NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the submitting standards web page.

NQF #: 1878 **NQF Project:** Cancer Project

(for Endorsement Maintenance Review)
Original Endorsement Date: **Most Recent Endorsement Date:** **Last Updated Date:** Mar 07, 2012

BRIEF MEASURE INFORMATION

De.1 **Measure Title:** Human epidermal growth factor receptor 2 (HER2) testing in breast cancer

Co.1.1 **Measure Steward:** American Society of Clinical Oncology

De.2 **Brief Description of Measure:** Percentage of adult patients (aged 18 or over) with invasive breast cancer who receive human epidermal growth factor receptor 2 (HER2) testing

2a1.1 **Numerator Statement:** HER2/neu testing performed

2a1.4 **Denominator Statement:** Adult women with invasive breast cancer

2a1.8 **Denominator Exclusions:** Patient history of metastatic cancer
Multiple primaries prior to or within the measurement period

1.1 **Measure Type:** Process
2a1. 25-26 **Data Source:** Electronic Clinical Data : Electronic Health Record, Paper Records
2a1.33 **Level of Analysis:** Clinician : Group/Practice, Clinician : Team

1.2-1.4 **Is this measure paired with another measure?** No

De.3 **If included in a composite, please identify the composite measure** (*title and NQF number if endorsed*)**:**
n/a

STAFF NOTES (*issues or questions regarding any criteria*)

Comments on Conditions for Consideration:

Is the measure untested? Yes☐ No☐ If untested, explain how it meets criteria for consideration for time-limited endorsement:

1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (*check De.5*)**:**
5. Similar/related endorsed or submitted measures (*check 5.1*)**:**
Other Criteria:

Staff Reviewer Name(s):

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1a. High Impact: **H☐ M☐ L☐ I☐**

(*The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact aspect of healthcare.*)

De.4 Subject/Topic Areas *(Check all the areas that apply)*: Cancer, Cancer : Breast
De.5 Cross Cutting Areas (*Check all the areas that apply)*:

1a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact *(Provide epidemiologic or resource use data)*:
• In 2011, an estimated 230,480 new cases of invasive breast cancer will be diagnosed among women, as well as an estimated 57,650 additional cases of in situ breast cancer.
• In 2011, approximately 39,520 women are expected to die from breast cancer. Only lung cancer accounts for more cancer deaths in women.

Based on growth and aging of the U.S. population, medical expenditures for cancer in the year 2020 are projected to reach at least $158 billion (in 2010 dollars) — an increase of 27 percent over 2010, according to a National Institutes of Health analysis.

The projections were based on the most recent data available on cancer incidence, survival, and costs of care. In 2010, medical costs associated with cancer were projected to reach $124.6 billion, with the highest costs associated with breast cancer ($16.5 billion), followed by colorectal cancer ($14 billion).

1a.4 Citations for Evidence of High Impact cited in 1a.3: American Cancer Society. Breast Cancer Facts & Figures 2011-2012. Atlanta: American Cancer Society, Inc.

Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, and Brown ML. Projections of the Cost of Cancer Care in the United States: 2010-2020. Jan 19, 2011, JNCI, Vol. 103, No. 2

1b. Opportunity for Improvement: H☐ M☐ L☐ I ☐
(*There is a demonstrated performance gap - variability or overall less than optimal performance*)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
We envision that use of this measure will improve concordance with recommendations for HER2 testing. HER2 overexpression is associated with clinical outcomes in patients with breast cancer. 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, including trastuzumab.

1b.2 Summary of Data Demonstrating Performance Gap *(Variation or overall less than optimal performance across providers)*:
[*For Maintenance* – *Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.***]**

1b.3 Citations for Data on Performance Gap: [*For Maintenance* – *Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included***]**

1b.4 Summary of Data on Disparities by Population Group: [*For Maintenance* –*Descriptive statistics for performance results for this measure by population group***]**

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For Maintenance* – *Description of the data or sample for measure results*

reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included]

1c. Evidence (*Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.*)
Is the measure focus a health outcome? Yes☐ No☐ **<u>If not a health outcome</u>, rate the body of evidence.**

Quantity: H☐ M☐ L☐ I ☐ Quality: H☐ M☐ L☐ I ☐ Consistency: H☐ M☐ L☐ I ☐

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	M-H	M-H	**Yes**☐
L	M-H	M	**Yes**☐ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise **No**☐
M-H	L	M-H	**Yes**☐ IF potential benefits to patients clearly outweigh potential harms: otherwise **No**☐
L-M-H	L-M-H	L	**No** ☐

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	**Does the measure pass subcriterion1c?** **Yes**☐ IF rationale supports relationship

1c.1 **Structure-Process-Outcome Relationship** *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
This is a process measure that is closely linked to outcomes for invasive breast cancer patients. HER2 overexpression is associated with clinical outcomes in patients with breast cancer. 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, including trastuzumab.

1c.2-3 **Type of Evidence** *(Check all that apply)*:
Clinical Practice Guideline

1c.4 **Directness of Evidence to the Specified Measure** *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
The primary outcome of interest was the correlation between HER2 status and benefit from anti-HER2 therapy. Other outcomes of interest included the positive predictive value (PPV) and negative predictive value (NPV) of fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) to determine HER2 status, alone and in combination; concordance across platforms; and accuracy in determining HER2 status, sensitivity, and specificity of specific tests.

1c.5 **Quantity of Studies in the Body of Evidence** *(Total number of studies, not articles)*: 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.

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.

1c.6 **Quality of Body of Evidence** (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events*): 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 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. Evidence tables were developed based on selected studies that met the criteria for inclusion.

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 countries, primarily the United States, and international organizations. Letters, commentaries, and editorials were reviewed for any new information. Case reports were excluded.

1c.7 **Consistency of Results across Studies** *(Summarize the consistency of the magnitude and direction of the effect):*

1c.8 **Net Benefit** *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**
Based on preliminary reports of three large RCTs, the addition of one year of trastuzumab, following a variety of adjuvant or neoadjuvant chemotherapy regimens, significantly improved the primary endpoint of DFS in patients with HER2/neu positive early breast cancer. Secondary endpoints of RFS, DDFS, and TTR in all studies, and OS in one combined study, were also significantly improved with the addition of trastuzumab. Those results are only applicable to women with HER2/neu overexpressing breast cancer who complete a minimum of four cycles of adjuvant or neoadjuvant chemotherapy. Although the majority of the patients in those studies had node-positive breast cancer, women with high-risk node-negative breast cancer were also included in HERA (32% were N0 but had tumours T1c) and NCCTG 9831 (11% were N0 but had tumours >1cm if ER negative, >2cm if ER positive). Therefore, those results are also generalizable to women with node-negative breast cancer meeting these criteria. The magnitude of incremental benefit conveyed by adjuvant trastuzumab well exceeds the gains accrued by over three decades of adjuvant chemotherapy use.

Based on experience in the metastatic setting, the concurrent use of trastuzumab and anthracyclines has prohibitive cardiac toxicity. Based on the current reports, the cardiac toxicity with adjuvant trastuzumab appears to be acceptable, although the reported rate of cardiac events was higher in the concurrent versus sequential trastuzumab arm (in NSABP B31 4.1% vs. 0.7%, HR of 7.2; in NCCTG 9831 3.3% vs. 2.2%). The toxicity is considered acceptable, given the increase in survival.

1c.9 **Grading of Strength/Quality of the Body of Evidence.** Has the body of evidence been graded? Yes

1c.10 **If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:** The guideline panel provided a narrative review of the body of evidence.

For the 2012 Update, in process, the new "ASCO Summary Ratings including the Assessment of Study Quality, Strength of Evidence, and Strength of Recommendations System" is being used.

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.

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.

1c.11 **System Used for Grading the Body of Evidence:** Other

1c.12 **If other, identify and describe the grading scale with definitions:** Detailed narrative description of the strength of each study.

1c.13 **Grade Assigned to the Body of Evidence:** n/a

1c.14 **Summary of Controversy/Contradictory Evidence:** A thorough discussion of the limitations of the literature and/or controversies is included in the "Summary and Recommendations" section in the body of the guideline, presented after each Clinical Question.

1c.15 **Citations for Evidence other than Guidelines*(Guidelines addressed below)*:**

1c.16 **Quote verbatim, the specific guideline recommendation** *(Including guideline # and/or page #)*:
Verbatim recommendations can be viewed at http://jco.ascopubs.org/content/25/1/118.full.pdf in Table 4 on page 123.

Summary of 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 bright?eld ISH, is proposed. Elements to reliably reduce assay variation (for example, specimen handling, assay exclusion, and reporting criteria) are speci?ed. An algorithm de?ning positive, equivocal, and negative values for both HER2 protein expression and gene ampli?cation is recommended: a positive HER2 result is IHC staining of 3 (uniform, intense membrane staining of 30% of invasive tumor cells), a ?uorescent 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 ?nal determination.

1c.17 **Clinical Practice Guideline Citation:** American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.
Wolff AC, Hammond ME, Schwartz JN, Hagerty KL, Allred DC, Cote RJ, Dowsett M, Fitzgibbons PL, Hanna WM, Langer A, McShane LM, Paik S, Pegram MD, Perez EA, Press MF, Rhodes A, Sturgeon C, Taube SE, Tubbs R, Vance GH, van de Vijver M, Wheeler TM, Hayes DF; American Society of Clinical Oncology/College of American Pathologists.
Arch Pathol Lab Med. 2007;131(1):18-43.
PMID:19548375

1c.18 **National Guideline Clearinghouse or other URL:** http://www.guideline.gov/content.aspx?id=10384&search=her2+testing AND http://jco.ascopubs.org/content/25/1/118.full.pdf+html

1c.19 **Grading of Strength of Guideline Recommendation.** Has the recommendation been graded? No

1c.20 **If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:**

1c.21 **System Used for Grading the Strength of Guideline Recommendation:** Other

1c.22 **If other, identify and describe the grading scale with definitions:** n/a

1c.23 **Grade Assigned to the Recommendation:** n/a

1c.24 Rationale for Using this Guideline Over Others: The systematic review and the process of rating the body and evidence and the strength of recommendations makes it a very transparent and credible document. The collaboration between ASCO and the College of American Pathologists (CAP) assured a consistent message to medical oncologists and pathologists. Other guidelines recommend HER2 testing for women with invasive breast cancer (e.g., National Comprehensive Cancer Network).

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High **1c.26** Quality: High**1c.27** Consistency**:** High

Was the threshold criterion, *Importance to Measure and Report*, met?
(*1a & 1b must be rated moderate or high and 1c yes*) Yes☐ No☐
Provide rationale based on specific subcriteria:

For a new measure if the Committee votes NO, then STOP.
For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. **(evaluation criteria)**
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

S.1 Measure Web Page *(In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained*). Do you have a web page where current detailed specifications for this measure can be obtained? No

S.2 If yes, provide web page URL:

2a. RELIABILITY. Precise Specifications and Reliability Testing: H☐ M☐ L☐ I☐

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome)***:**
HER2/neu testing performed

2a1.2 Numerator Time Window *(The time period in which the target process, condition, event, or outcome is eligible for inclusion)***:**
Within 4 weeks (28 days) of diagnosis

Date of diagnosis: [Refer to the pathology/hemato-pathology or cytology report and record the date of the report (not the date of the specimen). If there are multiple reports, enter the first date. In the absence of a pathology or cytology report, record any documentation regarding date of initial diagnosis (e.g., a practitioner's notation).]

2a1.3 Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses***:**
HER-2/neu status = HER2 positive
OR
HER-2/neu status = HER2 negative
OR
HER-2/neu status = Test ordered, results not yet documented
OR
HER-2/neu status = Test ordered, insufficient sample for results
OR

(HER-2 equivocal AND New test ordered within 10 days of report = Yes or N/A (patient died or transferred out of practice))

Numerator definitions:
Select 'Test ordered, results not yet documented' only if there is documentation in the chart that a test that reports HER-2/neu analyses was ordered.

In the absence of any documentation regarding HER-2/neu status, select 'Test not ordered/no documentation.'
Enter information from the most recent test report.
Patients are classified as having HER-2 positive disease based on positive results with either test.
If the most recent report indicates insufficient sample, select 'Test ordered, insufficient sample for results.'
If a physician note and the HER-2/neu report differ in results, report the status in the physician note if the note explains the discrepancy. Otherwise, report the status from the HER-2/neu report.

Use the following definitions to determine HER-2/neu status:
Positive:
- IHC 3+ cell surface protein expression (defined as uniform intense membrane staining of >30% of invasive tumor cells) or
- FISH ratio >2.2 or
- HER2 gene copy >6.0

Equivocal:
- Not positive according to any of the criteria above, AND
- (IHC with scores 2+ AND FISH ratio 1.8-2.2) or
- HER2 gene copy 4.0-6.0

Negative:
- Not positive according to any of the criteria above, AND
- IHC 0 or 1+ or
- FISH ratio 1.8 or
- HER2 gene copy <4.0

- If the results indicate 'non-amplified', choose HER-2/neu negative.
- If the results indicate 'weakly positive', choose HER-2/neu positive.

New test ordered within 10 days of report of equivocal result: Respond 'Yes' if a new test was ordered within 10 days of oncologist review of the report with inconclusive results. Choose 'N/A' if the patient died or transferred out of the practice within 10 days of review of the report with inconclusive results or fewer than 10 days have passed.
If the chart documents that the pathologist has ordered a new test, respond 'Yes.'

2a1.4 Denominator Statement *(Brief, narrative description of the target population being measured)*:
Adult women with invasive breast cancer

2a1.5 Target Population Category *(Check all the populations for which the measure is specified and tested if any)*: Adult/Elderly Care

2a1.6 Denominator Time Window *(The time period in which cases are eligible for inclusion)*:
None specified; should be specific to the periodicity of analysis.

2a1.7 Denominator Details (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
Female
And
2 or more encounters at the reporting site
And

Age at diagnosis greater than or equal to 18 years
And
Breast cancer diagnosis (174.xx)

Definitions
Encounter = new patient visit (CPT 99201 -99205) or established patient (CPT 99211-99215), not consult (CPT 99241-99245) office consult or inpatient consult (CPT 99251-99255)

2a1.8 Denominator Exclusions *(Brief narrative description of exclusions from the target population)*:
Patient history of metastatic cancer
Multiple primaries prior to or within the measurement period

2a1.9 Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
'Multiple primaries' is defined as two or more distinct cancer diagnoses. This includes patients with simultaneous bilateral breast cancer or two distinct cancers in one breast.

2a1.10 Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses)*:
n/a

2a1.11 Risk Adjustment Type *(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13)*: No risk adjustment or risk stratification **2a1.12 If "Other," please describe:**

2a1.13 Statistical Risk Model and Variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.)*:
n/a

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:

2a1.17-18. Type of Score: Rate/proportion

2a1.19 Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*: Better quality = Higher score

2a1.20 Calculation Algorithm/Measure Logic*(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*:
This measure is a proportion with exclusions; thus, the calculation algorithm is: Patients meeting the numerator/(Patients in the denominator – Patients with valid exclusions) x 100.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the

sample, conducting the survey and guidance on minimum sample size (response rate):
QOPI sampling methodology: sequentially select patients diagnosed in the past 2 years with an invasive malignancy, and who have had at least 2 encounters at the practice site in the preceding 6 months, until all patients are included or target abstraction sample sizes are met. Target sample size is based on the number of full-time physician full time equivalents . Details can be found at http://qopi.asco.org.

Alternate sampling methodology: Include all eligible patients (i.e., 100% sample).

2a1.25 Data Source *(Check all the sources for which the measure is specified and tested*). If other, please describe:
Electronic Clinical Data : Electronic Health Record, Paper Records

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*): QOPI data are entered via a case report form accessed via a secure web portal. The case report form includes logic and data validation.

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment: URL
http:/qopi.asco.org/

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested)*: Clinician : Group/Practice, Clinician : Team

2a1.34-35 Care Setting *(Check all the settings for which the measure is specified and tested)*: Ambulatory Care : Clinician Office

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
2008 IRR study: A sample of 300 records was planned for re-abstraction in four geographic regions: Midwest, Northeast, South, and West. 50 QOPI practices were randomly selected from the 4 geographic areas and invited to participate. Within each practice, six previously abstracted charts were selected randomly for re-abstraction from the population of 13,561 records submitted in spring 2007 round. Forty-four practices agreed to participate, and submitted 264 records (6 per practice).

2010-2011 audit: QOPI practices applying for the QOPI Certification Program are required to submit copies of documentation from 3-5 records which were previously abstracted. Trained ASCO auditors randomly select records within each relevant module for audit. Agreement at the data element level is documented. 426 audited records from 130 practices were complete in November 2011 and included in the concordance analysis.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
2008 IRR study: ASCO engaged the Virginia Quality Health Center to conduct an inter-rater reliability study of the QOPI case report form and measures. Trained, independent nurse abstractors served as the 'gold standard' against which practice abstractions were compared for accuracy. Sampling is described above. The 264 sampled records allowed for reliability analysis at a 95% confidence level with a +/- 3.88% marking of error.

Kappa statistics were used to analyze the reliability of the audit data set compared to the submitted data. Kappa statistics are the commonly accepted standard for determining inter-rater reliability in the healthcare setting (Allison, Calhoun, et al, 2000; Cassidy, Marsh, et al, 2002). The Kappa statistic is conceptually similar to the rate of agreement between two reviewers, but it imposes a more stringent standard that simple agreement and mismatch rates. The following standards were used (Cohen, 1960; Sim and

Wright, 2005; Feinstein and Cicchetti, 1990):
- Kappa > .0.75 denotes excellent reliability
- Kappa between 0.40 and 0.75 denotes good reliability
- Kappa less than 0.40 denote marginal reliability

2010-2011 audit: Agreement data from 426 records were imported into a formatted data table for analysis. First, agreement data were used to calculate concordance at the data element level. Second, by applying the measure analytic calculation, concordance at the measure level was calculated.

2a2.3 **Testing Results** *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)***:**
2008 IRR study: measure level Kappa 0.85 (excellent reliability)

2010-2011 audit: measure level concordance 98% (valid N=132 records)

2b. VALIDITY. Validity, Testing, <u>including all Threats to Validity</u>: H☐ M☐ L☐ I ☐

2b1.1 **Describe how the measure specifications** *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**
Measure specifications are consistent with the evidence cited. Members of the ASCO guideline development panel reviewed the measure specification to ensure consistency.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 **Data/Sample** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**

2b2.2 **Analytic Method** *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)***:**
In 2009, an ASCO steering group comprised of medical oncologists, health services researchers, and quality experts undertook an iterative, criteria-based assessment process to identify QOPI measures that are appropriate for use for accountability measurement. This measure was selected as appropriate for accountability.

Face validity of the measure score was assessed via survey of experts involved in ASCO committees in 2011. The survey explicitly asked whether the scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

2b2.3 **Testing Results** *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)***:**
Face validity survey results revealed that 95% of respondents 'strongly agree' or 'agree' that this measure provides an accurate reflection of quality and can be used to distinguish good and poor quality.

POTENTIAL THREATS TO VALIDITY. (*<u>All</u> potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 **Data/Sample for analysis of exclusions** *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)***:**
n/a

2b3.2 **Analytic Method** (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)***:**
n/a

2b3.3 **Results** *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
n/a

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
n/a

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
n/a

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
n/a

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: n/a

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Data reported are from the Fall 2011 QOPI round, reflecting data submitted October and November 2011. 208 practices reported this measure. Data from 7987 patient records were submitted for this measure.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
QOPI measure analytics at the practice level were generated. Practices with fewer than 5 records were not included in calculations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
For the Fall 2011 QOPI round: practice mean = 98%; practice minimum = 75%; practice maximum = 100%.

This measure has been implemented in QOPI for several years. In this self-selected group on oncology practitioners committed to quality assessment and improvement, concordance with this measure has been high overall; however, a proportion of practices (new and experienced) continue to demonstrate sub-optimal variation.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
n/a

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
n/a

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
n/a

2c. Disparities in Care: H☐ M☐ L☐ I☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: n/a

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
n/a

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. **(evaluation criteria)**

C.1 Intended Purpose/ Use *(Check all the purposes and/or uses for which the measure is intended)*: Professional Certification or Recognition Program, Public Reporting, Quality Improvement (Internal to the specific organization), Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

3.1 Current Use *(Check all that apply; for any that are checked, provide the specific program information in the following questions)*: Professional Certification or Recognition Program, Quality Improvement with Benchmarking (external benchmarking to multiple organizations), Quality Improvement (Internal to the specific organization)

3a. Usefulness for Public Reporting: H☐ M☐ L☐ I ☐
(*The measure is meaningful, understandable and useful for public reporting.*)

3a.1. Use in Public Reporting - disclosure of performance results to the public at large *(If used in a public reporting program, provide name of program(s), locations, Web page URL(s)).* If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: **[*For Maintenance* –** *If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.***]**
n/a

3a.2.Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results**:** Members of the Alliance of Dedicated Cancer Centers (ADCC), an organization whose members are PPS-exempt cancer centers, independently reviewed this measure and support its use for public reporting. The ADCC agrees that the measure is supported by the current literature, is important to measure, and reflects excellence in patient care.

3.2 Use for other Accountability Functions (payment, certification, accreditation). If used in a public accountability program, provide name of program(s), locations, Web page URL(s)**:** An ASCO steering group comprised of medical oncologists, health services researchers, and quality experts undertook an iterative, criteria-based assessment process to identify QOPI measures that are appropriate for the QOPI Certification Program. The QOPI Certification Program is a practice-level certification program based on exemplary performance against QOPI measures, compliance with standards for chemotherapy safety, and a rigorous site visit. http://qopi.asco.org/certification

3b. Usefulness for Quality Improvement: H☐ M☐ L☐ I ☐
(*The measure is meaningful, understandable and useful for quality improvement.*)

3b.1. Use in QI. If used in quality improvement program, provide name of program(s), locations, Web page URL(s):
[*For Maintenance* – *If not used for QI, indicate the reasons and describe progress toward using performance results for improvement*].
This measure has been used in ASCO's Quality Oncology Practice Initiative (QOPI) since 2007. QOPI is a medical oncology practice improvement initiative that was launched nationally in 2006. Registered practices can participate in QOPI data submission rounds twice per year. Abstracted data are submitted via a secure web portal. Reports generated at the close of data collection rounds provide practice-specific and national aggregate data. About 750 diverse oncology practices across the country are registered for the QOPI program. During each round of QOPI data collection within the past year, approximately 250 practices submitted about 25,000 patient records.

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated *(e.g., QI initiative),* describe the data, method and results:
QOPI is a program developed by and for medical oncologists to inform their quality improvement. Diverse committees (including quality committees and relevant guideline panels) and member groups select measurement areas and contribute to measure development. One of the explicit criteria for review of concepts for measure development and implementation in QOPI is whether the reported data will meaningful, understandable, and useful for practice-based QI.

This measure has been implemented in QOPI since 2007; data have been submitted over 9 rounds by hundreds of practice sites from thousands of patient records. QOPI measures are reviewed twice per year by committee experts to determine whether updates or 'sunsetting' are needed. All participating QOPI practices are surveyed twice per year with specific questions regarding abstraction feasibility and usefulness of reported measures. Data availability, interpretation, or other abstraction issues received via the QOPI HelpDesk are logged, organized, and analyzed. Ad hoc feedback is welcomed and received. QOPI participating practices convene yearly at ASCO's Annual Meeting to inform the program and its measures.

These formal and informal mechanisms for input from QOPI member practices provide evidence regarding the meaningfulness of QOPI measures for quality improvement. Her2 testing is a straightforward measure and is easy to interpret and understand; it is backed by strong evidence, and represents a clear, actionable target for improvement if a practice notes a gap.

Overall, to what extent was the criterion, *Usability*, met? H☐ M☐ L☐ I ☐
Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. **(evaluation criteria)**

4a. Data Generated as a Byproduct of Care Processes: H☐ M☐ L☐ I ☐

4a.1-2 How are the data elements needed to compute measure scores generated? *(Check all that apply).*
Data used in the measure are**:**
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

4b. Electronic Sources: H☐ M☐ L☐ I ☐

4b.1 Are the data elements needed for the measure as specified available electronically *(Elements that are needed to compute measure scores are in defined, computer-readable fields)***:** Some data elements are in electronic sources

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources: Electronic sources of all data elements are not always available.

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H☐ M☐ L☐ I ☐

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:

4d. Data Collection Strategy/Implementation: H☐ M☐ L☐ I ☐

A.2 **Please check if either of the following apply** (*regarding proprietary measures*)**:**

4d.1 **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues** *(e.g., fees for use of proprietary measures)***:**

During testing implementation in QOPI, the need for the addition of the response option, 'HER-2/neu status = Test ordered, insufficient sample for results OR HER-2 equivocal' was identified and this change was made.

Overall, to what extent was the criterion, *Feasibility*, met? H☐ M☐ L☐ I ☐
Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes☐ No☐
Rationale:

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure before a final recommendation is made.

5.1 **If there are related measures *(either same measure focus or target population)* or competing measures *(both the same measure focus and same target population*), list the NQF # and title of all related and/or competing measures:**

5a. Harmonization

5a.1 **If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s): Are the measure specifications completely harmonized?**

5a.2 **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:**

5b. Competing Measure(s)

5b.1 **If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures** *(e.g., a more valid or efficient way to measure quality)***; OR provide a rationale for the additive value of endorsing an additional measure.** *(Provide analyses when possible)***:**

CONTACT INFORMATION

Co.1 **Measure Steward (Intellectual Property Owner):** American Society of Clinical Oncology, 2318 Mill Road, Suite 800, Alexandria, Virginia, 22314

Co.2 **Point of Contact:** Thomas, Murray, Tom.Murray@asco.org, 571-483-1641-

Co.3 **Measure Developer if different from Measure Steward:** American Society of Clinical Oncology, 2318 Mill Road, Suite 800, Alexandria, Virginia, 22314

Co.4 **Point of Contact:** Thomas, Murray, Tom.Murray@asco.org, 571-483-1641-
Co.5 **Submitter:** Thomas, Murray, Tom.Murray@asco.org, 571-483-1641-, American Society of Clinical Oncology
Co.6 **Additional organizations that sponsored/participated in measure development:**
Co.7 **Public Contact:** Thomas, Murray, Tom.Murray@asco.org, 571-483-1641-, American Society of Clinical Oncology

ADDITIONAL INFORMATION
Workgroup/Expert Panel involved in measure development Ad.1 **Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.**
Ad.2 **If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward:**
Measure Developer/Steward Updates and Ongoing Maintenance Ad.3 **Year the measure was first released:** Ad.4 **Month and Year of most recent revision:** Ad.5 **What is your frequency for review/update of this measure?** Ad.6 **When is the next scheduled review/update for this measure?**
Ad.7 **Copyright statement:**
Ad.8 **Disclaimers:**
Ad.9 **Additional Information/Comments:**
Date of Submission (*MM/DD/YY*)**:** 01/13/2012