Recommendations from the EGAPP Working Group: can tumor gene expression profiling improve outcomes in patients with breast cancer?

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group*

Disclaimer: *This recommendation statement is a product of the independent EGAPP Working Group. Although the Centers for Disease Control and Prevention (CDC) provides support to the EGAPP Working Group, including staff support in the preparation of this document, recommendations made by the EGAPP Working Group should not be construed as official positions of the CDC or the U.S. Department of Health and Human Services.*

Summary of Recommendations: The EGAPP Working Group (EWG) found insufficient evidence to make a recommendation for or against the use of tumor gene expression profiles to improve outcomes in defined populations of women with breast cancer. For one test, the EWG found preliminary evidence of potential benefit of testing results to some women who face decisions about treatment options (reduced adverse events due to low risk women avoiding chemotherapy), but could not rule out the potential for harm for others (breast cancer recurrence that might have been prevented). The evidence is insufficient to assess the balance of benefits and harms of the proposed uses of the tests. The EWG encourages further development and evaluation of these technologies.

Rationale: The measurement of gene expression in breast tumor tissue is proposed as a way to estimate the risk of distant disease recurrence in order to provide additional information beyond current clinicopathological risk stratification and to influence decisions about treatment in order to improve health outcomes. Based on their review of the EGAPP-commissioned evidence report, Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes[1] and other data summaries, the EWG found no direct evidence linking tumor gene expression profiling of women with breast cancer to improved outcomes, and inadequate evidence to construct an evidence chain. However, further evaluation on the clinical utility of some tests and management algorithms, including well-designed randomized controlled trials, is warranted. **Analytic Validity:** Some data on technical performance of assays were identified for MammaPrint and Oncotype DX, though estimates of analytic sensitivity and specificity could not be made. Published performance data on the laboratory developed Quest H:I Test were limited. Overall, the EWG found the evidence to be inadequate. **Clinical Validity:** The EWG found adequate evidence regarding the association of the Oncotype DX Recurrence Score with disease recurrence and adequate evidence for response to chemotherapy. The EWG found adequate evidence to characterize the association of MammaPrint with future metastases, but inadequate evidence to assess the added value to standard risk stratification, and could not determine the population to which the test would best apply. The evidence was inadequate to characterize the clinical validity of the Quest H:I Test. **Clinical Utility:** The EWG found no evidence regarding the clinical utility of the MammaPrint and Quest H:I Ratio tests, and inadequate evidence regarding Oncotype DX. These technologies have potential for both benefit and harm. **Contextual Issues:** The EWG reviewed economic studies that used modeling to predict potential effects of using gene profiling, and judged the evidence inadequate. *Genet Med* 2009:11(1): 66–73.

Key Words: *breast cancer, tumor gene expression, recurrence*

*EGAPP Working Group: Chair: Alfred O. Berg, MD, MPH (University of Washington), Members: Katrina Armstrong, MD, MSCE (University of Pennsylvania School of Medicine); Jeffrey Botkin, MD, MPH (University of Utah); Ned Calonge, MD, MPH (Colorado Department of Public Health and Environment); James Haddow, MD (The Warren Alpert Medical School of Brown University); Maxine Hayes, MD, MPH (Washington State Department of Health); Celia Kaye, MD, PhD (University of Colorado School of Medicine); Kathryn A. Phillips, PhD (University of California, San Francisco); Margaret Piper, PhD, MPH (Blue Cross/Blue Shield Association Technology Evaluation Center); Carolyn Sue Richards, PhD, FACMG (Oregon Health & Science University); Joan A. Scott, MS, CGC (Johns Hopkins University); Ora L. Strickland, PhD, DSc (Hon.), RN, FAAN (Emory University); Steven Teutsch, MD, MPH (Merck & Co.).

E-mail: egappinfo@egappreviews.org.

CLINICAL CONSIDERATIONS
Definitions used by EGAPP

- Analytic validity refers to a test's ability to accurately and reliably measure the genotype or analyte of interest, in this case the expression of mRNA by breast cancer tumor cells.
- Clinical validity defines the ability of the test to accurately and reliably identify or predict the intermediate or final outcomes of interest. This is usually reported as clinical sensitivity and specificity.
- Clinical utility defines the balance of benefits and harms associated with the use of the test in practice, including improvement in measureable clinical outcomes and use-

fulness/added value in clinical management and decision-making compared with not using the test.

Patient population under consideration

These recommendations apply to individuals diagnosed with Stage I or Stage II, node-negative breast cancer. Tumors may be estrogen receptor (ER) positive or negative for MammaPrint testing, but must be estrogen receptor positive to be eligible for Oncotype DX or Quest H:I testing.

Considerations for practice

- Until more data are available, clinicians must decide on a case by case basis if the use of a gene expression profile test adds value beyond the use of the current prognostic markers (and how to weigh and combine these risks), and if each test's validation population is relevant to their patients' age, disease status, and race/ethnicity.
- If a clinician considers use of gene expression profiling in an individual with newly diagnosed breast cancer, provision of counseling and educational materials is suggested to inform the patient about both the potential benefits and harms associated with testing and discuss whether the test results are likely to change the patient's decision about therapy.

BACKGROUND AND CLINICAL CONTEXT FOR THE RECOMMENDATION

Breast cancer is the most common cancer and the second leading cause of cancer-related death in women in the United States, with 178,000 new cases and 40,000 deaths expected in 2007.[1] Treatment involves surgery, endocrine therapy for women with tumors expressing the ER, and/or chemotherapy or radiation. Prognostic decision-making algorithms (e.g., National Comprehensive Cancer Network guidelines, St. Gallen expert criteria, Adjuvant! Online)[2–5] support assessment of risk for breast cancer recurrence, and recommendations relevant to the decision about treatment options. Such algorithms have been based on risk factors such as patient age and menopausal status, comorbidities, tumor size and cancer grade, axillary lymph node involvement, and ER status,[1] and have limited effectiveness in predicting risk of recurrence.[1–5] Most women with early-stage breast cancer are offered chemotherapy.

In women with ER-positive breast cancer, 5-year postoperative treatment with tamoxifen reduces recurrence rates and improves survival, with one clinical trial reporting a 10-year recurrence rate of 15%.[6] This effect of tamoxifen is not seen in women with ER-negative tumors, but in ER-positive tumors is largely independent of other tumor characteristics, age, and chemotherapy treatment.[2,7,8] The side effects of tamoxifen therapy are relatively mild for most women (e.g., hot flashes, nausea/vomiting, gynecologic problems) and are infrequently severe enough to discontinue treatment.[9]

Adjuvant chemotherapy reduces the annual odds of recurrence and death for many women with breast cancer, especially those with ER-negative tumors. The size of the chemotherapy effect varies depending on the drug(s) used and the therapy regimen. Adjuvant chemotherapy considered for patients with early breast cancer includes CMF (cyclophosphamide, methotrexate, and 5-fluorouracil), AC (doxorubicin and cyclophosphamide), and anthracycline-based regimens.[8] Overall, node-negative women with early stage ER-positive tumors treated with tamoxifen

have the best prognosis, but this population also receives a small, but significant, benefit from chemotherapy.[7]

Adverse drug effects also vary by drug and regimen, but can have a well-described negative impact on patients' quality of life. Data on the number, severity, costs, and long-term sequelae of serious adverse effects in pre- and postmenopausal women undergoing chemotherapy were considered but not systematically reviewed. However, chemotoxicity-related deaths appear to be uncommon; a quoted estimate from 2000 was 1 in 200–500 for all women treated with adjuvant chemotherapy.[8] In addition, new medications have become available to minimize some side effects, and some new shorter treatment regimens reduce the duration of adverse effects. Consequently, a key question is how the relatively small absolute benefit of chemotherapy in node-negative, ER-positive, tamoxifen-treated women weighs against the harm of adverse drug effects, particularly when as many as 85% of such women who do not receive chemotherapy may remain disease free at 10 years.[6]

Descriptions of tests and intended use claims

The EGAPP-commissioned evidence report focused on three gene expression profiling tests for women with breast cancer that were clinically available in the United States at the time the review was initiated.[1,5] The intended uses of these three tests, and the performance claims made, are different. All claim to provide prognostic information (i.e., recurrence and survival rates) in specific subpopulations of women with early-stage breast cancer, and to identify women most likely to benefit from chemotherapy.

Agendia (Amsterdam, The Netherlands) materials state that the MammaPrint® Test is intended for use in women 61 years of age or younger with primary invasive (Stage I or II) breast cancer who are lymph node-negative and have a ≤5 cm, ER-positive *or* ER-negative tumor.[1,5,10] MammaPrint was cleared for marketing by the U.S. Food and Drug Administration (FDA) in 2007 for use as a prognostic test to be used along with other clinicopathologic factors, and is not intended "to predict or detect response to therapy, or to help select the optimal therapy for patients."[10] Test results are reported as *low risk* ("13% chance to develop distant metastases at 10 years without adjuvant treatment") or *high risk* ("56% chance to develop distant metastases at 10 years without adjuvant treatment").[11] The MammaPrint report also states that MammaPrint "provides independent prognostic information to clinicopathological risk assessment . . .," and that "its performance characteristics and clinical utility in the United States Population have not been established."[11]

Genomic Health Inc. (Redwood City, CA) states that the Oncotype DX® Breast Cancer Assay is intended for use with other conventional risk assessment approaches (e.g., tumor staging/grading, analysis of other markers) to predict the likelihood of distant breast cancer recurrence in women of any age with newly diagnosed Stage I or II breast cancer, lymph node-negative and ER-positive, who will be treated with tamoxifen.[12] Results are reported as a Recurrence Score™ (RS; scale of 0–100) that correlates to a patient-specific "Average Rate of Distant Recurrence" (with a 95% confidence interval). To determine prognosis, patients are categorized as low (RS <18), intermediate (RS 18–30), or high risk (RS ≥31). The *low-*, *intermediate-*, and *high-risk* categories are stated to correspond to 10-year distant recurrence rates after 5 years of tamoxifen therapy of <12%, from 12 to 21%, and from 21 to 33%, respectively.[12,13] Oncotype DX claims to provide information beyond conventional risk assessment tools, including how

likely the woman is to benefit from chemotherapy (CMF/MF) in addition to tamoxifen therapy.[12]

The Breast Cancer Gene Expression Ratio *(HOX13:IL17BR)* Assay (or H:I ratio test), was developed by Quest Laboratories based on licensed gene expression profiling technology from AviaraDx, Inc. (Carlsbad, CA). This test measures the ratio of the expression of the homeobox gene-B13 (*HOXB13*) and the interleukin-17B receptor gene (*IL17BR*). The test was originally designed to go beyond the current clinical standard (e.g., estrogen/progesterone receptor status) to predict tumor recurrence risk for women on tamoxifen monotherapy, for whom alternative therapies (e.g., aromatase inhibitors, chemotherapy) might be considered. Quest indicates that the H:I ratio is a "continuous marker of recurrence in untreated ER-positive/node-negative patients."[14] Results are reported as a normalized H:I expression ratio along with a categorization of *low* (roughly 10–27% based on Test Summary Figure) or *high* (roughly 28 to >60%) breast cancer recurrence risk at 5 years.[14] The Quest web page for this test states that clinical uses are to "predict breast cancer recurrence risk" and "determine appropriate therapy."[15]

The evidence report

In an attempt to understand the utility of tumor gene expression profiling in predicting the risk of breast cancer recurrence or in identifying the women most likely to benefit from chemotherapy, EGAPP commissioned an evidence-based review to address an overarching question regarding the following specific clinical scenario:

What is the direct evidence that gene expression profiling tests in women diagnosed with breast cancer, or any specific subset of this population, lead to improvement in outcomes?

REVIEW OF SCIENTIFIC EVIDENCE

This statement summarizes the supporting scientific evidence used by the EWG to make recommendations regarding the use of three specific tumor gene expression profiling tests in women with breast cancer.

Methods

EGAPP is a project developed by the National Office of Public Health Genomics at the CDC to support a rigorous, evidence-based process for evaluating genetic tests and other genomic applications that are in transition from research to clinical and public health practice in the U.S.[16] A key goal of the EWG is to develop conclusions and recommendations regarding clinical genomic applications, and to establish clear linkage to the supporting scientific evidence. The EWG members are nonfederal multidisciplinary experts convened to establish methods and processes, set priorities for review topics, participate in technical expert panels for commissioned evidence reviews, and develop and publish recommendations.

EGAPP commissioned an evidence review through the Agency for Healthcare Research and Quality (AHRQ); The Johns Hopkins University Evidence-Based Practice Center conducted the review. Since it was anticipated that data might not be available to directly answer the overarching question, the EWG constructed an analytic framework and key questions to address different components of evaluation (e.g., analytic and clinical validity, clinical utility) for the purpose of providing relevant indirect evidence of efficacy. Established AHRQ Evidence Based Practice Center methods were followed in conducting this review. A Technical Expert Panel that included three EWG members provided expert guidance during the course of the review. The final report "Impact of Gene Expression Profiling Tests on Breast Cancer Outcomes" is available online.[1] A peer-reviewed summary report of the evidence has also been published.[5]

EWG members reviewed the evidence report, key primary publications, other sources of information, and comments on the evidence report from the test developers and a group of eight peer reviewers. The process also included assessment of key gaps in knowledge and relevant contextual factors (e.g., availability of diagnostic or therapeutic alternatives, feasibility and practicality of implementation, cost-effectiveness). The final EWG recommendation statement was formulated based on magnitude of effect, certainty of evidence, and consideration of contextual factors.[17]

Technology

Gene expression profiling is an emerging technology that identifies expression or activity of genes that may be associated with disease prognosis by characterizing and quantifying cellular messenger RNA (mRNA) in tumor tissue. Methods used for gene expression analysis, such as real-time reverse transcription polymerase chain reaction (RT-PCR) and microarray technologies, have been widely used in research, and are now being used in clinical settings. Oncotype DX and the H:I ratio test are proprietary laboratory-developed tests, each offered by a single CLIA-certified laboratory. Both tests use RT-PCR for the detection and quantitation of mRNA in formalin fixed, paraffin-embedded (FFPE) breast cancer tissue. Oncotype DX analyzes expression of 21 genes, 16 cancer related, and 5 normative. The H:I ratio test measures the ratio of the expression of the *HOXB13* and *IL17BR* genes, along with expression of four normative genes.

The MammaPrint test is based on microarray technology (labeled patient mRNA hybridized to DNA sequences from known genes on a customized microarray chip), and is used for identification of mRNA in tumor tissue that is fresh (transported in a specific preservation solution) or frozen.[1] The MammaPrint "custom microarray" that is used clinically tests for 70 cancer related and about 1800 normative genes. This 70-gene profile is also proprietary, and tests are conducted in Agendia's CLIA-certified laboratory in the Netherlands. The MammaPrint assay has been cleared by the FDA as a class II, 510(k) product, which ensures independent review of data and labeling, and conformance of the device sponsor to good manufacturing practices.[10] However, the FDA did not evaluate treatment outcomes as a result of use of this "prognostic" device.

Defined protocols exist for each of these assays for evaluating tumor content of specimens undergoing analysis, preparing mRNA samples, normalizing expression data, and computing summary indices. In addition to the procedures involved in analyses, differences exist in gene panels utilized and other test characteristics.

Analytic validity

Analytic validity is defined as a test's ability to accurately and reliably measure the analyte or genotype of interest (in this case, expression of mRNA by breast cancer tumor cells), and usually includes measures of analytic sensitivity and specificity, as well as assay reproducibility, robustness (e.g., resistance to small changes in preanalytic or analytic variables), and quality control. In this clinical context, there is no "gold standard" against which gene expression profiling tests can be directly compared to compute estimates of analytic sensitivity and specificity. However, data were available to assess other measures of analytic validity.

MammaPrint

Glas et al.[18] reported a high correlation between results from the original gene signature report[19] and sample retesting using the MammaPrint custom mini-array. Studies of the custom mini-array technology included information on linearity, reportable range, and variability in repeat sample analyses (coefficients of variation <7%).[8,10,18] The FDA summary reported result accuracy of 98.5% and classification accuracy of 97.7% on repeat testing.[20] Reproducibility studies at three sites showed no significant differences in results between sites or testing on different days. RNA labeling appeared to be the largest contributor to interlaboratory variation, but accounted for differences in gene expression ratios of 5% or less.[21] A failure rate of 19.1% was reported in fresh tissue samples, all attributed to poor RNA samples.[22]

Oncotype DX

Cronin et al.[23] demonstrated that archived FFPE tumor specimens could be successfully used to measure gene expression levels, and reported amplification efficiencies of 88–96%, linearity across a wide range of RNA concentrations, and precision and accuracy of testing for individual gene components. Three studies reported that between-day, and between- and within-sample reproducibility was <2.5 RS units, but did not address the impact of RS variability on risk stratification.[23–25] Based on seven studies, testing initially failed in 14.5% of samples; 10.9% of failures were attributed to insufficient tumor content in samples, and 4.1% to poor RNA samples and RT-PCR assay failure.[1,25–30]

H:I ratio test

Ma et al.[31] supported the use of the microarray method in laser-micro-dissected FFPE specimens. Three studies reported initial failure of microarray testing in 11.7% of samples; 9.2% of failures were attributed to insufficient tumor content in samples, and 2.7% to poor RNA samples and RT-PCR assay failure.[31–33] Assay details and performance characteristics for the laboratory-developed H:I test offered by Quest have not been published.

Analytic validity conclusions

- More studies are available to assess analytic performance of the MammaPrint and Oncotype DX tests, though additional data on some points (e.g., impact of variability on risk classification) are needed.

- In the absence of a gold standard or referent technology, no estimates are available for analytic false positive or false negative rates.
- Testing on 12–19% of samples initially failed across the three tests, suggesting that tissue sampling and processing may be a potential issue for broad implementation of testing.

Ongoing monitoring of test performance and careful evaluation of the quality of submitted specimens are needed to ensure that technical performance of the assays in clinical practice is at least comparable to existing reported data. Since all three of these tests are proprietary (single clinical source) and external proficiency testing is not available at this time, reporting by the laboratories of quality control/quality assessment protocols and analytic performance data would provide additional information for clinicians and consumers.

Clinical validity

Clinical validity is defined as a test's ability to accurately and reliably identify or predict the disorder or phenotype of interest, in this case prediction of overall survival or recurrence-free survival 5–10 years after surgery versus avoidance of chemotherapy toxicity and quality of life. Clinical validity was documented to some degree for all three gene expression profiles.
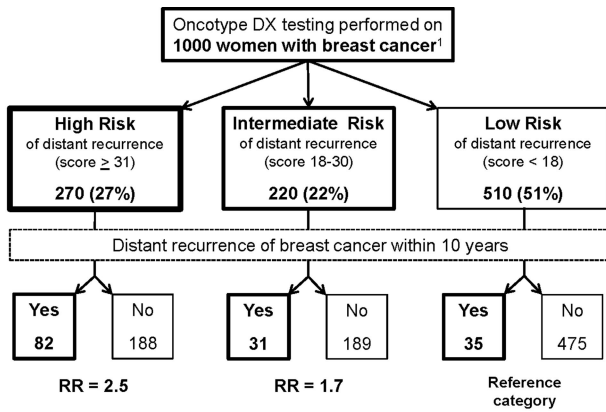
Oncotype DX

- A significant correlation was reported between RS and distant recurrence risk; data provided in three studies were sufficient to derive odds ratios and estimates of clinical sensitivity and specificity with 95% confidence intervals (Table 1).[24,25,28,34]
- A higher quality study conducted retrospectively on samples from an arm of the NSABP B-14 randomized controlled trial reported that 10-year distant recurrence rate estimates for tamoxifen-treated patients classified by RS as high, intermediate, and low risk were 30.5% (95% CI 23.6–37.4), 14.3% (95% CI 8.3–20.3), and 6.8% (95% CI 4.0–9.6), respectively.[25]
- Women in the high risk category have a relative risk of 10-year distant metastasis that is 2.5 times higher than women with low risk results.[25] Figure 1 provides a simple model showing the number of women in each risk group,

Table 1 Estimated clinical sensitivity, specificity, and odds ratios for Oncotype DX recurrence scores in women with lymph node-negative, estrogen receptor-positive breast cancer treated with tamoxifen, and subsequent breast cancer outcome

Publication	Study design	Total N (outcome)[a]	Primary outcome	Positive test defined as[b]	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Odds ratio (95% CI)
Habel, et al.[24]	Case-control	205 (55)	Death by 10 yrs	High+ IM	71 (57–82)	63 (55–71)	4.2 (2.1–8.7)
				High only	31 (19–45)	87 (80–92)	2.9 (1.3–6.5)
Paik, et al.[25]	Cohort	668 (99)	Distant recurrence by 10 yrs	High+ IM	77 (67–85)	55 (51–59)	4.1 (2.4–6.9)
				High only	56 (45–66)	78 (74–81)	4.4 (2.8–7.0)
Paik, et al.[34]	Cohort	227 (27)	Recurrence by 10 yrs	High+ IM	85 (66–96)	66 (58–72)	11 (3.4–39)
				High only	70 (50–86)	86 (80–90)	15 (5.4–41)

[a]Total number in the study (number with primary outcome).
[b]All studies categorized Oncotype DX recurrence scores into high (>31), intermediate-IM (18–30) and low- (<18) risk groups. For these calculations, "Positive" was defined twice for each study, with the intermediate risk group first combined with the high-risk group, and then with the low-risk group.

Fig. 1. Expected performance of the Oncotype DX test in a defined population of women with breast cancer. Test performance is derived from the report by Paik S et al.,[25] of 668 women from the NSABP trial B-14.[25] The risk ratios indicate the relative increase in the rate of distant recurrence of breast cancer within 10 years, and evidence suggests that chemotherapy might be more effective among the high-risk patients than among patients in the intermediate- or low-risk groups.

along with the recurrence risks for the three groups based on the results of this study.

- Using data from NSABP B-20, this group also reported a statistically significant interaction between RS and chemotherapy benefit ($P = 0.038$).[34] There is evidence that breast cancer patients in the high risk category show benefit from chemotherapy (limited to CMF/MF regimens), whereas patients with low risk results were reported to show little benefit from chemotherapy.[33] Estimated benefit was uncertain for women in the intermediate risk category.[34]
- Incremental value of the test over classical clinical factors has been reported, but the EWG was unable to determine the size of the potential benefit.[1,5]

MammaPrint

- MammaPrint has been validated in two studies[18,22] subsequent to the development of the 70-gene signature (Table 2).[35]

- Van Vijver reported 5-year recurrence rates for patients with a MammaPrint *poor* or *good* prognosis were 39% and 5%, respectively; 10-year recurrence rates were 49% and 15%, respectively.[15]
- Some data support the incremental value of the test over classical clinical factors, but the MINDACT study will provide further validation.[5,36]
- One study stratified by ER status for the outcome of 10-year survival, providing odds ratios of 3.5 (95% CI 1.6–7.8) for ER-positive women and 1.76 (95% CI 0.2–16.3) for ER-negative women.[22]

H:I ratio test

- The body of data on H:I ratio testing is quite heterogeneous; reports include different ratio formulations, normalizing factors, optimal cut points, and populations (e.g., mixed node-negative and -positive).[1] A consistent finding was that increasing H:I expression ratio is associated with increased recurrence and death in patients with node-negative, ER-positive breast cancer treated with tamoxifen.
- A single study[31] on 308 women with primary, untreated, ER-positive, node-negative breast cancer (testing not conducted at Quest) provides the basis for the H:I test offered by Quest, with an H:I ratio cut-point (defined as H:I ratio >1.0) separating low risk from high risk. This test is still clearly in a developmental phase.

Clinical validity conclusions

- For Oncotype DX, the EWG found adequate evidence from one higher quality study to support the association between RS and rates of 10-year distant metastasis,[25] and adequate evidence to support the association between RS and chemotherapy benefit.[34] Study subjects were mainly whites, and how characteristics of other demographic populations might affect test performance is not known.
- For MammaPrint, the EWG found that data were adequate to support an association between the MammaPrint Index and 5- or 10-year metastasis rates, but the relative efficacy of testing in ER-positive and -negative women is not clear. Study subjects were European, and how characteristics of other demographic populations might affect test performance is not known.
- For the H:I test, the EWG found that the evidence available to assess clinical validity is inadequate, with a small num-

Table 2 Estimated clinical sensitivity, specificity, and odds ratios for MammaPrint recurrence scores in women with lymph node-negative, estrogen receptor-positive or -negative Stage I or II breast cancer treated, and subsequent breast cancer outcome

Publication	Study design	Total N (outcome)[a]	Metastasis by (yrs)	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Odds ratio (95% CI)
Van't Veer, et al.[35]	Case-control, signature set	78 (34)	5	91 (76–98)	59 (43–74)	15 (3.6–72)
van de Vijver, et al.[19]	Cohort	67 (12)	5	92 (62–99)	58 (44–71)	15 (1.8–86)
Buyse, et al.[22]	Cohort, FDA submission	302 (48)	5	87 (75–95)	41 (35–48)	4.9 (1.9–13)
Summary			5			8.8 (3.9–19)
van de Vijver, et al.[19]	Cohort	151 (59)	10	86 (75–94)	57 (46–67)	8.3 (3.3–21)
Buyse, et al.[22]	Cohort, FDA submission	302 (48)	10	83 (72–91)	42 (36–49)	3.7 (1.8–7.9)
Summary			10			5.2 (2.4–12)

[a]Total number in the study (number with primary outcome).

ber of studies in a variety of heterogeneous populations, and only one study[31] that applies directly to the laboratory-developed test offered by Quest. Gaps in knowledge exist regarding the calibration of the expression-based score with the absolute observed risk, and how characteristics of different demographic populations might affect test performance.

Clinical utility

In this context, clinical utility is the likelihood that using a gene expression profiling test to guide management in patients with diagnosed early-stage breast cancer will significantly improve health-related outcomes. Clinical utility is assessed by investigating the balance of benefits (reduced adverse events due to low risk women avoiding chemotherapy) and harms (cancer recurrence that might have been prevented) associated with the use of the test, and how that compares to the use of alternative management strategies.

MammaPrint

- No direct evidence links use of the MammaPrint test to clinical outcomes.
- No studies evaluated MammaPrint for ability to predict benefit from other treatments (e.g., chemotherapy).
- No studies determined whether use of MammaPrint in place of, or in addition to, current clinicopathologic markers (e.g., Adjuvant! Online, St. Gallen) changes management, and/or improves outcomes based on management with clinicopathologic markers alone.

Oncotype DX

- Retrospective analysis of one arm of a prospective clinical trial showed that the chemotherapy benefit in ER-positive, node-negative patients, randomized to tamoxifen or to tamoxifen plus chemotherapy, was most convincing in women in the Oncotype DX RS high risk category (27% reduction in 10-year recurrence rate).[34] Many would likely have been offered chemotherapy without testing, but test results may influence patient-clinician decision making. This study provided the strongest evidence available addressing the clinical utility of the Oncotype DX test, but the study design was not optimal and prospective confirmation of these findings is needed.
- Three quarters of Oncotype DX results are in the intermediate and low risk ranges, where estimates of recurrence risk have wide and overlapping confidence intervals, and evidence of benefit of chemotherapy is inadequate to make decisions. The TAILORx trial will focus on these result groups,[37] but results will not be known for some time.

H:I ratio test

- No evidence was found that links use of the H:I test to clinical outcomes.

Clinical utility conclusions

The EWG found no direct evidence linking any of the three tests to improved outcomes, but also examined the components of clinical utility that might provide indirect evidence for clinical utility. The EWG found encouraging indirect evidence for Oncotype DX, and plausibility for potential use of MammaPrint and, possibly, the H:I ratio test.

It is possible that the harms associated with chemotherapy among women who will not have a distant recurrence outweigh the benefit of chemotherapy among women who are destined to have a distant recurrence. It seems plausible that more women will benefit (i.e., avoid unnecessary chemotherapy), but there is the potential for significant harms among a small number of low or intermediate risk women (who might have benefited from chemotherapy), possibly resulting in breast cancer recurrence or death. There are currently insufficient data to confidently estimate these risks and benefits. In addition, it is difficult to determine what proportion of women with moderate to high risk based on conventional risk assessments will have a "low enough" Oncotype DX RS score to affect their decision about chemotherapy.

Clinical trials

Two prospective randomized trials are in progress. The TAILORx trial is primarily designed to determine the benefit of chemotherapy for women with intermediate risk Oncotype DX results (results 2013).[37] However, RS cutpoints for the trial are much more conservative than those used for the commercially available test. In this trial, women in the low risk category (RS <11 rather than <18) will receive adjuvant hormonal therapy and be followed to determine 10-year distant disease-free survival. High risk women (RS >25 rather than ≥31) will receive hormonal therapy and chemotherapy. Women at intermediate risk (RS 11–25 rather than 18–30) will be randomized to hormonal therapy alone or hormonal therapy plus chemotherapy. Outcomes will be compared with RS, current clinicopathological criteria, and other prognostic indicators (e.g., HER2, estrogen and progesterone receptor status, other genes).

The MINDACT trial is designed to compare the effectiveness of MammaPrint test results versus clinical evaluation in predicting 15-year disease-free survival and overall survival.[36] This trial will compare clinical response to endocrine therapy alone and with chemotherapy regimens (anthracycline-based, docetaxel-capecitabine, Letrozole).

Contextual issues important to the recommendation

- As discussed above, the use of gene expression profiling tests in this clinical scenario provides the potential for significant benefit but also potential for harm. More work is needed to better understand the balance of benefits and harms.
- No firm guidance can be given to clinicians on how MammaPrint and Oncotype DX results can be acted upon until data are available from the TAILORx and MINDACT trials.
- Studies of changes to clinical management using the Oncotype DX test are minimally informative because they have not specified the information actually given to the patient or how clinicians combine test results with other risk factors to limit or expand therapeutic choices. Consequently, it cannot be determined whether documented changes in management are due to compliance with physician recommendations, with weighing of risks and benefits, or reflect the effects of test marketing.
- Better understanding of risk tolerance in women will be needed to identify patients who might benefit most from testing and to help direct discussion with women about the potential risks and benefits of the tests. What is the recurrence risk below which women are comfortable with a decision to decline chemotherapy? How does the presentation of risk affect choices?
- A future scenario with a proliferation of competing licensed products without comparative effectiveness data has potential to confuse patients and clinicians, and not

deliver on the potential improvement in clinical outcomes that the current evidence individually suggests.

Cost-effectiveness

Two of three studies addressing the potential cost-effectiveness of gene expression profile tests concluded that use of one gene expression profile test (Oncotype DX) would be "relatively cost-effective" for those defined as low risk, and cost-saving for those at high risk.[38,39] However, concerns about the parameter estimates, lack of sensitivity analyses to assess sources of bias, and changes in the National Comprehensive Cancer Network (NCCN) guidelines reduce the confidence and relevance of one of these studies.[37] The second study had substantial limitations in the descriptions of the model structure, assumptions and comparators, as well as deficiencies in data specification, utilities, and sensitivity analyses.[37] Both studies were sponsored by the manufacturer. The EWG judged this body of evidence to be inconclusive.

An earlier study, meeting most standards for appraising the quality of an economic analysis, projected that MammaPrint would result in an absolute 5% decrease in the proportion of distant recurrence cases prevented and would yield slightly fewer quality-adjusted life years, but would marginally lower total costs (USD 2882).[40] The authors suggested the need for further validation before use in clinical practice.

Research gaps

The EGAPP Working Group found the research literature insufficient, but encouraging in many respects, and recommends further studies that could address important gaps in knowledge.

- Clinical validation of gene expression tests must include examination of the tests as actually available in typical populations of patients, and assessment of test characteristics across relevant ethnic groups. The risk estimates that result must be calibrated against actual observed risk.
- Management algorithms that are used to interpret tests need to be further evaluated. In addition, algorithms that integrate clinicopathologic data and other risk factors or scores may improve predictive value.
- A high priority should be publication of better quality data regarding analytic validity of tests, including data from external proficiency testing, as well as the feasibility of providing these tests routinely to large numbers of women.
- Based on the reported rates of testing failures related to sample quality, preanalytic issues related to sample preparation, transport, and processing should be addressed in routine practice.
- The differential predictive ability of the MammaPrint test based on ER status should be further evaluated.
- The TAILORx and other clinical trials are needed to understand the impact of chemotherapy on outcomes of women classified into the high, intermediate, or low risk groups or based on risk as a continuous variable. Women reclassified as lower risk by testing are of most concern, because patients may forego chemotherapy based on limited data.
- Research is needed to understand how women understand and use the risk information and investigate tolerance levels for risk in decision making.
- Clinical trials are needed to address predictive value of tests for the benefit of other existing or future chemotherapy regimens.
- Additional cost-effectiveness analyses are needed.

Recommendations of other groups

NCCN, Clinical Practice Guidelines in Oncology™—Breast Cancer, 2008

BIN V-6: Category 2B recommendation, defined as "nonuniform NCCN consensus (but no major disagreement), based on lower-level evidence including clinical experience . . ."[41]

"For patients with hormone receptor-positive, HER2-negative tumors that are 0.6–1.0 cm and moderately/poorly differentiated or with unfavorable features, or >1 cm, the recommendation for use of a 21-gene RT-PCR assay (category 2B) was added to the systemic adjuvant treatment decision pathway as an option for guiding chemotherapy treatment decisions." Pending the results of prospective trials, the NCCN Breast Cancer Panel considered the 21-gene RT-PCR assay (Oncotype DX) as an option for patients described above. "The panel emphasizes that the recurrence score should be used for decision making only in the context of other elements of risk stratification for an individual patient."

And in regard to microarray-based assays: "While many of the DNA microarray technologies are able to stratify patients into prognostic and/or predictive subsets on retrospective analysis, the gene subsets differ from study to study and prospective clinical trials testing the utility of these techniques have yet to be reported."[41]

American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer

"In newly diagnosed patients with node-negative, estrogen-receptor positive breast cancer, the Oncotype DX assay can be used to predict the risk of recurrence in patients treated with tamoxifen. . . . (and) to identify patients who are predicted to obtain the most therapeutic benefit from adjuvant tamoxifen and may not require adjuvant chemotherapy . . . . . patients with high recurrence scores appear to achieve relatively more benefit from adjuvant chemotherapy (specifically (C)MF) than from tamoxifen . . . . ."[42]

- The recommendation notes that testing of retrospectively collected tissues from a prospectively collected arm of a clinical trial might be considered Level I (high quality) evidence to support use of this test.
- They add that "there are insufficient data at present to comment on whether these conclusions generalize to hormonal therapies other than tamoxifen, or whether this assay applies to other chemotherapy regimens."
- "The precise clinical utility and appropriate application for other multiparameter assays, such as the MammaPrint assay . . . .and the Breast Cancer Gene Expression Ratio [H:I ratio] are under investigation."[42]

ECRI Institute Target Policy Statement—August, 2007 (based on a systematic review of Oncotype DX and MammaPrint only)

"The key clinical issues for this technology included the following:

- Do women who have the gene assay performed and have assay-guided treatment have significantly less breast cancer recurrence than patients who do not have the assay performed and have a treatment program based on standard methods?

- Do patients who have assay-guided treatment have improved survival compared with patients who do not have assay-guided treatment?"[43]

"Existing studies provide clinical validation for the ability of the Oncotype DX assay and the MammaPrint assay to predict tumor recurrence and response to chemotherapy. However, the studies are insufficient to allow one to draw strong conclusions regarding the clinical utility of these assays for guiding treatment decisions for patients with early-stage invasive breast cancer."

REFERENCES

1. Marchionni L, Wilson RF, Marinopoulos SS, et al. Impact of gene expression profiling tests on breast cancer outcomes. Evidence Report/Technology Assessment No. 160. AHRQ Publication No. 08-E002. 2008. Rockville, MD: Agency for Healthcare Quality and Research, 2008.
2. Carlson RW, Brown E, Burstein HJ, et al; for the National Comprehensive Cancer Network. NCCN task force report: adjuvant therapy for breast cancer. *J Natl Compr Canc Netw* 2006;4(suppl 1):S1–S26.
3. Adjuvant! Online. Available at: https://www.adjuvantonline.com/faq.jsp. Accessed August 2008.
4. Harbeck N, Jakesz R. St. Gallen 2007: Breast Cancer Treatment Consensus Report. *Breast Care* 2007;2:130–134.
5. Marchionni L, Wilson RF, Wolff AC, et al. Systematic review: gene expression profiling assays in early-stage breast cancer. *Ann Intern Med* 2008;148:358–369.
6. Fisher B, Jeong JH, Bryant J, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet* 2004;364:858–868.
7. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005; 365:1687–1717.
8. Piper MA. Gene expression profiling of breast cancer to select women for adjuvant chemotherapy. Blue Cross and Blue Shield Association. *Technol Eval Cent Asses Program Exec Summ* 2008; 22:1–51. Available at: http://www.bcbs.com/ blueresources/tec/vols/22/22_13.pdf. Accessed December 11, 2008.
9. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: Current Status of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst* 2005;97:1652–1662.
10. US Food and Drug Administration. US Food and Drug Administration 510(k) Substantial Equivalence Determination Decision Summary for agendia MammaPrint®. US Food and Drug Administration 510(k). Available at: http://www.fda.gov/cdrh/reviews/K062694.pdf. Accessed April 2008.
11. MammaPrint sample patient report "Analysis Explanation." Agendia. Available at: http://usa.agendia.com/en/your_test_results_4.html. Accessed August 22, 2008.
12. Genomic Health Web site: Oncotype DX sample patient report "Clinical Experience." GenomicHealth. Available at: http://www.genomichealth.com/ oncotype/publications/oncotype_reportform.pdf. Accessed June 8, 2008.
13. Genomic Health Web site: OncotypeDx Breast Cancer Assay. Genomic Health. Available at: http://www.genomichealth.com/oncotypdx/hcphome.aspx. Accessed August 26, 2008.
14. Medical Comment for Breast Cancer Gene Expression Ratio (HOXB13: IL17BR), 19600X Intrepretive Guide (Test Summary and figure). Quest Diagnostics/Nichols Institute. Available at: http://www.questdiagnostics.com. Accessed June 8, 2008.
15. Quest Diagnostics Website: Test Menu Breast Cancer Gene Expression Ratio (code 19600). Quest Diagnostics. Available at: http://www.questdiagnostics.com. Accessed August 26, 2008.
16. Centers for Disease Control: Evaluation of Genomic Applications in Practice and Prevention (EGAPP). Available at: http://www.cdc.gov/genomics/gtesting/EGAPP/about.htm. Accessed June 6, 2008.
17. Teutsch SM, Bradley LA, Palomaki GE, et al. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: Methods of the EGAPP Working Group. *Genet Med*; 3–14.
18. Glas AM, Floore A, Delahaye LJ, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 2006;7:278.
19. van de Vijver MJ, He YD, Van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
20. 501(k) Submission for MammaPrint Service in the US Section 5: 501(k) Summary. Available at: http://www.fda.gov/cdrh/pdf7/k070675.pdf. Accessed December 11, 2008.
21. Ach RA, Floore A, Curry B, et al. Robust interlaboratory reproducibility of a gene expression signature measurement consistent with the needs of a new generation of diagnostic tools. *BMC Genomics* 2007;8:148.
22. Buyse M, Loi S, van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst* 2006;98:1183–1192.
23. Cronin M, Sangli C, Liu ML, et al. Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer. *Clin Chem* 2007;53:1084–1091.
24. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;8:R25.
25. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351: 2817–2826.
26. Chang JC, Makris A, Gutierrez MC, et al. Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. *Breast Cancer Res Treat* 2008;108:233–240.
27. Cobleigh MA, Tabesh B, Bitterman P, et al. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin Cancer Res* 2005;11:8623–8631.
28. Esteva FJ, Sahin AA, Cristofanilli M, et al. Prognostic role of a multigene reverse transcriptase-PCR assay in patients with node-negative breast cancer not receiving adjuvant systemic therapy. *Clin Cancer Res* 2005;11:3315–3319.
29. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol* 2005;23:7265–7277.
30. Mina L, Soule SE, Badve S, et al. Predicting response to primary chemotherapy: gene expression profiling of paraffin-embedded core biopsy tissue. *Breast Cancer Res Treat* 2007;103:197–208.
31. Ma XJ, Hilsenbeck SG, Wang W, et al. The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol* 2006;24:4611–4619.
32. Goetz MP, Suman VJ, Ingle JN, et al. A two-gene expression ratio of homeobox 13 and interleukin-17B receptor for prediction of recurrence and survival in women receiving adjuvant tamoxifen. *Clin Cancer Res* 2006;12: 2080–2087.
33. Jerevall PL, Brommesson S, Strand C, et al. Exploring the two-gene ratio in breast cancer-independent roles for HOXB13 and IL17BR in prediction of clinical outcome. *Breast Cancer Res Treat* 2008;107:225–234.
34. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3726–3734.
35. Van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–536.
36. European Organisation for Research and Treatment of Cancer. Microarray in node negative disease may avoid chemotherapy (MINDACT) Trial. European Organisation for Research and Treatment of Cancer. Available at: http://www.eortc.be/ services/unit/mindact/documents/MINDACT_trial_outline.pdf. Accessed August 25, 2008.
37. National Cancer Institute, National Institutes of Health. TAILORx Breast Cancer Trial. National Cancer Institute. Available at: http://www.cancer.gov/clinicaltrials/digestpage/Tailorx. Accessed June 8, 2008.
38. Hornberger J, Cosler LE, Lyman GH. Economic analysis of targeting chemotherapy using a 21-gene RT-PCR assay in lymph-node-negative, estrogen-receptor-positive, early-stage breast cancer. *Am J Manag Care* 2005;11:313–324.
39. Lyman GH, Cosler LE, Kuderer NM, Hornberger J. Impact of a 21-gene RT-PCR assay on treatment decisions in early-stage breast cancer: an economic analysis based on prognostic and predictive validation studies. *Cancer* 2007;109:1011–1018.
40. Oestreicher N, Ramsey SD, Linden HM, et al. Gene expression profiling and breast cancer care: what are the potential benefits and policy implications? *Genet Med* 2005;7:380–389.
41. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology™ - Breast Cancer, v. 2.2008. Available at: http://www.nccn.org/ professionals/physician_gls/PDF/breast.pdf. Accessed August 10, 2008.
42. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 2007;25:5287–5312.
43. ECRI Institute, Gene Expression Profiling of Breast Cancer to Predict the Likelihood of Recurrence, August 07. ECRI Institute. Available at: www.ecri.org to subscribers only. Accessed August 26, 2008.