REVIEW ARTICLE

A Systematic Review of the Impact of Molecular Diagnostics on Treatment Decisions for Patients with Breast Cancer

Michael S. Broder, M.D., and Amy P. Sing, M.D.

ABSTRACT

BACKGROUND

To reduce overtreatment of breast cancer without compromising outcomes, physicians use risk stratification, which can include the use of genomic assays. The impact of these assays on treatment recommendations has not been systematically reviewed.

METHODS

English literature (2004–2012) was searched in PubMed to identify clinical utility studies for 9 available multigene breast cancer assays. Articles were reviewed to determine whether the study reported the proportion of patients for whom treatment changed after the assay, or the proportion of patients who were recommended chemotherapy both before and after the assay.

RESULTS

After abstract and selected full-text review, data were extracted from 20 (3 abstracts and 17 full length) of 1226 originally identified articles. Study methodology, size, and quality varied widely. Fourteen Oncotype DX studies reported that 25–74% (median 49%, mean 50%) of the patients were given a preassay recommendation for chemotherapy and 17 Oncotype DX studies reported that 13–53% (median 28%, mean 32%) were given a postassay chemotherapy recommendation. A single MammaPrint study reported preassay chemotherapy recommendation in 44% and postassay recommendation in 51%. Twelve Oncotype DX studies reported that the recommendation for chemotherapy changed (either from chemotherapy to no chemotherapy, or vice versa) in 19–45% (median 31%, mean 32%) of the patients. The 2 MammaPrint studies reported 11% and 29% change.

CONCLUSION

We found published evidence of clinical utility only for Oncotype DX and MammaPrint. There is substantial evidence from both prospective and retrospective studies that Oncotype DX changes treatment decisions in about one-third of the patients and reduces chemotherapy use by more than 20%. Three studies provide evidence that the 70-gene assay changes treatment recommendations, but no evidence for an overall reduction in chemotherapy.

Keywords: clinical utility; genomics; biomarkers; recurrence risk; chemotherapy

Partnership for Health Analytic Research, LLC, Beverly Hills, CA, USA (M.S.B.); Genomic Health, Redwood City, CA, USA (A.P.S.). Correspondence: Amy P. Sing, M.D., Genomic Health, 301 Penobscot Drive, Redwood City, CA 94063 (ASing@genomichealth.com).

Conception and design: MSB, APS.
Collection and assembly of data: MSB.
Data analysis and interpretation: MSB, APS.
Manuscript writing: MSB, APS.
Final approval of manuscript: MSB, APS.

Submitted March 8, 2013; revised March 23, 2013; accepted May 30, 2013

TJOP 2013;1:131–143
DOI: 10.13032/tjop.2052-5931.100042.

BACKGROUND

Improvements in diagnosis and treatment have led to a reduction in breast cancer mortality over the last decade.[1] Nonetheless, the National Cancer Institute estimates that among US women there were more than 226,000 new cases and 39,000 deaths from breast cancer in 2012.[2] For most early-stage breast cancers, treatment has been designed to address both local (primary tumor) and systemic disease (micrometastatic). This has likely resulted in overtreatment for many women at low risk for distant recurrence. In an attempt to reduce overtreatment without compromising the outcome, many physicians incorporate risk stratification into their treatment recommendations.[3]

Risk stratification can take the form of qualitative assessment of clinical and pathologic factors such as invasion, lymph node status, and individual biomarkers, as well as the use of electronic decision tools such as Adjuvant! Online.[4] In addition, the analysis of epigenomic, genomic, and transcriptional changes has led to the development and use of multigene assays to assess recurrence risk in a variety of cancers.[5] These assays use a variety of techniques, including immunohistochemistry, microarrays (DNA or RNA), and real-time polymerase chain reaction (RT-PCR), to assign discrete or continuous risk scores.

There are multiple genomic breast cancer assays currently available for use in the United States, and clinicians have incorporated them into routine practice, although many in the oncology community remain unsure of the methods by which to judge and differentiate molecular diagnostic tests. A recent National Comprehensive Cancer Network (NCCN) task force report designed to educate clinicians describes several types of evidence that must be available in order for a test to be useful in routine clinical practice. These include evidence of analytic validity (that the assay accurately and reliably measures the marker[s] of interest), clinical validity (that the assay result is meaningfully associated with the outcome of interest), and clinical utility (the ability of the assay to improve clinical decision making and patient outcomes).[6]

The demonstration of analytic validity is required for Clinical Laboratory Improvement Amendment (CLIA) certification, and evidence of clinical validity has become a standard part of the process by which multigene assays are evaluated, but the evaluation of clinical utility is much less standardized. To further inform clinicians, we undertook a systematic review to assess the studies measuring clinical utility for each of the currently available breast cancer multigene assays.

METHODS

A systematic review was conducted by the authors, both of whom have experience in molecular diagnosis and systematic review. DistillerSR (2012, Ontario, Canada), a web-based application specifically designed to conduct systematic reviews, was used. The search was designed to include only multigene assays available in the United States at the time of the review, either as a result of clearance for use by the Food & Drug Administration (FDA) or because they were conducted in a CLIA-approved laboratory. A search of PubMed using Medical Subject Headings for "breast neoplasm" and "tumor marker, biological" or "gene expression profiling" was conducted. Based on preliminary results, searches were restricted by the keywords "gene expression" or "diagnostic test" and to publication dates between 2004 and 2012, human subjects, adults, and English language.

Titles and abstracts were screened by a single reviewer (MB) for evidence that the study included a multigene assay for breast cancer that met the specified search criteria (FDA-cleared, CLIA-accredited, predictive of treatment benefit, etc). At the time the review was conducted, the assays identified were BreastOncPx (US Labs), MammaPrint (Agendia), Mammostrat (Clarient), Molecular Grade Index (bioTheranostics), Oncotype DX® Recurrence Score® (Genomic Health), EndoPredict (Sividon), PAM50 (ARUP and Nanostring), NuvoSelect (Nuvera Biosciences), and IHC4 (UK laboratory and Genoptix). Titles and abstracts were further screened for the inclusion of patients with invasive breast cancer and original data on the impact of studied test(s) on treatment recommendations. Articles not meeting the above criteria were excluded.

Full-text articles were reviewed to determine whether the study reported either (1) the proportion of patients for whom treatment changed after the assay was obtained (eg who were

initially recommended chemotherapy and subsequently no chemotherapy or who were initially recommended no chemotherapy and subsequently chemotherapy) or (2) the proportion of patients who were recommended (or received) chemotherapy both before and after the assay was obtained. Included studies could have collected data prospectively or retrospectively, and could have used a variety of comparisons, including between preassay recommendation and postassay treatment (or recommendation), comparisons between postassay treatment and guidelines, or theoretical comparisons (eg scenarios presented to experts or decision-analytic models). References from included studies were reviewed using a similar process, as were abstracts for the annual meeting of the *American Society of Clinical Oncology*, the Breast Cancer Symposium, and the San Antonio Breast Cancer Symposium for 2010–2012. Review articles, treatment guidelines, editorials, and letters were excluded from the analysis, but their references were reviewed for relevant studies.

RESULTS

We reviewed 1226 abstracts, of which 1099 were excluded at the initial screen, the vast majority because they did not report data on one of the identified multigene assays. Of the remaining 127 studies, 86 did not report on the impact of an included assay on treatment recommendations, 15 did not report on comparison of chemotherapy recommendations, and one did not involve invasive breast cancer, leaving 25 studies. Data were extracted from the remaining 25. Of these, 2 were duplicates of subsequent reports (eg an abstract and subsequent full-length article), 2 were review articles, and one was a patient survey, leaving 20 publications (3 abstracts and 17 full-length articles) (Fig. 1).

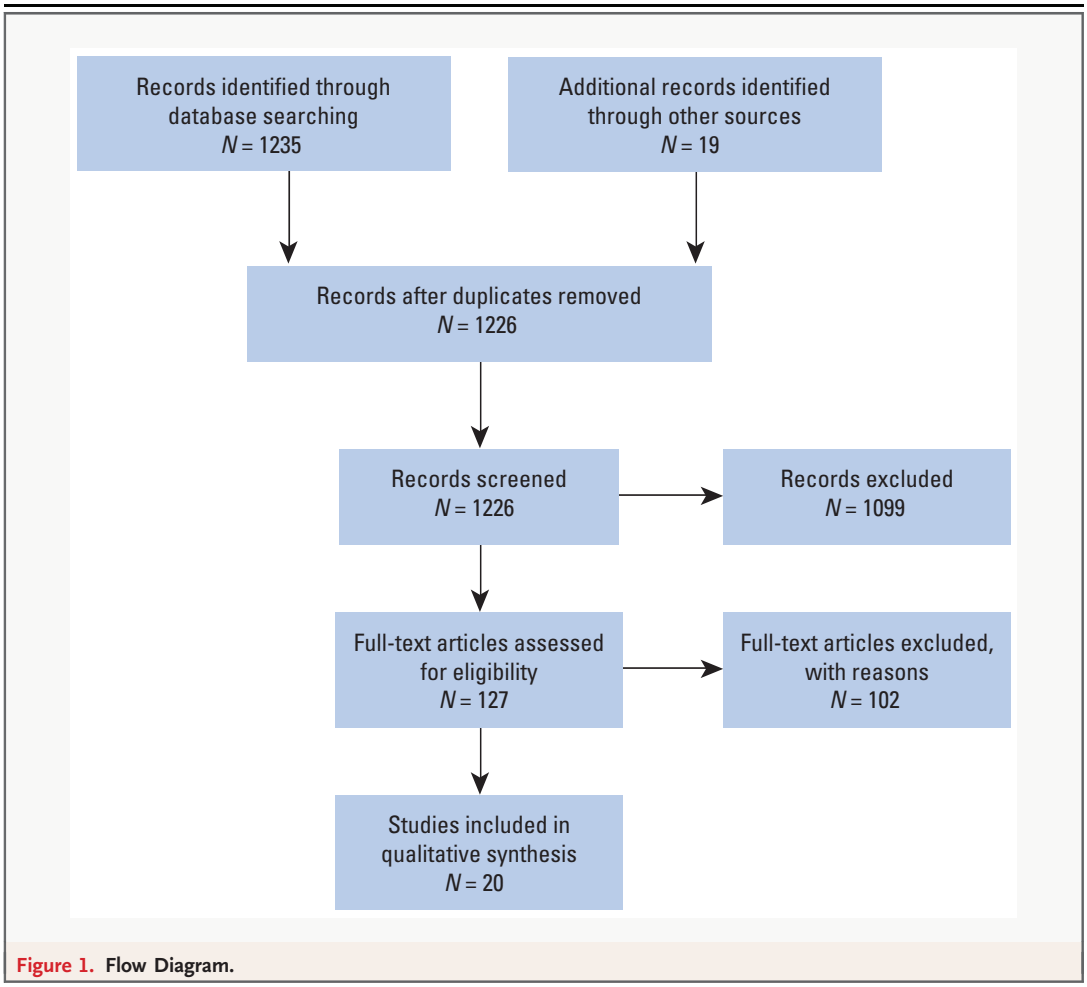

Figure 1. Flow Diagram.

Seventeen articles reported on the clinical utility of Oncotype DX, and 3 on MammaPrint. There were no clinical utility data on the other 7 assays. One article included patients tested with both assays; of its 1511 subjects, all but 6 used Oncotype DX.[7] It is therefore grouped as an Oncotype DX article in this review. Sixteen studies reported results from the United States (15 Oncotype DX, 1 MammaPrint), 3 Europe (1 Oncotype DX [Spain], 2 MammaPrint [1 Italy and 1 The Netherlands]), and one Israel (Oncotype DX). In all the studies, assays were performed between 2004 and 2010 (Table 1).

Table 1. Study Descriptions.

Study (year)	Setting	Data source	Years of assay use
Oncotype DX			
Prospective			
Albanell et al. (2012)[8]	Spain	Physician survey	2009–2010
Henry et al. (2009)[9,a]	USA	Medical records, physician survey	2004–2006
Lo et al. (2010)[10]	USA	Physician survey	2005–2006
Retrospective			
Ademuyiwa et al. (2011)[11]	USA	Medical records, physician survey	2005–2009
Asad et al. (2008)[12]	USA	Medical records	2006–2008
Erb et al. (2007)[13,b]	USA	Medical records	2005–2006
Gregg et al. (2009)[14,b]	USA	Medical records, registry	2005–2008
Haas et al. (2011)[15]	USA	Medical records, insurance claims	2006–2008
Hassett et al. (2012)[7]	USA	Registry	2006–2008
Joh et al. (2011)[16]	USA	Medical records, physician survey	2004–2009
Kamal et al. (2011)[17]	USA	Medical records, physician survey	pre-2006
Oratz et al. (2007)[18]	USA	Medical records	2004–2005
Rayhanabad et al. (2008)[19]	USA	Medical records	2006
Schneider et al. (2012)[20]	USA	Medical records, physician survey	2005–2010
Thanasoulis et al. (2008)[21,b]	USA	NR	pre-2008
Model			
Hornberger et al. (2011)[22]	USA	Insurance claims, published literature	2006–2010
Klang et al. (2010)[23]	Israel	Insurance claims, physician survey, published literature	pre-2010
MammaPrint			
Prospective			
Bueno-de-Mesquita et al. (2007)[24]	Netherlands	Study clinical registration form	2004–2006
Retrospective			
Bighin et al. (2010)[25]	Italy	Medical records	pre-2010
Model			
Chen et al. (2010)[26]	USA	Published literature, SEER registry	pre-2010

[a]Prospective collection of recommendations prior to Recurrence Score result. Postassay, medical record review to determine actual administration of adjuvant therapies; separately, an expert panel made adjuvant therapy recommendations after being presented with patient information without and with Recurrence Score results.
[b]Abstract only.
Note: NR, not reported; SEER, Surveillance Epidemiology and End Results.

STUDY DESIGNS

Four studies used a prospective design (3 Oncotype DX, 1 MammaPrint)[8–10,24]; none were randomized. There were 13 retrospective studies (12 Oncotype DX, 1 MammaPrint). Of the retrospective studies, 5 used medical records data only,[12,13,18,19,25] 4 a combination of medical records and physician survey,[11,16,17,20] one a combination of medical records and registry data,[14] one a combination of medical records and insurance claims,[15] one registry data,[7] and one did not identify the data source.[21] There were 3 decision models (2 Oncotype DX, 1 MammaPrint)[22,23,26] using a variety of data input sources (Table 1).

The primary outcome of interest in this review was clinical utility as assessed by change in treatment, which was measured in different ways. Six Oncotype DX studies[8–10,18,20,23] compared actual preassay treatment recommendations to actual postassay treatment or recommendations. In 3 of these studies, recommendations were prospectively collected[8–10] and in 3 retrospectively, either by medical record abstraction,[18] or by physician survey.[20,23] The remaining 14 studies used comparisons to guidelines, scenarios, models, and other techniques to assess utility. Specifically, 5 studies (3 Oncotype DX, 2 MammaPrint) compared recommendations incorporating multigene assay results to the recommendations that theoretically would have been made under guidelines (2 NCCN,[12,19] 1 Dutch guidelines[24]) or by "usual practice."[21,25] Four Oncotype DX studies compared recommendations for chemotherapy between patients who had the assay and historical or contemporaneous controls.[7,13–15] Three Oncotype DX studies involved showing patient scenarios to physicians, first without, then with, the assay results, and comparing the resulting recommendations.[9,16,17] Two studies (1 Oncotype DX, 1 MammaPrint) used model calculations based a range of data and assumptions to estimate the clinical impact of the assay.[22,26]

STUDY PATIENTS

Oncotype DX and MammaPrint have been validated for different subject populations. Oncotype DX is validated for newly diagnosed invasive breast cancer patients who are either (1) Stage I or II node-negative, estrogen-receptor-positive (ER+) or (2) postmenopausal, node-positive, hormone-receptor-positive.[27–29] MammaPrint was validated in a cohort of patients obtained from centers in the Netherlands and included a heterogeneous population of patients with Stage I or II node-negative invasive breast cancer with tumor size <5.0 cm and a mixture of hormone receptor status. MammaPrint was initially validated in women below the age of 61 who were not treated with hormonal therapy despite being ER+.[30] The Oncotype DX 21-gene assay was validated in NSABP B14 and B20, predating knowledge of the significance of human epidermal growth receptor (HER2) expression as a prognostic indicator.[27,28] Based on this, and the fact that the current treatment recommendation for HER2-positive disease is chemotherapy, Oncotype DX is generally not used in these patients. HER2 is not one of the 70 genes in the MammaPrint signature, and approximately 11% of the validation cohorts were HER2-positive.

In the studies of Oncotype DX, the median sample size (considering only patients on whom the assay was performed and whose results were evaluable) was 124 (range 29–1505); in the 2 MammaPrint studies, the sample size was 12 and 427. Two studies included fewer than 30 patients, 6 included between 31 and 100 patients, 7 between 101 and 300, and 4 between 301 and 1505. Subject ages were reported in a variety of ways (and unreported in 2 studies) with means ranging from 48 to 59 and medians from 49 to 61. All studies included patients with ER+ tumors and negative lymph nodes (N–). Two studies included patients regardless of estrogen receptor and node status, 3 included estrogen receptor negative (ER–) (but not node-positive [N+]) patients, and 2 N+ (but not ER–) patients. Of the 12 studies reporting human epidermal growth factor receptor 2 (HER2) status, 8 included patients with HER2+ and HER2– tumors while 4 included only HER2– tumors. Grade 1–3 tumors were considered in 14 studies and grade 1–2 in one study; grade was not reported in the remainder. Tumor stage was reported as 1–3 in 3 studies and 1–2 in 3 studies (Table 2).

CLINICAL UTILITY RESULTS

Utility was reported by providing the proportions of patients recommended chemotherapy before and after the assay, or by providing the proportion of patients with a change in chemotherapy recommendation from before the assay to after the assay. "Change in recommendation" included both a change to a recommendation *for* chemotherapy when none was initially offered

Table 2. Study Design and Patient Population.[a]

Study (Year)	Number tested[b] (Total study *N*)	Age (years) Mean	Median	Range	T stage	Tumor grade	ER +/−	N +/−	HER2 +/−
Oncotype DX									
Prospective									
Albanell et al. (2012)[8]	107 (107)	53.2[c]	NR	NR	T1, T2, T3	1, 2, 3	+	−	−
Henry et al. (2009)[9]	29 (139)	NR	51	31–74	NR	1, 2, 3	+	−	−
Lo et al. (2010)[10]	89 (93)	55	NR	35–77	NR	1, 2, 3	+	−	+, −
Retrospective									
Ademuyiwa et al. (2011)[11]	276 (276)	54.8	55	29–82	NR	1, 2, 3	+	−	−
Asad et al. (2008)[12]	85 (85)	54	NR	NR	NR	1, 2, 3	+	−	+, −
Erb et al. (2007)[13,d]	124 (1213)	NR	NR	NR	NR	NR	+, −	−	NR
Gregg et al. (2009)[14,d]	244 (973)	NR	NR	NR	T1, T2, T3	1, 2, 3	+	−	+, −
Haas et al. (2011)[15,d]	138 (534)	50.2[c]	NR	NR	NR	1, 2, 3	+, −	+, −	+, −
Hassett et al. (2012)[7,d]	1505 (7375)	56.9[c]	NR	NR	NR	1, 2, 3	+, −	+, −	+, −
		RS <18: 55.5	54	NR	NR	1, 2, 3	+	+, −	+, −
Joh et al. (2011)[16]	154 (154)	RS 18–30: 53.4	51						
		RS >30: 59.1	58.5						
Kamal et al. (2011)[17]	186 (186)	NR	53	42–82	NR	1, 2, 3	+	−	+, −
Oratz et al. (2007)[18]	68 (74)	NR	54	35–77	NR	1, 2, 3	+	−	NR
Rayhanabad et al. (2008)[19]	58 (1365)	54	NR	26—78	T1, T2	NR	+	−	NR
Schneider et al. (2012)[20]	89 (89)	57.4[c]	NR	NR	NR	1, 2, 3	+	−	NR
Thanasoulis et al. (2008)[21]	78 (78)	59	NR	33–82	NR	NR	+	−	NR
Model									
Hornberger et al. (2011)[22]	925 (925)	59	NR	NR	NR	NR	+	−	NR
Klang et al. (2010)[23]	313 (368)	57	57	29–81	NR	1, 2, 3	+	−	NR
MammaPrint									
Prospective									
Bueno-de-Mesquita et al. (2007)[24]	427 (812)	48	49	27–60	T1, T2, T3	1,2,3	+, −	−	+, −

Table 2 (*Continued*)

Study (Year)	Number tested[b] (Total study *N*)	Age (years) Mean	Median	Range	T stage	Tumor grade	ER +/−	N +/−	HER2 +/−
Retrospective									
Bighin et al. (2010)[25]	12 (21)	59	61	41–80	T1, T2	1,2	+	+, −	NR
Model									
Chen et al. (2010)[26]	NA	NR	NR	≤60	T1, T2	NR	+, −	−	−

[a]Characteristics of tested patients with evaluable results, unless otherwise noted.
[b]Number of patients tested who had evaluable results.
[c]Percentages of age groups were provided in the study. To estimate the mean age, we assumed the mean age in each age group is its midpoint and calculated a weighted average.
[d]Characteristics of all patients assessed in the study.
ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; N, node; NA, not applicable; NR, not reported; RS, Recurrence Score.

and a change to a recommendation *against* chemotherapy if it was initially offered. Fifteen studies (14 Onco*type* DX, 1 MammaPrint) reported the before and after proportions of patients receiving chemotherapy, 14 studies (12 Onco*type* DX, 2 MammaPrint) reported the proportion of patients with a change in recommendation, and 10 studies (all Onco*type* DX) reported both. The 14 Onco*type* DX studies reported 25–74% (median 49%, mean 50%) of patients were given a preassay recommendation for chemotherapy. The 17 Onco*type* DX studies that reported postassay recommendations reported 13–53% (median 28%, mean 32%) were given a chemotherapy recommendation postassay. The lone MammaPrint study reported preassay chemotherapy recommendation in 44%, postassay recommendation in 51% (and actual postassay receipt of chemotherapy in 47%). Of the studies reporting a change in the proportion recommended chemotherapy, the 12 Onco*type* DX studies reported 19–45% of patients (median 31%, mean 32%), compared to the 2 MammaPrint studies, which reported 11% and 29% change (Table 3).

UTILITY BY STUDY TYPE

The 3 prospective Onco*type* DX studies reported both the proportion recommended chemotherapy pre- and postassay and the proportion in whom the recommendation for chemotherapy changed after the assay. Albanell et al. included 107 evaluable patients,[8] with 36% recommended chemotherapy preassay and 27% postassay. Thirty-two percent had a recommendation change. In a study of 89 evaluable Onco*type* DX patients,[10] Lo reported 47% were recommended chemotherapy preassay and 26% postassay. In a study of 29 evaluable Onco*type* DX patients,[9] Henry reported chemotherapy recommendations were 45% preassay and 28% postassay. Both studies reported 31% of patients had a change in their recommendation for chemotherapy postassay. In the single prospective study of 427 evaluable MammaPrint patients,[24] Bueno-de-Mesquita reported 44% were recommended chemotherapy preassay and 47% received it postassay. Treatment change was not reported (Table 3).

The 3 prospective Onco*type* DX studies compared actual physician recommendations pre- and postassay, as did 3 retrospective Onco*type* DX studies[18,20,23] (none of the other MammaPrint studies had such a comparison). One of these studies[23] was published as a decision model, but the data on treatment recommendations were collected from the treating physician for preassay recommendations and from claims data for actual treatment. In this study, Klang reported 56% of patients with a preassay chemotherapy recommendation and 28% postassay. In the second study,[20] the rates were 61% and 26%, and in the third,[18] 49% and 49% (with postassay receipt of chemotherapy in 32%). These retrospective studies also reported the proportion of patients in whom recommendations changed postassay, and these were 21%,[18] 45%,[20] and 40%.[23]

All 3 studies using patient scenarios to assess both pre- and postassay recommendations reported the proportion with a change (all

Table 3. Clinical Utility Outcomes.

			Clinical utility	
Study (year)	**Pre-assay CTx (%)**	**Postassay CTx (%)**	**% Tx change**	**Notes**
Oncotype DX				
Prospective				
Albanell et al. (2012)[8]	36	27	32	Pre-compared to postassay recommendations
Henry et al. (2009)[9,a]	44.8	27.6[b]	31.0[b]	Pre-assay recommendation compared to postassay tx
	41.4	31.0	24.1	Scenario-based recommendations
Lo et al. (2010)[10]	47	25.8	31.5	Pre-compared to postassay recommendations
Retrospective				
Ademuyiwa et al. (2011)[11]	45.3	32[b]	38[b]	Pre-assay scenario-based recommendations compared to postassay tx
Asad et al. (2008)[12]	74	38	44	Guideline-based compared to postassay recommendations
Erb et al. (2007)[13]	55	25	NR	Historical controls who were not tested compared to postassay recommendations
Gregg et al. (2009)[14]	NR	27.6[b]	NR	Controls who were not tested compared to postassay tx
Haas et al. (2011)[15]	NR	50.0[b]	NR	Controls who were not tested compared to postassay tx
Hassett et al. (2012)[7]	54.7[b]	32.9[b]	NR	Controls who were not tested compared to postassay tx
Joh et al. (2011)[16]	43.7	35.1	24.9	Scenario-based recommendations
Kamal et al. (2011)[17]	25.3	22.6	18.8	Scenario-based recommendations
Oratz et al. (2007)[18]	48.5	32.3[b]	25[b]	Pre-assay recommendation compared to postassay tx
		48.5	20.6	Pre-compared to postassay recommendations
Rayhanabad et al. (2008)[19]	72	53.4[b]	26[b]	Guideline-based recommendations compared to postassay tx
Schneider et al. (2012)[20]	61	26[b]	45[b]	Pre-assay recommendations compared to postassay tx
Thanasoulis et al. (2008)[21]	48.7	12.8	41	''Usual practice''-based compared to postassay recommendations
Model				
Hornberger et al. (2011)[22]	NR	27[b]	26.8[b,c]	Model estimation based on data and assumptions
Klang et al. (2010)[23]	56	28[b]	40[b]	Pre-assay recommendation compared to postassay tx
MammaPrint				
Prospective				
Bueno-de-Mesquita et al. (2007)[24]	43.56	47.31[b]	NR	Guideline-based recommendations compared to postassay tx
		51.29	NR	Guideline-based compared to postassay recommendations

Table 3 (*Continued*)

			Clinical utility	
Study (year)	**Pre-assay CTx (%)**	**Postassay CTx (%)**	**% Tx change**	**Notes**
Retrospective				
Bighin et al. (2010)[25]	NR	NR	11	''Usual practice''-based compared to postassay recommendations
Model				
Chen et al. (2010)[26]	NR	NR	28.8	Decision model comparing management with Adjuvant! Online to management with assay

[a]Prospective collection of recommendations prior to Recurrence Score result. Postassay, medical record review to determine actual administration of adjuvant therapies; separately, an expert panel made adjuvant therapy recommendations after being presented with patient information without and with Recurrence Score results.
[b]Reflects actual treatment administered.
[c]Reflects only reduction in chemotherapy.
CTx, chemotherapy; NR, not reported; Tx, treatment.

Oncotype DX). Henry[9] used an expert panel of 5 breast oncologists to examine 29 cases in which Oncotype DX was used and reported 41% preassay and 31% postassay were recommended chemotherapy, with a change in recommendation in 24%. Joh[16] used a panel of 4 surgical oncologists, 4 medical oncologists, and 4 pathologists to examine 154 case histories and reported a preassay recommendation for chemotherapy in 44% and a postassay recommendation in 35%. Kamal[17] presented 31 case histories to 6 oncologists, who made treatment recommendations without and with knowledge of the Recurrence Score result. This study reported that 25% preassay and 23% postassay were recommended chemotherapy, with a change in recommendation in 19%. One additional study presented 276 case histories without Recurrence Score result to 2 medical oncologists and compared their preassay recommendations to the treatment actually received by the patient, reporting 45% and 32% were recommended chemotherapy pre- and postassay and 38% had a change in recommendation.[11]

DISCUSSION

Over the last decade and a half, a series of linked developments, including the sequencing of the human genome, the explosion of genome-wide association studies, and the development of molecular/genomic profiling technologies, led to a proliferation of research into patterns of gene expression in many cancers. Several important consequences of this research can be seen in our understanding and treatment of early-stage breast cancer. While surgical staging and histology remain the primary methods by which breast cancers are categorized, variation in specific gene expression (ie hormone receptor and HER2), measured using immunohistochemistry or fluorescence in situ hybridization, has become integral to classification and treatment decision making. Now, multigene assays using DNA/RNA microarrays or quantitative RT-PCR are being used with increasing frequency to further classify and risk-stratify these cancers.

The rapid introduction of these new genomic tests, with more than a dozen available or in development for breast cancer[4] may have outstripped the ability of clinicians to evaluate them, potentially hindering the uptake of useful tests, or encouraging the uptake of unproven ones. It may be unclear to physicians how to distinguish between tests that are prognostic and associated with an important outcome such as recurrence or survival, and those that are predictive of the impact of a given therapy. For early-invasive breast cancer, this distinction is crucial. A clinically useful assay must be able to segregate patients into those at low-enough recurrence risk that chemotherapy can be avoided and those in whom chemotherapy provides meaningful benefit.

Systems for grading the strength of evidence for clinical utility have been developed, but there is no generally accepted method. The NCCN Task Force Report on Evaluating the Clinical Utility of Tumor Markers in Oncology proposed

a system of grading[6] based on prior work by Hayes,[31] and a related system was proposed by Simon and Paik.[32] One accepted principle from these frameworks is that the strongest evidence of clinical utility is a high-quality study demonstrating that patients whose care was informed by the assay had longer survival (or fewer recurrences) than those whose care was not. None of the reviewed assays had such evidence, and all had limitations. It may not be feasible to conduct a large long-term study that determines the superiority of testing versus no testing by randomization of patients to test or no test. Ongoing prospective trials for Oncotype DX and MammaPrint are designed to examine different questions. The TAILORx study is designed to determine the effect of chemotherapy for patients with mid-range Oncotype DX Recurrence Score values. The MINDACT Study will examine the outcome of patients for whom the MammaPrint risk classification is discordant with risk as determined by clinical and pathologic features.[33] Thus, with the information that is available, it is reasonable to assess clinical utility on the strength of validation studies for the intended use of the assay[6] and the evidence of impact of these tests on clinical decision making.

Our systematic review identified 17 published studies on the impact of Oncotype DX on treatment decisions and 3 published studies of MammaPrint. We found no evidence for the clinical utility of any of the other approved multigene assays. Multiple study designs were employed in the 20 included studies, all subject to various limitations. Since none of the studies were randomized or prospective controlled trials (Category A or B[32]), all may have been subject to selection bias, with physicians ordering the assay only on patients for whom they were already contemplating altering the treatment plan. As a result, the assays' impact may be lessened if used less selectively. There were 3 prospective, observational studies of Oncotype DX and one of MammaPrint (Category C[32]). The Oncotype DX studies were consistent with one another, reporting a pre- to postassay reduction in chemotherapy recommendation from 36% to 27%,[8] 47% to 26%,[10] and 45% to 28%.[9] The MammaPrint study reported 44% preassay and 47% postassay.[24]

The remaining 16 studies were retrospective, observational trials and considered Category D evidence.[32] The patients examined in the Oncotype DX studies tended to be older, with 2 of the 3 MammaPrint studies restricted to patients under 60 years of age, consistent with its validation data, which covers this population. Although reporting was incomplete, the disease severity appeared similar across all studies. All studies included ER+/N− patients, and some included ER− and/or N+ patients. Fifteen of the 17 Oncotype DX studies were conducted in the United States, compared to one of the MammaPrint studies. One MammaPrint study compared postassay recommendations to "usual practice" in Italy, and another compared postassay recommendations to the 2004 Dutch guidelines which are no longer reflective of standard practice. In studies using hypothetical scenarios, physicians may be more likely to report an impact of the assay in a situation when actual care is not being affected. Studies comparing assay-driven results to guidelines implicitly assume that in the absence of the assay, care follows guidelines, yet there is some evidence refuting that assumption.[34]

Considering studies of any methodology, the use of Oncotype DX reduced chemotherapy recommendations from 25–74% of patients preassay to 13–53% postassay. The proportion with a change in recommendation ranged from 19% to 45%. Methodologic differences do not appear to explain the differences between studies. The chemotherapy change ranged from 24% to 32% in 3 prospective studies, and was 27% and 40% in 2 models. Studies using scenarios demonstrated treatment changes between 18.8 and 38%. The most important factor in the impact of the assay on chemotherapy recommendations appears to be whether the study was conducted before or during/after 2008, when NCCN guidelines incorporated use of Oncotype DX. Considering studies of assay use before 2008, the range of percent change in treatment recommendations was 18.8–31.5%, whereas in later studies, the range was 32–41%.

The use of MammaPrint resulted in an increase in chemotherapy recommendations from pre- to postassay from 44% and 51% and, in 2 additional studies, a change in recommendation of 11% and 29%. An earlier systematic review had similar results. Hornberger and colleagues searched for "clinical validity/utility, change in practice patterns, and economic implications" of risk-stratifiers in breast cancer (not limited to multigene assays). They identified a similar set of articles, with 10 of Oncotype DX, one of MammaPrint, and 4 of Adjuvant! Online.[4]

Applying the levels of evidence described by Simon,[32] they found Level I evidence for the Oncotype DX 21-gene Recurrence Score to predict chemotherapy response and Level III evidence for MammaPrint. The same group also published a meta-analysis of 7 studies and reported a 30% reduction in chemotherapy associated with the use of Oncotype DX.[35]

A recent NCCN panel conducted its own review in preparation for the development of its clinical practice guideline in breast cancer[3] and achieved consensus that Oncotype DX be used to estimate the benefit from chemotherapy in the context of other elements of risk stratification. The guidelines recognized the existence of other multigene assays, including MammaPrint, but felt only Oncotype DX had sufficient clinical validation and utility data to warrant inclusion. The MammaPrint test was developed using a largely untreated patient population that may not reflect the established standard of care. The Oncotype DX test was developed in a patient population receiving standard of care hormonal therapy which then allowed validation of the relative therapeutic benefit of adding chemotherapy.

A 2008 review by the US Agency for Healthcare Research and Quality (AHRQ)[36] came to a similar conclusion, finding justification for the use of Oncotype DX but not for MammaPrint. The AHRQ assessment included studies published through 2006 and did not include any of the MammaPrint studies included in the current review. As this article was being prepared for press, 5-year outcomes from the MicroarRAy PrognoSTics in Breast CancER (RASTER) study of MammaPrint became available online.[37] This report was not included in the full review as it was published after the prespecified time frame (2004–2012). It was conducted in Dutch community hospitals, using the 2004 Dutch guidelines as the underlying basis for decision making, modified by the data provided by the MammaPrint assay. The goal of our systematic review was to determine the extent to which multigene assay results affected treatment recommendations. The results of RASTER as they relate to chemotherapy recommendations were published by Bueno-de-Mesquita in 2007[24] and were included in our analysis. In that study, 44% of patients were recommended chemotherapy preassay and 51% postassay. A challenge in interpreting this result is the difference in the preassay basis for a chemotherapy recommendation. The 2004 Dutch guidelines are restrictive, recommending chemotherapy only if the absolute survival benefit was anticipated to be greater than 5% at 10 years. In this context, an increase in chemotherapy recommendations could represent an appropriate use of the assay.

The initial Oncotype DX validation data were published in 2004[27] and the test became commercially available that same year. The initial validation of the MammaPrint 70-gene signature was published in NEJM in 2002[38] and the test received FDA clearance in 2007 as a prognostic tool. Thus, it may be that with time, further evidence for the clinical utility of MammaPrint or other assays will emerge.

CONCLUSION

In a systematic review of the English language literature, we found published evidence of clinical utility only for the 21-gene and 70-gene assays. The methodologies, sample sizes, patient populations, and outcomes reported vary across studies, making aggregation of results within each assay difficult, and making direct comparisons between the assays impossible. The majority of studies were related to the Oncotype DX 21-gene assay, and there is substantial evidence from both prospective and retrospective studies that Oncotype DX changes treatment decisions in about one-third of patients, reduces chemotherapy use by more than 20%, and can predict treatment benefit. The impact on decision making was larger in studies done after NCCN incorporated the assay in its breast cancer guidelines. Three studies provide evidence that the MammaPrint 70-gene assay results in changed treatment recommendations, but as yet, no evidence that it leads to an overall reduction in chemotherapy, or that it predicts benefit from chemotherapy.

ACKNOWLEDGMENT

This study was funded by Genomic Health, Inc.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Employment: Michael Broder, Partnership for Health Analytic Research, LLC.
Employment: Amy Sing, Genomic Health, Inc.

REFERENCES

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Peto R, Davies C, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet.* 2012;379*(9814): 432–444.*

2. Breast Cancer Home Page – National Cancer Institute. http://www.cancer.gov/cancertopics/types/breast. Accessed February 26, 2013.

3. National Comprehensive Cancer Network. *NCCN Breast Cancer Guidelines 2012*. Fort Washington, PA: National Comprehensive Cancer Network; 2012.

4. Hornberger J, Alvarado MD, Rebecca C, Gutierrez HR, Yu TM, Gradishar WJ. Clinical validity/utility, change in practice patterns, and economic implications of risk stratifiers to predict outcomes for early-stage breast cancer: a systematic review. *J Natl Cancer Inst.* 2012;104 (14):1068–1079.

5. Griffith O, Gray J. Omic approaches to preventing or managing metastatic breast cancer. *Breast Cancer Res.* 2011;13(6):230.

6. Febbo PG, Ladanyi M, Aldape KD, et al. NCCN Task Force report: evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Canc Netw*. 2011;9 *(Suppl. 5):S1–S32; quiz S33.*

7. Hassett MJ, Silver SM, Hughes ME, et al. Adoption of gene expression profile testing and association with use of chemotherapy among women with breast cancer. *J Clin Oncol.* 2012;30(18):2218–2226.

8. Albanell J, González A, Ruiz-Borrego M, et al. Prospective transGEICAM study of the impact of the 21-gene Recurrence Score assay and traditional clinicopathological factors on adjuvant clinical decision making in women with estrogen receptor-positive (ER+) node-negative breast cancer. *Ann Oncol.* 2012;23 (3):625–631.

9. Henry LR, Stojadinovic A, Swain SM, Prindiville S, Cordes R, Soballe PW. The influence of a Gene expression profile on breast cancer decisions. *J Surg Oncol.* 2009;99(6):319–323.

10. Lo SS, Mumby PB, Norton J, et al. Prospective multicenter study of the impact of the 21-gene recurrence score assay on medical oncologist and patient adjuvant breast cancer treatment selection. *J Clin Oncol.* 2010;28(10):1671–1676.

11. Ademuyiwa FO, Miller A, O'Connor T, et al. The effects of oncotype DX recurrence scores on chemotherapy utilization in a multi-institutional breast cancer cohort. *Breast Cancer Res Treat.* 2011;126(3):797–802.

12. Asad J, Jacobson AF, Estabrook A, et al. Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? *AmJ Surg.* 2008;196(4):527–529.

13. Erb C, Fox K, Patel M, et al. Evaluation of practice patterns in the treatment of node-negative, hormonereceptor positive breast cancer patients with the use of the oncotype DX assay at the University of Pennsylvania. https://docs.google.com/viewer?a=v&q=cache:j_8GQzN_eXwJ:www.genomichealth.com/~/media/Files/Basic/Breast/Publications/Abstract_3082.ashx+&hl=en&gl=us&pid=bl&srcid=ADGEESgxUv7PG2BFGtJT4ShibpkQ6jsoT5KsW5W1QxoeJQC6igVHqyretGu_dnPFIMEw3j4TcBKDFEsp_BnYsPMaM2uBuyShtm8invjI6Vq1SnbQQJeO37wLe5riDpfb7RJAuCrLEgb2&sig=AHIEtbRJAMrcaVov3v4NSLJnMb7_CzOTdw. Accessed January 18, 2013.

14. Gregg X, Belnap T, Rowley B, Rees W. Experience with use of the oncotype DX gene assay test in a multicenter community-based healthcare system. http://cancerres.aacrjournals.org/cgi/content/abstract/69/24_MeetingAbstracts/6058. Accessed January 18, 2013.

15. Haas JS, Liang S-Y, Hassett MJ, Shiboski S, Elkin EB, Phillips KA. Gene expression profile testing for breast cancer and the use of chemotherapy, serious adverse effects, and costs of care. *Breast Cancer Res Treat.* 2011;130(2):619–626.

16. Joh JE, Esposito NN, Kiluk JV, et al. The effect of oncotype DX Recurrence Score on treatment recommendations for patients with estrogen receptor-positive early stage breast cancer and correlation with estimation of recurrence risk by breast cancer specialists. *Oncologist.* 2011;16(11):1520–1526.

17. Kamal AH, Loprinzi CL, Reynolds C, et al. Breast medical oncologists' use of standard prognostic factors to predict a 21-gene recurrence score. *Oncologist.* 2011;16(10):1359–1366.

18. Oratz R, Paul D, Cohn AL, Sedlacek SM. Impact of a commercial reference laboratory test Recurrence Score on decision making in early-stage breast cancer. *J Oncol Pract.* 2007;3(4):182–186.

19. Rayhanabad JA, Difronzo LA, Haigh PI, Romero L. Changing paradigms in breast cancer management: introducing molecular genetics into the treatment algorithm. *Am Surg.* 2008;74(10):887–890.

20. Schneider JG, Khalil DN. Why does Oncotype DX recurrence score reduce adjuvant chemotherapy use? *Breast Cancer Res Treat.* 2012;134(3):1125–1132.

21. Thanasoulis LC, Brown A, Frazier T. The role of oncotype DX TM assay on appropriate treatment for estrogen positive, lymph node negative invasive breast cancer. *American Society of Breast Surgeons Annual Meeting*. 2008; New York, NY [abstract].

22. Hornberger J, Chien R, Krebs K, Hochheiser L. US insurance program's experience with a multigene assay for early-stage breast cancer. *J Oncol Pract.* 2011;7(3S):e38s–e45s.

23. Klang SH, Hammerman A, Liebermann N, Efrat N, Doberne J, Hornberger J. Economic implications of 21-gene breast cancer risk assay from the perspective of an Israeli-managed healthcare organization. *Value in Health.* 2010;13(4):381–387.

24. Bueno-de-Mesquita JM, van Harten WH, Retel VP, et al. Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). *Lancet Oncol.* 2007;8 (12):1079–1087.

25. Bighin C, Del Mastro L, Canavese G, et al. Use in current clinical practice of 70-gene signature in early breast cancer. *Int J Cancer.* 2010;127(11):2736–2737.

26. Chen E, Tong KB, Malin JL. Cost-effectiveness of 70-gene MammaPrint signature in node-negative breast cancer. *Am J Manag Care.* 2010;16(12):e333–e342.

27. Paik S, Shak S, Tang G, Kim C, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004;351 (27):2817–2826.

28. 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(23):3726–3734.

29. Genomic Health, Inc. FAQs. http://www.oncotypedx.com/en-US/Breast/HealthcareProfessionalsInvasive/FAQ#2%20accessed%204/4/13. Accessed April 4, 2013.

30. Office of the Commissioner. *Consumer Updates – Test Determines Risk of Breast Cancer Returning*. http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm048477.htm. Accessed April 4, 2013.

31. Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst.* 1996;88 (20):1456–1466.

32. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101(21):1446–1452.

33. Feero WG, Guttmacher AE, McDermott U, Downing JR, Stratton MR. Genomics and the continuum of cancer care. *N Engl J Med.* 2011;364(4):340–350.

34. Malin JL. Results of the national initiative for cancer care quality: how can we improve the quality of cancer care in the United States? *J Clin Oncol.* 2006;24 (4):626–634.

35. Hornberger J, Chien R. Abstract P2-09-06: meta-analysis of the decision impact of the 21-gene breast cancer Recur-

rence Score in clinical practice. *Cancer Res*. 2010;70(S24): S210. [Abstract nr P2-09-06].

36. Marchionni L, Wilson RF, Marinopoulos SS, et al. *Impact of gene expression profiling tests on breast cancer outcomes*. Evidence Report/ Technology Assessment No. 160 (Prepared by The Johnd Hopkins University Evidence-base Practice Center under contract No. 290-02-0018). AHRQ Publication No. 08-E002. Rockville, MD: Agency for Healthcare Research and Quality; January 2008: 2–5. http://www.ncbi.nlm.nih.gov/books/NBK38448/. Accessed February 26, 2013.

37. Drukker CA, Bueno-de-Mesquita JM, Retèl VP, et al. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study [published online ahead of print January 31 2013]. *Int J Cancer*. 2013. http://doi.wiley.com/10.1002/ijc.28082. Accessed February 26, 2013.

38. Van de Vijver MJ, He YD, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 200219;347*(25):1999–2009*.