracking. *Arch Pathol Lab Med* 126:1036-1044, 2002

73. Dako A/S G, Denmark: HercepTest package insert. http://www.dakousa.com/prod_download-packageinsert.pdf?objectid=105073003

Guideline for HER2 Testing in Breast Cancer

74. Carlson RW, Anderson BO, Burstein HJ, et al: Breast Cancer. *J Natl Compr Canc Netw* 3:238-289, 2005
75. Perez EA, Suman VJ, Davidson NE, et al: HER2 testing by local, central, and reference laboratories in the NCCTG N9831 Intergroup adjuvant trial. *J Clin Oncol* 22:567, 2004
76. Mass RD, Press MF, Anderson S, et al: Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer* 6:240-246, 2005
77. Bartlett JM, Going JJ, Mallon EA, et al: Evaluating HER2 amplification and overexpression in breast cancer. *J Pathol* 195:422-428, 2001
78. Dybdal N, Leiberman G, Anderson S, et al: Determination of HER2 gene amplification by fluorescence in situ hybridization and concordance with the clinical trials immunohistochemical assay in women with metastatic breast cancer evaluated for treatment with trastuzumab. *Breast Cancer Res Treat* 93:3-11, 2005
79. Masood S, Bui MM: Assessment of Her-2/neu overexpression in primary breast cancers and their metastatic lesions: An immunohistochemical study. *Ann Clin Lab Sci* 30:259-265, 2000
80. Simon R, Nocito A, Hubscher T, et al: Patterns of HER-2/neu amplification and overexpression in primary and metastatic breast cancer. *J Natl Cancer Inst* 93:1141-1146, 2001
81. Hammond M, Taube S: Issues and barriers to development of clinically useful tumor markers: A development pathway proposal. *Semin Oncol* 29:213-221, 2002
82. Rhodes A, Borthwick D, Sykes R, et al: The use of cell line standards to reduce HER-2/neu assay variation in multiple European cancer centers and the potential of automated image analysis to provide for more accurate cut points for predicting clinical response to trastuzumab. *Am J Clin Pathol* 122:51-60, 2004
83. Gulmann C, Loring P, O'Grady A, et al: Miniature tissue microarrays for HercepTest standardisation and analysis. *J Clin Pathol* 57:1229-1231, 2004
84. Wick MR, Swanson PE: Targeted controls in clinical immunohistochemistry: A useful approach to quality assurance. *Am J Clin Pathol* 117:7-8, 2002
85. Moskaluk CA: Standardization of clinical immunohistochemistry: Why, how, and by whom? *Am J Clin Pathol* 118:669-671, 2002
86. Bloom K, Harrington D: Enhanced accuracy and reliability of HER-2/neu immunohistochemical scoring using digital microscopy. *Am J Clin Pathol* 121:620-630, 2004
87. Wang S, Saboorian MH, Frenkel EP, et al: Assessment of HER-2/neu status in breast cancer: Automated cellular imaging system (ACIS)-assisted quantitation of immunohistochemical assay achieves high accuracy in comparison with fluorescence in situ hybridization assay as the standard. *Am J Clin Pathol* 116:495-503, 2001
88. US Food and Drug Administration: Clinical Laboratory Improvement Amendments. www.fda.gov/cdrh/clia
89. Dowsett M, Hanby AM, Laing M, et al: HER-2 testing in the UK: Consensus from a national consultation. *J Clin Pathol*, in press

Acknowledgment

The panel is grateful to Deborah Collyar, Nancy Davidson, MD, Christopher Desch, MD, Cindy Geoghegan, David Hicks, MD, Michael Neuss, MD, Lisa Newman, MD, Frances O'Malley, Martine Piccart-Gebhart, MD, PhD, George Sledge, MD, and the American Society of Clinical Oncology Health Services Committee and the College of American Pathologists Council on Scientific Affairs for reviewing earlier drafts of the manuscript.

Appendix A

| Panel Member | Institution |
|---|---|
| M. Elizabeth H. Hammond, MD, Co-Chair | Intermountain Healthcare, University of Utah School of Medicine |
| Daniel F. Hayes, MD, Co-Chair | University of Michigan Comprehensive Cancer Center, University of Michigan Health and Hospital System |
| Jared N. Schwartz, MD, PhD, FACP, Co-Chair | Presbyterian Hospital |
| Antonio C. Wolff, MD, FACP, Co-Chair | The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins |
| D. Craig Allred, MD | Washington University School of Medicine in St Louis |
| Richard J. Cote, MD, FRCPath | University of Southern California Keck School of Medicine/Norris Comprehensive Cancer Center, Los Angeles |
| Mitchell Dowsett, PhD | Royal Marsden Hospital |
| Patrick L. Fitzgibbons, MD, FCAP | St Jude Medical Center |
| Wedad M. Hanna, MD, FRCPC(C) | Sunnybrook Health Sciences Centre/Women's College Hospital |
| Amy Langer | Patient advocate and educator |
| Lisa M. McShane, PhD | National Cancer Institute |
| Soonmyung Paik, MD | National Surgical Adjuvant Breast and Bowel Project |
| Mark D. Pegram, MD | University of California in Los Angeles, Geffen School of Medicine |
| Edith A. Perez, MD | Mayo Clinic Jacksonville |
| Michael F. Press, MD, PhD | University of Southern California |
| Anthony Rhodes, PhD | University of the West of England, Bristol |
| Catharine Sturgeon, Ph.D., MRCPATH, FRCB | The Royal Infirmary of Edinburgh |
| Sheila E. Taube, PhD | National Cancer Institute |
| Raymond Tubbs, DO | Cleveland Clinic Foundation |
| Gail H. Vance, MD | Indiana University |
| Marc van de Vijver, MD | Netherlands Cancer Institute |
| Thomas M. Wheeler, MD | Baylor College of Medicine |
| Steven Gutman, MD, MBA, <i>Ex-Officio</i> | Food and Drug Administration |
| Patricia Keegan, MD, <i>Ex-Officio</i> | Food and Drug Administration |
| Judy Yost, MA, MT (ASCP), <i>Ex-Officio</i> | Centers for Medicare & Medicaid Services |

Appendix B

| Guest | Affiliation |
|----------------------------|---|
| Kenneth J. Bloom, MD | Clariant Inc (Aliso Viejo, CA) |
| Jon Askaa, DVM, PhD | Dako A/S (Glostrup, Denmark) |
| Patrick C. Roche, PhD | Ventana Medical Systems Inc (Tucson, AZ) |
| Mary Hickman | Abbott (Vysis) Molecular Inc (Des Plaines, IL) |
| Allen M. Gown, MD | PhenoPath Laboratories (Seattle, WA) |
| Jess Savala Jr, MD | US LABS (LabCorp; Irvine, CA) |
| Arturo Anguiano, MD | Quest Diagnostics (Lyndhurst, NJ) |
| Robert D. Mass, MD | Genentech Inc (South San Francisco, CA) |
| Robert B. Jenkins, MD, PhD | Mayo Medical Labs (Rochester, MN) |

Appendix C**Evidence on HER2 Status and Trastuzumab Benefit**

HER2 testing in early trials. Approximately 6,000 tumor samples were centrally screened at LabCorp for eligibility to three of the early studies of trastuzumab in metastatic breast cancer.¹⁶⁻¹⁸ Two thirds lacked HER2 overexpression (defined as an IHC score of 0 or 1+) and were presumed unlikely to benefit from anti-HER2 therapy with trastuzumab, while the remaining one third was determined to be HER2-positive (defined as an IHC score of 2+ or 3+).⁶⁰ As the US Food and Drug Administration Center for Biologics Evaluation and Research required that a diagnostic assay be made available for clinical use in situations where the decision to use a therapeutic product is dependent on that information, an IHC assay (HercepTest; Dako A/S, Glostrup, Denmark) was approved along side trastuzumab in 1998.

The HercepTest uses the same scoring criteria as the CTA but a different (polyclonal) antibody,⁷³ and validation studies using clinical outcome (benefit from trastuzumab) as the end point of interest could not be done as available precut slides from the trastuzumab efficacy trials were considered technically unsuitable for this specific purpose.⁶⁰

An exploratory analysis of data from the randomized trial in metastatic disease suggested that trastuzumab benefit was greater among patients with an IHC score of 3+.¹⁶ While the US Food and Drug Administration considered those data insufficient to justify restricting the approval of trastuzumab to this patient subgroup, Genentech (South San Francisco, CA) was asked to further evaluate the clinical outcome of patients selected for trastuzumab therapy on the basis of the HercepTest and other HER2 assays.⁶⁰

Clinical algorithm and concordance studies. The early experience with IHC testing using the CTA for patient selection influenced the adoption in clinical practice of the testing algorithm most commonly used, in which HER2 status is determined by IHC or FISH and an IHC score of 2+ will trigger reflex FISH testing (Tables A3, A4, and A5).⁷⁴ Major deficiencies were observed in the determination of HER2 status among patients found to have HER2-positive disease based on local testing and referred for enrollment in two adjuvant trastuzumab trials in North America with a high rate of false-positive results observed among the first few hundred patients.^{30,40} An amendment mandated confirmatory testing in a central or reference laboratory before random assignment to reduce the chances of enrolling patients with a false-positive HER2 results on those studies, but it resulted in only a modest improvement in the final concordance rate between local and central laboratory (IHC 81.6% and FISH 88.1%; Table A3).³¹

Concerns about false-negative results were also raised in a study by Perez et al⁵⁵ where 12% of patients (n = 216) with an IHC score of 2+ had evidence of gene amplification by FISH. Similar concerns led to previous recommendations by CAP that a laboratory should perform reflex FISH for both IHC scores of 1+ and 2+ until it had shown that less than 5% of its IHC 1+ scores were FISH-positive.⁴⁷ Subsequent data from trial N9831 show that the concordance between local versus central concordance of both IHC and FISH continued to be suboptimal.³¹

Press et al⁶² compared the performance of several HER2 assays using 117 specimens with known gene amplification and protein overexpression levels and observed concordance rates ranging from 95.7% to 97.4% for two FISH assays and 88.9% to 95.7% for four IHC assays (two “home brews” and two commercially available). Bartlett et al⁷⁷ observed a concordance rate between FISH amplified and IHC 2+/3+ of 67% with the HercepTest and 83% with the CB11 antibody. These data showed that the commercially available, US Food and Drug Administration-approved IHC methods were statistically significantly less accurate than FISH at correctly characterizing these tumors with known HER2 status, and support a minority view within the panel that IHC is not a sufficiently desirable assay method and that FISH should be preferentially used. However, data from trial N9831 show that the concordance between local versus central HER2 testing using IHC or FISH remains suboptimal (Table A3).³¹

In contrast, a majority within the panel was concerned about existing gaps in our knowledge in regards to whether one assay is superior to the other in predicting benefit from anti-HER2 therapy. In addition, up to 3% have discordant IHC/FISH results (ie, IHC 3+/FISH nonamplified and IHC < 3+/FISH amplified), and the available clinical data are insufficient to determine whether these represent distinct biologic subgroups with therapeutic implications versus inaccurate IHC results.

A concordance study between the CTA IHC assay and FISH (PathVysion, Abbott-Vysis, Chicago, IL) done at the same laboratory (LabCorp) using tumor specimens from the early metastatic trials¹⁶⁻¹⁸ showed that HER2 amplification in the 0, 1+, 2+, 3+ groups was observed in 4.2%, 6.7%, 23.9%, and 89.3% of the samples, respectively, for an overall concordance rate of 82%. A subset of 488 samples was also tested with FISH in a second central laboratory at the University of Southern California Los Angeles, CA, for an overall concordance rate of 92% (HER2-positive results observed in 71% and 78% of specimens, respectively),⁷⁸ which is somewhat similar to the concordance data observed in the North American Intergroup adjuvant randomized trial for samples tested in two reference laboratories.³¹ Press et al⁵⁶ observed a concordance rate between local and central FISH testing of 92% among samples submitted for central testing in the Breast Cancer International Research Group trial 006.

Accuracy of IHC and FISH in predicting trastuzumab benefit. It is important to remember that as the entry criteria for initial studies of trastuzumab was an IHC score of 2+ or 3+ by real-time central testing using the CTA, clinical outcome data are unavailable for patients

Table A3. Time Evolution of HER2 Testing Concordance in NCCTG 9831: Concordance Central v Local Laboratory

| Method | Study | | |
|---------------------|----------------------------------|----------------------------------|------------------------------------|
| | Roche ⁴⁰ (N = 119) | Perez ⁷⁵ (N = 976) | Perez ³¹ (N = 2,535) |
| IHC 3+ (HercepTest) | 74% | 79.5% | 81.6% |
| FISH+ (PathVysion) | 67% | 85% | 88.1% |

Abbreviations: HER2, human epidermal growth factor receptor 2; NCCTG, North Central Cancer Treatment Group; IHC, immunohistochemistry; HercepTest, Dako A/S, Glostrup, Denmark; FISH, fluorescent in situ hybridization; PathVysion, Abbott/Vysis Molecular Inc, Des Plaines, IL.

Table A4. Objective Response Rate According to HER2 Status (FISH) in the Randomized Trial H648g

| HER2 Status | Therapy (N = 451) | | | | P |
|-------------|----------------------------|----|--------------------|----|---------|
| | Chemotherapy + Trastuzumab | | Chemotherapy Alone | | |
| | No. | % | No. | % | |
| FISH+ | 95/176 | 54 | 51/168 | 30 | < .0001 |
| FISH- | 19/50 | 38 | 22/57 | 39 | NS |

NOTE. Adapted from Mass et al.⁷⁶
Abbreviations: HER2, human epidermal growth factor receptor 2; FISH, fluorescent in situ hybridization; NS, not significant.

with an initial score of 0 or 1+ since they were considered ineligible. These algorithms also use HER2 amplification as the gold standard, when in fact the optimal outcome of interest should be clinical benefit from anti-HER2 therapy. Mass et al⁷⁶ reanalyzed the clinical outcome data from the randomized trial of chemotherapy ± trastuzumab (trial H648g) according to HER2 status by IHC (prospective central testing using the CTA) and FISH (retrospective central testing using PathVysion), and suggested that benefit was limited to patients with tumors showing gene amplification (HER2:CEP17 ratio ≥ 2; Table A4). Unfortunately, this analysis provided no outcome data for the subset of patients with IHC 2+/FISH-positive and IHC 3+/FISH-negative disease as the authors judged that statistical power was insufficient to adequately address the issue retrospectively.

An unplanned underpowered subset analysis from trial H0649g in 209 patients treated with trastuzumab after progression on chemotherapy for metastatic disease¹⁷ suggested a differential response rate observed in patients with HER2 protein overexpression and gene amplification (higher response in IHC 3+/FISH-positive as compared with IHC 2+/FISH-positive), but no responses were seen in patients whose tumors lacked evidence of gene amplification regardless of the IHC score (Table A5). A similarly unplanned, underpowered subset analysis from trial H648g (Table A6) shows conflicting data in the subset of patients with IHC 2+/FISH-positive. These patients appear to have benefited from the addition of trastuzumab to chemotherapy according to some parameters (time to tumor progression) but not others (response rate and overall survival).⁶⁰ These conflicting data highlight the existing gaps in our knowledge and are summarized in Table A7, but this retrospective analysis conducted by the US Food and Drug Administration lacks statistical power for definitive statements.⁶⁰

Little is known from randomized trials on the benefit of trastuzumab in patients who lack HER2 overexpression. This is further compounded by the fact that many of these patients had HER2 testing done locally by IHC only (IHC 0, 1+, and 2+ if negative FISH). The Cancer and Leukemia Group B trial C9840 is the only study that randomly assigned these patients who were then treated with paclitaxel with or without trastuzumab,²⁷ and a trend towards improved time to progression was observed among patients randomly assigned to trastuzumab (n = 113) versus not (n = 115). Possible explanations include: the play of chance; false negative local HER2 testing; and a conversion in HER2 status from the primary to a metastatic site.^{79,80} Further analysis correlating these data with central testing (including FISH) is ongoing. There are also no reported data on any proportional benefit from trastuzumab according to various levels of gene amplification, and such effort will likely require pooling of the data from the various randomized adjuvant trials.

Table A5. Overall Response Rate to Single-Agent Trastuzumab (Trial H649g) According to HER2 Status (FISH and IHC)

| HER2 Status | IHC (CTA) | | | | | |
|-------------------|-----------|----|--------|----|--------|----|
| | CTA 3+ | | CTA 2+ | | All | |
| | No. | % | No. | % | No. | % |
| FISH (PathVysion) | | | | | | |
| Positive | 30/136 | 22 | 3/27 | 11 | 33/163 | 20 |
| Negative | 0/21 | 0 | 0/25 | 0 | 0/46 | 0 |
| All | 30/157 | 19 | 3/52 | 6 | 33/209 | 14 |

NOTE. Patients with metastatic breast cancer who had progressed on prior chemotherapy. Analysis derived from an unplanned subset analysis of Cobleigh et al¹⁷ by the FDA.⁶⁰
Abbreviations: HER2, human epidermal growth factor receptor 2; PathVysion, Abbott/Vysis Molecular Inc, Des Plaines, IL; FISH, fluorescent in situ hybridization; IHC, immunohistochemistry; CTA, clinical trials assay.

Table A6. Outcome Observed With the Addition of Trastuzumab to Chemotherapy As First-Line Chemotherapy (Trial H648g) According to HER2 Status (FISH and IHC)

| FISH (PathVysion) | IHC (CTA) | | | | | | | | | | | | | |
|----------------------------------|----------------------------|----|--------------|----|----------------------------|----|--------------|---------|----------------------------|----|--------------|----|--|--|
| | 2+ and 3+ | | | | 3+ | | | | 2+ | | | | | |
| | Chemotherapy + Trastuzumab | | Chemotherapy | | Chemotherapy + Trastuzumab | | Chemotherapy | | Chemotherapy + Trastuzumab | | Chemotherapy | | | |
| | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | | |
| Response Rate | | | | | | | | | | | | | | |
| Any* (n = 469) | 106/235 | 45 | 67/234 | 29 | 87/176 | 49 | 46/173 | 27 | 19/59 | 32 | 21/61 | 34 | | |
| Positive (n = 325) | 89/164 | 54 | 49/161 | 30 | 81/148 | 55 | 40/145 | 28 | 8/16 | 50 | 9/16 | 56 | | |
| Negative (n = 126) | 25/62 | 40 | 24/64 | 38 | 13/21 | 62 | 12/22 | 55 | 12/41 | 29 | 12/42 | 29 | | |
| Median overall survival | | | | | n = 169 | | | n = 167 | | | n = 57 | | | |
| Positive, months (n = 325) | | | | | 26.3 | | | 18.8 | | | 21.4 | | | |
| Negative, months (n = 126) | | | | | 29.1 | | | 29.7 | | | 19.3 | | | |
| Median time to tumor progression | | | | | | | | | | | | | | |
| Positive, months (n = 325) | | | | | 7.3 | | | 4.6 | | | 9.4 | | | |
| Negative, months (n = 126) | | | | | 11.3 | | | 6.3 | | | 5.3 | | | |

NOTE. This unplanned, underpowered subset analysis of the data from Slamon et al¹⁶ conducted by the FDA⁶⁰ suggests that the benefit from adding trastuzumab to chemotherapy was restricted to patients with HER2 protein overexpression and gene amplification (IHC 3+/FISH+), and was not seen in patients with protein overexpression without gene amplification (IHC 3+/FISH-) or equivocal overexpression with gene amplification (IHC 2+/FISH+) disease. This was observed in regards to response rate and median overall survival, but not median time to tumor progression.

Abbreviations: HER2, human epidermal growth factor receptor 2; FISH, fluorescent in situ hybridization; IHC, immunohistochemistry; CTA, clinical trials assay.

*Data from trastuzumab package insert.

Appendix D

Evidence on HER2 Testing Variation

Accuracy of HER2 testing is critical to ensure that those patients most likely to benefit are offered trastuzumab while those unlikely to benefit are spared the cost and toxicity of this agent. Unfortunately several sources of variability may result in inaccurate HER2 results.

The optimal type of assay is controversial. Because trastuzumab binds exclusively to HER2 protein, in theory measures of this molecule would provide the best indication of whether patients should or should not receive it. In this regard, there are several methods to measure the expression of HER2 protein, including Western blotting, enzyme-linked immunoassay, and in situ immunostaining techniques, such as immunofluorescence and IHC. Of these, IHC has gained the widest acceptance in clinicopathological laboratories because it is used for many different types of proteins in these laboratories and can be performed on formalin-fixed, paraffin-embedded tissue. However, various factors discussed in this section may affect the performance of IHC and result in substantial variability (Table 3). For this reason, measures of *HER2* amplification have been proposed as more accurately performed assays. These can be performed as either brightfield or fluorescence ISH assays. Of these, FISH has gained widespread clinical use in most pathology departments.

Significant variation in HER2 testing practices has been documented or can be inferred from the current circumstances of testing. These sources of variation stem from nearly every step of the process (ie, preanalytic, analytic, and postanalytic; Table 3).

There is no standardized single method for immunohistochemical detection of HER2 antigens in breast cancer tissue, nor is there a single standardized method for detection of *HER2* genes in breast cancer tissue. For example, although methods of detection specify that tissue for this testing should be fixed only in 10% buffered formalin, many laboratories use a variety of other fixatives for breast cancer specimens that dramatically alters the results of testing.⁶¹ There are numerous commercial antibodies directed against various HER2 epitopes,⁵⁶ and these antibodies have documented differences in staining characteristics when applied to breast cancer tissue for immunohistochemical detection of HER2. Because standardized control material has been unavailable for

Table A7. HER2 As a Predictive Factor for Trastuzumab Benefit in Metastatic Breast Cancer

| IHC/FISH | 0 | | 1+ | | 2+ | | 3+ | |
|-------------------|---------------|----------|---------------|----------|---------------|-----------|-----------|----------|
| | Negative | Positive | Negative | Positive | Negative | Positive | Negative | Positive |
| Prevalence, %* | 97 | 3 | 93 | 7 | 76 | 24 | 11 | 89 |
| Clinical benefit† | Presumed none | Unknown | Presumed none | Unknown | Presumed none | Uncertain | Uncertain | Yes |

Abbreviations: HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization.

*Estimates of prevalence based on data from Dybdal et al.⁷⁸

†Estimates of clinical benefit based on metastatic trials¹⁶⁻¹⁸ and unplanned subset analyses.⁶⁰

Table A8. Probability (%) of Meeting or Exceeding Benchmark Concordance Rates of 80%, 85%, 90%, or 95% in Proficiency Testing As a Function of a Laboratory's True Concordance Rate and the Number of Cases in the Test Case Set

| Benchmark Concordance Rate (%) | True Concordance Rate (%) | No. of Cases in Test Set | | |
|--------------------------------|---------------------------|--------------------------|------|------|
| | | 20 | 40 | 80 |
| 80 | 80 | 41 | 44 | 46 |
| | 85 | 65 | 76 | 86 |
| | 90 | 87 | 96 | 99 |
| | 95 | 98 | > 99 | > 99 |
| | 98 | > 99 | > 99 | > 99 |
| 85 | 80 | 21 | 16 | 10 |
| | 85 | 40 | 43 | 45 |
| | 90 | 68 | 79 | 90 |
| | 95 | 92 | 99 | > 99 |
| | 98 | 99 | > 99 | > 99 |
| 90 | 80 | 7 | 3 | 1 |
| | 85 | 18 | 13 | 7 |
| | 90 | 39 | 42 | 45 |
| | 95 | 74 | 86 | 95 |
| | 98 | 94 | 99 | > 99 |
| 95 | 80 | 1 | < 1 | < 1 |
| | 85 | 4 | 1 | < 1 |
| | 90 | 12 | 8 | 4 |
| | 95 | 36 | 40 | 43 |
| | 98 | 67 | 81 | 92 |

HER2 immunohistochemical assays until recently, there has been no method to normalize these differences to a standard value.⁶¹ Furthermore, many of these antibodies require antigen retrieval steps to allow them to adequately detect the epitopes in paraffin-embedded tissue. This antigen retrieval can be done in a variety of ways and produces marked differences in staining patterns determined by the method used.⁶² These variations result in significant differences in staining reactions in tissue so that the values assigned for positive and negative tests become highly variable, but many laboratories are unaware of this variation. It has been estimated that only a small subset of laboratories doing HER2 testing (30%) participate in external proficiency testing programs and only a small fraction of laboratories doing such testing actually validate the test procedure before offering the test (Genentech phone survey data, 2001; provided as PowerPoint slide to M.E.H. Hammond, 2003). Although US Food and Drug Administration-approved methods are available for both immunohistochemical and FISH testing for HER2, only 48% of laboratories using US Food and Drug Administration approved testing methods indicate that they actually follow the US Food and Drug Administration-approved method in their laboratories (CAP survey results, 2004; personally reviewed by M.E.H. Hammond, 2005). Conflicting results are observed with the differing methods,⁶² and there is no clear recommendation about the best laboratory practice in these situations.⁸¹

Slamon et al²⁵ have shown that even breast cancers lacking *HER2* gene amplification have low expression of HER2 that is detected by IHC. Most published studies using paraffin-embedded breast cancers show a high frequency in the IHC 0 category.⁶² However, even though these low expression cases have only minimal membrane staining, the 0 category essentially does not exist in frozen tissue specimens and could be considered an artifact of tissue fixation and processing. They have also reported that a proportion of breast cancers known to have gene amplification and overexpression (validated by Northern hybridization, Western immunoblot, and frozen section IHC) may lose membrane staining after paraffin embedding.²⁵ This would suggest that most, and perhaps all, breast cancers experience a reduction in IHC immunostaining when they are fixed and paraffin embedded. However, there is no evidence to suggest that a HER2-amplified breast cancer that lost immunostaining on tissue processing is biologically different from the breast cancers that retain IHC staining after tissue processing, as Slamon et al observed that frozen section IHC was equally strong in cases losing and retaining IHC staining after paraffin-embedding. Therefore, loss of antigenicity resulting in a potential false-negative IHC (and unequal gain of antigenicity with antigen retrieval resulting in a potential false-positive IHC) can also be affected by poor standardization of fixative methods. Of note, experienced laboratories report a very low frequency of FISH amplification in the IHC 0 to 1+ subgroup.

Appendix E

Tissue Handling Requirement and Control Materials

Tissue handling requirement. Based on recommendations of the previous consensus conferences dealing with HER2 testing^{47,61} and US Food and Drug Administration approval recommendations, incisional and excisional biopsy samples used for HER2 testing of either type

should be fixed in 10% neutral buffered formalin for intervals ranging from at least 6 hours to no more than 48 hours. Fixation time alters protein antigen expression and also changes the requirements for enzymatic digestion that is part of the ISH protocol to detect gene amplification. Prolonged fixation, for example more than 48 hours, may result in false-negative results. Fixation times for needle biopsies have not been addressed. Time of fixation should be routinely recorded for troubleshooting purposes if that information is available. Breast specimens, after appropriate gross inspection and designation of margins, should be promptly sliced at 5 to 10 mm intervals and fixed in formalin (unsliced samples should not be fixed). The interval between tissue acquisition and fixation of breast specimens should be as short as possible. Samples fixed in formalin should be routinely processed into paraffin and cut onto glass slides within 48 hours. Prolonged storage of glass slides with cut sections of tissue should be avoided. Length of storage that does not compromise antigen preservation is variable, depending on the fixation conditions.⁸² Tissue microarray slides stored at 4°C have been used for up to 270 days for IHC testing.⁸³ However, sections should ideally not be used for HER2 IHC testing if cut more than 6 weeks earlier though this will vary depending on primary fixation and storage conditions. If slides have had prolonged storage before testing and a negative result is obtained, this storage condition should be noted. The effect of new rapid tissue processing protocols on HER2 testing is unknown. Any alteration of standard conditions, such as use of alternative fixatives, microwave fixation, or alternative processing methods, must be validated against standard methods of testing before a test routinely using these conditions is offered in a laboratory. Validation must consist of testing of the same samples with the alternative fixative buffered formalin using the same HER2 testing method to demonstrate concordance of the result.

Control materials. High quality control materials, such as cell line standards, will improve the quality of HER2 testing once they become widely available for use in day-to-day quality control of HER2 assays.^{84,85} This type of material has been developed at the European Collection of Cell Cultures using stringent quality assured conditions to produce quality control slides with standardized expression levels and will be made available to clinical laboratories in late 2006. Ongoing work at the NIST will provide similar cell line standards in the United States⁶¹ which can then be used by commercial companies to create secondary, more widely available control materials.

Appendix F

Statistical Requirements for Assay Validation

The purpose of a laboratory proficiency testing program is to identify those laboratories for which reported assay results differ significantly from established gold standard results, and more broadly, to identify factors responsible for discrepant findings.

A variety of factors can affect the outcome of an assay. These include inherent characteristics of the marker or specimen being assayed, for example the lability of the marker, the cellular heterogeneity of the specimen, or the degree of tumor necrosis. Specimen handling factors such as how rapidly after surgical excision the specimen is processed and the type and duration of fixation may impact on assay performance. Of particular concern for HER2 testing are variations in assay methods across laboratories and quality fluctuations within laboratories. Laboratories may use different reagents, different technical protocols, or different scoring systems. Random fluctuations even within a laboratory may arise due to factors such as reagent lots, batch effects, climatic effects, or technician effects. All of these variations can impact on the end result assay value that is reported by the laboratory.

Critical elements to consider when designing a laboratory proficiency testing program include the make-up of the test case set, specimen format, timing of the testing, specification of performance measures and benchmarks, and statistical operating characteristics. This analysis begins with a discussion of the first several elements, with a more detailed statistical discussion of performance measures and operating characteristics to follow. Ideally, the test case set should reflect as closely as possible the types of specimens that a laboratory will encounter in routine practice with respect to the range of inherent specimen characteristics and specimen handling and fixations conditions. If there are community accepted best practices for tissue adequacy, handling, and fixation, the specimens selected for the test case set should abide by these best practices. There are likely to be practical limitations on the range of inherent specimen characteristics that can be represented in the test case set. For example, the specimens will be biased toward larger specimens, and the specimens must be ones for which there is sufficient biologic homogeneity and staining or hybridization consistency so that an unambiguous gold standard value can be determined. Test specimens may be presented in either TMA format or as conventional full sections. Although conventional full sections are representative of what laboratories see in routine practice, the TMA format offers a more affordable and less resource-intensive way of testing a broader selection of specimens and should capture the majority of aspects relevant to a laboratory's assay performance. Proficiency testing is recommended two to four times per year, with the preferred frequency of testing dependent on the consistency of the within-laboratory performance. A laboratory with rigorously trained, experienced staff, minimal staff turnover, and stringent internal quality assurance (QA)/quality control (QC) procedures is likely to perform more consistently over time than a laboratory not having these attributes. For this reason, review of each laboratory's internal QA/QC procedures and staff training, experience, and stability should be considered a part of the proficiency assessment.

Standard measures of performance for diagnostic tests having binary outcomes include sensitivity, specificity, and overall accuracy. This discussion focuses on the analytic performance of a test (ie, how accurately it determines the true value of the analyte; in this case the binary HER2 status determined by a gold standard assay). Clinical performance of the test, or how well the test predicts who will respond to anti-HER2 therapy, can only be evaluated in an appropriately designed clinical study, and that is not

the focus of this current discussion on statistical requirements for assay validation. A precise definition of accuracy is how close the measured values are to a supposed true value, and it incorporates both variability (ie, precision) and bias (ie, a systematic difference between average measured value and true value). Implicit in this discussion is that a suitable gold standard has been established for purposes of determining true status for each specimen. The difficulties inherent in selecting an appropriate gold standard have been discussed in the Introduction of this guideline. As this section of the Appendix deals with the special case of binary measurements, accuracy is reduced to percent concordance between the assay being evaluated and the gold standard assay, and we will preferentially use the term concordance here. Further considerations are relevant when test results are reported as ordinal ratings or continuous measures, but for the present discussion we focus only on the simplest case in which the marker assay result is reported as a binary value. Sensitivity is defined as the percent of positive test results obtained when evaluating only specimens that are truly positive. Specificity is the percent of negative test results reported when only truly negative specimens are evaluated. Overall accuracy (concordance) combines sensitivity and specificity into a single measure of the percentage of cases (positive and negative) for which the assay result agrees with the true status. Note that overall accuracy (concordance rate) can be strongly influenced by the positive-negative mix of the test case set if the sensitivity and specificity rates are not similar. Generally, it is important to ensure adequate representation of both positive and negative test cases in order to reliably assess both specificity and sensitivity.

Having adopted specific performance measures, one must set a benchmark for what will be considered an acceptable level of performance for a laboratory to be declared proficient. An evaluation of the statistical operating characteristics of the proposed proficiency benchmark requirements is essential. To facilitate this discussion, we suppose that a randomly selected case from the set of cases used for proficiency testing can be viewed as having a certain probability of its HER2 status being reported correctly by the laboratory undergoing proficiency testing. This probability is influenced by inherent characteristics of the specimen, how the laboratory's specific assay method performs under those specimen characteristics, and other random influences. Further, we will assume that correct reporting outcomes can be viewed as essentially independent from specimen to specimen. While this independence assumption may not hold exactly, if, for example, there is a catastrophic failure for an assay batch containing multiple test cases, it will usually be a reasonable assumption and is a necessary one to make the problem statistically tractable. Under the above assumptions, the proportion of correct assay evaluations can be viewed as a binomial proportion of successes, and evaluation of the operating characteristics is possible based on statistical properties of binomial proportions. In particular, for a given number of test cases, an assumed true concordance rate, and a specified benchmark for achieving proficiency, we can calculate from a binomial distribution the probability that a laboratory will meet the proficiency benchmark on a test case set containing a specified number of cases. Examples of such probability calculations are presented in Table A8. This Table corresponds to benchmark concordance rates of 80%, 85%, 90%, and 95%, respectively. The rows of the table provide a selection of possible true (long-term average) laboratory concordance rates (80%, 85%, 90%, 95%, or 98%), and the columns correspond to different numbers of cases in the test set (20, 40, or 80). The first observation we make is that if the benchmark is set equal to the laboratory's true concordance rate, the probability that the laboratory will pass the test is only approximately 50%. This is because in a finite set of test cases, there will always be some variability of the laboratory's observed concordance rate around its true, long-term average, with approximately half of the time the observed concordance rate falling below the true value and half of the time falling above. (The probability is not exactly 50% due to the fact that the observed number of correctly reported assay results must be an integer value.) The greater the amount by which the laboratory's true concordance rate exceeds the benchmark, the higher the probability that the laboratory will meet or exceed the benchmark concordance rate on the test case set. For example, if the benchmark is set at 90% and a laboratory's true concordance rate is 95%, then the probability that the laboratory will meet or exceed the 90% concordance rate on a set of 40 test cases is 86%; whereas, if the laboratory's true concordance rate is 98%, it has 99% chance of meeting or exceeding the 90% concordance rate on a set of 40 test cases. Under the same proficiency testing plan (40 cases, 90% benchmark), a laboratory operating at 85% concordance rate would have only 13% chance of meeting or exceeding the 90% concordance benchmark on the set of test cases. In order to have even tighter discrimination, one must increase the number of cases in the test set. For example, if the test set is comprised of 80 test cases and the benchmark is maintained at 90%, then the probability that a laboratory operating with an average concordance rate of 85% will meet or exceed the 90% benchmark on the test case set is only 7%; whereas, the probability that a laboratory operating with an average concordance rate of 95% will meet or exceed the 90% benchmark on the test case set is 95%.

The term concordance rate has been used in a generic sense in the preceding discussion in this Appendix. It could refer to overall concordance (performance on combined positive and negative cases), sensitivity (performance on positive cases only), or specificity (performance on negative cases only). The sample sizes corresponding to the columns in Table A8 must be interpreted in the context of what concordance rate is being monitored. If the concordance rate being monitored is overall concordance, then the sample size refers to the total number of cases in the test set (positive and negative combined). If sensitivity is being monitored, then the sample size refers to the number of true-positive cases. If specificity is being monitored, then the sample size refers to the number of true-negative cases.

In summary, development of an effective and fair laboratory proficiency testing program requires careful consideration of the composition of the test case sets, the number of cases in the test sets, and the required performance benchmarks. Proficiency testing results should also be viewed in combination with a laboratory's standard QA/QC operating procedures. With due consideration given to these factors, it is possible to develop a laboratory proficiency testing program capable of distinguishing between laboratories with high concordance rates and laboratories with unacceptably low concordance rates with the overall goal of substantially improving the accuracy of HER2 testing conducted in research and community settings.

Appendix G

Interpretation Criteria and Test Reporting of HER2 Test

Interpretation criteria. In order to address this topic, panel members had to first assume that variation in testing would be minimized through mandatory requirements of laboratories to technically validate tests before offering them, use only standardized methods, and engage in ongoing internal and external quality assurance and laboratory accreditation processes. This would ensure that HER2 testing methods would be consistent and accurate over time and across various laboratories. The panel agreed that the interpretation criteria for each assay type (IHC and FISH) must be specifically defined and followed to assure improvement in assay results. Currently, other HER2 assays are not approved for clinical use in the United States.

For IHC assays of HER2 protein expression, the original US Food and Drug Administration-approved interpretation guidelines provide insufficient specificity. Several experts, including those serving as central reviewers on clinical trials, have specified that a threshold of more than 30% of tumor (rather than the originally specified 10%) should show strong circumferential membrane staining for a positive result. This means that according to this guideline, strong circumferential staining of 30% or less of cells would be considered equivocal and be subjected to confirmatory FISH testing. A cutoff of more than 30% reflects the cumulative experience of panel members that usually a high percentage of the cells will be positive if it is a true IHC 3+, published reports using cutoff values higher than 10%,⁵² and the goal of the panel to decrease the incidence of false positive 3+. Those with lower percentage will then fall into equivocal range and be subjected to FISH confirmation. Other criteria strongly correlated with positive assays of protein expression include uniformity of circumferential dark membrane staining creating the impression of a pattern commonly referred to as chicken wire.⁴⁷ If both uniformity and a homogeneous, dark circumferential pattern are seen, the resultant cases are likely to be amplified by FISH as well as positive for HER2 protein expression. Strong staining associated only with tissue borders or staining found in regions of crush artifact should be ignored. Interpretation of assays exhibiting strong staining of normal breast ducts should be avoided. Results must be interpreted in the context of the positive, equivocal, and negative controls run with each assay. Image analysis improves consistency of interpretation (Table 7).^{86,87}

If image analysis is used, the method must be validated and thresholds of positive, equivocal, and negative must be defined using samples with known FISH amplification levels performed on the same samples. The negative category must contain no validating samples which are FISH amplified, and the positive category must have amplification in 100% of samples. Because the analytic tolerance of the equipment is very tight, higher concordance levels with FISH can be required. Annual rechecking of samples (11 to 26) against FISH is required to assure that thresholds are valid.

For FISH assays of *HER2* gene amplification, the most important criterion is the documentation that the counting of signals occurs in at least two areas known to be areas of invasive tumor, either on sequential sections stained with hematoxylin and eosin or for HER2 protein expression by IHC. Areas of in situ carcinoma should not be counted. Counts of 20 cells should be done and involve at least two observers. A pathologist must verify that the counting occurs in areas of invasive carcinoma and must survey entire sample for genomic heterogeneity (Table 8). Results must be interpreted in the context of the positive, equivocal, and negative controls run with each batch of tests. If signals are weak in more than 25% of cells, if nuclei are indistinct, if greater than 10% of signals occur over cytoplasmic areas of cells, or if background staining obscures signals, the assay should be repeated. The number of *HER2* signals/nucleus as well as the number of chromosome 17 (CEP 17) signals per nucleus should be recorded so that monoallelic deletions as well as monosomy or polysomy of chromosome 17 can be described (Table 8). If genomic heterogeneity of *HER2* gene amplification is found, it must be specifically reported.⁶⁶ No consensus recommendations exist at this time for handling of genomic heterogeneity.

Test reporting. Standardized report formats will help to insure that reports contain all necessary elements and provide them in a clear and concise format. Such a format can most easily be thought of as a list of elements. The required elements include information about the patient, physician, facility, specimen, date of service, assay method specifics and results, and interpretation. The specific list of elements recommended by the panel is provided in Tables 9 and 10. Elements are identical for both assays except for specifics of results, scoring system, and final interpretations.

Appendix H

Regulatory Requirements for Laboratories

Any laboratory routinely offering HER2 assays should follow good laboratory practices as defined in the CLIA 88 regulations and promulgated by laboratory accrediting organizations. Equipment used in the assays must be calibrated and placed into service with regular maintenance and internal quality control measures. Test procedures must be standardized and validated before being

put into routine use. The panel agreed that it was preferable to use the US Food and Drug Administration-approved methods for either IHC HER2 tests of protein expression or FISH tests for gene amplification detection. If these US Food and Drug Administration-approved methods are used, the procedures thus defined must be strictly followed. Because of cost constraints, some laboratories in other countries have found it impossible to require US Food and Drug Administration-approved methods, but have carefully validated alternative methods before use. If such alternative methods are employed, it is the responsibility of the laboratory director to ensure that the method offered is validated and provides accurate consistent test results comparable to those provided by US Food and Drug Administration-approved methods. The use of methods that are not US Food and Drug Administration approved should be disclosed in assay reports. Similarly, internally validated method modifications must be disclosed in assay reports.

CLIA 88

CLIA was passed by Congress in 1988 for the purpose of ensuring accurate, reliable testing regardless of location. CLIA 88⁸⁸ regulates all testing on humans for health purposes using minimum quality standards. The standards set regulations based on test complexity. All predictive cancer factor testing is deemed highly complex, which requires adherence to the most stringent quality standards and biannual surveys by deemed agencies to assess the level of adherence to these standards.

Quality standards are subject to review by surveying agencies and include: personnel qualifications and responsibilities; quality control (ie, a mechanism to ensure that the test is working reliably daily); specimen integrity and record keeping; proficiency testing (external testing for concordance); and quality assessment, which is ongoing. Quality assessment includes a system with a comprehensive plan to monitor and ensure quality results and to communicate and solve any problems. It is assumed that any test offered has been validated by the laboratory before offering the test clinically.

On inspection, laboratory tests are observed, personnel are interviewed, records (including reports) are reviewed, and outcomes are evaluated, including the outcome of proficiency testing. If the laboratory is found to be deficient in performance, the accrediting agency will review a plan of corrective action and, if warranted, stop the laboratory from performing further testing.

Laboratory Accreditation as Interpreted by CAP

CAP and the Joint Commission on Accreditation of Healthcare Organizations are two organizations which enjoy deemed status by the US Department of Health and Human Services to inspect laboratories and accredit them under the aegis of CLIA 88. HER2 testing is highly complex and, as such, can be subjected to the highest level of scrutiny by laboratory inspection. The analyte can be required by accrediting agencies to be evaluated regularly by mandated proficiency testing. Furthermore, the laboratory accreditation inspection criteria can be modified to be more stringent, within the bounds of the CLIA 88 regulation.

US Food and Drug Administration Regulatory Requirements

US Food and Drug Administration regulates the medical devices used in the testing for HER2 through regulations put into place as a result of the Medical Device Amendments of 1976. A variety of general controls include requirements for registration and listing of HER2 diagnostic materials, production following good manufacturing practices, and postmarket reporting of adverse events. The 1976 law also put into effect requirements for premarket review. Because HER2 is involved in making drug treatment choices with information having a potential direct impact on patient morbidity and mortality, this product is considered a high risk, or class 3, device. It is subject to review under the provisions of the premarket approval application section of US Food and Drug Administration law. US Food and Drug Administration premarket review for HER2 testing requires demonstration of safety and effectiveness of testing and is based on information characterizing both analytic and clinical performance. US Food and Drug Administration premarket review of HER2 tests also entails an analysis of labeling and a premarket assessment of the quality system planned for test production.

US Food and Drug Administration does not regulate in-house or “home brew” tests for HER2, tests developed and used at unique or individual laboratory sites. Validation of these tests occurs under the laboratory’s CLIA requirements. However, if these tests are made using commercially purchased active ingredients (also known as ASRs), these important components of the test are subject to the general controls noted above. In addition, HER2 test results when reported require a clear statement that the test has not been US Food and Drug Administration approved and that the laboratory reporting the results takes responsibility for test performance.

Appendix I

International External Quality Assessment Initiatives

Quality assessment in the United Kingdom. Guidelines in the United Kingdom have been published outlining the main criteria for laboratories testing for HER2 in clinical laboratories. A two-phase testing algorithm based on IHC assay as the primary screen with reflex to FISH reserved for equivocal cases is currently recommended.⁴⁸ This is based on evidence showing very good concordance between IHC and FISH results on breast carcinomas from 37 laboratories when tested in experienced reference centers.⁶⁴ Emphasis continues to be placed on the standardization of methodology, assessment, and strategies to achieve high quality performance.⁴⁸ Recommendations in particular include participation in external quality assessment, as there is evidence to show increased reproducibility of results by laboratories over time when participating in external quality assessment for HER2.^{68,82} Full results of a national consultation that included

about 200 United Kingdom oncologists and pathologists on HER2 testing will soon be released.⁸⁹ This consultation revealed a consensus on several areas, such as: the need for generalized screening of all invasive breast cancers for HER2; maintenance of expertise by restricting immunohistochemical HER2 testing to laboratories with a minimum annual caseload of 250 cases; FISH centers should test a minimum of 100 per year and preferably at least 150; and recommendations that other than in exceptional circumstances all HER2 testing services should be accredited by Clinical Pathology Accreditation (CPA UK Ltd, Sheffield, England).

Quality assessment initiatives in continental Europe and Australia. A nonsystematic review of efforts in Continental Europe and Australia conducted by panel members revealed that, in most countries, the testing algorithm is based on IHC and in situ hybridization.⁵⁰ The generally accepted guideline is that an IHC score of 3+ is regarded as HER2 positive while IHC 0 or 1+ is regarded as HER2 negative. IHC 2+ is regarded as HER2 equivocal and in such cases retesting with FISH or CISH is recommended. In some countries, like Belgium and Finland, all IHC 2+ and 3+ cases are retested by FISH or CISH, and only a combination of 2+ or 3+ IHC and gene amplification by ISH is regarded as HER2 positive. The guidelines for HER2 testing are captured in a consensus document in most countries. These documents provide guidelines for handling of the tissue before testing, the IHC methods to be used, and the ISH protocols to be used. In general, there is also advice on the minimum number of cases that should be tested in each laboratory, though most countries do not specify how many cases an individual pathologist should review. Most countries have a quality assurance program consisting of proficiency testing exercises (ring trials) for IHC or ISH, where unstained slides are circulated to all participating laboratories. The participating laboratory performs IHC staining or ISH and reports back the result of the staining. Twelve French laboratories reported a multicentric calibration test of in-house IHC assays with two separate HER2 antibodies, and reported a 95% IHC accuracy rate when using FISH as gold standard for 116 of 119 of the samples and by considering IHC results as a continuous variable and taking 60% invasive stained cells as the cut off for HER2 overexpression.⁵²

Quality assurance efforts in Canada. A consensus for a testing algorithm was reached by Canadian pathologists. The algorithm recommended IHC testing at first, using one or two antibodies, with equivocal cases tested with FISH. The publication of the testing algorithm in 2002⁴⁹ included a list of guidelines for proper fixation, interpretation criteria, and a list of surrogate indicators for reflex ISH testing. In order to increase the accuracy rate of IHC testing, the group recommended reflex FISH testing for the IHC 2+ category and for any cases where clinical circumstances suggested that HER2 positivity was likely.

In 2001, the Ministry of Health, laboratories branch in Ontario and the Quality Management Program for Laboratory Services supported a QA program and funded HER2 testing in 14 labs including two reference labs in Toronto, Ontario (Mount Sinai Hospital and Sunnybrook Hospital). Each testing site was requested to send 5% negative and 5% positive cases or five positive and five negative cases every 6 months to their designated reference laboratory. The reference labs retested with IHC using their routine protocols. Retesting with ISH (FISH or CISH) was done for discrepant cases.³⁷ The data collected to date from the two reference labs showed that the false positive rate varied from 5% to 7% and the false negative was less than 1% in 300 cases assessed in the QA program. A study by O'Malley et al⁶⁹ for HER2 testing in Ontario and British Columbia for metastatic breast cancer showed that the concordance with centralized testing for negative cases was 97.6% and for positive cases was 87.4%. This QA program is supplemented by frequent teaching sessions, workshops, and educational material provided by the reference laboratory pathologist.

Quality assurance efforts by CAP. CAP has developed and implemented several laboratory proficiency surveys for use by laboratories doing HER2 clinical testing.^{46,47} These surveys include the general immunohistochemistry survey (MK Survey), the tissue microarray-based HER2 survey (HER2 Survey), the interphase fluorescence in situ hybridization survey (CYH Survey), the bright field in situ hybridization survey (ISH Survey), and the IHC and image analysis survey (IMG Survey). Table A9 summarizes the format, number of participants, and number of cases in each survey program. Table A10 summarizes concordance data from the programs having the largest number of participants for which consecutive sections from the same cases were shared. Variability in quantification of HER2 signals by CYH survey participants for *HER2* amplified cases is summarized in Table A11.

Table A9. CAP Laboratory Proficiency Surveys for Use by Laboratories Doing HER2 Clinical Testing

| Survey Program | Method | Participants Enrolled in 2005 | Current Format | Format Changes Planned |
|----------------|------------------------|-------------------------------|--|--|
| MK | IHC | 826 | Educational program of eight cases of whole sections annually (four cases twice annually); one breast cancer case annually | None |
| HER2 | IHC | 155 | 80 cases (40 cases in 10-core TMAs twice annually) | None |
| CYH | FISH | 191 | Two cases of whole sections once annually | 20 cases (10 cases in 5 core TMAs twice annually) |
| ISH | BRISH | 95 | One case of whole sections once annually | 80 cases (40 cases in 10-core TMAs twice annually) |
| IMG | IHC and image analysis | 52 | One case of whole sections once annually | 80 cases (40 cases in 10-core TMAs twice annually) |

Abbreviations: HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization; BRISH, brightfield in situ hybridization; TMAs, tissue microarrays.

The results of these surveys have shown that there is variation in laboratory performance, particularly for immunohistochemical assays with equivocal positive results or FISH testing with borderline results. The variation in immunohistochemical testing is greater than the variation in FISH testing. These results need to be interpreted with caution; however, as far fewer labs perform FISH than perform IHC, and FISH testing tends to be done in higher volume laboratories. The voluntary enrollment in this educational program leads to only a minority of laboratories doing this testing. Although the exact number of laboratories involved in HER2 testing is unknown, it is thought that about 2,000 laboratories perform HER2 IHC and about 500 laboratories perform HER2 FISH testing in the United States.

Table A10. Cases Shared Between MK (IHC) and CYH (FISH) CAP Surveys

| MK | Cases | | | Concordance (%) | |
|---------|-------|--------|---------------|-----------------|-----|
| | IHC | CYH | FISH | MK | CYH |
| 2000-04 | 3+ | 2000-2 | AMP | 96 | 100 |
| 2002-02 | 3+ | 2002-2 | AMP | 98 | 100 |
| 2004-08 | 0* | 2004-1 | AMP | 87* | 96 |
| 2003-05 | 1+ | 2003-2 | Low level AMP | 73 | 56 |
| 2003-05 | 1+ | 2002-1 | Low level AMP | 73 | 88 |
| 2001-01 | 0 | 2000-1 | NONAMP | 72 | 100 |
| 2002-05 | 2+ | 2003-1 | NONAMP | 86 | 97 |

Abbreviations: CAP, College of American Pathologists; IHC, immunohistochemistry; FISH, fluorescent in situ hybridization; AMP, *HER2* gene amplification present; NONAMP, absence of *HER2* gene amplification; Low level AMP, low level *HER2* gene amplification.

*Invasive ductal carcinoma demonstrating gene/protein discordance; 87% of MK participants interpreted as 0 or 1+, and staining by College of American Pathologists Cell Markers Committee Experts confirmed negative IHC result.

Table A11. CYH Survey (FISH)

| CYH | Genotype* | Enrollment | Concordance (%) | Ratio (mean ± SD) |
|--------|-----------|------------|-----------------|-------------------|
| 2000-1 | NONAMP | 35 | 100 | 0.42 ± 0.19 |
| 2000-2 | AMP | 35 | 100 | 4.53 ± 3.38 |
| 2001-1 | NONAMP | 48 | 100 | 1.05 ± 0.14 |
| 2001-2 | AMP | 59 | 100 | 9.51 ± 4.20 |
| 2002-1 | LLAMP | 89 | 56 | 2.0 ± 0.66 |
| 2002-2 | AMP | 90 | 100 | 10.42 ± 6.07 |
| 2003-1 | NONAMP | 127 | 97 | 1.11 ± 0.17 |
| 2003-2 | LLAMP | 124 | 88 | 2.56 ± 0.72 |
| 2004-1 | AMP | 139 | 96 | 3.63 ± 1.34 |
| 2004-2 | NONAMP | 136 | 97 | 1.15 ± 0.20 |
| 2005-1 | NONAMP | 151 | 99 | 1.12 ± 0.36 |
| 2005-2 | AMP | 152 | 99 | 8.53 ± 3.54 |

Abbreviations: FISH, fluorescent in situ hybridization; SD, standard deviation; AMP, *HER2* gene amplification present; NONAMP, absence of *HER2* gene amplification; LLAMP, low level *HER2* gene amplification.

*Whole sections.