Washington State Health Care Authority

Gene expression profile testing of cancer tissue

Draft report – comment and response

February 16, 2018

Health Technology Assessment Program (HTA)

Washington State Health Care Authority
PO Box 42712
Olympia, WA 98504-2712
360-725-5126
www.hca.wa.gov/hta/
shtap@hca.wa.gov

Gene Expression Profile Testing of Cancer Tissue

Draft Report

Public Comment and Response

Provided by:

Center for Evidence-based Policy
Oregon Health & Science University

February 16, 2018

Responses to public comments on draft key questions

The Center for Evidence-based Policy is an independent vendor contracted to produce evidence assessment reports for the Washington Health Technology Assessment (HTA) program. For transparency, all comments received during the public comment period are included in this response document. Comments related to program decisions, process, or other matters not pertaining to the evidence report are acknowledged through inclusion only.

Public comments were received from these individuals and organizations:

- Jeffrey Evans, MD
- Johnathan Lancaster, MD, PhD, Chief Medical Officer, Myriad Genetic Laboratories regarding EndoPredict
- Johnathan Lancaster, MD, PhD, Chief Medical Officer, Myriad Genetic Laboratories regarding Prolaris

Specific responses pertaining to comments are shown in **Table 1.**

The full text of all public comments and included references and attachments follows the tables.

Table 1. Responses to comments on draft report for gene expression profile testing of cancer tissue

Comments	Response
Commenter: Jeffrey Evans, MD	
Specific comments:	
Urologists periodically utilize these advanced diagnostic tests when deciding treatment options for patients with adenocarcinoma of the prostate. The 2017 NCCN guidelines allow for consideration of tumor-based molecular assays in men with clinically localized disease. Retrospective case cohort studies have shown that molecular assays provide prognostic information independent of other NCCN risk groups. These include, but are not limited to, likelihood of death with conservative management and likelihood of biochemical progression after radical prostatectomy or radiation treatment.	Thank you for your comment. We have noted the NCCN guidelines in the report.
We are seeing an increasing trend toward active surveillance in low risk prostate cancer and these tests are particularly helpful in deciding if avoidance of therapy or intervention is warranted. Medicare has reviewed the literature and has made a positive coverage decision. The "Policy Context" on the HCA website gets it right stating "potential benefits of these types of tests include more appropriate treatment decisions and better patient outcomes, including avoiding....unnecessary treatments".	These are potential benefits, and confirmation of the clinical utility of these tests depends upon the availability of appropriate studies.
We can all agree that prospective studies are the most meaningful study design, but I caution your committee to consider the strong retrospective data available. Prospective data accumulation for prostate cancer is somewhat unrealistic given the long clinical course of the disease for most patients. If you base your determination on the lack of prospective studies it will penalize current patients with the disease. Please take this into consideration as you are making your decision for patients with prostate cancer in Washington.	The review did not exclude retrospective studies, as outlined in the Methods section of the report. The Albala et al., Dall-Era et al., and Eure et al. studies included for evaluation of clinical utility of genetic expression profile tests for prostate cancer were retrospective in nature. The report authors searched for studies on outcomes such as mortality and morbidities, which the commenter noted might not be available because of the usual long course of the disease. However, outcomes from use of these types of tests, such as impact on treatment decisions, could be more easily available for shorter-term studies.
Washington state urologists currently utilize this class of tests for select patients with prostate cancer and recommend they continue to be covered. We are not recommending any particular company's assay. Medicare patients already have access to these tests and we feel that the Medicaid beneficiaries deserve the same coverage.	Thank you for your comment regarding current use of these tests.

Comments	Response
Commenter: **Johnathan Lancaster, MD, PhD, Chief Medical Officer, Myriad Genetic Laboratories (EndoPredict)**	
Specific comments:	
EndoPredict is a 2nd generation breast cancer gene expression assay that was introduced in the United States in 2017 as an alternative to other breast cancer assays on the market. The intended use population for EndoPredict is the same as that for Oncotype DX, that is, patients with ER-positive, HER2-negative early breast cancer, with negative or positive lymph nodes. The test is used by physicians in the same manner as Oncotype DX, to determine whether the predicted 10-year distant recurrence risk warrants the use of adjuvant chemotherapy. Oncotype DX classifies patients into low, moderate and high risk groups; EndoPredict classifies patients into low and high risk groups. Published studies have shown EndoPredict to be at least as effective as Oncotype DX. (See below.) As an alternative test with demonstrated equivalency, we believe that EndoPredict should be supported for clinical use if the data on Oncotype DX are determined to support its clinical use.	Thank you for your comment. The purpose of the WA HTA report was to evaluate the strength of evidence on clinical utility outcomes for each test, and not to determine whether any test was "equivalent" to another test on clinical validity. As noted in the draft report (Appendix E, Table 25), the strength of evidence for the outcome of patient management decisions was moderate for Oncotype DX and very low for EndoPredict. There is very low evidence for Oncotype DX for other clinical outcomes, and there were no eligible studies for these outcomes for EndoPredict.
The publication by Buus et al. in 2016[1], which was excluded from the Committee's review by a limitation of the scope, included a comparative study of gene expressions tests comparing EndoPredict ("EPclin") and Oncotype DX ("RS"). The study included 928 patients from the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial and reported the following: • EndoPredict provided better separation of low and high risk breast cancer, with a hazard ratio of 5.99 for EPclin compared to 2.73 for RS (low vs non-low risk). • EndoPredict showed good prognostic strength with a 10 year distant recurrence (DR) rate of 5.8% in the low risk group (compared to 10.1% for Oncotype DX low risk group) and 28.8% in the high risk group. • The prognostic ability of EndoPredict was strong for node negative disease (5.9% DR for low risk vs. 20.0% DR for high risk) and even stronger for node positive disease (5.0% vs. 36.9%). In comparison, the Oncotype DX observed DR rate in low risk node positive patients was 25%. • When compared to Oncotype DX, EndoPredict provided a classification more closely aligned with patient outcomes. This was observed with patients who were classified differently by the two tests: ○ Of the 13% of patients who were classified as low risk by EPclin and non-low risk (intermediate or high risk) by RS, the observed rate of DR was 10.2%, i.e., more closely aligned with low risk. ○ Of the 15% of patients who were classified as high risk by EPclin and low risk by RS, the observed rate of DR was 26.9%, i.e. EndoPredict correctly identified these patients as high risk. • The authors concluded that the superior performance of EndoPredict was partly due to the inclusion of clinical variables (nodal status and	Thank you for your comment. The study by Buus et al. was located in the search and was excluded because it did not report on any clinical utility outcomes and was therefore not within the scope of the report. As the commenter notes, both the Oncotype DX and EndoPredict tests are prognostic in that they predict clinical outcomes of interest. This type of clinical validity information is included in the Background section of the draft report. However, the scope of this report was to evaluate whether the use of these types of tests changes outcomes, and the study by Buus and colleagues did not provide evidence of clinical utility.

Comments	Response
tumor size) in the EPclin score, but also due to a molecular signature with improved prediction of late events (in years 5-10).	
In 2016, Sestak et al. reported the first comprehensive comparison of the performance of six prognostic signatures, including four gene expression assays, in estrogen receptor positive breast cancers from the TransATAC cohort.[17] Publication of the peer-reviewed manuscript is anticipated to occur on February 15, 2018, in JAMA Oncology. The authors found that EndoPredict provided greater prognostic value than Oncotype DX.	Thank you for alerting us about this upcoming publication. The Sestak et al. reference is for a conference abstract, which is a type of document that is excluded from our search because these types of materials do not provide sufficient detail for a full evaluation (see the Methods section of the report for details). The article by Sestak et al. was published on February 15, 2018, after the public comment period on this report had ended. This article would be excluded from the evidence review for the same reasons as the Buus et al. study was excluded. Although it might have provided some assessment of the clinical validity of these tests, it did not provide any evidence for clinical utility outcomes.
The systematic review by Blok et al.[2] is described in the draft report. As stated in the draft, a reduction in invasive treatments and a corresponding increase in less invasive treatments was observed with use of all of the reviewed tests, including EndoPredict. This confirms equivalency in terms of impact on treatment decisions.	Thank you for your comment. As noted in the draft report (Appendix E, Table 25), the strength of evidence for the outcome of patient management decisions was moderate for Oncotype DX and very low for EndoPredict. There is very low evidence for Oncotype DX for other clinical outcomes, and there were no eligible studies for these outcomes for EndoPredict.
Simon et al. published a well-accepted framework for establishing clinical utility of tumor biomarkers.[3] The authors acknowledge that, while the gold standard is a prospective randomized clinical trial, an alternative framework is needed when logistical and cost barriers exist for such trials, particularly for prognostic biomarkers. According to the Simon framework, LOE IB confirms clinical utility and is achieved when at least two independent clinical validation studies performed on archived samples from previously performed prospective trials ("prospective-retrospective" studies) demonstrate consistent results. The ability of EndoPredict to separate ER-positive, HER2-negative early breast cancer into low and high risk for distant recurrence has been demonstrated in three separate cohorts (independent of the training cohort): ABCSG-6, ABCSG-8 and ATAC.[4,1] Thus, EndoPredict is confirmed to have clinical utility based on LOE IB, according to the Simon framework. The evidence-based reviews of the American Society of Clinical Oncology (ASCO),5 Blue Cross Blue Shield Evidence Street6 and the European Group on Tumor Markers[7] concluded that EndoPredict and Oncotype DX have both achieved similar LOE of IB and are	Thank you for your comment. This methodological article by Simon et al. was submitted during public comment on the Key Questions for this report. The report authors took it under advisement and allowed inclusion of retrospective studies that demonstrated clinical utility as detailed in the Methods section of the report. However, the report did not employ the Simon framework for grading evidence, but instead used the GRADE system, which is the international standard. The report noted that both the TAILORx and MINDACT studies will have forthcoming additional publications.

Comments	Response
therefore demonstrated to have clinical utility. Note that the TAILORx and MINDACT trials are prospective trials involving Oncotype DX and MammaPrint respectively, but only subsets of data have been reported so far, and only with 5-year outcomes as opposed to 10-year outcomes.[8,9] Therefore, LOE IB is the highest LOE achieved by any of the breast cancer assays to date.	We have reported only the outcomes and assessed the risk of bias of the currently available articles for both of these studies.
As stated in the draft report, EndoPredict is recommended by the American Society of Clinical Oncology (ASCO) to guide decisions on the need for adjuvant systemic chemotherapy.[5] The draft rates the guideline published by this group as having "good methodological quality". Although the Blue Cross Blue Shield HTA program was cited as a data source, the draft report does not mention the report from this group. Importantly, the thorough, evidence-based review by Blue Cross Blue Shield Evidence Street concluded in 2016 and reaffirmed in 2017 that EndoPredict has sufficient evidence "to determine that the technology results in a meaningful improvement in net health outcome."[6]	Thank you for your comment. As clarification, the methodological quality of a guideline or systematic review does not indicate that the underlying evidence has the same methodological quality. ASCO does recommend EndoPredict for LN-negative tumors; Oncotype DX, Prosigna, and BCI are also recommended for use in LN-negative breast cancer; MammaPrint is recommended for both LN-negative and LN-positive tumors; and the Mammostrat test is not recommended for use. The ASCO guideline also rates the evidence quality and strength of recommendations for each of these recommendations. EndoPredict and the BCI test recommendations for use among LN-negative patients are both rated as having intermediate quality of evidence and a moderate strength of recommendation, and Oncotype DX and Prosigna are rated as having high-quality evidence and a strong rating of recommendation strength. The source listed in the Methods section referred to the Blue Cross/Blue Shield Technology Evaluation Center, and we did not locate this technology assessment because reports produced by this program have changed recently from being publicly available to being proprietary. The disclaimer on the BCBS Association Evidence Street report states that "USE SUBJECT TO SUBSCRIPTION AGREEMENT AND TERMS OF USE Evidence Street is a proprietary, subscription-based web platform . . . " Sources included in this report were required to be publicly available for review.
Palmetto GBA's MolDX program reviewed EndoPredict and determined that the evidence was sufficient for coverage by the Centers for	As stated in the report, there is no CMS NCD for these tests. The Noridian LCD

Comments	Response
Medicare & Medicaid Services (CMS). Noridian has now published a Local Coverage Determination (LCD) for EndoPredict, effective 1/30/2018.[10]	was updated after this draft report was submitted, and the final report will be updated with this newer information.
Regarding the guidelines of the National Comprehensive Cancer Network (NCCN),[11] the draft report correctly quotes this guideline as stating that "other gene expression profile tests [besides Oncotype DX] can be considered to assess risk of cancer recurrence, but that they have not been validated to predict response to chemotherapy." The ability to assess risk of cancer recurrence versus predict response to chemotherapy is an important distinction. Physicians typically use the risk of cancer recurrence to determine the risk/benefit ratio for adjuvant chemotherapy in the treatment of early breast cancer. All of the breast cancer assays were developed to predict risk of cancer recurrence (that is, they are prognostic) and the claim that Oncotype DX is also able to predict response to chemotherapy (predictive) is based on subset analyses of older data, which have been called into question.[12,13]	Thank you for your comment. The Michiels and Ioannidis articles are commentaries about what might be required to demonstrate the clinical predictive ability of a test. Both refer to the Paik et al. and Albain et al. studies, as does the commenter later in the comment. Although they were both identified in the search, neither the Paik et al. nor the Albain et al. studies were included in the report because of lack of data on clinical utility (Albain et al.) and not meeting the date cutoff (Paik et al.). Neither were included in the clinical utility section of the Blok et al. systematic review. Any assessment we made of the Oncotype DX test did not rely on either of these studies for information. Although genetic expression profile tests are initially developed with studies that are designed to demonstrate clinical validity (prognostic ability), to ensure the actual usefulness of the test, clinical utility (predictive ability) must also be demonstrated.
The ASCO expert panel that reviewed the literature on breast cancer biomarkers concluded that data are inadequate to support a chemotherapy predictive claim, stating the following: "However, the B20 data [14] are confounded by the data set originally used to generate the 21-gene RS algorithm. The results from SWOG S8814 [15] must be considered hypothesis generating because the number of samples analyzed in each RS subgroup was small, there was no additional prediction beyond 5 years, and the risks of systemic recurrence continues to be high for patients with node-positive disease."[5] In any case, physicians using Oncotype DX usually base treatment decisions on its prognostic function, not the claimed predictive function. This is evidenced by the fact that patients reported to be Oncotype DX "intermediate risk" are frequently recommended for chemotherapy,[16] despite the publications used to support the predictive claim suggesting that intermediate risk patients do not benefit from chemotherapy.[14,15] In summary, gene expression assays were developed and are utilized for their prognostic ability, that is, their ability to predict distant recurrence, rather than predict response to a particular treatment.	Various guideline groups have reached different conclusions regarding tests in part because of whether the group adopts a stance of requiring demonstration of clinical utility. The Washington HTA was interested in whether each test demonstrated clinical utility.

Comments	Response
For clarity, we respectfully request replacement of the words "not recommended" for EndoPredict under NCCN in the draft's Table ES-2 with "can be considered." This would more accurately reflect what is stated in the NCCN guideline. We further request removal of the following statement in the draft's conclusion on pages 26-27: "However, NCCN has adopted a more rigorous standard and advises the exclusive use of Oncotype DX for both LN-negative and LN-positive patients" and replacement with the following: "NCCN has incorporated Oncotype DX into the treatment algorithm with a footnote stating that 'other prognostic multigene assays may be considered to help assess risk of recurrence but have not been validated to predict response to chemotherapy.'" We respectfully suggest that this would more fairly represent what is stated by NCCN.	Thank you for your comment. In the draft report, the footnote on Table ES-2 and Table 4 has been changed to "NCCN guidelines state that prognostic multigene assays other than Oncotype DX may be considered to help assess risk of recurrence but have not been validated to predict response to chemotherapy."
Currently, there is broad coverage for EndoPredict across the United States with more than 90% of breast cancer patients having access through their insurance. (Myriad, internal data.)	Thank you for your comment.
The Washington Medicare LCD covering EndoPredict for node negative and node positive patients, effective 1/30/18, was not included in the draft report.[10] In addition to coverage by Aetna and Regence, as reported in the draft, EndoPredict has positive coverage policies by numerous other commercial payers, including Premera and the majority of Blues plans, United Healthcare, Anthem and Humana, among others. Washington Medicaid currently covers EndoPredict, and other state Medicaid programs that cover include, but are not limited to, AK, AZ, CA, CT, DE, MD, MN, NH, NJ, NM, OH, OK, PA and WV.	Thank you for your comment. The Medicare LCD covering EndoPredict effective 1/30/18 was not published in time to be included in the draft report, but it has been included in the final report. Per the predetermined methods requested by the Washington HTA program, Medicare policies applying only to Washington and private payer policies for only Aetna, Cigna, and Regence were included in the report. We noted in the section on payer policies that Aetna covers EndoPredict, in addition to Oncotype DX, MammaPrint, Prosigna, and the BCI test for eligible patients who meet the clinical criteria (e.g., EndoPredict is covered only for patients with LN-negative tumors). Although the Cigna policy allows coverage for Oncotype DX, MammaPrint, and Prosigna under particular clinical conditions, EndoPredict is not covered. Regence covers Oncotype DX and BCI in addition to EndoPredict for patients with LN-negative cancers. Medicaid coverage was not included according to the predetermined methods requested by the Washington HTA program for this report.
The EndoPredict rate (CPT code: 81599-QP) has been set at the same rate as Oncotype DX (CPT code 81519) for Washington State Medicaid. This ensures budget neutrality for the state since providers will utilize	Thank you for information about current Washington Medicaid pricing for EndoPredict and Oncotype DX tests.

Comments	Response
one test or the other for the same intended use population. As stated in the draft, the systematic review by Blok et al.[2] reported a reduction in invasive treatments and a corresponding increase in less invasive treatments with use of all of the reviewed tests, including EndoPredict. The reduction in invasive treatments is expected to result in cost savings for the state.	However, the budget impact for each or both of these tests requires additional factors such as test uptake and will ultimately need to be based on empirical data for both Medicaid and other populations covered by the Washington HTA Clinical Committee's decisions.
References: 1. Buus R, Sestak I, Kronenwett R, et al. Comparison of EndoPredict and EPclin With Oncotype DX Recurrence Score for Prediction of Risk of Distant Recurrence After Endocrine Therapy. *J Natl Cancer Inst*. 2016 Jul 10;108(11). pii: djw149. doi: 10.1093/jnci/djw149. Print 2016 Nov. 2. Blok EJ, Bastiaannet E, van den Hout WB, et al. Systematic review of the clinical and economic value of gene expression profiles for invasive early breast cancer available in Europe. *Cancer Treat Rev*. 2018 Jan;62:74-90. 3. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101:1446–52. 4. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res.* 2011;17(18):6012–6020. 5. Harris LN, Ismaila N, McShane LM, et al; American Society of Clinical Oncology. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol.* 2016 Apr 1;34(10):1134-50. 6. Blue Cross Blue Shield Association. Assays of Genetic Expression in Tumor Tissue as a Technique to Determine Prognosis in Patients With Breast Cancer. December 2016 and December 2017. 7. Duffy MJ, Harbeck N, Nap M, et al. Clinical use of biomarkers in breast cancer: updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer*. 2017;75:284-298. 8. Sparano JA, Gray RJ, Makower DR, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med.* 2015;373;2005-14. 9. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med.* 2016;375(8):717-729. 10. MolDX: EndoPredict® Breast Cancer Gene Expression Test (L37311). Noridian Healthcare Solutions, LLC. Effective 1/30/2018. 11. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: breast cancer-version 3.2017. 2017; https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf.	Thank you for providing references for your comments.

Comments	Response
12. Michiels S, Ternès N, Rotolo F. Statistical controversies in clinical research: prognostic gene signatures are not (yet) useful in clinical practice. *Ann Oncol.* 2016 Dec;27(12):2160-2167. 13. Ioannidis JP. Is molecular profiling ready for use in clinical decision making? *Oncologist.* 2007 Mar;12(3):301-11. 14. 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 Aug 10;24(23):3726-34. 15. Albain KS, Barlow WE, Shak S, et al. Breast Cancer Intergroup of North America. Prognostic and predictive value of the 21-gene	
Commenter: **Johnathan Lancaster, MD, PhD, Chief Medical Officer, Myriad Genetic Laboratories (Prolaris)**	
Specific comments:	
New biomarker tests can achieve the "gold standard" LOE I for clinical utility by demonstrating improved outcomes through prospective randomized controlled trials; however, challenges such as variable medical care, small treatment effects, and long outcome timeframes can present obstacles to timely validation of prognostic tumor biomarkers for clinical utility. The lack of prospective trials and an alternative evaluation framework result in limited patient access to testing that can inform medical management, improve outcomes, and reduce costs. The American Medical Association (AMA) recently presented new policy regarding genetic and genomic testing, encouraging transparent coverage and payment policies "that are evidence-based and take into account the unique challenges of traditional evidence development through RCTs, and work with test developers and appropriate clinical experts to establish clear thresholds for acceptable evidence."[2] Furthermore, the NCCN guideline for prostate cancer treatment states, with regard to prognostic biomarkers, that prospective randomized clinical trials are "unlikely to be done," and that "men with clinically localized disease may consider the use of tumor-based molecular assays at this time."[3] Addressing the challenges of evaluating prognostic tumor biomarkers, Simon et al. presented an alternative framework for using archived tumor specimens to establish LOE I or II, with studies that meet LOE II serving as adequate evidence of clinical utility in "particularly compelling circumstances."[1] Validation studies using archived specimens from previous prospective studies with known outcomes can serve to confirm clinical utility, provided that the studies meet certain requirements and that the study cohort represents a defined medical indication for use of the particular biomarker. Numerous published validation studies (not included in the current report due to a limitation of the scope, but some included in the review by Sommariva et al.[4]) have shown that Prolaris adds new information to standard clinico-pathologic parameters and reliably predicts long term oncologic outcomes.[5-11] Notably, two of the studies included conservatively managed cohorts,[5,6] supporting Prolaris' ability to predict prostate	The methodological article by Simon et al. was submitted during public comment on the Key Questions for this report. The WA HTA report inclusion criteria allowed retrospective studies. Although the commenter notes that prospective RCTs are unlikely to be done, and indeed our search did not identify any, we identified eight studies for inclusion related to Key Question 1, two of which pertained to the Prolaris test and reported treatment recommendation outcomes. Both of these studies were before-after studies and both were judged to be at high risk of bias. The commenter refers to the prognostic ability of the test, which is a clinical validity outcome. The review in the WA HTA report was performed to evaluate whether there was evidence of clinical utility.

Comments	Response
cancer death when interventional therapy is not immediately selected, that is, the exact intended use for the test. The studies show consistent and statistically significant results across different populations, treatments and endpoints, thereby reducing play of chance with the results. Each study utilized archived specimens from prospective observational registries that satisfy the requirements of Simon Category C studies: 1) adequate amounts of archived tissue were available from enough patients for analyses to have adequate statistical power and for the patients included in the evaluation to be clearly representative of the patients in the trial; 2) the test was analytically and preanalytically validated for use with archived tissue; 3) the plan for biomarker evaluation was completely specified in writing before the performance of the assays on archived tissue and focused on evaluation of a single completely defined classifier; and 4) the results were validated using specimens from more than one similar, but separate, studies. Thus, Prolaris can be placed firmly within LOE II according to the Simon framework. We suggest that the significant overtreatment problem resulting from a lack of confidence in the ability of clinico-pathologic features to risk stratify adequately, combined with the long natural history of prostate cancer precluding prospective studies, present "compelling circumstances" to consider LOE II as practice changing for prostate cancer.	
Despite recent trends towards increased utilization of active surveillance, in the U.S. more than 80% of men with low risk prostate cancer, and more than 95% of those with intermediate risk prostate cancer undergo active treatment.[12] As discussed in the draft report, two published studies confirmed that use of Prolaris was associated with decreased active treatment.[13,14] Of note, these studies did not use historical controls, as stated in the draft report. Due to known variability in the standard of care and the inability of clinico-pathologic parameters to risk stratify and match cases and controls, each patient was used as his own control, by assessing planned treatment without the Prolaris result compared to actual treatment with the Prolaris result. Actual treatments were assessed, contrary to the conclusion in the draft report: In the study by Crawford, the stated treatment by the treating physician in the survey was compared to the actual treatment received by an independent chart audit performed a minimum of 45 days after the post-test survey response.[13] In 19.8%, there was a discrepancy between the post-test recommended treatment and the actual treatment administered, with one-third of those representing a change from "watchful waiting" to "active surveillance," one-third representing a change from active surveillance to an interventional therapy, and one-third representing a change from radiation to either active surveillance, prostatectomy or cryoablation.	The WA HTA draft report, both in the narrative on page 62, as well as Table 9, noted that the Crawford et al. study reported both recommended treatments for 305 subjects and actual treatments received for 113 of tested subjects (38% of the original group) based on chart review. In addition to the study's design as a before-after study with comparison to a historical group of patients who had not received testing, the very low follow-up for actual treatment received raises concerns regarding bias. Other features of the study that are concerning for bias are listed in Table 9 and Table 18.
In the study by Shore, the final measure was the actual treatment received a minimum of three months after the test result was reported.[14] Therefore, both studies assessed actual treatment administered, and showed furthermore that the actual treatment pursued was directionally aligned with the Prolaris result. The results demonstrate clearly that Prolaris influenced the treatment. With regard	As the commenter notes, and as is detailed in Table 9 of the draft report, the Shore et al. study enrolled 1,206 patients (from a registry containing 1,596 patients) of 124 treating physicians in multiple U.S. states. The

Comments	Response
to the criticism of bias in these studies, these represent real world use of Prolaris by numerous physicians in multiple states across the U.S. (31 states in the Crawford study; 21 states in the Shore study) who were familiar with the test and the sample collection process. The publication by Shore states that physicians were encouraged to enroll consecutive patients that were newly diagnosed (6 months or less), had clinically localized prostate cancer and had not received any treatment. Therefore, results are expected to represent the real world use of Prolaris and its impact on management decisions. The incorporation of the patient decision points in the study by Shore also reveals how patients in the real world are likely to respond to their physicians' treatment recommendations based on the Prolaris result.[14]	commenter does not respond to the veracity of the report's designation of this study as being rated at a high risk of bias except to say that it represents a "real world" examination of the test. However, the study sites are not specified in the Shore et al. article, and no demographic or socioeconomic description of the study population aside from age (mean 65.9 years) and racial/ethnic category (77% Caucasian) was given. It is therefore difficult to examine this claim fully. However, the biases noted included study funding by the test manufacturer, multiple study authors with financial relationships to the test manufacturer, and selection of a group of physicians specifically because of their prior experience with the test. This was a before-after study with patients acting as their own "controls." These biases are serious, and the our rating of the Shore et al. study remains the same, having a high risk of bias.
The outcome of Prolaris on prostate cancer mortality, morbidity and quality of life can be inferred from a chain of evidence, which links separate pieces of published evidence to prove clinical utility. This is an acceptable approach when a randomized controlled trial is not possible, according to the Effectiveness Guidance Document published by the Center for Medical Technology Policy (CMTP).[15,16] As discussed, two prospective clinical utility studies demonstrated a net reduction in prostatectomy and radiation therapy and an increase in active surveillance when Prolaris is used.[13,14] Prostatectomy and radiation have been associated with treatment-related morbidities and a reduction in quality of life[17-20] without providing a mortality benefit for low risk prostate cancer compared to active surveillance.[21-23] Since Prolaris results in fewer men receiving an invasive treatment, fewer men will experience the treatment related morbidity, resulting in a net health benefit. Therefore, a chain of evidence allows us to conclude that Prolaris can reduce treatment-related morbidities and their negative impact on quality of life, while achieving at least the same mortality outcomes.	Thank you for your comment. As described earlier in response to comments received on the EndoPredict test, in the absence of clinical endpoint data (e.g., mortality, morbidity, quality of life), it is not at all certain that a change in clinical decision making is actually improving care.
Clinical validation studies consistently demonstrate that the prognostic information provided by Prolaris is superior to that provided by standard clinico-pathologic parameters.[5-11] Therefore, treatment decisions that incorporate Prolaris results are expected to be more suitable for the individual patient and consequently safer than standard care without Prolaris. Adverse events related to over and under treatment are known to occur in the current setting of prostate cancer	Thank you for your comment. References 5 through 11 in the comment refer to clinical validation (prognostic ability) studies and do not address issues of clinical utility (predictive ability). The assertion that treatment decisions based on Prolaris

Comments	Response
treatment in the absence of Prolaris use, in which treatment decisions are being made with less-precise prognostic information. There are no data to suggest that adverse events are more common in the setting of Prolaris use. Since Medicare began covering Prolaris testing in 2015, a Medicare-mandated registry of tested patients has been maintained, to monitor for adverse events in patients reported to be low risk according to Prolaris. To date, no adverse events (prostate cancer death or metastasis) have been reported. (Myriad internal data.)	are more suitable for individual patients cannot be demonstrated with the studies cited, and no other evidence was provided to justify this assertion. Whether or not a test performs better or worse than various clinical and pathological factors is not the subject of the WA HTA report. We cannot independently verify the assertion that adverse events have not been reported within the Medicare-mandated Prolaris registry.
We respectfully suggest that the economic study published by Health Quality Ontario would have little relevance to the economics of healthcare in the U.S., given the different methods of funding in the two countries. As described in that document, treatment of prostate cancer is more conservative in Canada. Therefore, we would expect that Prolaris would have a more favorable economic impact in the U.S.	Thank you for your comment. We discuss the limitations of available studies to answer this question, including that this economic study was conducted to inform policy decisions in a Canadian province rather than in the U.S.
Because the use of Prolaris has been demonstrated to reduce overtreatment of prostate cancer, there are immediate cost savings that occur when Prolaris is utilized. Reductions in RP and RT are in the range of 30-50%.[13-14] An independent health economic model demonstrated that Prolaris can reduce costs by up to $2,850 per patient tested over 10 years, after accounting for test cost and also taking into account the number of men on active surveillance who opt for treatment at some point.[24] These cost savings include low, intermediate and high risk men in the analysis. For a health plan with 10 million members, this translates to savings of more than $16 million, with two-thirds of those savings realized in the first year following diagnosis and testing. When only low and intermediate risk men are included in the analysis, per patient savings in the initial year of diagnosis and treatment equals $7,510 after accounting for the cost of the test. These projected savings do not take into consideration the reduction in costs related to treatment-related morbidities; for example, Nam et al. demonstrated that 22.2% of patients who underwent either RP or RT for prostate cancer were admitted to the hospital within 5 years for a treatment-related complication.[25]	The 2015 poster presentation by Crawford et al. was not located by our search. We would have excluded it because of publication type; conference presentations were excluded for the report based on the predefined methods.
References: 1. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101:1446–52. 2. American Medical Association (AMA). Joint report of the Council of Medical Service and the Council on Science and Public Health: Payment and coverage for genetic/genomic precision medicine. Presented at the AMA Interim Meeting of the House of Delegates, November 2017.	Thank you for providing references for your comments.

Comments	Response
3. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: prostate cancer-version 2.2017. 2017; https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf 4. Sommariva S, Tarricone R, Lazzeri M, et al. Prognostic value of the cell cycle progression score in patients with prostate cancer: a systematic review and meta-analysis. *Eur Urol.* 2016;69(1):107-115. 5. Cuzick J, Berney DM, Fisher G, et al. Transatlantic Prostate Group. Prognostic value of a cell cycle progression signature for prostate cancer death in conservatively managed needle biopsy cohort. *Br J Cancer* 2012 Mar; 106(6):1095-9. 6. Cuzick J, Stone S, Fisher G, et al. Validation of an RNA cell cycle progression score for predicting death from prostate cancer in a conservatively managed needle biopsy cohort. *Br J Cancer* 2015 Jul; 113(3):382-9. 7. Bishoff JT, Freedland SJ, Gerber L, et al. Prognostic utility of the CCP score generated from biopsy in men treated with prostatectomy. *J Urol* 2014 Aug; 192(2):409–14. 8. Cooperberg MR, Simko JP, Cowan JE, et al. Validation of a cell-cycle progression gene panel to improve risk stratification in a contemporary prostatectomy cohort. J *Clin Oncol* 2013 Apr; 31(11):1428-34. 9. Freedland SJ, Gerber L, Reid J, et al. Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. *Int J Radiat Oncol Biol Phys* 2013 Aug; 86(5):848-53. 10. Tosoian JJ, Chappidi MR, Bishoff JT, et al. Prognostic utility of biopsy-derived cell cycle progression score in patients with National Comprehensive Cancer Network low-risk prostate cancer undergoing radical prostatectomy: implications for treatment guidance. *BJU Int* 2017 Dec;120(6):808-14. 11. Koch MO, Cho JS, Kaimakliotis HZ, et al. Use of the cell cycle progression (CCP) score for predicting systemic disease and response to radiation of biochemical recurrence. *Cancer Biomark* 2016 Jun; 17(1):83-8. 12. Maurice MJ, et al. Research letter: Current status of prostate cancer diagnosis and management in the United States. *JAMA Oncol* 2016; 2(11):1505-7. 13. Crawford ED, Scholz MC, Kar AJ, et al. Cell cycle progression score and treatment decisions in prostate cancer: results from an ongoing registry. *Curr Med Res Opin* 2014 Jun; 30(6):1025-31. 14. Shore N, Kella N, Moran B, et al. Impact of the cell cycle progression test on physician and patient treatment selection for localized prostate cancer. *J Urol* 2016 Mar; 195(3):612-8. 15. Center for Medical Technology Policy (CMTP) Effectiveness Guidance Document: Evaluation of Clinical Validity and Clinical Utility of Actionable Molecular Diagnostic Tests in Adult Oncology. May 1,	

Comments	Response
2013. Accessed January 29 2018 at http://www.cmtpnet.org/docs/resources/MDX_EGD.pdf 16. Deverka P, Messner DA, McCormack R, et al. Generating and evaluating evidence of the clinical utility of molecular diagnostic tests in oncology. *Genet Med.* 2016 Aug;18(8):780-7. 17. Donovan JL, Hamdy FC, Lane JA, et al; ProtecT Study Group. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *N Engl J Med.* 2016 Oct 13;375(15):1425-1437. 18. Jeldres C, Cullen J, Hurwitz LM, et al. Prospective quality-of-life outcomes for low-risk prostate cancer: Active surveillance versus radical prostatectomy. *Cancer* 2015;121:2465-73. 19. Barocas DA, Alvarez, J, Resnick MJ, et al. Association Between Radiation Therapy, Surgery, or Observation for Localized Prostate Cancer and Patient-Reported Outcomes After 3 Years. *JAMA* 2017;317(11):1126-40. 20. Chen RC, Basak R, Meyer AM, et al. Association Between Choice of Radical Prostatectomy, External Beam Radiotherapy, Brachytherapy, or Active Surveillance and Patient-Reported Quality of Life Among Men With Localized Prostate Cancer. *JAMA* 2017;317(11):1141-50. 21. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med* 2012; 367:203-13. 22. Wilt TJ, Jones KM, Barry MF, et al. Follow-up of prostatectomy versus observation for early prostate cancer. *N Engl J Med.* 2017;377:132-42. 23. Hamdy FC, Donovan JL, Lane JAet al; ProtecT Study Group. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *N Engl J Med.* 2016 Oct 13;375(15):1415-1424. 24. Crawford DE, Gustavsen G, Brawer MK, et al. Evaluation of the economic impact of the CCP assay in localized prostate cancer. Presented at the Genitourinary Cancers Symposium, San Francisco, CA, February 2015. 25. Nam RK, Cheung P, Herschorn S, et al. Incidence of complications other than urinary incontinence or erectile dysfunction after radical prostatectomy or radiotherapy for prostate cancer. *Lancet Oncol* 2014; 15:223-31.	

From:	Jeff Evans
To:	HCA ST Health Tech Assessment Prog
Subject:	Gene expression profile testing of cancer tissue
Date:	Friday, February 2, 2018 9:04:40 AM
Attachments:	EvidenceMatters_Reproducible Results.pdf

To whom it may concern,

My name is Dr. Jeff Evans. I am writing this comment on behalf of the Washington State Urology Society regarding the HCA review of "Gene Expression Profile Testing of Cancer Tissue". Our organization asks that your committee continue to allow access to this technology for our patients with prostate cancer.

Urologists periodically utilize these advanced diagnostic tests when deciding treatment options for patients with adenocarcinoma of the prostate. The 2017 NCCN guidelines allow for consideration of tumor-based molecular assays in men with clinically localized disease. Retrospective case cohort studies have shown that molecular assays provide prognostic information independent of other NCCN risk groups. These include, but are not limited to, likelihood of death with conservative management and likelihood of biochemical progression after radical prostatectomy or radiation treatment.

We are seeing an increasing trend toward active surveillance in low risk prostate cancer and these tests are particularly helpful in deciding if avoidance of therapy or intervention is warranted. Medicare has reviewed the literature and has made a positive coverage decision. They "Policy Context" on the HCA website gets it right stating "potential benefits of these types of tests include more appropriate treatment decisions and better patient outcomes, including avoiding....unnecessary treatments".

We can all agree that prospective studies are the most meaningful study design, but I caution your committee to consider the strong retrospective data available. Prospective data accumulation for prostate cancer is somewhat unrealistic given the long clinical course of the disease for most patients. If you base your determination on the lack of prospective studies it will penalize current patients with the disease. Please take this into consideration as you are making your decision for patients with prostate cancer in Washington.

Washington state urologists currently utilize this class of tests for select patients with prostate cancer and recommend they continue to be covered. We are not recommending any particular companies assay. Medicare patients already have access to these tests and we feel that the Medicaid beneficiaries deserve the same coverage.

We appreciate your review of our comments and addressing the issues that affect men with prostate cancer in our state. Feel free to contact me with any questions.

Respectfully,

Jeffrey Evans, MD

https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf
https://prolaris.com/publications/

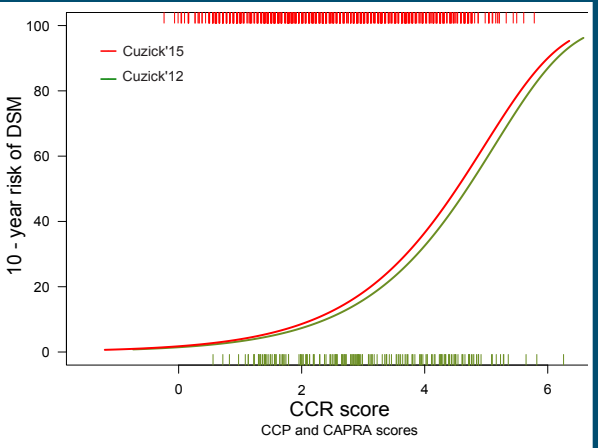
EVIDENCE MATTERS

Prolaris®

In 2016, a survey of 1,500 researchers found that more than 70% had tried and failed to reproduce another scientist's results, and more than half had failed to reproduce even their own experiments.[1]

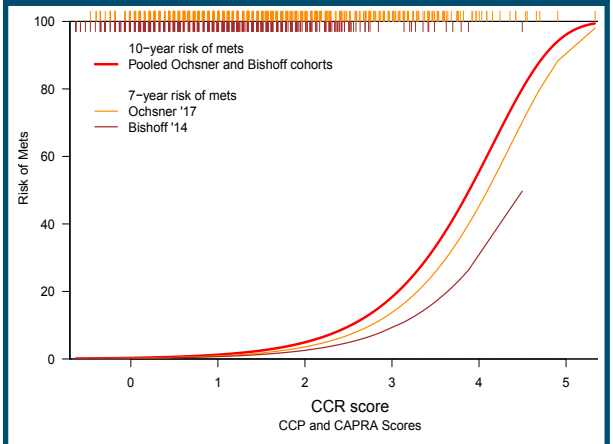
Prolaris provides a level of confidence and quality with reproducible and double-validated endpoints.

Disease Specific Mortality with **Conservative Management** Studies[2,3]



- Two studies with 1,110 patients in a conservatively managed cohort evaluating disease specific mortality.
- Both studies showed consistent 10-year risk curves of prostate cancer death.
- In both studies, Prolaris proved to be highly predictive of disease specific mortality with conservatively managed patients.

Metastatic Disease with **Definitive Treatment** Studies[4,5]



- Two studies with over 1,000 patients who were definitively treated with RP or EBRT and followed for metastatic disease for 10 years.
- Both studies showed consistent risk curves indicating that Prolaris is a strong predictor of metastatic disease and provides significant and independent prognostic information about prostate cancer outcome.

Two Clinical Utility Studies That Resulted In 48-65% Change in Management [6,7]

Prolaris® Prostate Cancer
www.prolaris.com

myriad® WHEN DECISIONS MATTER
Myriad Genetic Laboratories, Inc.

MYRIAD myRisk® Hereditary Cancer
www.prolaris.com/myrisk

PROLARIS® IS CLINICALLY RELEVANT ACROSS ALL RISK CATEGORIES AND POST RADICAL PROSTATECTOMY[4,8]

LOW RISK	INTERMEDIATE RISK		HIGH RISK	POST RP
3 + 3	3 + 4	4 + 3	All High Risk Gleason	Post Surgical
Prolaris®	Prolaris®	Prolaris®	Prolaris®	Prolaris®RP

PROLARIS CLINICAL VALIDATION PUBLICATIONS

PUBLICATION	SAMPLE TYPE	TREATMENT PARADIGM	ONCOLOGIC ENDPOINTS	AUA RISK GROUPS STUDIED
Cuzick 2011[8]	RP	Radical Prostatectomy	Biochemical Recurrence	Low, Intermediate, and High
	TURP	Conservatively Managed	Disease-Specific Mortality	Low, Intermediate, and High
Cuzick 2012[2]	Biopsy	Conservatively Managed	Disease-Specific Mortality	Low, Intermediate, and High
Cuzick 2015[3]	Biopsy	Conservatively Managed	Disease-Specific Mortality	Low, Intermediate, and High
Cooperberg 2013[9]	RP	Radical Prostatectomy	Biochemical Recurrence	Low, Intermediate, and High
Freedland 2013[10]	Biopsy	Primary External Beam Radiation	Biochemical Recurrence	Low, Intermediate, and High
Bishoff 2014[4]	Biopsy	Radical Prostatectomy	Biochemical Recurrence and Metastatic Disease	Low, Intermediate, and High
	Biopsy	Radical Prostatectomy	Biochemical Recurrence and Metastatic Disease	Low, Intermediate, and High
	Biopsy	Radical Prostatectomy	Biochemical Recurrence and Metastatic Disease	Low, Intermediate, and High
Koch 2016[11]	RP	Radical Prostatectomy	Metastatic Disease and Resposne to Salvage Radation after BCR	Low, Intermediate, and High
Tosion 2017[12]	Biopsy	Radical Prostatectomy	Biochemical Recurrence	Low
				TOTAL PATIENTS STUDIED: 2,979

1. 1,500 scientists lift the lid on reproducibility. (2016, May 25). Retrieved June 29, 2017, from http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970
2. Cuzick J, et al. Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. Br J Cancer 2012; 106:1095–1099.
3. Cuzick J, et al. Validation of an RNA cell cycle progression score for predicting death from prostate cancer in a conservatively managed needle biopsy cohort. Br J Cancer 2015; 113:382– 389.
4. Bishoff JT, et al. Prognostic utility of the cell cycle progression score generated from biopsy in men treated with prostatectomy. J Urol 2014; 192:409–414.
5. Bardot S, et al. Comparing the prognostic utility of the CCP Score for predicting metastatic disease in African American and Non-African American men with prostate cancer. Poster session presented at: American Urological Association; 2017 May 12-16; Boston, MA
6. Shore, N D., et al. Impact of the cell cycle progression test of physician and patient treatment selection for localized prostate cancer. The Journal of Urology 2016;195(3),612-618.
7. Crawford, E D, et al. Cell cycle progression score and treatment decisions in prostate cancer: Results from an ongoing registry. Current Medical Research and Opinion 2014:30(6), 1025-1031.
8. Cuzick J, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol 2011; 12:245– 255.
9. Cooperberg MR, et al. Validation of a cell-cycle progression gene panel to improve risk stratification in a contemporary prostatectomy cohort. J Clin Oncol 2013; 31:1428–1434.
10. Freedland SJ, et al. Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. Int J Radiat Oncol Biol Phys 2013; 86:848–853.
11. Koch, Michael O., et al. Use of the cell cycle progression (CCP) score for predicting systemic disease and response to radiation of biochemical recurrence. Cancer Biomarkers 2016; 17.1:83-88.
12. Tosoian J, et al. Prognostic Utility of Biopsy-Derived Cell Cycle Progression Score in Patients with NCCN Low-Risk Prostate Cancer Undergoing Radical Prostatectomy: Implications for Treatment Guidance. BJU International 2017; doi:10.1111/bju.13911

Comments submitted by:
Johnathan Lancaster, MD PhD
Chief Medical Officer
Myriad Genetic Laboratories
320 Wakara Way
Salt Lake City, Utah 84108
jlancaster@myriad.com
Tel: 801-505-5090

January 30th, 2018

Dear members of the Health Technology Clinical Committee,

We respectfully submit the enclosed comments on the draft evidence report, **Gene Expression Profile Testing of Cancer Tissue. These comments pertain specifically to EndoPredict for breast cancer.** A separate document is submitted with comments pertaining to Prolaris for prostate cancer.

We appreciate your attention to this important topic and the opportunity to submit comments. Please contact me if I can provide any additional information. Thank you.

TABLE OF CONTENTS

ENDOPREDICT

The draft report concludes that only Oncotype DX (breast) and Mammaprint have adequate evidence to support their use. We respectfully suggest that EndoPredict should also be supported for clinical use for the following reasons: 1) Equivalency has been demonstrated with Oncotype DX, which is determined to have adequate evidence. 2) EndoPredict has achieved a level of evidence (LOE) IB, similar to Oncotype DX, which confirms clinical utility. 3) Technical assessments and professional society guidelines recommend EndoPredict coverage, given sufficient levels of evidence of clinical utility and improvement in net health outcomes. 4) Current insurance coverage for EndoPredict is broad. 5) The addition of EndoPredict is budget neutral to the state.

1) Equivalency has been demonstrated with another breast cancer gene expression assay determined to have adequate evidence.

EndoPredict is a 2nd generation breast cancer gene expression assay that was introduced in the United States in 2017 as an alternative to other breast cancer assays on the market. The intended use population for EndoPredict is the same as that for Oncotype DX, that is, patients with ER-positive, HER2-negative early breast cancer, with negative or positive lymph nodes. The test is used by physicians in the same manner as Oncotype DX, to determine whether the predicted 10-year distant recurrence risk warrants the use of adjuvant chemotherapy. Oncotype DX classifies patients into low, moderate and high risk groups; EndoPredict classifies patients into low and high risk groups. Published studies have shown EndoPredict to be at least as effective as Oncotype DX. (See below.) As an alternative test with demonstrated equivalency, we believe that EndoPredict should be supported for clinical use if the data on Oncotype DX are determined to support its clinical use.

The publication by Buus et al. in 2016[1], which was excluded from the Committee's review by a limitation of the scope, included a comparative study of gene expressions tests comparing EndoPredict ("EPclin") and Oncotype DX ("RS"). The study included 928 patients from the Arimidex, Tamoxifen, Alone or in Combination (ATAC) trial and reported the following:

- EndoPredict provided better separation of low and high risk breast cancer, with a hazard ratio of 5.99 for EPclin compared to 2.73 for RS (low vs non-low risk).
- EndoPredict showed good prognostic strength with a 10 year distant recurrence (DR) rate of 5.8% in the low risk group (compared to 10.1% for Oncotype DX low risk group) and 28.8% in the high risk group.
- The prognostic ability of EndoPredict was strong for node negative disease (5.9% DR for low risk vs. 20.0% DR for high risk) and even stronger for node positive disease (5.0% vs. 36.9%). In comparison, the Oncotype DX observed DR rate in low risk node positive patients was 25%.
- When compared to Oncotype DX, EndoPredict provided a classification more closely aligned with patient outcomes. This was observed with patients who were classified differently by the two tests:
    - Of the 13% of patients who were classified as low risk by EPclin and non-low risk (intermediate or high risk) by RS, the observed rate of DR was 10.2%, i.e. more closely aligned with low risk.

    - Of the 15% of patients who were classified as high risk by EPclin and low risk by RS, the observed rate of DR was 26.9%, i.e. EndoPredict correctly identified these patients as high risk.
- The authors concluded that the superior performance of EndoPredict was partly due to the inclusion of clinical variables (nodal status and tumor size) in the EPclin score, but also due to a molecular signature with improved prediction of late events (in years 5-10).

In 2016, Sestak et al reported the first comprehensive comparison of the performance of six prognostic signatures, including four gene expression assays, in estrogen receptor positive breast cancers from the TransATAC cohort.[17] Publication of the peer-reviewed manuscript is anticipated to occur on February 15, 2018, in JAMA Oncology. The authors found that EndoPredict provided greater prognostic value than Oncotype DX.

The systematic review by Blok et al.[2] is described in the draft report. As stated in the draft, a reduction in invasive treatments and a corresponding increase in less invasive treatments was observed with use of all of the reviewed tests, including EndoPredict. This confirms equivalency in terms of impact on treatment decisions.

2) Level of evidence (LOE) IB confirms clinical utility.

Simon et al. published a well-accepted framework for establishing clinical utility of tumor biomarkers.[3] The authors acknowledge that, while the gold standard is a prospective randomized clinical trial, an alternative framework is needed when logistical and cost barriers exist for such trials, particularly for prognostic biomarkers. According to the Simon framework, LOE IB confirms clinical utility and is achieved when at least two independent clinical validation studies performed on archived samples from previously performed prospective trials ("prospective-retrospective" studies) demonstrate consistent results. The ability of EndoPredict to separate ER-positive, HER2-negative early breast cancer into low and high risk for distant recurrence has been demonstrated in three separate cohorts (independent of the training cohort): ABCSG-6, ABCSG-8 and ATAC.[4,1] Thus, EndoPredict is confirmed to have clinical utility based on LOE IB, according to the Simon framework. The evidence-based reviews of the American Society of Clinical Oncology (ASCO),[5] Blue Cross Blue Shield Evidence Street[6] and the European Group on Tumor Markers[7] concluded that EndoPredict and Oncotype DX have both achieved similar LOE of IB and are therefore demonstrated to have clinical utility. Note that the TAILORx and MINDACT trials are prospective trials involving Oncotype DX and Mammaprint respectively, but only subsets of data have been reported so far, and only with 5-year outcomes as opposed to 10-year outcomes.[8,9] Therefore, LOE IB is the highest LOE achieved by any of the breast cancer assays to date.

3) Technical assessments and professional society guidelines recommend EndoPredict coverage, given sufficient level of evidence of clinical utility.

- As stated in the draft report, EndoPredict is recommended by the American Society of Clinical Oncology (ASCO) to guide decisions on the need for adjuvant systemic chemotherapy.[5] The draft rates the guideline published by this group as having "good methodological quality".
- Although the Blue Cross Blue Shield HTA program was cited as a data source, the draft report does not mention the report from this group. Importantly, the thorough, evidence-based

review by Blue Cross Blue Shield Evidence Street concluded in 2016 and reaffirmed in 2017 that EndoPredict has sufficient evidence "to determine that the technology results in a meaningful improvement in net health outcome."[6]

- Palmetto GBA's MolDX program reviewed EndoPredict and determined that the evidence was sufficient for coverage by the Centers for Medicare and Medicaid Services (CMS). Noridian has now published a Local Coverage Determination (LCD) for EndoPredict, effective 1/30/2018.[10]
- Regarding the guidelines of the National Comprehensive Cancer Network (NCCN),[11] the draft report correctly quotes this guideline as stating that "other gene expression profile tests [besides Oncotype DX] can be considered to assess risk of cancer recurrence, but that they have not been validated to predict response to chemotherapy." The ability to assess risk of cancer recurrence versus predict response to chemotherapy is an important distinction. Physicians typically use the risk of cancer recurrence to determine the risk/benefit ratio for adjuvant chemotherapy in the treatment of early breast cancer. All of the breast cancer assays were developed to predict risk of cancer recurrence (that is, they are prognostic) and the claim that Oncotype DX is also able to predict response to chemotherapy (predictive) is based on subset analyses of older data which have been called into question.[12,13] The ASCO expert panel that reviewed the literature on breast cancer biomarkers concluded that data are inadequate to support a chemotherapy predictive claim, stating the following: "However, the B20 data [14] are confounded by the data set originally used to generate the 21-gene RS algorithm. The results from SWOG S8814 [15] must be considered hypothesis generating because the number of samples analyzed in each RS subgroup was small, there was no additional prediction beyond 5 years, and the risks of systemic recurrence continues to be high for patients with node-positive disease"[5] In any case, physicians using Oncotype DX usually base treatment decisions on its prognostic function, not the claimed predictive function. This is evidenced by the fact that patients reported to be Oncotype DX "intermediate risk" are frequently recommended for chemotherapy,[16] despite the publications used to support the predictive claim suggesting that intermediate risk patients do not benefit from chemotherapy.[14,15] In summary, gene expression assays were developed and are utilized for their prognostic ability, that is, their ability to predict distant recurrence, rather than predict response to a particular treatment. For clarity, we respectfully request replacement of the words "not recommended" for EndoPredict under NCCN in the draft's Table ES-2 with "can be considered". This would more accurately reflect what is stated in the NCCN guideline. We further request removal of the following statement in the draft's conclusion on pages 26-27: "However, NCCN has adopted a more rigorous standard and advises the exclusive use of Oncotype DX for both LN-negative and LN-positive patients" and replacement with the following: "NCCN has incorporated Oncotype DX into the treatment algorithm with a footnote stating that 'other prognostic multigene assays may be considered to help assess risk of recurrence but have not been validated to predict response to chemotherapy'". We respectfully suggest that this would more fairly represent what is stated by NCCN.

4) Current insurance coverage for EndoPredict is broad.

Currently, there is broad coverage for EndoPredict across the United States with more than 90% of breast cancer patients having access through their insurance. (Myriad, internal data.) The Washington Medicare LCD covering EndoPredict for node negative and node positive patients, effective 1/30/18, was not included in the draft report.[10] In addition to coverage by Aetna and Regence, as reported in

the draft, EndoPredict has positive coverage policies by numerous other commercial payers, including Premera and the majority of Blues plans, United Healthcare, Anthem and Humana, among others. Washington Medicaid currently covers EndoPredict, and other state Medicaid programs that cover include, but are not limited to, AK, AZ, CA, CT, DE, MD, MN, NH, NJ, NM, OH, OK, PA and WV.

5) The addition of EndoPredict coverage is budget neutral to the state.

The EndoPredict rate (CPT code: 81599-QP) has been set at the same rate as Oncotype DX (CPT code 81519) for Washington State Medicaid. This ensures budget neutrality for the state since providers will utilize one test or the other for the same intended use population. As stated in the draft, the systematic review by Blok et al.[2] reported a reduction in invasive treatments and a corresponding increase in less invasive treatments with use of all of the reviewed tests, including EndoPredict. The reduction in invasive treatments is expected to result in cost savings for the state.

6) Summary

In summary, we respectfully request that the report acknowledge that published data support at least equivalency of EndoPredict to Oncotype DX as well as a similar LOE IB that supports clinical utility. This would be consistent with the conclusions of other technology assessments, professional societies and the vast majority of U.S. payers. We respectfully request continued coverage of EndoPredict by Washington Medicaid to allow Washington healthcare providers access to an alternative, equivalent test under the "any willing provider" provision. By crosswalking the price of EndoPredict to CPT code 81519, continuing coverage of EndoPredict as an alternative option is budget neutral for the state.

REFERENCES

1. Buus R, Sestak I, Kronenwett R, et al. Comparison of EndoPredict and EPclin With Oncotype DX Recurrence Score for Prediction of Risk of Distant Recurrence After Endocrine Therapy. *J Natl Cancer Inst.* 2016 Jul 10;108(11). pii: djw149. doi: 10.1093/jnci/djw149. Print 2016 Nov.
2. Blok EJ, Bastiaannet E, van den Hout WB, et al. Systematic review of the clinical and economic value of gene expression profiles for invasive early breast cancer available in Europe. *Cancer Treat Rev.* 2018 Jan;62:74-90.
3. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101:1446–52.
4. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res.* 2011;17(18):6012–6020.
5. Harris LN, Ismaila N, McShane LM, et al; American Society of Clinical Oncology. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol.* 2016 Apr 1;34(10):1134-50.
6. Blue Cross Blue Shield Association. Assays of Genetic Expression in Tumor Tissue as a Technique to Determine Prognosis in Patients With Breast Cancer. December 2016 and December 2017.

7. Duffy MJ, Harbeck N, Nap M, et al. Clinical use of biomarkers in breast cancer: updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer*. 2017;75:284-298.
8. Sparano JA, Gray RJ, Makower DR, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med*. 2015;373;2005-14.
9. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med.* 2016;375(8):717-729.
10. MolDX: EndoPredict® Breast Cancer Gene Expression Test (L37311). Noridian Healthcare Solutions, LLC. Effective 1/30/2018.
11. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: breast cancer-version 3.2017. 2017; https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf.
12. Michiels S, Ternès N, Rotolo F. Statistical controversies in clinical research: prognostic gene signatures are not (yet) useful in clinical practice. *Ann Oncol*. 2016 Dec;27(12):2160-2167.
13. Ioannidis JP. Is molecular profiling ready for use in clinical decision making? *Oncologist.* 2007 Mar;12(3):301-11.
14. 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 Aug 10;24(23):3726-34.
15. Albain KS, Barlow WE, Shak S, et al. Breast Cancer Intergroup of North America. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, estrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomized trial. *Lancet Oncol*. 2010 Jan;11(1):55-65.
16. Petkov VI, Miller DP, Howlader N, et al. Breast-cancer-specific mortality in patients treated based on the 21-gene assay: a SEER population-based study. *NPJ Breast Cancer*. 2016 Jun 8;2:16017.
17. Sestak I, Buus R, Cuzick J, et al. Comprehensive comparison of prognostic signatures for breast cancer in TransATAC. Presented December 9th, 2016, San Antonio Breast Cancer Symposium.

OXFORD

JNCI J Natl Cancer Inst (2016) 108(11): djw149

doi: 10.1093/jnci/djw149
First published online July 10, 2016
Article

ARTICLE

Comparison of EndoPredict and EPclin With Oncotype DX Recurrence Score for Prediction of Risk of Distant Recurrence After Endocrine Therapy

Richard Buus, Ivana Sestak, Ralf Kronenwett, Carsten Denkert, Peter Dubsky, Kristin Krappmann, Marsel Scheer, Christoph Petry, Jack Cuzick, Mitch Dowsett

Affiliations of authors: The Breast Cancer Now Toby Robins Research Centre, Institute of Cancer Research, London, UK (RB, MD); Academic Department of Biochemistry, Royal Marsden Hospital, London, UK (RB, MD); Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK (IS, JC); Sividon Diagnostics GmbH, Cologne, Germany (RK, KK, MS, CP); Institute of Pathology, Charité University Hospital and German Cancer Consortium (DKTK), Berlin, Germany (CD); Department of Surgery and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria (PD)

Correspondence to: Mitch Dowsett, The Breast Cancer Now Toby Robins Research Centre, Institute of Cancer Research and Academic Department of Biochemistry, Royal Marsden Hospital, Fulham Road, London, SW3 6JJ, UK (e-mail: mitch.dowsett@icr.ac.uk).

Abstract

Background: Estimating distant recurrence (DR) risk among women with estrogen receptor–positive (ER+), human epidermal growth factor receptor 2 (HER2)–negative early breast cancer helps decisions on using adjuvant chemotherapy. The 21-gene Oncotype DX recurrence score (RS) is widely used for this. EndoPredict (EPclin) is an alternative test combining prognostic information from an eight-gene signature (EP score) with tumor size and nodal status. We compared the prognostic information provided by RS and EPclin for 10-year DR risk.

Methods: We used likelihood ratio χ^2 and Kaplan-Meier survival analyses to compare prognostic information provided by EP, EPclin, RS, and the clinical treatment score (CTS) of clinicopathologic parameters in 928 patients with ER+ disease treated with five years' anastrozole or tamoxifen. Comparisons were made for early (0-5 years) and late (5-10 years) DR according to nodal status. All statistical tests were two-sided.

Results: In the overall population, EP and EPclin provided substantially more prognostic information than RS (LRχ^2: EP = 49.3; LRχ^2: EPclin = 139.3; LRχ^2: RS = 29.1), with greater differences in late DR and in node-positive patients. EP and EPclin remained statistically significantly prognostic when adjusted for RS (ΔLRχ^2: EP+RS vs RS = 20.2; ΔLRχ^2: EPclin+RS vs RS = 113.8). Using predefined cut-offs, EPclin and RS identified 58.8% and 61.7% patients as low risk, with hazard ratios for non-low vs low risk of 5.99 (95% confidence interval [CI] = 3.94 to 9.11) and 2.73 (95% CI = 1.91 to 3.89), respectively.

Conclusions: EP and EPclin were highly prognostic for DR in endocrine-treated patients with ER+, HER2-negative disease. EPclin provided more prognostic information than RS. This was partly but not entirely because of EPclin integrating molecular data with nodal status and tumor size.

Breast cancer is the most common cancer in women. About 80% of primary breast cancers are estrogen receptor (ER)–positive disease. Patients with ER-positive disease receive adjuvant endocrine therapy after surgery that markedly improves their prognosis (1). A large proportion of patients receiving endocrine therapy have sufficiently low risk to safely avoid chemotherapy. Differentiating these patients from higher-risk patients who may benefit from adjuvant chemotherapy is a priority for breast cancer management (2).

Received: December 23, 2015; **Revised:** March 25, 2016; **Accepted:** May 9, 2016

Multigene expression prognostic assays may be used to estimate residual risk of recurrence following surgery and endocrine treatment to aid decisions on the appropriateness of chemotherapy treatment. The most widely used test is the Oncotype DX 21-gene recurrence score (RS) (3). Other prognostic scores to estimate residual risk in endocrine-treated patients include the PAM50 risk of recurrence (ROR) score (4), the Breast Cancer Index (BCI) (5), and the IHC4 test that is immunohistochemically based and is combined with the clinical treatment score (CTS) to integrate clinicopathological parameters (6). The amount of prognostic information provided for early (0-5 years) and late (beyond five years) recurrence varies across these tests (7).

The EndoPredict (EP) assay combines the expression of three proliferative and five ER-signaling/differentiation-associated genes and is normalized by three housekeeping genes (8). EP may be measured in formalin-fixed, paraffin-embedded tissue sections by quantitative real-time polymerase chain reaction (qRT-PCR) in decentralized laboratories (9) and provides a score that ranges between 0 and 15 after scaling. EPclin was derived from EP by incorporating nodal status and tumor size to create an integrated diagnostic algorithm for clinical decisions (8). Both EP and EPclin were trained on a cohort of 964 patients with ER-positive, human epidermal growth factor receptor 2 (HER2)–negative carcinomas treated with adjuvant endocrine therapy only (8). Thresholds for EP and EPclin to differentiate between patients at low or high risk corresponding to a 10% probability of distant recurrence (DR) at 10 years were set at 5 and 3.3, respectively. Both EP and EPclin were shown to be prognostic for early and late distant recurrence in the Austrian Breast and Colorectal Cancer Study Group (ABCSG)-6 and -8 trials (10).

TransATAC, the translational substudy of the Arimidex, Tamoxifen, Alone or in Combination trial (ATAC), served as a validation study for the Oncotype DX RS (11), PAM50 ROR (12), and BCI (13) scores and as a training set for a definition of PAM50 ROR cut-off values and for CTS and IHC4 scores (6).

Our aims were to assess the prognostic value of EP and EPclin for DR in postmenopausal women with hormone receptor–positive, HER2-negative primary breast cancer in TransATAC and to compare their prognostic ability with that of the Oncotype DX RS.

ARTICLE

Methods

Patient Cohort, RNA Extraction

The ATAC trial evaluated efficacy and safety of anastrozole vs tamoxifen given for five years in postmenopausal women with localized primary breast cancer (14). TransATAC draws upon formalin-fixed, paraffin-embedded tumor samples from a subset of women randomized to the monotherapy arms. RNA was extracted by Genomic Health Inc. (GHI) (11), and residual RNA was available for 928 ER-positive, HER2-negative women. For this analysis, eligibility required hormone receptor–positive, HER2-negative, chemotherapy-naive disease where RS and at least 350 ng residual RNA were available. A pilot study was conducted that confirmed the suitability of TransATAC samples for EP assessment (described in the Supplementary Methods, available online). This study was approved by the South-East London Research Ethics Committee, and all patients included gave informed consent.

Procedures

EP genes' analysis by qRT-PCR was performed by Sividon, who were blinded to all clinical data. Fifty to 100 ng RNA was used to quantitate the eight cancer-related genes of interest (BIRC5, UBE2C, DHCR7, RBBP8, IL6ST, AZGP1, MGP, and STC2) and three reference genes (CALM2, OAZ1, and RPL37A). EP and EPclin scores were determined as previously described (8). The predefined cut-offs for diagnostic decisions corresponding to a 10% DR rate at 10 years were applied to stratify patients into low- or high-risk groups: EP low risk (<5), EP high risk (≥ 5); EPclin low risk (<3.3), EPclin high risk (≥ 3.3) (8). RS risk groups were determined as previously described, where cut-offs of 18 and 31 in the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14 trial cohort corresponded to approximately 11% and 20% of 10-year risk of DR (3). In addition to these predefined diagnostic cut-points, we compared DR between tertiles based on the genomic assays to allow a more detailed comparison. CTS was derived as reported previously (6) and calculated with the prespecified algorithm: $CTS = 100x\{0.417N_{1\text{-}3} + 1.566N_{4+} + 0.930(0.497T_{1\text{-}2} + 0.882T_{2\text{-}3} + 1.838T_{>3} + 0.559Gr_2 + 0.970Gr_3 + 0.130Age_{\geq 65} - 0.149Ana)\}$.

Study Endpoints

The primary endpoint was distant relapse–free survival (DRFS), which was the time from diagnosis until DR. DR was defined as metastasis from the primary tumor at distant organs, excluding contralateral disease and locoregional and ipsilateral recurrences. Death before DR was treated as a censoring event.

Statistical Analysis

Our stepwise primary objectives were to assess whether EPclin had statistically significant prognostic information for 10-year DR in postmenopausal women with breast cancer given either Tamoxifen or Anastrozole monotherapy. If so, we would test if EPclin or EP added statistically significant prognostic information to RS and whether EP/EPclin provided statistically significant additional information to CTS. Secondary analyses included determining the prognostic ability of EP and EPclin in early (0-5 years) and late (>5 years) settings, in patients divided into subgroups by nodal status, and the additional prognostic information provided by tests in multivariable comparisons.

The statistical analysis plan was approved by the Long-term Anastrozole vs Tamoxifen Treatment Effects (LATTE) committee and Sividon before data analysis took place and is described in the Supplementary Methods (available online). All statistical tests were two-sided, and a P value of less than .05 was regarded as statistically significant. All statistical analyses were performed with STATA version 13.1 (College Station, TX).

Results

Sample availability is shown in Figure 1. Values for RS, EP, and EPclin scores were calculated for 928 patients. Demographics of the population are shown in Supplementary Table 1 (available online). A total of 128 DRs was recorded within the 10-year median follow-up period. In node-negative women ($n = 680$), there were 59 DRs; in node-positive women ($n = 248$), 69 DRs were recorded.

Univariate Analyses

Results for EP, EPclin, RS, and CTS are presented in Table 1. Both EP and EPclin were highly prognostic across 10 years (LRχ^2:

EP = 49.3; LRχ^2: EPclin = 139.3), with EPclin being statistically significantly more prognostic than EP in all time windows and subgroups, except for node-negative patients in years 0 to 5. Both EP and EPclin provided substantially more information than RS in years 0 to 10 (LRχ^2: RS = 29.1). EP had similar prognostic power to RS in years 0 to 5 in all subgroups. In node-negative patients, both EP and EPclin performance were very similar to that of RS (LRχ^2: EP = 15.5; LRχ^2: EPclin = 17.0; LRχ^2: RS = 18.7). In contrast, in node-positive patients, EPclin outperformed RS. EP and EPclin were superior to RS in years 5 to 10, where RS was particularly weak regardless of nodal status (LRχ^2: EP = 23.6; LRχ^2: EPclin = 59.3; LRχ^2: RS = 5.6).

Figure 2 shows the DR rate over 10 years for each of EP, EPclin, and RS for the overall population when divided into tertiles of their respective scores. The hazard ratio (HR) for the comparison between the lowest and highest tertiles of each score was 4.72 (95% confidence interval [CI] = 2.78 to 8.02), 18.01 (95% CI = 7.87 to 41.19), and 2.41 (95% CI = 1.59 to 3.64), respectively. For EPclin, the lowest tertile had a DR rate of only 2.1% (95% CI = 1.0 to 4.7) at 10 years while the highest tertile had a DR rate of 31.5% (95% CI = 26.4 to 37.4). Similar plots of EP, EPclin, and RS score tertiles for the separate node-negative and node-positive populations are shown in Supplementary Figures 2 and 3 (available online). EPclin identified a third of patients in the node-negative population, in which only one of 227 patients had a DR over 10 years, corresponding to a 10-year relapse rate of 0.5% (95% CI = 0.1 to 3.4). For EP and RS, the equivalent tertiles had DRs of 1.5% (95% CI = 0.5 to 4.7) and 7% (95% CI = 4.2 to 11.5), respectively, over the same time period.

Supplementary Figure 4 (available online) shows the continuous relationship between EPclin score and the estimated 10-year DR rate in TransATAC according to the proportion of the scores contributed by each nodal group. In this cohort, about half of patients (52.6%) with EPclin scores of 3.3 or higher (high-risk) were node-positive; only 8.6% of patients with scores of less than 3.3 were node-positive.

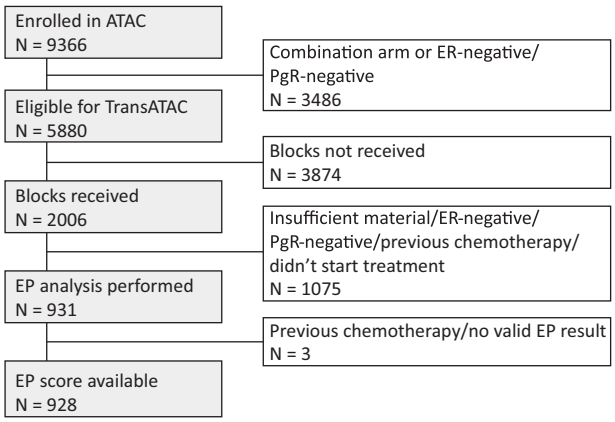


Figure 1. CONSORT diagram of the availability of samples for analysis from the Arimidex, Tamoxifen, Alone or in Combination trial. ATAC = Arimidex, Tamoxifen, Alone or in Combination; ER = estrogen receptor; PgR = progesterone receptor.

Multivariable Analyses

Multivariable comparisons are shown in Table 1. Both EP and EPclin provided statistically significant prognostic value when added to the RS across 10 years (LRχ^2: RS = 29.1; ΔLRχ^2: EP+RS vs RS = 20.2; ΔLRχ^2: EPclin+RS vs RS = 113.8) (Table 1). For EP, this was because of its additional information beyond RS in five to 10 years only. EPclin added statistically significant prognostic information to RS both before and beyond five years, except in the node-negative subgroup of patients in years 0 to 5.

For the overall population, statistically significant prognostic information beyond that of the CTS was provided in years 0 to 10 by EP, EPclin, and RS; however, it was greater for EP and EPclin than for RS. Similar results were observed within node-negative and -positive subgroups (Table 1). The better performance of EP and EPclin in years 0 through 10 was because of its greater prognostic value in years 5 to 10, where RS added no statistically significant prognostic information to CTS (LRχ^2: CTS = 64.7; ΔLRχ^2: EP+CTS vs CTS = 9.8; ΔLRχ^2: EPclin+CTS vs CTS = 9.9; ΔLRχ^2: RS+CTS vs CTS = 2.3).

Table 1. Likelihood (χ^2) for distant recurrence for all prognostic scores in all patients and subgroups*

Patient group	No. of patients	No. of DRs	EPclin		EP		RS		EPclin + RS vs RS†		EP + RS vs RS†		CTS		EPclin + CTS vs CTS†		EP + CTS vs CTS†		RS + CTS vs CTS†	
			LRχ^2	P	LRχ^2	P	LRχ^2	P	ΔLRχ^2	P	ΔLRχ^2	P	LRχ^2	P	ΔLRχ^2	P	ΔLRχ^2	P	ΔLRχ^2	P
All patients																				
0–10 y	928	128	139.3	<.001	49.3	<.001	29.1	<.001	113.8	<.001	20.2	<.001	149.8	<.001	20.3	<.001	16.4	<.001	12.8	<.001
0–5 y	928	61	80.0	<.001	25.7	<.001	26.1	<.001	54.0	<.001	3.1	.08	85.0	<.001	10.5	.001	6.9	.009	11.8	<.001
5–10 y	820	67	59.3	<.001	23.6	<.001	5.6	.02	59.6	<.001	21.6	<.001	64.7	<.001	9.9	.002	9.8	.002	2.3	.13
Node-negative patients																				
0–10 y	680	59	39. 7	<.001	30.8	<.001	21.3	<.001	18.3	<.001	9.7	.002	35.6	<.001	12.5	<.001	11.9	<.001	8.4	.004
0–5 y	680	24	17.0	<.001	15.5	<.001	18.7	<.001	1.6	.2	0.7	.4	19.0	<.001	3.6	.06	5.2	.02	8.1	.004
5–10 y	623	35	22.7	<.001	15.5	<.001	4.8	.03	20.9	<.001	12.4	<.001	16.9	<.001	9.0	.003	6.6	.01	1.4	.24
Node-positive patients																				
0–10 y	248	69	48.3	<.001	14.5	<.001	8.0	.005	44.8	<.001	6.5	.01	61.6	<.001	8.3	.004	5.4	.02	4.1	.04
0–5 y	248	37	32.2	<.001	7.9	.005	8.0	.005	25.9	<.001	0.9	.33	35.2	<.001	6.4	.01	2.3	.13	3.7	.05
5–10 y	197	32	16.1	<.001	6.6	.01	1.0	.32	18.3	<.001	7.1	.008	26.4	<.001	2.3	.13	3.4	.06	0.7	.39

*Both univariate and multivariable analyses are presented for years 0 to 10, years 0 to 5, and years 5 to 10 separately. Likelihood ratio test based on Cox proportional hazard models for univariate and multivariable analyses. Differences in likelihood ratio values (ΔLRχ^2) were used. CTS = clinical treatment score; DR = distant relapse; EP = EndoPredict; LR = likelihood ratio; RS = recurrence score.

†Denotes multivariable comparisons; eg, the EPclin + RS vs RS comparison assesses the extra prognostic information that EPclin contributes when combined with the RS. All statistical tests were two-sided. All scores are continuous variables.

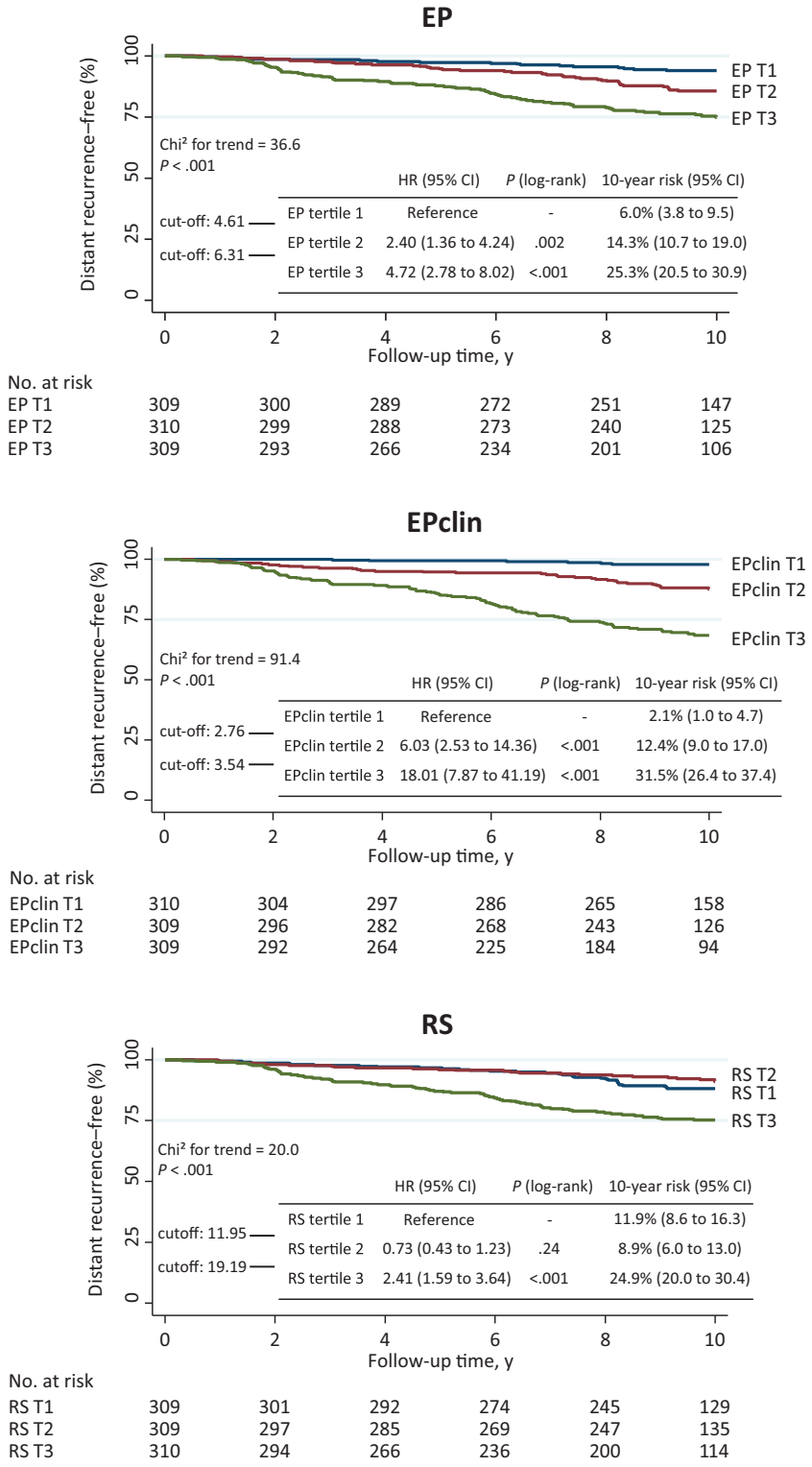


Figure 2. Kaplan-Meier estimates for 10-year distant recurrence according to EP, EPclin, and recurrence score, split into tertiles in all patients. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given **below each graph**. All statistical tests were two-sided. CI = confidence interval; EP = EndoPredict; HR = hazard ratio; RS = recurrence score.

Risk Stratification

For RS, the percentage of patients recurring over 10 years was 5.3% (95% CI = 3.5 to 8.2), 14.3% (95% CI = 9.8 to 20.6), and 25.1% (95% CI = 15.8 to 38.3) for the low-, intermediate-, and high-risk groups in node-negative patients and 25.1% (95% CI = 18.2 to 33.9), 34.8% (95% CI = 24.9 to 47.2), and 48.6% (95% CI = 31.4 to 69.2) for the node-positive group (Supplementary Figure 5,

available online). These are similar to rates observed over years 0 through 9 in 1178 TransATAC patients in our earlier report of RS' performance (11). To compare directly the recurrence rates in these categories with the low-/high-risk categories of EP and EPclin, we pooled the RS intermediate- and high-risk groups to create an RS non-low-risk group. More patients were stratified to the low-risk group by RS and EPclin than by EP (573 vs 546 vs 386 corresponding to 61.7%, 58.8%, and 41.6% of the cohort). The hazard ratio between the high-/non-low- vs low-risk groups was marginally greater for EP (HR = 2.98, 95% CI = 1.94 to 4.58, $P < .001$) than for RS (HR = 2.73, 95% CI = 1.91 to 3.89, $P < .001$) and substantially greater for EPclin (HR = 5.99, 95% CI = 3.94 to 9. 11, $P < .001$) (Figure 3).

EPclin's superior ability to classify patients as low risk was further demonstrated by the similar number of patients classified as low risk by RS coupled with a substantially lower 10-year recurrence rate (EPclin: 5.8%, 95% CI = 4.0 to 8.3; RS: 10.1%, 95% CI = 7.7 to 13.1) (Figure 3). A greater absolute separation of the DR rate was found between the risk groups for EPclin (23.0%) than for RS (13.4%). EPclin performed particularly well at stratifying node-positive patients where absolute separation at 10 years for DR rate was 31.9% compared with the 14.1% in node-negative patients (Supplementary Figures 6 and 7, available online).

For most cases, EPclin and RS categorization of risk agreed; however, 117 (12.6%) cases were EPclin low/RS non-low and 144 (15.5%) were EPclin high/RS low (kappa = 0.41, $P < .001$). Classification by EPclin aligned more closely with the observed risks: Pairwise comparison of EPclin high/RS low vs EPclin low/RS non-low (HR = 2.75, 95% CI = 1.39 to 5.44, $P = .002$) (Figure 4). The Net Reclassification Index (NRI) for EPclin vs RS was 17.5% ($P < .001$). In recurrent cases, the EPclin upgraded three times more cases into high-risk groups than the RS (McNemar's odds ratio = 3.00, 95% CI = 1.16 to 7.89, $P = .01$) whereas for noncases upgrading/downgrading was similar for these two scores.

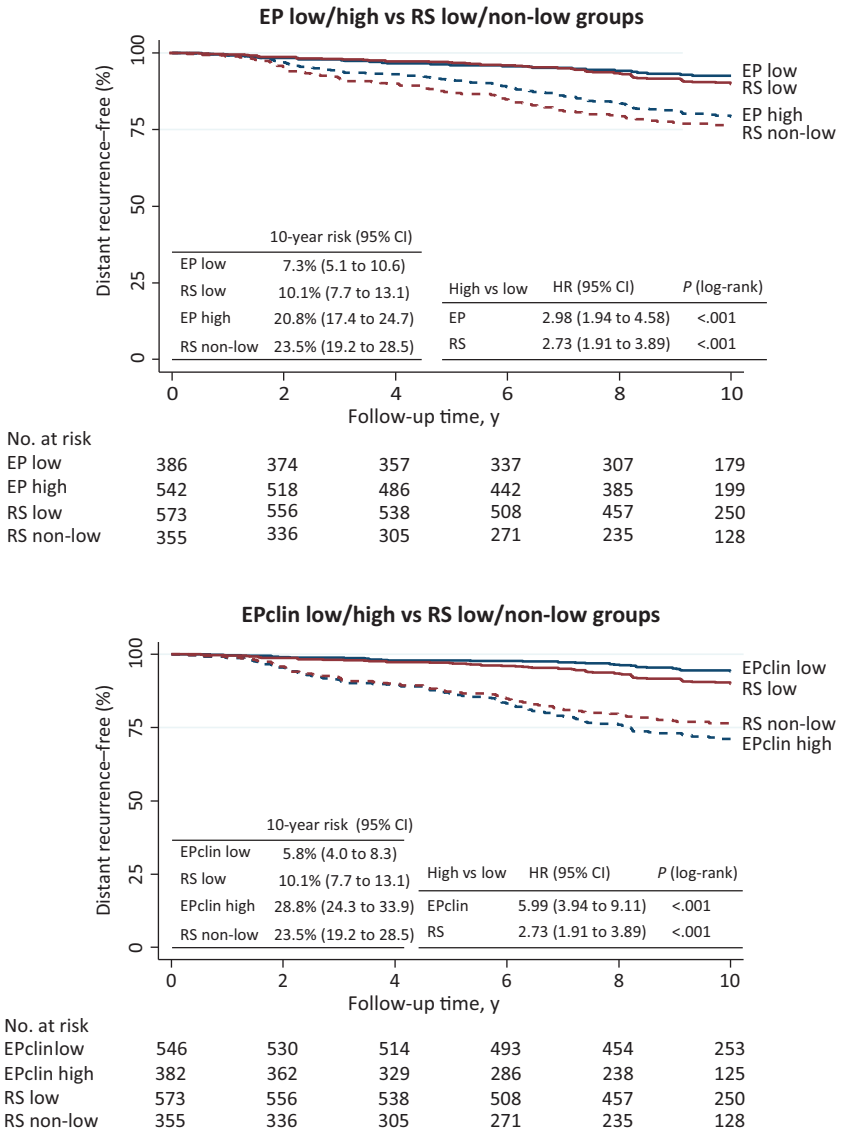


Figure 3. Kaplan-Meier plots for 10-year distant recurrence according to EP, EPclin, and recurrence score in all patients, stratified by cut-offs used for clinical decision-making. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given **below each graph**. All statistical tests were two-sided. CI = confidence interval; EP = EndoPredict; HR = hazard ratio; RS = recurrence score.

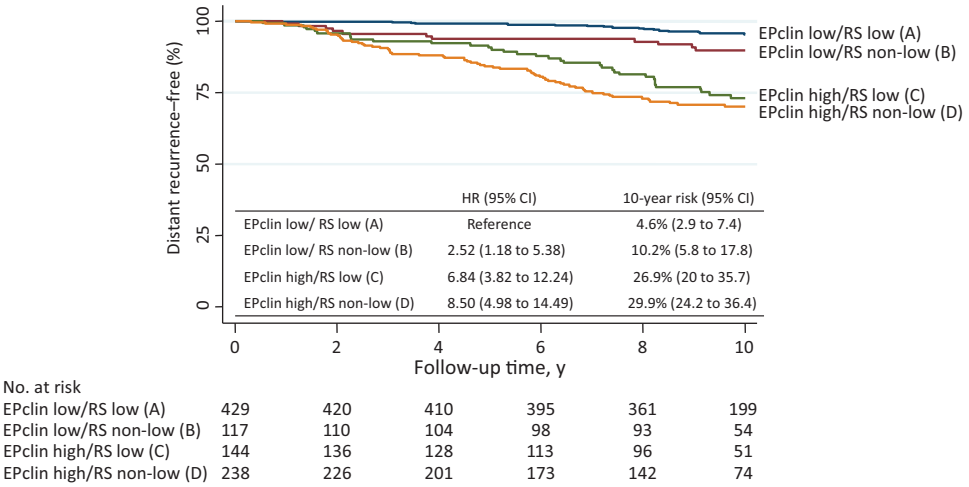


Figure 4. Kaplan-Meier plot of risk groups classified by EPclin and recurrence score for 10-year distant recurrence in all patients. Kaplan-Meier curves were calculated and tested for equality using the log-rank test. The numbers of patients at risk in each group at various time points are given **below each graph**. All statistical tests were two-sided. CI = confidence interval; EP = EndoPredict; HR = hazard ratio; RS = recurrence score.

Discussion

In this TransATAC population, we found that both EP and EPclin were highly prognostic across the 10 years of follow-up and both scores also identified early and late relapse events. This is in agreement with previous reports in ER-positive, HER2-negative patient cohorts from the ABCSG-6 and -8 trials (8,10). Moreover, EP and EPclin were prognostic in all assessed subgroups.

We also compared the prognostic information provided by EP and EPclin with that of the widely used Oncotype DX RS. This study is the first direct comparison of the clinical performance of EP/EPclin with RS. EPclin, as opposed to RS, includes information from clinical factors, making it more clinically useful but also making fair comparisons with RS complicated. Therefore, as well as direct comparisons, we conducted analyses to determine how much information was added by the respective scores to CTS.

We found that EP was similar to RS in years 0 to 5 but was superior in years 5 to 10. EPclin markedly outperformed RS across the 10-year follow-up period and also in all additional univariate analyses, except in node-negative patients in the early time window. These findings suggest that: 1) In years 0 to 5, EPclin predicts recurrence better in the overall population than RS because of the clinical components included in EPclin; and 2) in years 5 to 10, the superior performance of EPclin compared with RS is partly because of the inclusion of clinical variables in EPclin but also because of a molecular component that predicted late recurrences better. The latter is also reinforced by the very similar prognostic value of EP in the early and late follow-up periods, in marked contrast with RS, where performance diminished beyond five years.

EP's overall better performance over RS might be attributed to the differences in the training populations. The EP algorithm was trained on a HER2-negative, mixed node-negative and -positive population, unlike RS, which was optimized on a mixed HER2-negative and -positive, node-negative population. These differences may explain the better prognostic ability of EP, in particular in node-positive patients and in patients at risk of a late relapse.

A previous analysis of EP components in ABCSG-6 and -8 trial samples showed that proliferative genes contributed to early prediction and ER-signaling genes provided prognosis beyond five years (10). Recently, we reported our analysis of RS components in ER+/HER2- TransATAC patients that pointed to the loss of prediction by the ER module as the main reason for its weak performance after five years while the proliferation RS module was prognostic throughout the 10 years (15). The different behavior of the proliferation and ER-associated genes in the two scores may be because of the different identity of genes used and their weighting in the respective algorithms. Further analysis is necessary to understand fully the differing behavior of these prognostic scores.

The integration of nodal status into the EPclin score allows the algorithm to be used in both node-negative and node-positive patients, supported by the observed DR rates in the populations identified as low risk in the respective nodal groups. It was notable, however, that when the algorithm was applied as a continuous variable in the node-negative population it identified one-third of the node-negative population with an extremely low DR of just 0.5% at 10 years. Categorization of a patient in such a low-risk group could be highly reassuring. Our earlier publication showed the differing relationship of RS with risk of DR according to nodal status (11). The current data emphasize that RS should not be used in node-positive patients to estimate recurrence risk without appropriate calibration of the relationship of RS with DR for such patients.

Of note, a recent report from the TAILORx trial described the very low risk of DR rate in patients with RS of 10 or lower (16); this was, however, only over the first five years of follow-up. Generally, patients in a low-risk group would not be recommended to receive chemotherapy treatment because of their perceived low recurrence risk. Previously, ABCSG-6 and -8 observed 10-year DR rates by EPclin classification that were 4% in the low groups for both studies, 28% and 22% in the high-risk groups for the two trials, respectively (8). Our analysis showed similar 10-year recurrence rates at 5.8% and 28.8% in the low and high EPclin groups of TransATAC, respectively, in contrast with 10.1% and 23.5% observed for the RS low and non-low groups. An NRI favorable to EPclin indicated that EPclin classification aligned better with observed risk than RS and therefore provided superior risk stratification when compared with RS. If results are available from both assays yet disagree with one

another, more weight should be assigned to the EPclin risk estimate.

Previously, the importance of integrating clinical and molecular variables to create a more accurate prognostic index for RS (17) and for IHC4 (6) was reported. The superiority of EPclin over EP that resulted from such integration is probably best demonstrated by the DR rates in the highest and lowest tertiles of the respective scores.

Recently, GHI began providing an online Recurrence Score Pathology-Clinical (RSPC) calculator for use in node-negative patients that combines RS with clinicopathological variables including age, tumor size, grade, and planned adjuvant hormonal therapy. Tang et al. reported a greater separation of low- and high-risk patients and reduced number of patients in the intermediate-risk group when classified by RSPC (17). Nevertheless, GHI recommends that RSPC should only be used as an "educational tool" together with RS result to enhance the understanding of the score in the assessment of DR risk (18). It should be noted that while integration of clinicopathological factors with molecular features greatly enhances the prognostic power of risk assessments, this has not been shown to increase predictive information regarding chemotherapy benefit (17).

Our study has strengths and limitations. Strengths include the large patient cohort with long-term follow-up from a well-documented clinical trial and well-characterized set of samples. For this comparison, the same batch of RNA was used, reducing intrasample variation. EP data was obtained by personnel blinded to the clinical data and the results of previous assays in TransATAC. Nonetheless, limitations included the low event rate in ER-positive, endocrine-treated patients and CTS trained on the TransATAC cohort, slightly overestimating its performance compared with what we would expect in independent validation cohorts. Generalizability of the results may be limited by the analysis of patients from the United Kingdom only. Additionally, an unintended sample selection bias might have occurred as the assessment of samples for EP could only be performed where larger amounts of residual RNA were available. Although this might have been expected to relate to fewer smaller tumors in this study than in our earlier report on RS (11), the proportion of tumors 2 cm or smaller was identical at 67% in both. By necessarily restricting the performance of the scores to patients not receiving chemotherapy, this cohort is likely to be biased toward lower risk in the spectrum of ER-positive patients. Lastly, in a number of cases, multiple comparisons were made and caution is needed in interpreting those results. However, for our primary sequential objectives, all tests and comparisons were highly statistically significant at the 1% level, even after correction for multiple comparisons (nominal P value $< .001$). For subgroup analyses, heterogeneity tests are more important (19) and no heterogeneity was observed between subgroups.

In summary, this study has confirmed the independent prognostic ability of EP and EPclin in postmenopausal women with ER+/HER2- primary disease. EPclin provided more prognostic information than RS partly because of its integration with node and tumor size information but also because of a superior molecular component able to predict late events better than RS. Our data highlights the importance of the inclusion of clinicopathological factors in overall estimate of risk assessments.

Funding

This work was supported by the Royal Marsden National Institutes of Health Biomedical Research Centre and the Breast Cancer Now grant awarded to MD (CTR-Q4-Y1) and the Cancer Research UK grant awarded to JC (C569/A16891).

Notes

The study sponsor had no role in design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

This project was developed during discussions at the Biedenkopf meeting October 31, 2013, through November 2, 2013, supported by the BANSS foundation, Biedenkopf/Lahn, Germany.

References

1. Dowsett M, Forbes JF, Bradley R, et al. Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet*. 2015;386(10001):1341–1352.
2. Dowsett M, Goldhirsch A, Hayes DF, et al. International Web-based consultation on priorities for translational breast cancer research. *Breast Cancer Res*. 2007;9(6):R81.
3. 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(27):2817–2826.
4. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2010;16(21):5222–5232.
5. Jerevall PL, Ma XJ, Li H, et al. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer*. 2011;104(11):1762–1769.
6. Cuzick J, Dowsett M, Pineda S, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol*. 2011;29(32):4273–4278.
7. Sestak I, Dowsett M, Zabaglo L, et al. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst*. 2013;105(19):1504–1511.
8. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res*. 2011;17(18):6012–6020.
9. Denkert C, Kronenwett R, Schlake W, et al. Decentral gene expression analysis for ER+/Her2- breast cancer: results of a proficiency testing program for the EndoPredict assay. *Virchows Arch*. 2012;460(3):251–259.
10. Dubsky P, Brase JC, Jakesz R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer*. 2013;109(12):2959–2964.
11. Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol*. 2010;28(11):1829–1834.
12. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol*. 2013;31(22):2783–2790.
13. Sgroi DC, Sestak I, Cuzick J, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol*. 2013;14(11):1067–1076.
14. Cuzick J, Sestak I, Baum M, et al. Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol*. 2010;11(12):1135–1141.
15. Dowsett M, Sestak I, Buus R, et al. Estrogen Receptor Expression in 21-Gene Recurrence Score Predicts Increased Late Recurrence for Estrogen-Positive/HER2-Negative Breast Cancer. *Clin Cancer Res*. 2015;21(12):2763–2770.
16. Sparano JA, Gray RJ, Makower DF, et al. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2015;373(21):2005–2014.
17. Tang G, Cuzick J, Costantino JP, et al. Risk of recurrence and chemotherapy benefit for patients with node-negative, estrogen receptor-positive breast cancer: recurrence score alone and integrated with pathologic and clinical factors. *J Clin Oncol*. 2011;29(33):4365–4372.
18. Genomic Health Inc. Oncotype DX® tools for Genomic Education. https://breast-cancer.oncotypedx.com/en-US/Professional-Invasive/Resources/ODX-Tools-RSPC.aspx. Accessed December 23, 2015.
19. Cuzick J. Forest plots and the interpretation of subgroups. *Lancet*. 2005;365(9467):1308.

Use of Archived Specimens in Evaluation of Prognostic and Predictive Biomarkers

Richard M. Simon, Soonmyung Paik, Daniel F. Hayes

The development of tumor biomarkers ready for clinical use is complex. We propose a refined system for biomarker study design, conduct, analysis, and evaluation that incorporates a hierarchal level of evidence scale for tumor marker studies, including those using archived specimens. Although fully prospective randomized clinical trials to evaluate the medical utility of a prognostic or predictive biomarker are the gold standard, such trials are costly, so we discuss more efficient indirect "prospective–retrospective" designs using archived specimens. In particular, we propose new guidelines that stipulate that 1) adequate amounts of archived tissue must be available from enough patients from a prospective trial (which for predictive factors should generally be a randomized design) for analyses to have adequate statistical power and for the patients included in the evaluation to be clearly representative of the patients in the trial; 2) the test should be analytically and preanalytically validated for use with archived tissue; 3) the plan for biomarker evaluation should be completely specified in writing before the performance of biomarker assays on archived tissue and should be focused on evaluation of a single completely defined classifier; and 4) the results from archived specimens should be validated using specimens from one or more similar, but separate, studies.

J Natl Cancer Inst 2009;101:1446–1452

Many cancer patients do not benefit from the systemic treatments they receive. For example, adjuvant chemotherapy that is considered highly effective may often improve the disease-free or overall survival rate by only 5–10 percentage points. Also, chemotherapy for metastatic disease often provides sustained benefit for a small portion of the patients treated. Therefore, the practice of oncology has been very inefficient, with exposure of far more patients than will benefit to the cost and toxicity of these agents . Although this overtreatment is understandable in dealing with life-threatening diseases, the ability to better "personalize" treatment decisions could have important benefits for patients as well as medical costs. In spite of developments in biotechnology and genomics, the pace of acceptance of new markers to inform treatment decisions for patients with cancer has been slow. The limited introduction of effective biomarkers is partly because of the substantially lower reimbursement for tumor marker tests, as compared with therapeutics by health insurers, but is also because of a shortage of prospective studies of marker utility and the lack of reproducibility and reliability among the many published retrospective studies of prognostic and predictive markers (1,2).

Several committees and authors have proposed specific guidelines that might be used to evaluate and report a given marker. For example, in 1996, the members of the American Society of Clinical Oncology Tumor Markers Guidelines Committee recommended five Levels of Evidence (LOEs) that might be used to determine the clinical utility of a tumor marker (3). This LOE scale has been widely cited and used as a template for deciding whether to recommend the use of a tumor marker in clinical practice and for design and conduct of tumor marker studies (4,5). The criteria for reporting the results of marker studies (designated the REMARK criteria) have been published in several journals, and at least a few journals have incorporated REMARK into the required submission format (6,7).

In this article, we will address the nature of the methodological difficulties involved in studying tumor markers, both prognostic (ie, predictive of prognosis, independent of treatment) and predictive (ie, in terms of best choice of therapy). We will also propose that there are conditions in which archived specimens can be used to provide reliable evaluations of the clinical validity or medical utility of prognostic and predictive biomarkers.

Prospective Randomized Trials to Address Tumor Marker Utility

The gold standard for establishing clinical utility of a new medical intervention is the prospective randomized clinical trial. Several authors have proposed prospective randomized clinical trial designs for evaluation of prospective or predictive diagnostic markers (8–13). In the latter circumstance, the medical utility of the candidate predictive biomarker can be established by evaluating the benefit of the new drug according to marker status (positive or negative) in adequately sized patient subgroups using a prospectively specified analysis plan within a randomized clinical trial that compares a regimen containing the new drug to a control.

One might consider a prospective clinical trial in which the test itself is the investigational intervention to be the ultimate validation

Affiliations of authors: Biometric Research Branch, National Cancer Institute, Bethesda, MD (RMS); Division of Pathology, National Surgical Adjuvant Breast and Bowel Project, University of Pittsburgh, Pittsburgh, PA (SP); Breast Oncology Program, University of Michigan Comprehensive Cancer Center, Ann Arbor, MI (DFH).

Correspondence to: Richard M. Simon, DSc, Biometric Research Branch, National Cancer Institute, 9000 Rockville Pike, Bethesda, MD 20892-7434 (e-mail: rsimon@nih.gov).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djp335

Published by Oxford University Press 2009.

Advance Access publication on October 8, 2009.

of a prognostic or predictive tumor marker. That is, a trial may be designed so that a patient's care would be determined based on random assignment to use the test or not, as referred to as the marker strategy design by Simon and Wang (14). In such a trial, treatment decisions are made for patients who are randomly assigned to the control group using standard prognostic factors and practice guidelines. For patients who are randomly assigned to the investigational group, the test, or marker, is used in treatment determination, perhaps in conjunction with standard prognostic factors. The test would be performed only for patients who are randomly assigned to the test group, and the trial would be evaluated by comparing outcomes overall for the two randomization groups. The outcomes must be compared overall because the new test is not used for the "control" group. In many cases, this restriction seriously limits the information that can be gleaned from the design. Results can be particularly confounded and diluted in cases where the standard of care is variable among physicians.

The marker strategy design is also generally very inefficient in terms of the number of patients required for randomization. Sample size requirements for randomized clinical trials are often proportional to the reciprocal of the square of the size of the treatment effect to be detected with a specified statistical power. For the marker strategy design, only the overall treatment effect between the two randomized groups can be evaluated, and the size of that effect is generally quite small because many patients will receive the same treatment regardless of the group to which they are randomized. If the analysis is to demonstrate that withholding a standard therapy for test-negative patients is not inferior, then sample size problems are compounded, and even with a huge sample size, the results are unlikely to be convincing.

An alternative approach requires that all patients be tested for marker status "upfront." In this case, the evaluation can be focused on subsets of patients for whom the treatment assignment that is based on the test differs from treatment assignment that is based on standard of care. For example, suppose the standard of care is to use chemotherapy for stage II patients but not for stage I patients and the test purports to identify patients who are likely to benefit from chemotherapy regardless of stage; test-positive patients will receive chemotherapy and test-negative patients will not. In this case, the only patients randomly assigned are stage I patients with a positive test and stage II patients with a negative test. The design enables the effectiveness of chemotherapy to be evaluated separately for these subsets of patients. This design presumes, however, that the standard of care, as a function of standard prognostic variables, is determined.

This strategy of testing all patients up-front is used by two current clinical trials, the Microarray in Node-Negative Disease may Avoid Chemotherapy (MINDACT) study in Europe (15) and the Trial Assigning Individualized Options for Treatment (Rx) (TAILORx) study in North America (16). Although the designs of both trials are complex and somewhat different, they both address the medical utility of withholding standard of care chemotherapy from women with node-negative estrogen receptor–positive breast cancer who have a predicted low risk of recurrence, based on a predefined gene expression–based risk score. The MINDACT study evaluates a 70-gene classifier, and the TAILORx study evaluates a 21-gene classifier. Even though these designs are more efficient than the randomized marker strategy trial design, both of these studies will require many thousands of patients, and nearly a decade each from the time, accrual was begun until the first results are anticipated. The TAILORx and MINDACT studies will cost millions of dollars or Euros to conduct, and with the current speed of the evolution of technology, the test being evaluated may have become obsolete by the time such studies are completed.

It is common for a new marker to be identified after the definitive trials have demonstrated benefit for a specific agent or class of agents or even type of modality (such as chemotherapy in general). We maintain that, in many cases, it may be possible to use archived specimens collected in the past from appropriate previously conducted therapeutic trials and to preserve the focus, control of type I error, and statistical power of properly designed fully prospective studies. Indeed, when there is substantial preliminary evidence that a new marker predicts benefit from a specific drug, it may sometimes be possible to assay the marker in archived specimens from randomized clinical trials that were conducted to evaluate the drug, as was done for *KRAS* in colorectal cancer (17,18).

When suitable archived tissue is available and can be used reliably, it can facilitate and expedite delivery of valuable cancer diagnostics that may be of considerable benefit to patients. Nonetheless, there are certainly also risks to patients from the unreliable use of archived tissues. We have tried here to clarify the key features involved in using these resources in a reliable manner, and we propose a refinement to the previously published LOE scale that permits a more critical analysis of the quality of tumor marker studies using archived specimens.

Prospective vs Retrospective Studies: A Matter of Semantics

Although biomedical scientists and biostatisticians are taught that "prospective" studies are preferable to "retrospective" studies, the distinction between prospective and retrospective is often confused with the distinction between "experimental" and "observational." We propose that for studies of prognostic and predictive biomarkers in oncology, the term retrospective is in some cases misleading.

In cancer epidemiology, both retrospective case–control studies and prospective cohort studies are observational, rather than experimental, studies. Neither type of study involves random assignment of exposure, and hence, observed associations between exposures and disease do not provide as strong a basis for claims of causality as in experimental studies. The most serious limitation of epidemiological studies is their nonexperimental nature, not whether they are retrospective or prospective.

In therapeutics, many retrospective analyses are also nonexperimental, with treatment selection based on patient factors and referral pattern rather than on randomization. Such studies are also often conducted without a written protocol and are unfocused, with numerous patient subsets and endpoints compared without control for the overall chance of a false-positive conclusion. In contrast, prospective randomized clinical trials contain internal control of treatment assignment, careful and proscribed data collection (including outcomes and endpoints), and a focused analysis plan that is developed before the data are examined.

Many biomarker studies are conducted with convenience samples of specimens, which just happen to be available and are assayed for the marker, with no prospectively determined subject eligibility, power calculations, marker cut-point specification, or analytical plans. Such studies are very likely to result in highly biased conclusions and truly deserve to be pejoratively labeled as "retrospective." However, if a "retrospective" study is designed to use archived specimens from a previously conducted prospective trial, and if certain conditions are prospectively delineated in a written protocol before the marker study is performed, we argue that it might be considered a "prospective–retrospective" study. Such a study should carry considerably more weight toward determination of clinical utility of the marker than a simple study of convenience, in which specimens and an assay happen to be available. Having multiple studies of different candidate biomarkers based on archived tissues from the same prospective trial would, however, present a greater opportunity for false-positive conclusions than a single fully prospective trial focused on a specific biomarker. Consequently, independent confirmation of findings for specific biomarkers in multiple prospective–retrospective studies is important (see below).

Using Archived Tissue to Establish the Medical Utility of a Marker

In assessing the use of archived specimens in the evaluation of prognostic and predictive biomarkers, it is useful to consider the three requirements for clinical acceptance of a tumor marker that were first proposed by Henry and Hayes (2): 1) the specific setting and utility of the marker must be clear, 2) the magnitude in either outcomes or treatment effects between those patients who are "positive" for a marker must be sufficiently different from those who are "negative" for that marker that the clinician and/or patient would accept different treatment strategies for the two patients, and 3) the estimates of that magnitude must be reliable.

These criteria relate to establishing the clinical utility of the marker. It is useful to clarify the use of the term "validation" as applied to diagnostic tests. Hunter et al. (19) distinguished three types of validity in terms of genetic tests: "First, there is the question of a test's analytic validity, its ability to accurately and reliability measure the genotype of interest . . .. Second, one must consider clinical validity, or the ability of the test to detect or predict the associated disorder . . .. Finally, there is the issue of the test's clinical utility, or the balance of its associated risks and benefits if it were to be introduced into clinical practice." Clinical utility requires that the test is "actionable," that the clinical context and medical indication for use of the test is clear, and that the magnitude of outcomes or treatment effects associated with different results of the test are sufficiently great as to influence treatment decisions. A serious defect of most retrospective studies of prognostic markers is that the patients are not selected for addressing a defined medical indication for use of the marker. Such studies may establish a correlation with clinical outcome but not the medical utility of the marker.

The consideration of reliably establishing the magnitude of marker effect may be further divided into the following three conditions: 1) the technical and analytical properties of the marker assay must be accurate and/or robust and reproducible; 2) the clinical study design and analysis must be appropriate and adequate to address the utility of a precise intended clinical use; and 3) the results should be verified, or validated, in more than one study set, with similar estimates of the magnitude in separate populations of patients that resemble each other. Each of these conditions is potentially subject to considerable bias in most retrospective studies using archived specimens, especially those of convenience. Even if the investigation is a prospective–retrospective study, careful attention to each of these concerns will reduce the bias and inconsistent results obtained with studies of convenience, and we believe that it will further hasten the introduction of useful tumor markers into clinical practice.

Analytical Concerns

"Analytical validation" generally refers to reproducibility and robustness of the test or assay value. This generally includes minimizing variation with regard to both preanalytical factors, such as tissue collection, processing, storage, and preparation, as well as analytical factors, such as reagent choice, incubation time and conditions, and method of readout (including cut-point determination) (20,21).

For a clinical biomarker evaluation using archived tissues to be interpretable, it is necessary that the assay results from the archived sample reflect what would happen in a true clinical setting. The following are examples of how archived tissue might differ from true clinical specimens.

1) Preanalytical issues. It is possible that samples collected in the past, and specifically for the bank in hand, might be handled differently than they are in current practice. Examples of differences might include whether a precollection diagnostic biopsy was performed (which might affect various gene expression and tissue processes), the time after the sample was removed from the patient and processed (fixed, frozen, etc), procedures for fixation or freezing, how the sample was stored (temperature, exposed to room air, as a tissue block or a section on a slide, etc), and how many cycles it was frozen and thawed.

2) Analytical issues. For a tumor marker study to be sufficient to change clinical practice, the test itself should be ready for clinical practice. For studies to change clinical practice, the investigator should carefully and prospectively plan to use reagents, conditions, and cut points that have been previously determined to be accurate and reproducible. These considerations include fixed reagent supply sources, concentrations, and incubation times among many other possible variables. In addition, the investigator should have demonstrated with statistical confidence the analytical concordance of results between archived specimens and clinical samples for that specific assay. Examples of these concerns include whether the sample was prepared for analysis in a tissue microarray or as a whole section, and whether and how it was subjected to antigen retrieval.

As a precaution against bias that may result from incomplete analytical and preanalytical validation, marker studies using archived specimens should have the assays performed blinded to all clinical data, including treatment and patient outcome.

Clinical Study Design

As noted in the first required condition, the investigator should have a clear idea of the specific intended use for the assay. In general, this will be as a prognostic factor to decide if any further treatment is necessary or as a predictive factor to determine whether a particular type of therapy is likely to be effective. To establish medical utility of a prognostic marker, a randomized trial is sometimes not necessary. For example, a prospective single-arm trial in which chemotherapy is withheld from patients at a low risk of recurrence is used in the portion of the TAILORx clinical trial designed to validate the very favorable prognostic outcomes in the low recurrence score population. Assuming that preanalytical factors are well controlled and match current practice activities and that the clinical data are collected in a fashion typical of a clinical trial, archived tissue from a sufficiently large population of untreated patients may be adequate to permit accurate estimates of recurrence based on tumor marker subgroups for determination of clinical utility of the marker.

Tumor response data from a single-arm phase II clinical trial of a specified treatment can be used to establish the clinical validity of a biomarker for predicting response to that treatment, but a larger randomized trial with a survival or progression-free survival endpoint is generally required to establish the medical utility of the predictive marker.

Suggested Revision of LOEs

In the original American Society of Clinical Oncology LOE scale, "retrospective studies" were determined to be LOE II or worse (3). We now propose an updated revision of the LOE scale, in which more precise definitions are provided for the types of studies that might be used to analyze the clinical utility of a biomarker and in which retrospective studies using archived specimens might reach level I evidence. The LOE for the medical utility of a biomarker relates to key factors involving patients, specimens, assays, and statistical analysis plans (Tables 1 and 2).

Scientifically, the clinical utility of a biomarker in a particular situation is best addressed by a prospective randomized clinical trial (Table 1, category A). Patients are entered, treated, and followed

Table 1. Elements of tumor marker studies that constitute Levels of Evidence determination*

Category Element	A Prospective	B Prospective using archived samples	C Prospective/ observational	D Retrospective/ observational
Clinical trial	PCT designed to address tumor marker	Prospective trial not designed to address tumor marker, but design accommodates tumor marker utility Accommodation of predictive marker requires PRCT	Prospective observational registry, treatment and follow-up not dictated	No prospective aspect to study
Patients and patient data	Prospectively enrolled, treated, and followed in PCT	Prospectively enrolled, treated, and followed in clinical trial and, especially if a predictive utility is considered, a PRCT addressing the treatment of interest	Prospectively enrolled in registry, but treatment and follow-up standard of care	No prospective stipulation of treatment or follow-up; patient data collected by retrospective chart review
Specimen collection, processing, and archival	Specimens collected, processed, and assayed for specific marker in real time	Specimens collected, processed, and archived prospectively using generic SOPs. Assayed after trial completion	Specimens collected, processed, and archived prospectively using generic SOPs. Assayed after trial completion	Specimens collected, processed and archived with no prospective SOPs
Statistical design and analysis	Study powered to address tumor marker question	Study powered to address therapeutic question and underpowered to address tumor marker question	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study	Study not prospectively powered at all. Retrospective study design confounded by selection of specimens for study
		Focused analysis plan for marker question developed before doing assays	Focused analysis plan for marker question developed before doing assays	No focused analysis plan for marker question developed before doing assays
Validation	Result unlikely to be play of chance	Result more likely to be play of chance that A but less likely than C	Result very likely to be play of chance	Result very likely to be play of chance
	Although preferred, validation not required	Requires one or more validation studies	Requires subsequent validation studies	Requires subsequent validation

* PCT = prospective controlled trial; PRCT = prospective randomized controlled trial; SOPs = standard operating practices.

Table 2. Revised determination of Levels of Evidence using elements of tumor marker studies*

Level of evidence	Category from Table 1	Validation studies available
I	A	None required
I	B	One or more with consistent results
II	B	None or inconsistent results
II	C	2 or more with consistent results
III	C	None or 1 with consistent results or inconsistent results
IV–V	D	NA†

* Levels of Evidence (LOEs) revised from those originally proposed by Hayes et al. (3).

† NA = not applicable because LOE IV and V studies will never be satisfactory for determination of medical utility.

prospectively according to a prewritten protocol; the study is prospectively powered specifically to address the tumor marker question; and specimens are collected, processed, and assayed for the marker in real time. The randomized trial will generally not use a "marker strategy design" as described above, however, because of the serious limitations of that design. Although further confirmation in a separate trial of the results gained from a category A prospective trial is always welcome, compelling results from such a trial would be considered definitive and no other validating trial would be required. This strategy was included in the original LOE scale proposed by American Society of Clinical Oncology as LOE I and continues to be the "gold standard."

In the revised LOE scale, a second strategy to obtain level I data would be to perform a tumor marker study using archived specimens from a prospective trial that addresses a therapeutic question (or another marker question) and accommodates the current marker question (Table 1, category B). To evaluate prognostic markers that are intended to identify patients for whom prognosis is so good that further therapy would be withheld, the clinical trial in some cases may not need to be randomized. For example, in the TAILORx study, the low recurrence score group receives only endocrine therapy and is followed to determine if risk of recurrence is as low as predicted by the 21-gene recurrence score. To evaluate a predictive marker, the prospective trial would generally need to be a randomized trial that compares the treatment with an appropriate control treatment. As in study design A, patients are prospectively enrolled, treated, and followed, and specimens are prospectively collected, processed, and archived using generic standard operating procedures. The tumor marker question might be identified during the conduct of the trial or after its completion, but the specification of the tumor marker hypothesis should be based on results completely external to the trial. In fact, tissues archived from the trial should not be assayed until a new protocol has been written that focuses on the evaluation of the specified new marker with a completely specified statistical analysis plan. Before undertaking the study, the assay should be analytically and preanalytically validated for use with archived tissue, and the assay should be performed blinded to the clinical data. Because the trial was designed to address the therapeutic question, it will often be underpowered to establish the statistical significance of treatment by marker interaction (22). It may, however, be adequately sized to reliably identify a large treatment effect in "test-positive" patients, as might be expected for a predictive biomarker. Nevertheless, even with these caveats, results from such a study will be more likely to arise from chance than those from a fully prospective approach.

It is clearly desirable that the available specimens from the archived bank should be representative of the patients who were accrued to the study as a whole, although there is no guarantee that the study patients are themselves representative of the general population of patients. Although there are no minimal requirements that can be universally applicable, we suggest that the correlative study should include at least two-thirds of the total accrued patients or that the patients be selected in a way that strives to avoid selection bias. For example, if the investigator wishes to minimize resource utilization, or wishes to use intrastudy specimen sets for test and validation, one might use a mathematical randomization scheme to select a sample of specimens for study that mirror the known important prognostic and predictive factors of the population as a whole (5).

For a category B study to be sufficient to change practice, we maintain that the results must be confirmed using specimens from a second category B study based on archived tissue from a different trial that has been designed, conducted, and analyzed in a similar, if not identical, manner. The results of these two studies must be equally compelling to change clinical practice. Furthermore, these validation studies need to be performed using the same assay or similar assays that clearly identify the same marker. For example, different investigators have used several different assays for p53 status, including direct sequencing for genetic abnormalities, immunohistochemistry to determine protein expression, or even functional assays. These assays provide very different indications of p53, and therefore, the available data are very difficult to interpret (5). Validation studies must also address the same endpoint and that endpoint should reflect medical utility.

Using nearly 1500 archived specimens collected within a prospective randomized clinical trial, Hayes et al. (23) reported that node-positive, estrogen receptor–positive, and human epidermal growth factor receptor 2–negative patients did not appear to benefit from addition of adjuvant paclitaxel chemotherapy after four cycles of doxorubicin and cyclophosphamide. Although these observations were provocative, results from a completely separate, but similarly designed, prospective randomized clinical trial did not confirm these findings (24), and the question regarding selection of patients for adjuvant paclitaxel remains open (25). Thus, this issue is still considered to be LOE II in Table 2. By contrast, the recently observed association of presence of *KRAS* mutations with lack of benefit from monoclonal antibodies directed against the epidermal growth factor receptor , such as cetuximab and panitumumab (17,18), provides an example of successful use of category B archived samples to establish medical utility. Several prospective randomized trials have demonstrated a small but statistically significant benefit from these antibodies, either alone or in combination with chemotherapy, for treatment of patients with advanced colorectal cancer (26). Preliminary, LOE II or III studies suggested that cetuximab and panitumumab are only active in

patients whose cancers carry a wild-type *KRAS* (27). These data have now been validated in a retrospectively performed study using archived samples from large prospectively randomized clinical trials and therefore would achieve LOE I in our modified scale (Tables 1 and 2) (28).

Category C (Table 1) biomarker studies use prospective patient registries in which subjects are treated and followed according to standards of care. Specimens are collected, processed, and archived prospectively, using generic standard operating procedures, but are assayed after the study has completed patient accrual. Tumor marker studies conducted using these specimens are often not prospectively powered at all. Because of the lack of control of treatment assignment, specimen collection, and data collection, such settings are generally more susceptible to selection biases for patients, specimens, and clinical data that include outcomes. This concern may not be the case in some tightly controlled population-based registries. Category C studies are more likely confounded by unrecognized biases, and their results are more likely to result from chance than those of categories A and B. Category C studies may be validated to LOE II if two or more subsequent studies provide similar results (Table 2). However, it is unlikely that category C studies would ever be sufficient to change practice, except under particularly compelling circumstances.

Category D studies (Table 1) are the most common type of reported tumor marker analyses: studies of convenience in which specimens were collected for unknown reasons, processed and stored in a variety of ways, and happen to be available for assay. The results from these types of studies are highly unstable and likely to be because of chance alone.

Summary

Ideally, any new medical intervention will be adopted into clinical practice only in the setting of level I evidence, and ideally, such evidence is generated in a prospective randomized clinical trial. However, such trials are not always practical. In the case of tumor markers, practice guidelines and the availability of other diagnostic procedures can sometimes make it very difficult to perform new clinical trials because such trials may involve withholding of therapy that is considered standard of care. Even when they are considered ethical, such trials usually require many years to conduct and are quite expensive. For new drug development, in many cases, an analytically validated companion diagnostic test will not be available or the appropriate biological measurement may not be clear at the time that the pivotal trials of the drug are initiated, as for the use of *KRAS* mutation as a predictive biomarker for EGFR inhibitors in colorectal cancer (17,18,28).

Archived tissue specimens from high-quality datasets can therefore be of great importance for establishing the medical utility of a prognostic or predictive biomarker. We argue that it is appropriate to use archived tissue specimens from large prospective clinical trials to do so. For such an evaluation to be more useful than just for generating hypotheses, however, several conditions must be satisfied:

1) Archived tissue, adequate for a successful assay, must be available on a sufficiently large number of patients from the pivotal trials to permit appropriately powered analyses and to ensure that the patients included in the biomarker evaluation are clearly representative of the patients in the pivotal trials. Although no minimal requirement can be stated as universally applicable, we would suggest that samples from at least two-thirds of the patients be available for analysis.

2) Substantial data on analytical validity of the test must exist that ensure that results obtained from the archived specimens will closely resemble those that would have been obtained from analysis of specimens collected in real time. Assays should be conducted blinded to the clinical data.

3) The analysis plan for the biomarker evaluation must be completely developed before the performance of the biomarker assays. Both the analysis plan for the biomarker study and the design of the trial(s) whose samples were selected for analysis should be appropriate for the evaluation of a companion diagnostic had it been undertaken at the outset. The analysis should be focused on a single, completely defined, diagnostic classifier. For multigene classifiers, the mathematical form of combining the individual components, weights, and cut points should be specified beforehand. In general, the analysis should not be exploratory, and practices that might lead to a false-positive conclusion should be avoided.

4) The results must be validated in at least one or more similarly designed studies using the same assay techniques.

Physicians need improved tools for selecting treatments for individual patients. Cancers of the same primary site are in many cases heterogeneous in molecular pathogenesis, clinical course, and treatment responsiveness. Current approaches for treatment development, evaluation, and use result in treatment of many patients with ineffective drugs. Advances in cancer genomics and biotechnology are providing increased opportunities for development of more effective therapeutics and prognostic and predictive biomarkers to inform their use. These opportunities have enormous potential benefits for patients and for containing health-care costs. However, the complexity of cancer biology and the increased complexity of development of biomarkers with drugs offer formidable challenges to the transition to a more predictive oncology. In some cases, it is either ethically or practically impossible to evaluate the medical utility of prognostic and predictive biomarkers in a fully prospective manner.

It is essential to ensure that cancer patients are offered the benefits of valuable prognostic and predictive tests as soon as they are rigorously and reliably evaluated. In this article, we have tried to clarify some of the uncertainty in the field about the validation of prognostic and predictive biomarkers and to propose an update of a LOE schema that has been widely used for evaluating the medical utility of biomarkers in oncology. We believe that this update is important for improving the conduct of validation studies and, in some cases, for expediting the adoption of important diagnostic tools.

References

1. Ransohoff DE. Rules of evidence for cancer molecular-marker discovery and validation. *Nat Rev Cancer.* 2004;4(4):309–314.
2. Henry NL, Hayes DF. Uses and abuses of tumor markers in the diagnosis, monitoring and treatment of primary and metastatic breast cancer. *Oncologist.* 2006;11(6):541–552.

3. 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.
4. Locker GY, Hamilton S, Harris J, et al. ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol.* 2006;24(33):5313–5327.
5. 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(33):5287–5310.
6. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst.* 2005;97(16):1180–1184.
7. Hayes D, Ethier S, Lippman M. New guidelines for reporting tumor marker studies in breast cancer research and treatment: REMARK. *Breast Cancer Research and Treatment.* 2006;100(1):237–238.
8. Simon R. A roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol.* 2005;23(29):7332–7341.
9. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol.* 2005;23(9):2020–2027.
10. Freidlin B, Simon R. Adaptive signature design: an adaptive clinical trial design for generating and prospectively testing a gene expression signature for sensitive patients. *Clin Cancer Res.* 2005;11(21):7872–7878.
11. Jiang W, Freidlin B, Simon R. Biomarker adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst.* 2007;99(13):1036–1043.
12. Simon R. Using genomics in clinical trial design. *Clin Cancer Res.* 2008;14(19):5984–5993.
13. Simon R. Designs and adaptive analysis plans for pivotal clinical trials of therapeutics and companion diagnostics. *Expert Rev Mol Diagn.* 2008; 2(6):721–729.
14. Simon R, Wang SJ. Use of genomic signatures in therapeutics development. *Pharmacogenomics J.* 2006;6(3):166–173.
15. Bogaerts J, Cardoso F, Buyse M, et al. Gene signature evaluation as a prognostic tool: challenges in the design of the MINDACT trial. *Nat Clin Pract Oncol.* 2006;3(10):540–551.
16. Sparano JA. TAILORx: Trial Assigning Individualized Options for Treatment(Rx). *Clin Breast Cancer.* 2008;7(4):347–350.
17. Amado RG, Wolf M, Peeters M, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol.* 2008;26(10):1626–1634.
18. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N Engl J Med.* 2008;359(17):1757–1765.
19. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle—will we get our wish. *N Engl J Med.* 2008;358(2):105–107.
20. Goldstein NS, Hewitt SM, Taylor CR, Yaziji H, Hicks DG. Recommendations for improved standardization of immunohistochemistry. *Appl Immunohistochem Mol Morphol.* 2007;15(2):124–133.
21. Goldstein NS, Hunter S, Forbes S, Odish E, Tehrani M. Estrogen receptor antibody incubation time and extent of immunoreactivity in invasive carcinoma of the breast: the importance of optimizing antibody avidity. *Appl Immunohistochem Mol Morphol.* 2007;15(2):203–207.
22. Peterson B, George SL. Sample size requirements and length of study for testing interaction in a 1 x k factorial design when time-to-failure is the outcome. *Control Clin Trials.* 1993;14(6):511–522.
23. Hayes DF, Thor AD, Dressler LG, et al. HER2 and response to paclitaxel in node-positive breast cancer. *N Engl J Med.* 2007;357(14):1496–1506.
24. Martin M, Rodriguez-Lescure A, Ruiz A, et al. Randomized phase 3 trial of fluorouracil, epirubicin, and cyclophosphamide alone or followed by paclitaxel for early breast cancer. *J Natl Cancer Inst.* 2008;100(11): 805–814.
25. Hudis C, Dang C. The taxane limbo: how low can we go? *J Natl Cancer Inst.* 2008;100(11):761–763.
26. Labianca R, La Verde N, Garassino MC. Development and clinical indications of cetuximab. *Int J Biol Markers.* 2007;22(1 suppl 4):S40–S46.
27. Messersmith WA, Ahnen DJ. Targeting EGFR in colorectal cancer. *N Engl J Med.* 2008;359(17):1834–1836.
28. Allegra C, Jessup JM, Somerfield MR, et al. American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy [published online ahead of print]. *J Clin Oncol.* 2009;27(12):2091–2096.

Funding

Supported in part by the Fashion Footwear Charitable Foundation of New York/QVC Presents Shoes on Sale (D.F.H.).

Notes

The authors take full responsibility for the analysis and interpretation of the data, the writing of the manuscript, and the decision to submit the manuscript for publication.

The authors would like to acknowledge very helpful discussions with their colleagues and members of the American Society of Clinical Oncology Tumor Markers Guidelines Committee that contributed to much of this work. D. F. Hayes reports research and other support from Astra Zeneca, GlaxoSmithKline, Pfizer, Novartis, Ayerst-Wyeth, Veridex (Johnson & Johnson), and Sanofi-Aventis, Predictive Biosciences, Curalogie SAB, Incyte Corporation, DNA Repair, and Compendia Bioscience.

Manuscript received December 5, 2008; revised August 4, 2009; accepted August 27, 2009.

BlueCross BlueShield Association

EVIDENCE SUMMARY

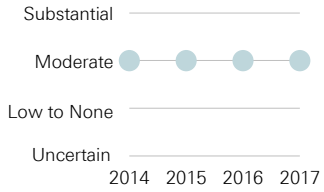
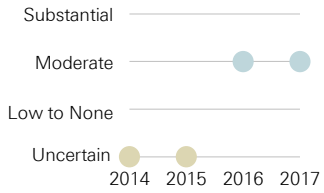
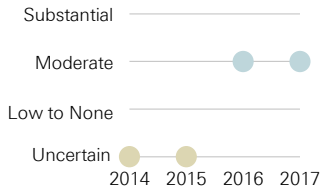
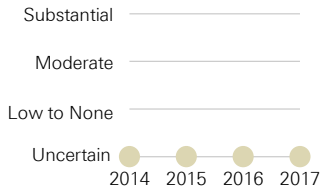
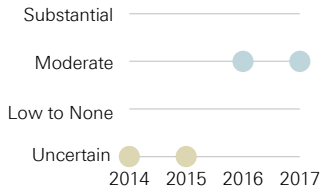
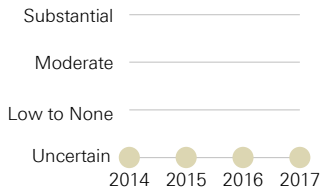
Laboratory tests have been developed that detect the expression, via messenger RNA, of many different genes in breast tumor tissue and combine the results into a prediction of distant recurrence risk for women with early-stage breast cancer. Test results may help providers and patients decide whether to include adjuvant chemotherapy in the postsurgical management of breast cancer, to alter treatment in patients with ductal carcinoma in situ, or to recommend extended endocrine therapy in patients who are recurrence-free at 5 years. This report summarizes the evidence of 5 tests and is organized by indication.

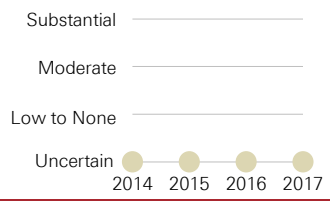
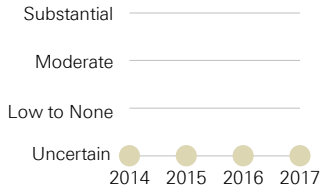
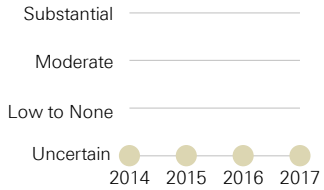
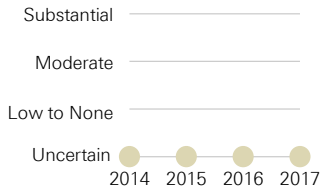
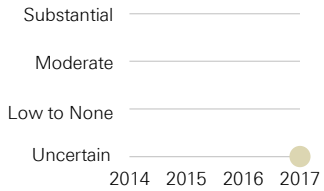
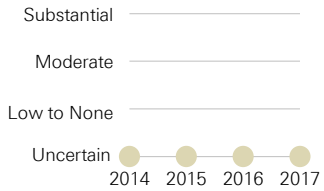
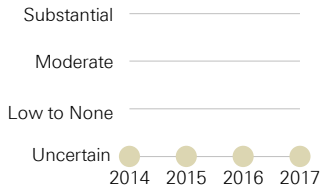
The objective of this evidence review is to determine whether Oncotype DX, EndoPredict, Breast Cancer Index, MammaPrint, and Prosigna testing have clinical utility in aiding decisions about breast cancer treatment.

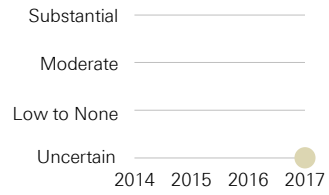
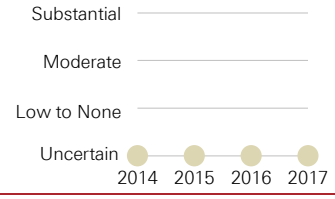
Populations	Interventions	Comparators	Outcomes
Individuals: • With early-stage node-negative invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with Oncotype DX (21-gene signature)	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with EndoPredict	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with the Breast Cancer Index	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with MammaPrint (70-gene signature)	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with Prosigna	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-positive invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with Oncotype DX (21-gene signature)	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-positive invasive breast cancer	Interventions of interest are: • Gene expression	Comparators of interest are: • Clinical risk	Relevant outcomes include: • Disease-specific survival

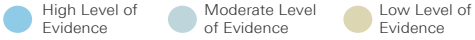
considering adjuvant chemotherapy	profiling with EndoPredict	prediction algorithms	• Change in disease status
Individuals: • With early-stage node-positive invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with MammaPrint (70-gene signature)	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-positive invasive breast cancer considering adjuvant chemotherapy	Interventions of interest are: • Gene expression profiling with Prosigna	Comparators of interest are: Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With ductal carcinoma in situ considering radiotherapy	Interventions of interest are: • Gene expression profiling with the Oncotype DX Breast DCIS Score	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy	Interventions of interest are: • Gene expression profiling with Oncotype DX (21-gene signature)	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy	Interventions of interest are: • Gene expression profiling with EndoPredict	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy	Interventions of interest are: • Gene expression profiling with the Breast Cancer Index	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy	Interventions of interest are: • Gene expression profiling with MammaPrint (70-gene signature)	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status
Individuals: • With early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy	Interventions of interest are: • Gene expression profiling with Prosigna	Comparators of interest are: • Clinical risk prediction algorithms	Relevant outcomes include: • Disease-specific survival • Change in disease status

Overview by Evidence Review Indications

Indication 1: Individuals with early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with Oncotype DX (21-gene signature). **The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.**	
Indication 2: Individuals with early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with EndoPredict. **The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.**	
Indication 3: Individuals with early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with the Breast Cancer Index. **The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.**	
Indication 4: Individuals early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with MammaPrint (70-gene signature). **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 5: Individuals with early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with Prosigna. **The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.**	
Indication 6: Individuals with early-stage node-positive invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with the 21-gene signature (Oncotype DX). **The evidence is insufficient to determine the effects of the technology on health outcomes.**	

Indication 7: Individuals with early-stage node-positive invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with EndoPredict. **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 8: Individuals with early-stage node-positive invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with MammaPrint (70-gene signature). **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 9: Individuals with early-stage node-positive invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with Prosigna. **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 10: Individuals with ductal carcinoma in situ considering radiotherapy who receive gene expression profiling with the Oncotype DX Breast DCIS Score. **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 11: Individuals with early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy, who receive gene expression profiling with Oncotype DX (21-gene signature). **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 12: Individuals with early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy, who receive gene expression profiling with EndoPredict. **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 13: Individuals with early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy, who receive gene expression profiling with the Breast Cancer Index. **The evidence is insufficient to determine the effects of the technology on health outcomes.**	

Indication 14: Individuals with early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy, who receive gene expression profiling with MammaPrint (70-gene signature). **The evidence is insufficient to determine the effects of the technology on health outcomes.**	
Indication 15: Individuals with early-stage node-negative invasive breast cancer, recurrence-free at 5 years, considering extended endocrine therapy, who receive gene expression profiling with Prosigna. **The evidence is insufficient to determine the effects of the technology on health outcomes.**	



BACKGROUND

NEWLY DIAGNOSED BREAST CANCER

Most women with newly diagnosed breast cancer in the United States present with early-stage or locally advanced (ie, nonmetastatic) disease. However, almost a third of women who are disease-free after initial local and regional treatment develop distant recurrences during follow-up.[1] Current breast cancer treatment regimens involve systemic adjuvant chemotherapy, hormonal therapy, biologic therapy, or a combination, depending on patients' baseline level of recurrence risk, hormonal markers, and risk tolerance.

Women whose tumors are positive for human epidermal growth factor receptor 2 (*HER2*) should receive adjuvant therapy with a *HER2*-directed therapy (trastuzumab with or without pertuzumab). Decision-making about adjuvant biologic therapy for women with *HER2*-positive cancer is not discussed here. This review focuses on 3 decision points:

1. ***The decision to pursue adjuvant chemotherapy following locoregional therapy, with or without neoadjuvant chemotherapy, based on the predicted risk of recurrence, for women who are hormone receptor–positive but HER2-negative.*** The use of adjuvant chemotherapy reduces the risk of breast cancer recurrence but carries risks of systemic toxicity. The risk:benefit ratio must be balanced for each patient, with a higher likelihood of net health benefits for patients with a greater baseline predicted the risk of recurrence. Some of the individual considerations are discussed below. *HER2* expression independently confers an unfavorable prognosis, but assessing the independent effects of *HER2* is complicated in the presence of targeted therapy; therefore, we focus specifically on patients without *HER2* expression.
2. ***The decision to pursue adjuvant endocrine therapy from 5 to 10 years for women who are hormone receptor–positive but HER2-negative and who have survived without recurrence for 5 years.*** For patients with hormone receptor–positive tumors, the use of adjuvant endocrine therapy (tamoxifen and/or an aromatase inhibitor, with or without ovarian suppression) for 5 to 10 years after an initial diagnosis has support in clinical practice. The 2017 guidelines from the

National Comprehensive Cancer Network recommend extended endocrine therapy.[2] The American Society for Clinical Oncology's 2014 focused update to its guidelines on adjuvant endocrine therapy for women with hormone receptor–positive breast cancer have recommended 10 years of tamoxifen for pre- or perimenopausal women, and a total of seven to ten years of endocrine therapy, following 1 of 4 regimens that include tamoxifen with or without an aromatase inhibitor for postmenopausal women.[3,4]

3. ***The decision to pursue adjuvant radiotherapy in women with ductal carcinoma in situ (DCIS).*** Adjuvant radiotherapy reduces the risk of local recurrences but has not been shown to change the risk of distant recurrence or mortality. There may be a group of patients for whom the reduction in risk for local recurrence may not be large enough to justify the risks of radiotherapy.

SELECTION OF ADJUVANT CHEMOTHERAPY BASED ON RISK OF RECURRENCE

An important part of treatment planning for women with breast cancer involves determining which patients could benefit from adjuvant cytotoxic chemotherapy. For example, for women with early-stage invasive breast cancer (ie, cancer extending beyond the basement membrane of the mammary ducts into adjacent tissue), adjuvant cytotoxic chemotherapy consistently provides approximately a 30% relative risk reduction in 10-year breast cancer mortality regardless of patients' baseline prognosis. However, the absolute benefit of chemotherapy depends on the underlying or baseline risk of recurrence. Women with the best prognosis have tumors that are small, early-stage, estrogen receptor–positive, and lymph node to negative (Table 1 shows recurrence risk for estrogen receptor–positive cancers for patients followed in the International Breast Cancer Study Group).[1] Patients may have received no adjuvant treatment, or adjuvant tamoxifen and/or adjuvant chemotherapy. These women have an approximately 15% 10-year risk of recurrence with tamoxifen alone; approximately 85% of these patients could avoid the toxicity of adjuvant cytotoxic chemotherapy if they could be accurately identified. Conventional risk classifiers (eg, Adjuvant! Online) estimate recurrence risk by considering criteria such as tumor size, type, grade, and histologic characteristics; hormone receptor status; and number of affected lymph nodes. Consensus guidelines for defining receptor status exist[5]; however, no single classifier is considered a criterion standard. As a result, a substantial number of patients are treated with chemotherapy who fail to benefit. Better predictors of recurrence risk could help women's decision-making, some of whom may prefer to avoid chemotherapy if assured their risk is low.

Table 1. Effect of Nodal Involvement, Tumor Size, and Grade on Annual Recurrence Hazard in Estrogen Receptor–Positive Breast Cancers

	Recurrence, Hazard[a] (SE), %				
	Years				
Nodes	0-5	5-10	10-15	15-20	20-25
0	5.8 (0.5)	3.3 (0.4)	2.0 (0.4)	2.1 (0.4)	1.1 (0.4)
1 to 3	9.5 (0.6)	5.8 (0.6)	3.0 (0.5)	3.5 (0.7)	1.5 (0.6)
≥4	17.2 (0.9)	10.9 (1.2)	5.9 (1.2)	3.8 (1.2)	1.3 (0.9)
Size					
≤2 cm	7.0 (0.4)	4.8 (0.4)	2.9 (0.4)	2.7 (0.5)	1.5 (0.5)
>2 cm	12.9 (0.6)	6.1 (0.6)	2.9 (0.5)	2.7 (0.5)	1.1 (0.5)
Grade					
1	5.8 (0.6)	4.9 (0.7)	3.6 (0.7)	4.0 (0.9)	0.7 (0.5)
2	9.6 (0.5)	6.3 (0.5)	2.8 (0.4)	2.7 (0.5)	1.8 (0.5)
3	14.1 (0.8)	4.1 (0.6)	2.5 (0.6)	2.4 (0.7)	0.4 (0.4)

Adapted from Colleoni et al (2016).[1]
[a] Number of events occurring within a time interval divided by the total years of follow-up during the interval accrued by patients at risk during the interval. Patients may have received no adjuvant treatment or have been treated with adjuvant tamoxifen and/or adjuvant chemotherapy.

SELECTION OF EXTENDED ENDOCRINE THERAPY

Randomized controlled trials have established that 5 years of tamoxifen improves mortality in women with hormone receptor–positive breast cancer. A 2011 individual patient data meta-analysis by the Early Breast Cancer Trialists' Collaborative Group, including 20 trials (total N=21,457 patients) found that 5 years of tamoxifen in estrogen receptor–positive disease reduced the risk of recurrences by almost 50% over 10 years on the relative scale; breast cancer mortality was decreased by 29% through 15 years.[6]

For patients with early-stage, invasive breast cancer that is hormone receptor–positive, the use of endocrine therapy (tamoxifen and/or an aromatase inhibitor, with or without ovarian suppression) for 5 to 10 years following initial diagnosis has support in national guidelines.[2,3,7] However, the regimens available and the evidence to support them vary.

Randomized controlled trials published recently have shown that extended endocrine therapy decreases the risk of recurrence. The American Society for Clinical Oncology and the National Comprehensive Cancer Network guidelines were informed primarily by results of the ATLAS trial, which compared 5 and 10 years of tamoxifen[8] and the subsequent aTTom trial (reported in abstract form).[9] In both trials, in women who were hormone receptor–positive and had completed 5 years of tamoxifen, 5 years of extended tamoxifen was associated with improvements in breast cancer–specific mortality; ATLAS showed improvements in overall survival (see Table 2).

Three previously reported randomized trials of extended tamoxifen treatment had mixed findings: Tormey et al (1996; total N=194 patients),[10] the National Surgical Adjuvant Breast and Bowel Project (Fisher et al, 2001; total N=1172 patients),[11] and the Scottish Cancer Trials Breast Group (Stewart et al, 2001; total N=342 patients)[12] (see Table 2).

Overall, the available trial evidence would suggest that 10 years of tamoxifen in pre- or postmenopausal women can be linked with improved survival while trials of extended aromatase inhibitors in different populations of hormone receptor–positive patients have had more mixed results.

Table 2. Randomized Trials Evaluating Adjuvant Extended Endocrine Therapies for Hormone Receptor–Positive Breast Cancer

Study	Population	Comparators	Breast Cancer–Specific Mortality		Overall Mortality	
			Event RR (95% CI)	p	Event RR (95% CI)	p
Extended tamoxifen						
ATLAS (2013)[8]	6846 women with ER-positive, early breast cancer, after 5 y of tamoxifen	Continue tamoxifen to 10 y (n=3428) vs stop tamoxifen at 5 y (n=3418)	0.83 (0.72 to 0.96) (331/3428 vs 397/3418)	0.01	0.87 (0.78 to 0.97) 722 (639/3428 vs 722/3418)	0.01
aTTom (2013)[9]	6953 women with ER-positive or untested breast cancer, after 5 y of tamoxifen	Continue tamoxifen to 10 y (n=3468) vs stop tamoxifen at 5 y (n=3485)	**10 years** 392/3468 intervention vs 442/3485 control **Years 5-9** 1.03 (0.84 to 1.27)	0.05	**10 years** 849/3468 intervention vs 910/3485 control **Years 5-9** 1.05 (0.90 to 1.22)	0.1

Study	Population	Comparators	Breast Cancer–Specific Mortality		Overall Mortality	
			After year 9 0.77 (0.64 to 0.92)		**After year 9** 0.86 (0.75 to 0.97)	
Extended aromatase inhibitor						
ABCSG (2007)[13]	856 post-menopausal women with ER- and/or PR-positive breast cancer, after 5 y of tamoxifen	Anastrozole for 3 y (n=386) vs no further therapy (n=466)			**5 years** 10.3% anastrozole vs 11.7% control **Event HR (95% CI)** 0.89 (0.59 to 1.34)	0.57
			Breast Cancer–Specific Survival		Overall Survival	
NCIC CTG MA.17 trial (2003, 2005)[14,15]	5187 post-menopausal women with ER- and/or PR-positive early breast cancer, after 5 y tamoxifen	Continue letrozole to 10 y (n=2593) vs stop tamoxifen at 5 y (n=2594)	**48 Months** 94.4% letrozole vs 89.8% placebo **Event HR** 0.58 (0.45 to 0.76)	<0.001	**48 Months** 96% letrozole vs 94% placebo **Event HR** 0.76 (0.48 to 0.21) **40 Months** 95.4% letrozole vs 95% placebo **Event HR** 0.82 (0.57 to 1.19)	0.25 0.3
NSABP (2008)[16]	1598 post-menopausal women with ER- and/or PR-positive early breast cancer, after 5 y of tamoxifen	Planned comparison: 5 y exemestane vs 5 y placebo. Accrual stopped (n=1598 randomized), and crossover allowed after results of NCIC CTG available: • Exemestane: 783 randomized, 560 continued after unblinding) • Placebo: 779 randomized, 334 crossed over to exemestane after unblinding	**48 Months** ITT: 91% exemestane vs 89% placebo	0.07		

ABCSG: Austrian Breast and Colorectal Cancer Study Group; CI: confidence interval; ER: estrogen receptor; HR: hazard ratio; ITT: intention to treat; NCIC CTG: National Cancer Institute Clinical Trials Group; NSABP: National Surgical Adjuvant Breast and Bowel Project; PR: progesterone receptor; RR: rate ratio.

In addition to the trials published in full-length form, 3 trials were presented in early 2017 evaluating extended endocrine therapy in postmenopausal women (NSABP-42 [NCT00382070]: 10 years vs 5 years of letrozole; DATA [NCT00301457]: 6 years vs 3 years of anastrozole; and IDEAL [NTR3077] 10 years vs 7.5 years of letrozole) did not meet their primary end points.

CLINICAL USES OF GENE EXPRESSION SIGNATURES FOR BREAST CANCER

In other clinical scenarios involving breast cancer, accurate assessment of prognosis may affect the decision to offer certain treatments. Recently, several groups have identified panels of gene expression markers ("signatures") that appear to predict the baseline risk of invasive breast cancer recurrence after surgery, radiotherapy, and endocrine therapy (for hormone receptor–positive tumors). Several gene expression tests commercially available in the United States are listed in Table 3. If these panels are more accurate risk predictors than current conventional classifiers, they could be used to aid decision-making on adjuvant treatments without greatly affecting disease-free survival and overall survival. This review focuses on gene expression profiling panels that have prognostic or predictive ability in individuals with early-stage, invasive breast cancer with known estrogen receptor and progesterone receptor and *HER2* status. The proposed clinical utility of these tests varies by the clinical context; these specific indications are discussed in this review:

1. Prognosis and/or prediction of treatment response in patients with node-negative, early-stage, hormone receptor–positive, *HER2*-negative invasive breast cancer who will receive adjuvant hormonal therapy for the purpose of determining whether patients can avoid adjuvant cytotoxic chemotherapy.
2. Prognosis and/or prediction of treatment response in patients with node-positive (1-3 nodes), hormone receptor–positive, early-stage, *HER2*-negative invasive breast cancer who will receive adjuvant hormonal therapy for the purpose of determining whether patients can avoid adjuvant cytotoxic chemotherapy.
3. Prognosis and/or prediction of treatment response in patients with DCIS for the purpose of determining whether patients can avoid radiotherapy.
4. Prognosis and/or prediction of treatment response in patients with node-negative, early-stage, hormone receptor–positive, *HER2*-negative invasive breast cancer, receiving adjuvant hormonal therapy, who have survived without progression to 5 years postdiagnosis, for the purpose of determining whether patients will continue adjuvant hormonal therapy.

For each of these indications, clinical trials have shown that there is some clinical benefit to receiving the additional therapy under consideration. However, each additional treatment has potential adverse effects. If a patient subgroup can be defined that has an extremely low risk of distant recurrence, or a subgroup can be defined that does not respond to the treatment, then the additional treatment can be forgone with little effect on cancer outcome due to the low risk of poor outcome or lack of response to treatment.

Table 3. Gene Expression Tests Reporting Recurrence Risk for Breast Cancer Considered Herein

Test	Manufacturer	Description
Oncotype DX®	Genomic Health (Redwood City, CA)	21-gene RT-PCR; identifies 3 groups as low, intermediate, and high risk for distant recurrence
EndoPredict®	Sividon Diagnostics (acquired by Myriad [Salt Lake City, UT] in 2016)	12-gene real-time RT-PCR; gene expression molecular score alone (EP) or EP is combined with the clinical parameters of tumor size and number positive lymph nodes (EPclin), resulting in classifications of EP low, EP high, EPclin low, or EPclin high risk for distant recurrence
Breast Cancer	Biotheranostics (San Diego, CA)	Combines MGI and the HOXB13:IL17BR Index

Test	Manufacturer	Description
Index[SM] Prognostic		measured using RT-PCR; identifies 2 groups as low or high risk for distant recurrence
MammaPrint®	Agendia (Amsterdam, The Netherlands)	70-gene DNA microarray; identifies 2 groups as low or high risk for distant recurrence
Prosigna®	NanoString Technologies (Seattle, WA)	Gene expression profile is assessed by the nCounter digital platform system to determine similarity with prototypic profiles of PAM50 genes for breast cancer; identifies 3 categorical ROR groups (ROR-low, ROR-intermediate, ROR-high)

MGI: Molecular Grade Index; PAM50: prediction analysis of microarray 50-gene set; ROR: risk of relapse; RT-PCR: reverse transcriptase polymerase chain reaction; EP: expression profile.

Additional commercially available tests may provide some prognostic or predictive information for breast cancer. Tests intended to assess estrogen receptor, progesterone receptor, and *HER2* status, such as TargetPrint® (Agendia; via quantitative microarray), are outside the scope of this review. In addition, tests that do not provide a specific recurrence risk are outside the scope of this review.

Other commercially available biomarkers are designed to provide information about tumors' molecular subtypes (ie, luminal A, luminal B, *HER2* type, and basal type). Prosigna was initially offered as a molecular subtype test. The BluePrint® 80-gene molecular subtyping assay is offered in combination with MammaPrint to augment predictive data about response to chemotherapy.

DECISION FRAMEWORK FOR EVALUATING BREAST CANCER BIOMARKERS

Simon et al Framework

Many studies have investigated individual biomarkers or combinations of biomarkers associated with breast cancer outcomes. Determining which studies constitute sufficient evidence that the test or biomarker is likely to be clinically useful depends on attributes of the test such as its performance and the quality of the study generating the results. Simon et al (2009) have described a framework to evaluate prognostic biomarker evidence.[17] Study designs, such as prospective clinical trials or previously conducted clinical trials with archived tumor samples, constitute stronger evidence than studies with less planned and systematic patient recruitment and data collection. Randomized trials allow determination of treatment-biomarker interactions that may be clinically important. In some clinical scenarios, demonstration of a treatment-biomarker interaction is not critical, because the decision to withhold chemotherapy in a low-risk group (to avoid chemotherapy-related morbidity) does not require the presence of a biomarker-treatment interaction. The study must generate an absolute estimate of outcomes in the patient group of interest that would result in a change in management (eg, withholding of chemotherapy), and the study must have sufficient precision (narrow confidence intervals). Results of the same test across studies should show the consistency of results and more than 1 study demonstrating the desired result should be available. Simon has proposed that at least 2 Simon category B studies showing results consistent with clinical utility are necessary to demonstrate adequate evidence of a biomarker.[17]

Breast Cancer–Specific Outcomes

The main outcome of interest for this review is 10-year distant recurrence-free survival. Distant recurrence is a hallmark of advanced breast cancer and thus more informative of overall survival than

disease-free survival. Disease-free survival also includes local recurrence, which has a much better treatment prognosis than distant disease. For the extended endocrine indications in this review, the main outcome of interest is 10-year distant recurrence-free survival conditional on recurrence-free survival for 5 years.

Decisions to undergo or forgo adjuvant therapy (chemotherapy or endocrine) depend on how a woman values the potential benefit of lower recurrence risk relative to the harms of treatment. The balance of benefits and harms determines the thresholds that inform decisions.[18,19] Most women will accept substantial adverse events for even modest benefit. For example, Simes et al (2001) interviewed 104 Australian women with breast cancer treated with cytotoxic chemotherapy and elicited preferences to undergo chemotherapy according to probable gain in survival.[20] With an expected survival of 5 years without chemotherapy, 73% said they would accept chemotherapy for an increased survival of 6 months or less; with an expected survival of 15 years, 39% would accept treatment for a gain of 6 months. Duric et al (2005) found 64% to 84% of 97 women expressing a willingness to undergo chemotherapy for a 1-year improvement in life expectancy or 3% increase in survival rates.[21] About half felt a single day would justify adjuvant chemotherapy. A major difference between the 2 studies was that the chemotherapy regimen in Duric et al was less toxic. Thewes et al (2005) adopted the same approach for adjuvant endocrine therapy preferences in 102 premenopausal women with early-stage breast cancers.[22] Among women having a baseline life expectancy of 5 years, 61% said they would accept endocrine therapy for a 6-month increase in life expectancy and 79% for 1 year; rates were similar if the baseline life expectancy was 15 years. These proportions are close to those for adjuvant chemotherapy found by Duric.

How these estimates correspond to the distant recurrence rates reported in prognostic studies is imprecise, but Henderson (2015) has suggested that below a recurrence threshold of 10% many patients will not elect adjuvant chemotherapy owing to the small absolute benefit.[23] He also noted that a majority of those patients are older with small node-negative tumors. That interpretation is consistent with a recent study of 81 women by Hamelinck et al (2016) who found that 78% of women ages 40 to 49 years, 88% ages 50 to 59, 59% ages 60 to 69, and 40% age 70 or older would accept adjuvant chemotherapy for a 0% to 10% absolute decrease in recurrence risk (see Table 4).[24] There was a wide range of minimally required absolute benefits, with the majority accepting chemotherapy for an absolute benefit of 1% to 5%. At a given age range, fewer women expressed a willingness to accept adjuvant endocrine therapy than chemotherapy for a given mortality benefit.

Table 4. Patient Preferences for Undergoing Adjuvant Therapy for <10% Reduction in Recurrence Risk

Age Range, y	Proportion That Would Accept 1% to 10% Benefit	
	Chemotherapy, %	Endocrine, %
40-49	78	78
50-59	88	44
60-69	59	63
≥70	40	46

Adapted from Hamelinck et al (2016).[24]

REGULATORY STATUS

Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; laboratory-developed tests must meet the general regulatory standards of the Clinical Laboratory Improvement Amendments. Oncotype DX® and other tests listed herein are available under the auspices of the Clinical Laboratory Improvement Amendments. Laboratories that offer laboratory-developed tests must be licensed by the Clinical Laboratory Improvement Amendments for high-complexity testing. To date, the U.S. Food and Drug Administration (FDA) has chosen not to require any regulatory review of this test.

In February 2007, MammaPrint® (Agendia) was cleared for marketing by the FDA through the 510(k) process for the prediction of breast cancer metastasis. In January 2015, MammaPrint® was cleared for marketing by the FDA through the 510(k) process for use in fresh-frozen, paraffin-embedded breast cancer tissue.

In September 2013, Prosigna® was cleared for marketing by the FDA through the 510(k) process. Moreover, the FDA determined that Prosigna® was substantially equivalent to MammaPrint®.

FDA product code: NYI.

Currently, the Breast Cancer Index℠ (Biotheranostics) and EndoPredict® (distributed by Myriad) are not FDA-approved.

RATIONALE

This evidence review was created in July 2004 and has been updated regularly with searches using MEDLINE database. The most recent literature update for all indications was performed through September 11, 2017 (see Appendix Table 1 for genetic testing categories).

Validation of the clinical use of any genetic test focuses on 3 main principles: (1) analytic validity, which refers to the technical accuracy of the test in detecting a variant that is present (or in excluding a variant that is absent); (2) clinical validity, which refers to the diagnostic performance of the test (sensitivity, specificity, positive and negative predictive values) in detecting clinical disease; and (3) clinical utility (ie, how the results of the diagnostic test will be used to change management of the patient and whether these changes in management lead to clinically important improvements in health outcomes).
The following is a summary of the key literature.

ASSAYS OF GENETIC EXPRESSION IN TUMOR TISSUE

Clinical Context and Test Purpose

The purpose of assays of genetic expression in tumor tissue in patients with early-stage node-negative or node-positive invasive breast cancer considering adjuvant chemotherapy; in patients with ductal carcinoma in situ (DCIS) considering radiotherapy; and in patients with early-stage node-negative invasive breast cancer, recurrence-free at 5 years considering extended endocrine therapy, is to determine risk of recurrence, which informs decisions about potential breast cancer treatment.

The question addressed in this evidence review is: Does the use of assays of genetic expression in tumor tissue improve the net health outcome in individuals with breast cancer?

The following PICOTS were used to select literature to inform this review.

Patients
The populations of interest include: patients with early-stage node-negative or node-positive invasive breast cancer considering adjuvant chemotherapy; patients with DCIS considering radiotherapy; and patients with early-stage node-negative invasive breast cancer, recurrence-free at 5 years considering extended endocrine therapy.

Interventions
The interventions of interest are assays of genetic expression in tumor tissue (Oncotype DX, EndoPredict, Breast Cancer Index [BCI], MammaPrint, Prosigna).

Comparators
The comparators of interest for all assays are clinical risk prediction algorithms.

Outcomes
Outcomes of interest for all assays are disease-specific survival and change in disease status.

If patients with early-stage invasive breast cancer are classified as low risk for distant recurrence, patients may be able to forgo adjuvant chemotherapy safely.

If patients with DCIS are classified as low risk for distant recurrence, they may be able to safely forgo radiotherapy.

If patients with invasive breast cancer who are recurrence-free for 5 years are classified as low risk for distant recurrence, patients may be able to safely forgo extended endocrine therapy.

Timing
The assays would be performed following the diagnoses of early-stage node-negative or node-positive invasive breast cancer, because patients are considering adjuvant chemotherapy.

The assays would be performed following the diagnosis of DCIS, because patients are considering RT.

The assays would be performed after 5 years of no recurrence of early-stage node-negative invasive breast cancer because patients are considering extended endocrine therapy.

Setting
The setting is a laboratory meeting general regulatory standards of the Clinical Laboratory Improvement Amendments.

EARLY-STAGE NODE-NEGATIVE INVASIVE BREAST CANCER CONSIDERING ADJUVANT CHEMOTHERAPY

Oncotype DX (21-Gene Assay)

Clinical Validity
We identified 4 studies meeting selection criteria (see Appendix 1).[25-28] The studies derive from 3 completed randomized trials and thus are all Simon category B studies. The study by Paik et al (2006) evaluated patients from a trial in which the subjects were part of the training set used to develop the Oncotype algorithm, so its results might be biased.[27] The study by Tang et al (2011)[28] represents the

same results as Paik et al (2004),[26] but categorized by the Adjuvant! Online clinical risk stratifier (see Table 5).

Across all 3 studies in which patients were solely classified by Recurrence Score (RS), the 10-year risk of distant recurrence was low in the RS low category. Ten-year distant recurrence rates were all below the 10% threshold suggested by Henderson (2015),[23] and the upper limit of the 95% confidence intervals (CIs) were also below 10%. In the study by Tang et al (2011), which categorized patients by both clinical risk and RS category, the RS provided further risk stratification within clinical risk categories. The recurrence rates for each clinical risk and RS group, although they showed that each characteristic provides some predictive capability, are somewhat arbitrary because the cutoffs used to categorize clinical risk were simply based on creating classes similar in size to RS categories. Different cutoffs for the clinical risk categories would render different recurrence rates.

A prospective trial of Oncotype DX evaluating prognosis was published by Sparano et al (2015).[29] Although the trial only evaluated outcomes at 5 years, it is among the few Simon category A studies available. In it, women with node-negative, estrogen receptor–positive, human epidermal growth factor receptor 2 (*HER2*)–positive breast cancer were evaluated with Oncotype DX. Depending on the RS, women were assigned to endocrine therapy alone (low RS), randomized to adjuvant chemotherapy or no chemotherapy (middle category RS), or assigned to adjuvant chemotherapy (high RS). The published trial only reported the findings of the group at low risk of recurrence assigned to endocrine therapy. Of 10,253 subjects, 1629 patients had a RS of 0 to 10 and did not receive adjuvant chemotherapy (it should be noted that the cutoff score of 10 is lower than that for other studies evaluating Oncotype DX and thus evaluates a group at lower predicted risk of distant recurrence than other Oncotype DX studies, which typically used a cutoff of 18). Consequently, only 15.9% of the study population was judged low risk, which is much lower than other studies. At 5 years, the distant recurrence rate was 0.7% (95% CI, 0.4% to 1.3%). Other outcomes at 5 years were rate of invasive disease-free survival (93.8%; 95% CI, 92.4% to 94.9%), rate of freedom from recurrence (98.7%; 95% CI, 97.9% to 99.2%), and overall survival (OS; 98%; 95% CI, 97.1% to 98.6%). Results from the randomized subjects in the trial are not available. The outcomes of these subjects, who were at higher predicted risk of recurrence, would demonstrate the risk of outcomes of subjects with higher scores and perhaps determine the magnitude of benefit of chemotherapy in these subjects.

Clinical Utility

No decision-impact studies were identified that provide clinical outcomes such as survival or recurrence. However, evidence for clinical validity has shown that Oncotype DX is able to identify women who can safely forgo adjuvant chemotherapy with tight precision, and thereby avoid negative effects of the therapy.

Section Summary: Oncotype DX (21-Gene Assay)

Multiple studies derived from archived samples of previously conducted randomized controlled trials (RCTs) have shown that a low RS is associated with a low absolute risk of 10-year distant recurrence with an upper 95% CI bound not exceeding 10% in any study. These low absolute risks would translate to small absolute benefit from adjuvant chemotherapy. In these studies, over half of patients were classified at low risk. The 2015 prospective study by Sparano et al, although reporting results only at 5 years and using a more stringent cutoff to define a low-risk score, showed very low distant recurrence rates and is consistent with the previously reported studies.

Table 5. Ten-Year Distance Recurrence by Oncotype DX Risk Score Group

Study (Source of Patients)	N	Risk Score Group by % Patients in Risk Group			10-Year Distant Recurrence (95% Confidence Interval), %		
		Low	Int	High	Low	Int	High
Paik et al (2004)[26] (TAM arm of NSABP B-14 trial)	668	51	22	27	6.8 (4.0 to 9.6)	14.3 (8.3 to 20.3)	30.5 (23.6 to 37.4)
Paik et al (2006)[27] (TAM arm of NSABP B-20 trial)	227	59	20	21	3.2 (0.1 to 6.3)	9.1 (0.6 to 17.5)	39.5 (25.2 to 53.8)
Buus et al (2016)[25] (ATAC trial)	680	64	27	10	5.3 (3.5 to 8.2)	14.3 (9.8 to 20.6)	25.1 (15.8 to 38.3)
Tang et al (2011)[28] (TAM arm of NSABP B-14 trial)	668	Clin low/RS low: 32 Clin low/RS int-high: 21 Clin int-high/RS low: 18 Clin int-high/RS int-high: 29				5.6 (2.5 to 9) 12.9 (7 to 19) 8.9 (4 to 14) 30.7 (24 to 38)	

ATAC: Arimidex, Tamoxifen, Alone or in Combination; Clin: Clinical; Int: intermediate; NSABP: National Surgical Adjuvant Breast and Bowel Project; RS: Recurrence Score; TAM: tamoxifen.

EndoPredict

Clinical Validity

We identified 2 studies with 3 sets of findings that met selection criteria (see Table 6). The study by Filipits et al (2011) assessed patients from two previously conducted clinical trials.[30] We selected the study even though it included patients with positive nodes (32% of patients) because the expected effect of inclusion of these patients is to increase the recurrence rates and result in a conservative (biased to be high) estimate of distant recurrence. Buus et al (2016) studied patients from the ATAC trial, which evaluated the efficacy and safety of anastrozole vs tamoxifen in postmenopausal women with localized breast cancer.[25] In both studies, risk scores were defined as high and low based on a predefined cut point corresponding to a 10% risk of distant recurrence. EndoPredict provides an expression profile (EP) score based solely on the gene expression assay; the EPclin score incorporates the EP score plus clinical data on tumor size and nodal status. Results of the subgroup of node-negative patients in both studies were only reported in supplementary materials because the main report focused on combined node-positive and node-negative results. Node-negative patients constituted 73% of the subjects included in Buus et al and 68% in Filipits et al.

All 3 sets of findings showed that a low EP score is associated with a low absolute risk of 10-year distant recurrence. In 1 study the confidence interval exceeded 10%, but this was the smallest study (n=378 subjects). When the EP score incorporates tumor size and nodal status, a low EPclin score is also associated with a low absolute risk of 10-year distant recurrence. A higher proportion of subjects were classified as low risk (55%-73%) using EPclin, but the 10-year distant recurrence rates in the low-risk group were similar to rates in the EP low-risk group. This demonstrated that EPclin discriminates outcomes better than EP; it also suggests that using EPclin would result in fewer patients choosing chemotherapy than using EP alone. Subgroup analyses in Filipits et al including only patients with node-negative cancers showed an absence of distal recurrence of 95.0% (95% CI, 93.2% to 97.6%) in the EPclin low-risk group and 83.6% (95% CI, 77.2% to 90.0%) in the EPclin high-risk group. Subgroup analyses in Buus et al reported distant recurrence-free rates of 94.1% in the EPclin low-risk group and 80.0% in the EPclin high-risk group.

Clinical Utility

No decision-impact studies were identified that provide clinical outcomes such as survival or recurrence. However, evidence for clinical validity has shown that EndoPredict is able to identify women who can safely forgo adjuvant chemotherapy with tight precision, and thereby avoid negative effects of the therapy.

Section Summary: EndoPredict

Three sets of findings, derived from archived samples of previously conducted RCTs, have shown that a low EP or low EPclin score is associated with a low absolute risk of 10-year distant recurrence with an upper 95% CI bound generally below 10%, except in a small study. These low absolute risks would translate to small absolute benefit of adjuvant chemotherapy. In these studies, over half of the patients were classified at low risk. The EPclin score classified a higher proportion of patients as low risk than the EP score.

Table 6. Ten-Year Distance Recurrence by EndoPredict Risk Group

Study (Source of Patients)	N	Risk Score Group by % Patients in Risk Group				10-Year Distant Recurrence (95% Confidence Interval), %			
		EP Low	EP High	EPclin Low	EPclin High	EP Low	EP High	EPclin Low	EPclin High
Filipits et al (2011)[30,a] (ABCSG-6 trial)	378	51	49	55	45	8 (3 to 13)	22 (15 to 29)	4 (1 to 8)	28 (20 to 36)
Filipits et al (2011)[30,a] (ABCSG-8 trial)	1324	48	52	65	35	6 (2 to 9)	15 (11 to 20)	4 (2 to 5)	22 (15 to 29)
Buus et al (2016)[25] (ATAC trial)	680	43	57	73	27	3.0 (2 to 6)	14.6 (11 to 19)	5.9 (4 to 9)	20.0 (15 to 27)

ABCSG: Austrian Breast and Colorectal Cancer Study Group; ATAC: Arimidex, Tamoxifen, Alone or in Combination; EP: expression profile score; EPclin: EndoPredict score.

[a] ABCSG-6 and ABCSG-8 studies included a combined 32% node-positive patients.

Breast Cancer Index

Clinical Validity

We identified 2 studies with 3 sets of findings of the BCI that met selection criteria (see Table 7).[31,32] Some *HER2*-positive patients were included in both studies, but the number was not provided. Sgroi et al (2013) analyzed patients receiving anastrozole or tamoxifen in the ATAC trial.[31] This trial constitutes a Simon category B study. Two versions of the BCI score were generated in the study: (1) the BCI-C, based on cubic combinations of the variables, and (2) the BCI-L, based on linear combinations of the variables. The second study, by Zhang et al (2013), reported 2 sets of findings, one deriving from a clinical trial and another from patient registries.[32] Patients from the registry were only included if tissue samples were available.

In all sets of findings, the BCI classified more than half of the patients as low risk, and these patients had low risk of disease recurrence at 10 years. Sgroi et al report that the patients categorized as low risk by BCI-C and BCI-L experienced a low risk of disease recurrence, with the CIs not exceeding 10%. In the Zhang et al study, patients in BCI low-risk categories also showed a low risk of distant disease recurrence, with CIs not exceeding 10%.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. However, evidence for clinical validity has shown that the BCI is able to identify women who can safely forgo adjuvant chemotherapy with tight precision, and thereby avoid negative effects of the therapy.

Section Summary: Breast Cancer Index

Three sets of findings for the BCI have shown a low risk of 10-year distant recurrence among patients classified at low risk. Two sets of findings have been derived from clinical trials and are categorized as Simon category B. The findings from the multicenter registry are Simon category C.

Table 7. Ten-Year Distance Recurrence by Breast Cancer Index Risk Group

Study (Source of Patients)	N	Risk Score Group by % Patients in Risk Group			10-Year Distant Recurrence (95% Confidence Interval), %		
		BCI Low	**BCI Int**	**BCI High**	**BCI Low**	**BCI Int**	**BCI High**
Zhang et al (2013)[32] (multicenter registry)	358	55	22	23	6.6 (2.9 to 10)	23.3 (12.3 to 33)	35.8 (24.5 to 45.5)
Zhang et al (2013)[32] (Stockholm trial)	317	64	20	16	4.8 (1.7 to 7.8)	11.7 (3.1 to 19.5)	21.1 (8.5 to 32.0)
		BCI-C Low	**BCI-C Int**	**BCI-C High**	**BCI-C Low**	**BCI-C Int**	**BCI-C High**
Sgroi et al (2013)[31] (ATAC trial)	665	58	25	17	6.8 (4.4 to 10)	17.3 (12.0 to 24.7)	22.2 (15.3 to 31.5)
		BCI-L Low	**BCI-L Int**	**BCI-L High**	**BCI-L Low**	**BCI-L Int**	**BCI-L High**
		59	25	16	4.8 (3.0 to 7.6)	18.3 (12.7 to 25.8)	29.0 (21.1 to 39.1)

ATAC: Arimidex, Tamoxifen, Alone or in Combination; BCI-C: Breast Cancer Index using cubic form of variables; BCI-L: Breast Cancer Index using linear form of variables.

MammaPrint (70-Gene Signature)

Clinical Validity

We identified 2 studies using MammaPrint that met selection criteria (see Table 8). Several studies could not be included due to mixed populations, including node-positive patients, mixed node-positive, and node-negative patients, or patients receiving chemotherapy.

The study by Bueno-de-Mesquita et al (2011) evaluated a mixed node-positive and node-negative population, but subgroup results were also calculated.[33] The study sample was derived from 3 separate cohorts in cancer registry studies (Simon category C). For this evidence review, we present only the results for estrogen receptor–positive cancers. Risk groups were based on multiple clinical classification methods and the gene expression profile. Three clinical classification methods were used, and the results of any 2 clinical methods were classified as concordant low risk, discordant, and concordant high risk. Because the patterns were very similar across all 3 combinations of 2 clinical classification methods, only the results for combining Adjuvant! Online and Nottingham Prognostic Index are presented.

Only patients with both clinical low-risk scores and a MammaPrint low-risk score had 10-year distant recurrence risk below 10%. All other combinations of clinical risk and MammaPrint risk had 10-year recurrence risks greater than 10%. This pattern would suggest that a clinical strategy of using MammaPrint only in those with 2 clinical risk scores indicating low risk would identify patients with low absolute risk of recurrence.

In the van 't Veer et al (2017) study, analyses were conducted on the Stockholm tamoxifen (STO-3) trial, which randomized patients with node-negative breast cancer to 2 years of tamoxifen, followed by an optional randomization for an additional 3 years to tamoxifen or no treatment.[34] Both 10-year distant metastases-free survival (DMFS) and 20-year breast cancer–specific survival (BCSS) rates were calculated, by low-risk and high-risk groups, and by treatment group (tamoxifen vs no treatment). Patients receiving tamoxifen experienced longer DMFS and BCSS in both the low- and high-risk groups compared with patients not receiving tamoxifen.

Table 8. Ten- and 20-Year Follow-up Results by MammaPrint Risk Group

Study (Source of Patients)	N	MP Risk Score Group, n (%)	10-Year DMFS, % (95% CI)	20-Year BCSS, % (95% CI)
Van 't Veer et al (2017)[34,a]	538	Low risk, with tamoxifen: 199 (37)	93 (88 to 96)	90 (84 to 94)
		Low risk, without tamoxifen: 172 (32)	83 (76 to 88)	80 (72 to 86)
		High risk, with tamoxifen:82 (15)	85 (75 to 91)	83 (72 to va90)
		High risk, without tamoxifen: 85 (16)	70 (58 to 79)	65 (53 to 75)
		Clinical Risk Score Group and MP Risk Score Group N (%)	**10-Year Distant Recurrence % (95% CI)**	
Bueno-de-Mesquita et al (2011)[33] (3 combined cohorts)	139	Clin low/low MP low: 24	3 (0 to 9)	
		Clin low/low MP high: 10	34 (9 to 59)	
		Clin discordant MP low: 22	11 (0 to 22)	
		Clin discordant MP high: 9	31 (6 to 56)	
		Clin high/high MP low: 9	23 (0 to 46)	
		Clin high/high MP high: 26	47 (31 to 63)	

BCSS: breast cancer–specific survival; CI: confidence interval; Clin: clinical; DMFS: distant metastases-free survival; MP: MammaPrint.

[a] Confidence intervals provided by the manufacturer in October 2017.

Clinical Utility

The MINDACT trial (Cardoso et al, 2016) is a prospectively designed trial evaluating MammaPrint, with additional randomized components.[35] Currently, 5-year results are available. In this trial, women with early-stage breast cancer were evaluated with both MammaPrint and a clinical risk estimator. Women at low risk with both methods did not receive chemotherapy. Women with discordant risks were randomized to chemotherapy or to no chemotherapy. Women at high risk with both methods received chemotherapy.

Although parts of the study are an RCT, the end point for this particular analysis was the distant recurrence rate among patients with high-risk clinical and low-risk genetic profile who did not receive chemotherapy. Investigators prespecified that the upper bound of the 95% CI for distant recurrence was 8%, which they stated would be a sufficiently low risk that such patients could reasonably avoid chemotherapy. Declaring this to be the main end point implies a clinical strategy of using MammaPrint only in patients at high clinical risk, and deferring chemotherapy in those tested patients who have low genetic risk scores. In this strategy, patients at low clinical risk are not tested with MammaPrint.

Trial entry criteria included patients with either node-positive, estrogen receptor–positive, or *HER2*-positive breast cancer. However, these patients constituted a minority of those in the study. The main results included these patients. The authors conducted supplemental analyses of various subgroups, including the subset who were node-negative, estrogen receptor–positive, or *HER2*-negative. To report results of patients most comparable with the other studies discussed herein, BCBSA staff abstracted the

results of these supplemental analyses (see Table 9). The results are qualitatively similar to the published main results.

In the main article, the principal objective of the study was met. The group at high clinical risk and low genomic risk who did not receive chemotherapy had a distant recurrence rate of 5.3% (95% CI, 3.8% to 7.5%). In the node-negative, estrogen receptor–positive, or *HER2*-negative subgroup analysis, this group had a distant recurrence rate of 4.5% (95% CI, 3.8% to 8.4%).

In the group with clinical low risk and high genomic risk, who were not considered in the main outcome, in both the main analysis and in the node-negative, estrogen receptor–positive, or *HER2*-negative subgroup, the results would indicate that the risk of distant recurrence is not low enough to avoid chemotherapy (main analysis distant recurrence, 5% [95% CI, 3% to 8.2%]; hazard ratio subgroup distant recurrence, 6.1% [95% CI, 3.9% to 9.4%]). In the testing strategy implied in this study, by not testing for genomic risk in the low clinical risk group, these patients would not be identified.

The groups randomized to chemotherapy showed no significant difference in 5-year distant recurrence, but the CIs were wide and thus less informative regarding whether chemotherapy is or is not beneficial in these patient groups. In the main study, the hazard ratio (HR) for chemotherapy in the high clinical risk/low genomic risk was 0.78 (95% CI, 0.5 to 1.21). The HR for chemotherapy in the low clinical risk/high genomic risk group was 1.17 (95% CI, 0.59 to 2.28).

Table 9. MINDACT Trial 5-Year Distant Recurrence for the Node-Negative, Estrogen Receptor–Positive, or *HER2*-Negative Subgroup

Study (Trial)	N	Risk Score Group by % Patients in Risk Group	5-Year Distant Recurrence % (95% Confidence Interval)
Cardoso et al (2016)[35] (MINDACT trial)	4225	Clin low/MP low: 58	2.4 (1.8 to 3.1)
		Clin low/MP high: 11	6.1 (3.9 to 9.4)
		Clin high/MP low: 17	4.5 (2.4 to 8.4)
		Clin high/MP high: 14[a]	9.1 (6.8 to 12)

Clin: clinical; *HER2*: human epidermal growth factor receptor 2; MINDACT: Microarray In Node-negative and 1 to 3 positive lymph node Disease may Avoid ChemoTherapy; MP: MammaPrint.
[a] All clin high/MP high subjects received chemotherapy.

Section Summary: MammaPrint (70-Gene Signature)

One Simon category C study and 1 Simon category B study have been evaluated MammaPrint and provided 10-year distant recurrence outcomes. In the category C study, only subjects with both low clinical risk and low gene profiling risk have absolute rates of recurrence low enough to consider deferring chemotherapy. The sample size was small, and the proportion of patients identified at low risk was a small proportion (24%) of the study sample. The category B study showed that receiving tamoxifen improved recurrence and survival, in both low- and high-risk groups. The Simon category A study of MammaPrint has currently provided only 5-year distant recurrence outcomes. The principal result of the clinical high-risk plus MammaPrint low-risk patients may not be a low enough risk to defer chemotherapy because these 5-year recurrence rates will probably be much higher at 10 years. A group that may ultimately be identified as having sufficiently low absolute risk (but was not highlighted in the published study) is the group at clinical low risk and MammaPrint low risk, which at 5 years had a low absolute risk of distant recurrence of 2.4%.

Prosigna

Clinical Validity

Two studies that met selection criteria were identified (both studies are classed as Simon category B).[36,37] However, the distant recurrence rates from the study by Dowsett et al (2013) were not directly reported in the published article. As a result, rates cited in Table 10 are based on visual estimates of the graphic results; CIs are not available.).[36] Both studies reported distant recurrence rates below 5%, with the CIs for the 2 studies reporting them not exceeding 8%. In the 2 studies reporting the proportion of patients classified as low risk, more than 47% of patients were classified at low risk.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. However, evidence for clinical validity has shown that the assay is able to identify women who can safely forgo adjuvant chemotherapy with tight precision, and thereby avoid negative effects of the therapy.

Section Summary: Prosigna

Two category Simon B studies of Prosigna have shown absolute risks of 10-year distant recurrence that are sufficiently low for consideration of avoiding adjuvant chemotherapy. However, these results should be viewed cautiously because they may be due to variation in the tests used in these different studies.

Table 10. Ten-Year Distance Recurrence by Prosigna Recurrence Score Group

Study (Trial)	N	Risk Score Group (% Patients in Risk Group)			10-Year Distant Recurrence (95% Confidence Interval), %		
		Low	Int	High	Low	Int	High
Gnant et al (2014) (ABCSG-8 trial)[37]	1047	47	32	22	3.4 (2.1 to 5.6)	9.6 (6.7 to 13.7)	15.7 (11.4 to 21.6)
Dowsett et al (2013)[36] (ATAC trial)	739	59	33	8	4.8 (NR)	13.8 (NR)	30.2 (NR)

ABCSG: Austrian Breast and Colorectal Cancer Study Group; ATAC: Arimidex, Tamoxifen, Alone or in Combination; Int: intermediate; NR: not reported.

EARLY-STAGE NODE-POSITIVE INVASIVE BREAST CANCER CONSIDERING ADJUVANT CHEMOTHERAPY

Five studies that met selection criteria were identified (see Appendix 1), all prospective-retrospective designs, examining the prognostic value of gene expression profiling tests in patients with early-stage node-positive breast cancer receiving only endocrine therapy. Oncotype DX RS was evaluated in 2 studies,[38,39] Prosigna ROR (risk of recurrence)[40] in 1 study, and EndoPredict in 2 studies. Albain et al (2010) also explored a possible role for Oncotype DX in predicting chemotherapy benefit.[38] We also discuss results from the MINDACT trial, a prospectively designed trial evaluating MammaPrint. Table 11 displays the characteristics of patients assessed across the prospective-retrospective analyses. Almost all cancers were estrogen receptor–positive and *HER2*-negative, most patients had three or fewer positive lymph nodes, and all women were postmenopausal.

Table 11. Characteristics of Patients Included in Node-Positive Prospective-Retrospective Studies

				Tumor Size, n (%)			Nodes, n (%)		Adjuvant Chemo	
Study	**N**	**ER+**	***HER2*+**	**≤2 cm**	**2-5 cm**	**>5 cm**	**1-3**	**≥4**	**n (%)**	**Trial/Study**
Oncotype DX										
Albain (2010)[38,a]	148	145 (98)	13 (9)	46 (31)	94 (64)	8 (5)	94 (64)	54 (36)	0 (0)	SWOG-8814
Albain (2010)[38,b]	219	210 (96)	30 (14)	74 (34)	136 (62)	9 (4)	133 (61)	86 (39)	219 (100)	
Dowsett (2010)[39]	306	306 (100)	NR for node-positive patients				243 (79)	63 (21)	0 (0)	TransATAC
EndoPredict										
Filipits (2011)[30]	537	537 (100)	0 (0)	NR for node-positive patients			454 (85)	83 (15)	0 (0)	ABCSG6, ABCSG8
Buus (2016)[25]	248	248 (100)	0 (0)	NR for node-positive patients			198 (80)	50 (20)	0 (0)	TransATAC
Prosigna										
Gnant (2015)[40]	543		28 (5)	314 (58)			229 (42)	0 (0)	543 (100)	0 (0)

ABCSG: Austrian Breast and Colorectal Cancer Study Group; ATAC: Arimidex, Tamoxifen, Alone or in Combination; chemo: chemotherapy; ER: estrogen receptor; *HER2*: human epidermal growth factor receptor 2; NR: not reported.
[a] Tamoxifen.
[b] Cyclophosphamide, doxorubicin, and fluorouracil chemotherapy followed by tamoxifen.

Table 12 displays 10-year event rates reported by risk categories. Distant recurrence rates were not reported by Albain et al, but the 60% ten-year disease-free survival in the low-risk group would suggest substantial event rates in patients not receiving adjuvant chemotherapy. Confidence intervals were not reported, but given the small number of low-risk patient intervals, would likely include a large range of plausible estimates. Dowsett et al (2010) reported a 17% distant recurrence rate (death was considered a censoring event) in the low-risk category. Finally, Gnant et al (2015) reported 10-year distant recurrence rates in the Prosigna low-risk group with a single positive node of 6.6% (as much as 2-fold greater than for Prosigna-classified low-risk node-negative patients; see Table 11) with an upper bound of the 95% CI of 12.8%.[40] None of the studies reported the ability of tests to reclassify after assigning risk based on clinical predictors.

Table 12. 10-Year Event Rates According to Risk Categories in Identified Prospective-Retrospective Studies

				Risk Category									
				Low			Intermediate			High			
Test	Study	N	Endpoint	n	Outcome	95% CI	n	Outcome	95% CI	n	Outcome	95% CI	Simon
Oncotype Dx	Albain 2010, Tamoxifen	148	DFS	55	60%	NR	46	49%	NR	47	43%	NR	B
			OS		77%	NR		68%	NR		51%	NR	
	Dowsett 2010 (9-year outcomes)	296	DR[a]	150	17%	(12–24)	94	28%	(20–49)	52	49.0%	(35–54)	B
			OS		74%			69%			54%		
EndoPredict													
	Filipits 2011 (EP score)	537	DR	240	15%	NR				297	27%	NR	B
	Buus 2016 (EP score)	248	DR[a]	94	21.3%	(13.9–31.9)				154	36.4%	(28.9–45.2)	B
	(EPclin score)	248	DR[a]	47	5.0%	(1.2–18.9)				201	36.9%	(30.2–44.5)	B
Prosigna	Gnant 2015, 1 N+	331	DR[b]	132	6.6%	(3.3–12.8)	106	15.5%	(9.5–25.5)	93	25.5%	(17.5–36.1)	B
	Gnant 2015, 2-3 N+	212	DR[b]				83[c]	12.5%	(6.6–22.8)	129	33.7%	(25.5–43.8)	B

CI: confidence interval; DFS: disease-free survival; DR: distant recurrence; NR: not reported; OS: overall survival.
[a] Death from any cause considered a censoring event.
[b] Death from breast cancer included as a distant recurrence.

[c] Combined low- and intermediate-risk categories.

Oncotype DX (21-Gene Assay)

Clinical Validity

Albain et al (2010) analyzed data from the Southwest Oncology Group Trial 8814, an RCT that enrolled estrogen receptor–positive postmenopausal women and compared cyclophosphamide, doxorubicin, and fluorouracil chemotherapy followed by tamoxifen (CAF-T) for 5 years with tamoxifen alone.[38] Archived samples from 41% (n=148) and 39% (n=219) of the 2 trial arms, respectively, were available for analysis, and patients included in the analyses had fewer positive nodes and smaller tumors than those in the overall trial. Based on the RS results (includes *HER2* assay), about 1 in 10 patients had a *HER2*-positive tumor. The primary end point was disease-free survival (time from enrollment to locoregional or distant recurrence, new primary cancer, or any cause of death). Neither distant disease-free survival nor distant recurrence rates were available for analysis.

In addition to examining the prognostic value of the RS in node-positive patients, its potential predictive ability was also analyzed (see Table 13). While the hazard ratios appeared to vary with time, the magnitude differed by RS category, raising the possibility that adjuvant chemotherapy might not benefit those with low-risk scores. However, the CIs for the low-risk group include HRs consistent with benefit, and the small number of patients studied precludes drawing conclusions.

Table 13. Hazard Ratios for Chemotherapy Benefit of Sequential CAF-T vs Tamoxifen Alone by Oncotype DX RS

Variables	OS, HR (95% CI)		DFS, HR (95% CI)	
	10 Years	10 Years	≤5 Years	≥5 Years
Parent trial	0.78 (0.63 to 0.97)	0.69 (0.56 to 0.84)	0.68 (0.53 to 0.86)	0.72 (0.51 to 1.00)
RS sample[a]	0.77 (0.52 to 1.14)	0.72 (0.51 to 1.00)	0.79 (0.53 to 0.86)	0.63 (0.39 to 1.04)
Low RS			1.34 (0.47 to 3.82)	0.88 (0.38 to 1.92)
Intermediate RS			0.95 (0.43 to 2.14)	0.52 (0.20 to 1.52)
High RS			0.59 (0.32 to 1.11)	0.60 (0.22 to 1.62)

Adapted from Albain et al (2010).[38]
CAF-T: cyclophosphamide, doxorubicin, and fluorouracil chemotherapy followed by tamoxifen; CI: confidence interval; DFS: disease-free survival; HR: hazard ratio; RS: Recurrence Score; OS: overall survival.
[a] Adjusted for number of positive nodes.

Oncotype DX risk score appears to be associated with 10-year distant recurrence-free survival in patients with node-positive disease, although, as expected, the recurrence rates for the node-positive disease are higher than for node-negative (ie, 10-year distant recurrence-free survival in Albain et al). Overall, there is significant uncertainty in the estimates, and only 1 Simon category B study has reported on point-estimates for 10-year distant recurrence-free survival with CIs.

Dowsett et al (2010) examined a sample of node-negative and node-positive patients from the ATAC trial (Simon category B).[39] Archived samples were available for 306 node-positive patients of whom 243 (80%) had 1 to 3 involved nodes. The 9-year distant recurrence rate (censoring for any cause of death) in low-risk node-positive patients was 17% (95% CI, 12% to 24%) compared with 4% (95% CI, 3% to 7%) for the low-risk node-negative group. OS rates by risk group were similar to those reported by Albain et al. Dowsett et al fitted a model to recurrence rates using a continuous risk score and number of nodes,

which suggested considerably lower recurrence rates with 1 to 3 nodes compared with 4 or more. A potential predictive effect was not examined and OS not reported.

Although the RS appears to have some prognostic ability across the risk categories for node-positive disease, the absolute distant recurrence rates in the low-risk group were considerably higher than those proposed to be low enough to lead patients to forgo adjuvant chemotherapy in low-risk node-negative patients. There is a wide range of survival improvements over which individual patients would elect or refuse adjuvant chemotherapy, but accurate risk estimates are needed so that patients can make informed decisions. Given that patients would typically elect adjuvant chemotherapy for a modest improvement in survival (almost 50% reported that they would choose it for even a 1% gain)[20,23] raises a question whether in practice the RS offers sufficient prognostic information to inform decisions.

Nitz et al (2017) conducted a phase 3 Plan B trial with a mixed population of women with node-negative and node-positive breast cancer.[41] The trial was initially designed to compare anthracycline-containing chemotherapy with anthracycline-free therapy. An amendment was made to recommend endocrine therapy alone for patients with pN0/pN1 breast cancer and an RS of 11 or less. A total of 2642 patients were included in the trial. Median age was 56 years, 59% were node-negative, 35% were pN1, and 6% were pN2-3. Details of subgroup analyses of node-positive patients were limited. The authors stated that 5-year OS in patients with an RS between 12 and 25 was significantly higher than in patients with an RS greater than 25 within all nodal subgroups and that 5-year OS in low RS patients was higher compared with high RS patients in all nodal subgroups, but rates and CIs were not provided.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. Studies providing evidence for the clinical validity of Oncotype DX for patients with node-positive breast cancer have reported imprecise estimates of survival improvements in patients classified as low risk.

Section Summary: Oncotype DX (21-Gene Assay)

Results from prospective-retrospective Simon category B studies have suggested uncertainty in the estimates of the distant recurrence-free survival risk for patients in different Oncotype DX RS categories. One study did not report CIs for the estimates of survival and, in the other, the CIs were very wide. Another study mentioned that OS was significantly higher in patients with a low RS, but rates were not provided. Although it is expected that the distant recurrence-free survival estimates will be lower than those experienced by patients with node-negative disease, more certain estimates of risk are needed before a reasonable discussion about whether patients would or should decline adjuvant chemotherapy can occur. Albain et al (2010) suggested the test might also be predictive, albeit based on a small sample. Although there has been substantial adoption of the RS to inform adjuvant chemotherapy choices in node-positive patients,[42,43] convincing evidence that decisions based on test results will improve outcomes is lacking, and guidelines do not offer support.[44] The ongoing RxPONDER trial is randomizing patients with early-stage estrogen receptor–positive, *HER2*-negative breast cancer and 1 to 3 positive nodes, stratified by RS (0 to 13, 14 to 25) to adjuvant chemotherapy or no adjuvant chemotherapy. Results of that trial will most likely define the clinical utility of the RS in node-positive patients.

EndoPredict

Clinical Validity

Filipits et al (2011) evaluated the potential prognostic value of the EndoPredict EP and EPclin risk scores among node-positive patients in a combined analysis of ABCSG-6 and ABCSG-6 trial samples (Simon category B).[30] Of the 537 node-positive patients, 85% had a single positive node, 240 were classified as EP low risk, and 297 were classified as EP high risk. The 10-year absence of distant recurrence for node-positive patients was shown in a Kaplan-Meier curve in the article supplement. The 10-year absence of distance recurrence estimate for node-positive patients appears to be about 85% in EP low-risk and 73% in EP high-risk patients based on visual inspection; CIs were not provided. The 10-year absence of distance recurrence estimates for the EPclin low-risk group and EPclin high-risk group were 94.9% (95% CI, 90.8% to 99.0%) and 72.2% (95% CI, 65.6% to 78.8%), respectively.

Buus et al (2016) also reported on the prognostic value of EndoPredict among node-positive patients from ATAC in the article supplement (Simon category B).[25] Of the 248 node-positive patients, 80% had a single positive node, 94 were classified as EP low risk, and 154 were classified as EP high risk; 47 were classified as EPclin low risk, and 201 were classified as EPclin high risk. The 10-year distant recurrence-free survival for EP low and high risk were 21.3% (95% CI, 13.9% to 31.9%) and 36.4% (95% CI, 28.9% to 45.2%), respectively. The 10-year distant recurrence-free rate for EPclin low and high risk were 5.0% (95% CI, 1.2% to 18.9%) and 36.9% (95% CI, 30.2% to 44.5%), respectively.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. One of the 2 Simon category B studies provided evidence for clinical validity with tight precision, which would allow for the identification of women who can safely forgo adjuvant chemotherapy. The second study also reported a low point estimate; however, the wide CIs exceeded 10%.

Section Summary: EndoPredict

Two Simon category B studies, which met inclusion criteria, were identified. For node-positive, EPclin low-risk patients, the 10-year distant recurrence rate estimates was 5% (it should be noted that 1 study had a precise estimate while the other study had wide CIs, and the upper bound for the 95% CI was well above the range judged clinically informative in node-negative patients).

Breast Cancer Index

No studies were identified that met inclusion criteria in node-positive study populations for the BCI test.

70-Gene Signature (MammaPrint)

Clinical Utility

The previously described MINDACT study (Simon category A) initially enrolled only patients with node-negative disease but began including women with 1 to 3 positive nodes in 2009. Subgroup results were reported from the randomized MINDACT comparison of adjuvant chemotherapy with no chemotherapy in node-positive patients who were classified as high-risk based on clinical criteria and low-risk based on genetic risk with MammaPrint.[35] Overall, the study included 1404 node-positive patients; 296 (16%) with 1 positive node, 114 (6%) with 2 positive nodes, 65 (4%) with 3 positive nodes, and 2 (0.1%) with 4 or more positive nodes. In the high clinical risk and low genetic risk group, 353 node-positive patients

were randomized to chemotherapy, and 356 node-positive patients were randomized to no chemotherapy. The 5-year distant recurrence was 3.7% (95% CI, 1.9% to 6.9%) in the chemotherapy group and 4.4% (95% CI, 2.6% to 7.3%) in the no chemotherapy group (HR=0.88; 95% CI, 0.42 to 1.82; p=0.72). MINDACT has currently only provided 5-year distant recurrence outcomes; high clinical risk, low genetic risk patients may not be at low enough risk to defer chemotherapy because these 5-year recurrence rates will probably be higher at 10 years.

Mook et al (2009) evaluated the prognostic value of MammaPrint in patients with node-positive breast cancer.[45] Patients were selected from consecutive series of breast cancer patients from 2 institutions (Simon category C). A total of 241 patients were included, 99 were classified as low risk, and 142 were classified as high risk. Fifty-one percent of the patients had 1 positive node, 32% had 2 positive nodes, and 17% had 3 positive nodes. Median follow-up was 7.8 years. Ten-year BCSS was 96% (standard error [SE], 2%) for the low-risk group and 76% (SE=4%) for the high-risk group. The probability of remaining distant metastases-free at 10 years was 91% (SE=4%) for the low-risk group and 76% (SE=4%) for the high-risk group.

Section Summary: MammaPrint

One Simon category A study and 1 Simon category C study have investigated the use of MammaPrint to assess distant recurrence risk in women with node-positive breast cancer. The category C study reported 10-year follow-up results, which showed that patients categorized as low risk experienced better survival and recurrence rates than patients categorized as high risk. However, the recurrence rate with standard error did not meet the threshold benefit of less than 10%. The Simon category A study found 5-year distant recurrence rates for treated and untreated women are similar, which would indicate that the low-risk patients can safely forgo adjuvant chemotherapy. Longer follow-up is necessary for confirmation of the category A study results.

Prosigna

Clinical Validity

Gnant et al (2015) examined the potential prognostic value of the PAM50 ROR score, including clinical predictors, among node-positive patients in a combined analysis of the ABCSG-8 and ATAC trial samples.[40] Samples from 543 patients treated with endocrine therapy alone were included, and 10-year distant recurrence (the primary end point) analyzed. Among patients with a single positive node and a low-risk score, a 10-year distant recurrence occurred in 6.6% (95% CI, 3.3% to 12.8%). In all other risk categories or with 2 to 3 positive nodes, distant recurrence rates were considerably higher with upper bounds for the 95% CIs of 25% or more. OS was not included in the report.

Clinical Utility

No decision-impact studies were identified that provide clinical outcomes such as survival or recurrence. One study provided evidence for clinical validity. The point estimate for the 10-year distant recurrence rate was 7%however, the CI was large and did not meet the threshold benefit of less than 10%.

Section Summary: Prosigna

One Simon category B study meeting inclusion criteria was identified. The 10-year distant recurrence rate in patients with a single positive node and low-risk ROR scores is about 2-fold the rate in node-negative patients with low-risk ROR scores. The 10-year distant recurrence rate estimate for node-

positive, low-risk patients had an upper bound for the 95% CI approaching the range judged clinically informative in node-negative patients. Additional studies are needed to confirm the magnitude and precision of the estimates.

DUCTAL CARCINOMA IN SITU CONSIDERING RT

Oncotype DX Breast DCIS Score

Clinical Validity

DCIS is breast cancer located in the lining of the mammary ducts that has not yet invaded nearby tissues. It may progress to invasive cancer if untreated. The incidence of DCIS diagnosis in the United States has increased in tandem with the widespread use of screening mammography, accounting for about 20% of all newly diagnosed invasive plus noninvasive breast tumors. Recommended treatment is lumpectomy or mastectomy with or without radiotherapy; postsurgical tamoxifen treatment is recommended for estrogen receptor–positive DCIS, especially if excision alone is used. Because the overall rate of ipsilateral tumor recurrence (DCIS or invasive carcinoma) is approximately 25% at 10 years, it is believed many women are overtreated with radiotherapy. Thus, accurate prediction of recurrence risk may identify those women who can safely avoid radiation. The Oncotype DX Breast DCIS Score uses information from 12 of the 21 genes assayed in the standard Oncotype DX test for early breast cancer to predict 10-year risk of local recurrence (DCIS or invasive carcinoma). The stated purpose is to help guide treatment decision-making in women with DCIS treated by local excision, with or without adjuvant tamoxifen therapy.

In a retrospective analysis of data and samples from patients in the prospective Eastern Cooperative Oncology Group E5194 study, Solin et al (2013) compared the Oncotype DX Breast DCIS Score with 10-year local recurrence risk in a subset of DCIS patients treated only with surgery or with tamoxifen (n=327).[46] This study is Simon category B. The continuous Oncotype DX Breast DCIS Score was significantly associated with developing either a local recurrence or invasive carcinoma (HR=2.31; 95% CI, 1.15 to 4.49; p=0.02) whether or not patients were treated with tamoxifen. The prespecified DCIS risk groups of low, intermediate, and high had 10-year risks of developing either a local recurrence or invasive carcinoma of 11%, 27%, and 26%, respectively. This study addressed the development of the Oncotype DX Breast DCIS Score and clinical validity (association of the test result with local recurrence outcomes). Whether women are better categorized as to their local recurrence risk by Oncotype DX Breast DCIS Score compared with standard clinical indicators of risk has not been addressed.

In another retrospective analysis, Rakovitch et al (2015) evaluated 571 tumor specimens with negative margins from a convenience cohort of patients with DCIS treated by breast-conserving surgery (lumpectomy) alone.[47] Patients were drawn from a registry of 5752 women in Ontario, Canada, who were diagnosed with DCIS between 1994 and 2003. This study is Simon category C. Median follow-up for the 571 women was 9.6 years. There were 100 local recurrence events—43 were DCIS, and 57 were invasive cancer. The Oncotype DX Breast DCIS Score was significantly associated with local recurrence outcomes (HR=2.15; 95% CI, 1.43 to 3.22). Sixty-two percent of patients were classified as low risk, 17% as intermediate risk, and 21% as high risk. Corresponding 10-year local recurrence estimates were 13% (95% CI, 10% to 17%), 33% (95% CI, 24% to 45%), and 28% (95% CI, 20% to 38%), respectively. Corresponding 10-year estimates for DCIS recurrence (5% [95% CI, 3% to 9%]; 14% [95% CI, 8% to 24%]; 14% [95% CI, 9% to 22%], respectively) and for invasive breast cancer recurrence (8% [95% CI,

6% to 12%]; 21% [95% CI, 13% to 33%]; 16% [95% CI, 9% to 25%], respectively) were based on small numbers of events.

Clinical Utility

No decision-impact studies were identified that provide clinical outcomes such as survival or recurrence. Two studies provided evidence for the clinical validity of the Oncotype DX DCIS score; however, the recurrence risk estimates for the low-risk group were not low enough or precise enough (did not meet the threshold of 10%).

Section Summary: Oncotype DX Breast DCIS Score

Evidence consists of 1 Simon category B study and 1 Simon category C study. Based on the Oncotype DX Breast DCIS Score of low risk for recurrence, it is unclear whether estimated recurrence risks for this group are low enough or estimated with sufficient precision (point-estimates and CIs included the threshold of10%) to meaningfully affect the decision to have or to forgo radiotherapy.

EndoPredict, Breast Cancer Index, MammaPrint, and Prosigna

We did not identify studies evaluating the EndoPredict, BCI, MammaPrint, or Prosigna tests for patients with DCIS.

EXTENDED ADJUVANT ENDOCRINE THERAPY BEYOND 5 YEARS

Multiple randomized controlled trials have demonstrated improvements in overall and BCSS outcomes with 5 to 10 years of tamoxifen for estrogen receptor–positive tumors. However, extended adjuvant endocrine therapy may be associated with serious adverse events, including pulmonary embolism, endometrial cancer, osteoporosis, and fractures. Common side effects—hot flashes, sexual dysfunction, and musculoskeletal symptoms—often lead to poor compliance, with as many as 40% of patients discontinuing treatment after 3 years.[48] Accurately identifying low-risk patients who might obtain little benefit from extended endocrine therapy could allow patients to make treatment decisions consistent with how they value the potential benefits and harms.

In the absence of direct evidence that gene expression profiling tests improve outcomes in women considering extended endocrine therapy, the following need to be considered: (1) the expected magnitude and certainty of benefit from extended endocrine therapy, (2) how women value harms relative to benefit, and the range of thresholds in risk that a test is likely to change decisions, (3) whether a test accurately discriminates good from poor outcomes (ie, prognostic value for recurrences) at those thresholds, and (4) whether the test provides incremental improvement over clinical risk prediction algorithms or tools.

Seven studies (see Table 14) meeting selection criteria (see Appendix 1) were identified that examined the prognostic value of a gene expression profiling test for late recurrences after 5 years of endocrine therapy.[31,32,49-53] All 7 studies were prospective-retrospective designs of patients with early-stage node-negative or node-positive breast cancer receiving up to 5 years of endocrine therapy. One study (2013) examining prognosis[32] and an additional nested case-control study (Sgroi et al, 2013)[54] analyzed the potential predictive value of the HOXB13/IL17BR (H/I) index included in the BCI test. All but 1 cohort analyzed in Zhang et al (2013)[32] included only postmenopausal women. In addition, samples overlapped across some studies, as shown in the table by the trials used for analysis. Tables 15-19 display results from studies of prognosis subsequently discussed.

Table 14. Characteristics of Patients in Extended Endocrine Therapy Studies of Prognosis or Predicting Treatment Benefit

Study	N	Tumor Size, n (%)		Nodes, n (%)			Adjuvant Chemo, n (%)	Trial
		≤2 cm	>2 cm	None	1-3	≥4		
Oncotype DX								
Sestak (2013)[52]	940			683 (73)	257 (27)		0 (0)	TransATAC
EndoPredict								
Dubsky (2013)[49,a]	1702	1136 (67)	563 (33)	1165 (68)	454 (27)	83 (5)	0 (0)	ABCSG-6, ABCSG-8
Breast Cancer Index								
Zhang (2013)[32]	285	259 (82)	55 (17)	285 (100)	0 (0)	0 (0)	0 (0)	Stockholm Trial TAM-treated
	358	237 (66)	121 (34)	358 (100)	0 (0)	0 (0)	115 (32)	2-institution cohort
Sgroi (2013)[31]	597	442 (74)	155 (26)	597 (100)	0 (0)	0 (0)	0 (0)	TransATAC
Sgroi (2013)[54]	249	110 (44)	139 (56)	94 (38)	146 (59)		148 (59)	Nested case-control in MA.17
MammaPrint								
Esserman (2017)[53]	652	499 (77)	145 (22)	652 (100)	0 (0)	0 (0)	0 (0)	Stockholm Trial TAM-treated
Prosigna								
Filipits (2014)[50]	1246	NR (see below)		919 (74)	327 (26)		0 (0)	ABCSG-8
Sestak (2013)[52]	940			683 (73)	257 (27)		0 (0)	TransATAC
Sestak (2015),[51] all patients	862	587 (68)	275 (32)	647 (75)	180 (21)	35 (4)	0 (0)	TransATAC
Sestak (2015),[51] node-negative	1275	938 (74)	337 (26)	933 (73)	307 (24)	35 (3)	0 (0)	ABCSG-8

ABCSG: Austrian Breast and Colorectal Cancer Study Group; Chemo: chemotherapy; NR: not reported; TAM: tamoxifen; TransATAC: translational substudy of the Arimidex, Tamoxifen, Alone or in Combination.
[a] Sample size and characteristics represent patients at enrollment for Dubsky et al (2013).

Table 15. Prognosis for Late Distant Recurrence Based on Gene Expression Profiling Test Results

Test	Study	N	Distant Recurrence Years	Risk Category: Low n	Low DR	Low 95% CI	Intermediate n	Intermediate DR	Intermediate 95% CI	High n	High DR	High 95% CI	Simon
EndoPredict	Dubsky 2013 (EP)	998	5–10	503	3.7%	(0.9–6.5)				495	9.0%	NR	B
	Dubsky 2013 (EPclin)			642	1.8%	(0.1–3.5)				356	13.0%	NR	
BCI	Zhang 2013 (Stockholm TAM)	285	5–10	184	2.8%	(0.3–5.2)	58	7.2%	(0.1–13.8)	43	10.1%	(0.2–19.1)	B
	Zhang 2013 (Cohort study)	312	5–10	181	2.5%	(0.0–5.0)	70	16.9%	(6.5-26.2)	61	15.0%	(5.5–23.6)	C
	Sgroi 2013	596	5–10	366	3.5%	(2.0–6.1)	146	13.4%	(8.5–20.5)	84	13.3%	(7.4–23.4)	B
Prosigna	Filipits 2014	1246	5–15	460	2.4%	(1.1–5.3)	416	9.1%	(5.8–14.1)	370	17.5%	(12.9–25.2)	B
	Sestak 2013	940	5–10	NR	4.1%	NR		NR		NR	19.0%	NR	B
	Sestak 2015, all patients	2137	5–10	1183	2.4%	(1.6–3.5)	538	8.3%	(6.1–11.2)	416	16.6%	(13.1–20.9)	B
	Sestak 2015, node negative	1580	5–10	983	2.0%	(1.3–3.2)	344	9.0%	(6.3–13.0)	122	11.5%	(6.8–19.0)	
Oncotype DX	Sestak 2013	940	5–10	NR	7.6%	NR				NR	17.6%	NR	B

Simon refers to study category as discussed in the Background.
Sestak et al (2015) include samples from Sestak et al (2013) and Filipits et al (2014).
Note that, except for Filipits et al (2014), recurrences are over 5-year periods, or shorter than reported for adjuvant chemotherapy.
BCI: Breast Cancer Index; CI: confidence interval; DR: distant recurrence; EP: expression profile; EPclin: EndoPredict with clinical factors; NR: not reported; TAM: tamoxifen.

Oncotype DX (21-Gene Assay)

Clinical Validity

Sestak et al (2013) (previously discussed with the TransATAC study) also displayed late distant recurrences for risk categories of Oncotype DX in a Kaplan-Meier curve without confidence intervals.[52] The cumulative distant recurrence rate in the low-risk group between 5 and 10 years was estimated at 7.6%, or considerably higher than for any of the other tests considered. That result was consistent with the higher annualized hazard found in those years compared with PAM50 ROR. These limited results do not suggest a role for Oncotype DX for predicting late recurrences.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. No studies comparing genetic test classifications with clinical risk prediction tools were identified. One study provided evidence for clinical validity; the limited results did not support the clinical utility of Oncotype DX for this indication.

Section Summary: Oncotype DX

Evidence for the use of Oncotype DX for the prognosis of risk recurrence in women considering extending endocrine therapy beyond 5 years consists of a single study. The point estimate of risk recurrence was high, and CIs were not provided. Additional evidence would be needed to consider this indication.

EndoPredict

Clinical Validity

Dubsky et al (2013) analyzed late recurrences from patients in the ABCSG6 and ABCSG8 trials (see Table 14) treated with 5 years of endocrine therapy (tamoxifen for 5 years or tamoxifen for 2 years followed by anastrozole for 3 years).[49] Although 32% of patients were node-positive, none received adjuvant chemotherapy. Of the 1702 enrolled patients with estrogen receptor–positive *HER2*-negative cancers, follow-up was analyzed for 998 patients free of recurrence over 5 years and untreated with extended endocrine therapy. Risk categories were assigned based on gene expression profile (EP) alone and combined with a score that included nodal status and tumor size (EPclin). In the EP low-risk group, between 5 and 10 years the cumulative late distant recurrence rate was 3.7% (95% CI, 0.9% to 6.5%) (see Table 15). The distant recurrence rate in the EP high-risk group was 9% (CIs not reported). Adding clinical predictors suggested fewer late distant recurrences in the low-risk group (see Table 15). The risk of late distant recurrence in the node-negative patients (from digitized supplemental figure) was 3.6% or comparable with the overall EP low-risk group (n=503).

EP and EPclin appear to be able to identify a group at low risk of distant recurrence from years 5 to 10 in this prospective-retrospective study (Simon category B) of patients untreated with adjuvant chemotherapy enrolled in the ABCSG-6 and -8 trials. In the current environment, a significant proportion of high-risk patients would have been treated with adjuvant chemotherapy based on a gene expression profiling result. C statistics (area under the receiver operating characteristic curve) were reported to support incremental improvement with the EP or EPclin over Adjuvant! Online or nodal status, tumor

size, or grade. However, they appeared to include EP and EPclin as continuous variables and not threshold cutoffs for those tests that would inform decisions.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. No studies comparing genetic test classifications with clinical risk prediction tools were identified. One study provided evidence for clinical validity, showing that EP and EPclin scores adequately predicted the risk of distant recurrence, which would allow for the identification of women who can safely forgo extended endocrine therapy.

Section Summary: EndoPredict

One Simon category B study with some limitations found EndoPredict (EP and EPclin) prognostic for late distant recurrences. At least 2 Simon category B studies showing results consistent with clinical utility are necessary to demonstrate adequate evidence. In addition, studies comparing genetic test classifications with clinical risk prediction tools are needed.

Breast Cancer Index

Breast Cancer Index Prognosis

The prognostic component of BCI is based on the combination of an endocrine response biomarker H/I and a proliferation biomarker (Molecular Grade Index). These indices are used to categorize patients into groups of high and low risk for distant recurrence.

Clinical Validity

Incorporating the BCI as a continuous variable, Zhang et al (2013) developed an "optimized model" to predict early and late distant recurrences.[32] Patient samples from 2 studies were used (see Table 14): the Stockholm trial (Simon category B), which compared 2 or 5 years of tamoxifen with no treatment in early-stage breast cancer; and a cohort (Simon category C) of estrogen receptor–positive lymph node–negative patients retrospectively identified from a U.S. university medical center and a hospital (patients were treated between 1990 and 2000). Most patients were *HER2*-negative, with 5% of the Stockholm trial *HER2*-positive, and 10% of the cohort *HER2*-positive. Data from patients in the untreated arm of the Stockholm trial were used for model development; the tamoxifen arm of the trial and the 2-institution cohort were used for validation. The primary end point was distant recurrence-free survival (censoring for any cause of death). The Stockholm trial enrolled postmenopausal women who did not receive adjuvant chemotherapy; the 2-institution cohort included premenopausal and postmenopausal women of whom one-third received adjuvant chemotherapy (see Table 14). A median follow-up of 10 years was analyzed with distant recurrences occurring in 16% of all patients over 10 years. In the validation tamoxifen-treated arm of the Stockholm trial, there were 20 late distant recurrences and 65% of patients were classified as low risk; in the 2-institution cohort, there were 23 late distant recurrences, and 58% of patients were classified as low risk.

From years 5 to 10, distant recurrence rates were low in the low-risk groups of the validation samples (see Table 15). The results support the prognostic value of the BCI for late recurrences in node-negative patients. About one-third (32%) of the cohort received adjuvant chemotherapy, but whether any of those patients were at low BCI risk was not noted. However, the authors reported chemotherapy was not associated with a lower risk of late recurrence.

Sgroi et al (2013) examined late distant recurrences among 597 estrogen receptor–positive, *HER2*-negative, node-negative patients from the ATAC trial (Simon category B) not treated with adjuvant chemotherapy.[31] Patients who died were censored in the analysis of distant recurrences. In the analytic sample, distant recurrences occurred among 4% of patients in years 0 to 5 and among 7% in years 5 to 10. From years 5 to 10, in the BCI low-, intermediate-, and high-risk groups' distant recurrence rates were 3.5% (95% CI, 2.0% to 6.1%), 13.4% (95% CI, 8.5% to 20.5%), and 13.3% (95% CI, 7.4% to 23.4%), respectively. But when examined as a continuous predictor for late recurrence (using the model developed by Zhang et al[32]), at a value of 5 (which is categorized as low risk), the predicted distant recurrence rate was 6.8% (95% CI, 4.7% to 9.1%) (CIs were provided by the manufacturer in October 2017).

The authors concluded: "…our results suggest that BCI might have the potential to influence two important decisions in the management of postmenopausal patients with oestrogen-receptor-positive, N0 breast cancer: first at the time of diagnosis and second at 5-year disease-free follow-up." These results would suggest that the BCI has prognostic value for late distant recurrences over a 5- to 10-year period. Among the higher risk patients, none received adjuvant chemotherapy or therapy not consistent with test results; the accuracy of late recurrence predictions in those patients is uncertain.

Schroeder et al (2016)[55] calculated distant recurrence-free survival (DRFS) rates following 5 years of endocrine therapy among the subset of patients with clinically low-risk (T1N0) breast cancer from the 2 populations studied by Zhang et al (2017). The Stockholm trial had 237 patients, and the U.S. medical center cohort contributed 210 patients that were T1N0. The BCI classified 68% (160/237) and 64% (135/210) of the Stockholm population and the medical center population as low risk, respectively. Median follow-up was 17 years for the Stockholm study and 10 years for the medical center cohort. Table 16 lists the 5- to 15-year distant recurrence-free survival rates (as categorized by BCI risk) for the 2 trial populations.

Table 16. Five to 15-Year DRFS by Breast Cancer Index Risk Stratification

Study	Population	N	Low Risk, % (95% CI)	High Risk, % (95% CI)
Schroeder et al (2016)[55]	Stockholm T1N0 total	237	95.4 (92.1 to 98.8)	86.7 (78.9 to 95.3)
	Stockholm T1N0 *HER2*-negative	225	95.2 (91.9 to 98.8)	86.9 (78.8 to 95.9)
	Stockholm T1N0 *HER2*-negative, G1 & G2	204	95.7 (92.5 to 99.1)	90.4 (82.8 to 98.8)
	Multi-institutional T1N0 total	210	98.4 (96.3 to 100)	89.6 (82.4 to 97.4)
	Multi-institutional T1N0 *HER2*-negative	190	98.4 (96.1 to 100)	87.5 (79.1 to 96.9)
	Multi-institutional T1N0 *HER2*-negative, G1 & G2	173	98.2 (95.8 to 100)	87.6 (78.5 to 97.7)

CI: confidence interval; DRFS: distant recurrence-free survival; *HER2*: human epidermal growth factor receptor 2.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. Evidence for clinical validity has shown that the BCI is able to identify women who can safely forgo extended endocrine therapy with tight precision, and thereby avoid negative effects of the therapy. However, no studies comparing genetic test classifications with clinical risk prediction tools were identified.

Breast Cancer Index Prediction

The endocrine predictive component of the BCI is based on the H/I ratio alone, in which a high H/I ratio predicts the likelihood of benefit from extended endocrine therapy.

Clinical Validity

Sgroi et al (2013) conducted a prospective-retrospective, nested case-control study within the MA.17 trial that compared extended endocrine therapy (letrozole) with placebo in postmenopausal women with hormone receptor–positive cancers.[54] The trial randomized 5157 women recurrence-free at 5 years to letrozole or placebo. A case-control design was adopted owing to challenges in obtaining archived tumor samples. An eligible case (319 of which 83 were examined) was one that experienced a local, regional, or distant recurrence and had an available tumor sample. Two controls free of recurrence longer than cases were matched to each case based on age, tumor size, node status, and prior chemotherapy. Any recurrence (locoregional or distant) was used as the end point; patients with contralateral or unknown recurrences were excluded. Using the 2-gene expression H/I ratio, which is obtained from the BCI, there was a 42% relative risk reduction in the low-risk group vs a 77% reduction in the high-risk group. Although statistical significance was lacking in the low-risk group, the CIs were wide and included values consistent with those observed in the high-risk group (see Table 16).

Zhang et al (2013) also reported a larger potential relative risk reduction in the high-risk group of the Stockholm trial, with similar uncertainty reflected in the CIs (see Table 17).[32]

Table 17. Predictive Effect of the H/I Index in the BCI for Extended Endocrine Therapy Benefit

			Low Risk		High Risk		
Study	**N**	**Comparators**	**HR (95% CI)**	**ARR**	**HR (95% CI)**	**ARR**	**Note**
Sgroi et al (2013)[54]	249	Letrozole vs placebo	0.58 (0.25 to 1.36)	4%	0.33 (0.15 to 0.73)	16.5%	Nested matched CC study; 83 recurrences in 166 controls; 5-y ARRs reported
Zhang et al (2013)[32]	600	Tamoxifen vs placebo	0.67 (0.36 to 1.24)	4.9%	0.35 (0.19 to 0.65)	19.6%	Stockholm trial, 15-y results

ARR: absolute risk reduction; BCI: Breast Cancer Index; CC: case-control; CI: confidence interval; H/I test: HOXB13/IL17BR; HR: hazard ratio.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. No studies comparing genetic test classifications with clinical risk prediction tools were identified. Two studies provided evidence for the clinical validity of the BCI Prediction. Wide CIs in the results do not support the clinical utility of this test in identifying women who can safely forgo extended endocrine therapy.

Section Summary: Breast Cancer Index (Prognosis and Prediction)

Three studies analyzing data from 2 Simon category B studies and 1 Simon category C study evaluated the BCI Prognosis for women who are recurrence-free for 5 years considering extended endocrine therapy. The 10-year distant recurrence rates were significantly low, and the 10-year distant recurrence-free survival estimates were significantly high for patients identified by the BCI as low risk. The studies evaluating the BCI Prediction reported results with wide CIs, indicating uncertainty in distinguishing between the low- and high-risk groups.

MammaPrint (70-Gene Signature)

Clinical Validity

Esserman et al (2017) conducted a secondary analysis on data from women who were node-negative, participating in an RCT of tamoxifen vs no systemic therapy, with over 20 years of follow-up (Stockholm tamoxifen trial, STO-3).[53] This is a Simon category B study. A total of 652 tissue samples from the trial underwent MammaPrint risk classification, 313 from the tamoxifen arm and 339 from the no therapy arm. The primary outcome was 20-year BCSS. Initial classification by MammaPrint identified 58% of the patients as low risk for distant recurrence and 42% as high risk. Twenty-year BCSS rates were 85% and 74% (p<0.001), respectively. Analysis was conducted on a subgroup of the low-risk group, considered ultralow risk. The tamoxifen-treated ultralow-risk group did not experience any deaths at 15 years. Survival rates were high for all patients in the ultralow-risk group, 97% for those treated with tamoxifen and 94% for those untreated. Table 18 details survival rates for the initial low- and high-risk groups, and for the subgroup analysis that separated an ultralow-risk group.

Table 18. Ten- and 20-Year Follow-up Results by MammaPrint Risk Group

Study (Source of Patients)	N	MP Risk Score Group, N (%)	10-Year BCSS, % (95% CI)	20-Year BCSS, % (95% CI)
Esserman et al (2017)[53]	652	Low risk: 377 (58)	90 (87 to 93)	85 (80 to 89)
		High risk: 275 (42)	81 (74 to 86)	74 (66 to 80)
Esserman et al (2017)[53]	652	Ultralow risk: 98 (15)	99 (92 to 100)	95 (86 to 99)
		Low but not ultralow risk: 279 (43)	88 (83 to 91)	82 (76 to 86)
		High risk: 275 (42)	80 (75 to 85)	73 (67 to 79)

BCSS: breast cancer–specific survival; CI: confidence interval; MP: MammaPrint.

Clinical Utility

No decision-impact studies were identified that reported clinical outcomes such as survival or recurrence. One study provided evidence for the clinical validity of MammaPrint when a subgroup of the low-risk group, an ultralow-risk group, was identified, that can safely forgo extended endocrine therapy. However, no studies comparing genetic test classifications with clinical risk prediction tools were identified.

Section Summary: MammaPrint

One Simon category B study meeting inclusion criteria was identified. A subgroup of the low-risk patients was identified, and it showed high 10-year BCSS rates. Additional studies are needed to confirm the benefit of MammaPrint for identifying women who may forgo extended endocrine therapy. Studies comparing the genetic test to clinical prediction models are also needed.

Prosigna

Clinical Validity

Filipits et al (2014) analyzed data from patients in the ABCSG-8 trial (5 years of adjuvant tamoxifen vs tamoxifen for 2 years followed by anastrozole).[50] Adjuvant chemotherapy was not administered. The PAM50 ROR predecessor test of Prosigna was obtained from archival samples using the NanoString nCounter device. At 5 years, 1246 patients free of recurrence were included in the analyses (74% node-negative). Almost all patients (97%) classified as low risk were node-negative. Between years 5 and 15,

there were 7 distant recurrences in the low-risk group (n=460) and none recorded among the 12 low-risk node-positive patients. The cumulative risk of late distant recurrence was 2.4% (95% CI, 1.1% to 5.3%). However, as of year 11, 59% of the low-risk group was being followed and at risk, and at year 14 just 11%. The authors also evaluated a clinical linear predictor score (age, grade, nodal status, endocrine treatment) but did not present recurrence rates by clinical risk categories (eg, low, intermediate, high).

Sestak et al (2013) reported limited results concerning late recurrences obtained from patients in the ATAC trial who received anastrozole with tamoxifen alone or in combination.[52] From a subset of women in the monotherapy arms with archived tissue (a sample forming the TransATAC study), a total of 940 U.K. women from the study were analyzed. Distant recurrence was the primary end point (censored at death). The sample included patients with node-positive and node-negative cancers, but proportions were not reported. There were 83 distant recurrences from years 5 to 10. A clinical treatment score derived from age, node status, treatment, stage, and grade was examined but its prognostic value not reported. Annualized hazards (distant recurrence rates) were consistent with a lower late recurrence risk for node-negative tumors 2 cm or smaller and among those with a low PAM50 ROR score. From a Kaplan-Meier plot, the late distant recurrence risk in the PAM50 ROR low-risk group was estimated at 4.1% (CIs were not displayed). The absence of CIs and comparison or reclassification of clinical predictors' prognosis limits any conclusions.

A subsequent publication by Sestak et al (2015)[51] combined samples of women with hormone receptor–positive, *HER2*-negative cancers from the ABSCG-8 and TransATAC studies included in the 2 prior publications.[50,52] Risk was determined using both a Clinical Treatment Score (CTS; treatment received, positive nodes, tumor size, age, and grade) and the PAM50 ROR. As in the prior studies, death was considered a censoring event; women with recurrences through 5 years were excluded, and the median follow-up was 10 years. Approximately 25% of patients had positive nodes. Both the ROR and CTS were prognostic, but cumulative event rates reported only for the ROR (see Table 15). In the ROR low-risk group, the distant recurrence rate was 2.4% (95% CI, 1.6% to 3.5%) in all women and 2.0% (95% CI, 1.3% to 3.2%) when only node-negative patients were examined. Finally, the authors compared the ability of the ROR to reclassify patients with the CTS. From a reclassification analysis (see Table 19), assuming a selective as opposed to a treat-all strategy and that only low-risk women would not be treated: (1) adding the ROR to the CTS would have resulted in 5 (3.4%) more of 148 patients experiencing distant recurrence being treated, and (2) 60 (3.0%) of 1989 additional patients not experiencing a recurrence would have been incorrectly treated. The reclassification results would suggest caution when interpreting prognostic estimates without considering clinical predictors.[25]

Table 19. Classification and Reclassification Achieved by Adding ROR Score to the CTS

Distant Recurrence		CTS					CTS			
		Low	Int	High	Total		Low	Int	High	Total
ROR	Low	18	14	0	32	ROR + CTS	25	3	0	28
	Intermediate	7	31	7	45		8	53	0	61
	High	8	17	46	71		0	6	53	59
	Total	33	62	53	148		33	62	53	148
No Distant Recurrence		**CTS**					**CTS**			
		Low	Int	High			Low	Int	High	
ROR	Low	837	273	41	1151	ROR + CTS	1030	136	0	1166
	Intermediate	209	221	63	493		76	448	25	549
	High	60	137	148	345		0	47	227	274
	Total	1106	631	252	1989		1106	631	252	1989

CTS: Clinical Treatment Score; Int: intermediate; ROR: risk of recurrence.

Clinical Utility

No decision-impact studies were identified that report clinical outcomes such as survival or recurrence. Limitations (eg, lack of reporting recurrence rates by ROR categories, lack of CIs) in the studies that evaluated clinical validity preclude any conclusions for clinical utility of this test for this indication. One study compared genetic test classifications with a clinical risk prediction tool and reported minimal improvement of the test over the clinical prediction tool.

Section Summary: Prosigna

Studies obtained from 2 completed trials analyzed in different publications (2 Simon category B studies) have found that the PAM50 ROR can identify patients at low risk of late distant recurrence. However, a reclassification result suggested that the test may offer little improvement over clinical predictors alone.

Section Summary: Extended Endocrine Therapy Beyond 5 Years

At least 3 randomized controlled trials have demonstrated survival improvements with extended tamoxifen. While the evidence for extended aromatase inhibitor is moreso mixed than the other trials, guidelines have recommended extended endocrine therapy with tamoxifen or an aromatase inhibitor in all hormone receptor–positive women. However, 3 trials completed and presented in 2017 but not yet published (described in the Background section) may challenge a "treat-all" approach. Results of these trials may affect the uncertainty in possible benefit and the impact on treatment strategies.

Compared with the choice of adjuvant chemotherapy depending on baseline recurrence risk, there is less empirical research on women's threshold for decision-making to forgo extended endocrine therapy based on recurrence risk. To be clinically useful, a test should be able to predict accurately a cumulative lifetime recurrence rate in a range that would be meaningful for decision-making.

If one assumes, as suggested by the studies reviewed, that the predicted 10-year or later distant recurrence rates would be sufficiently lower than 10%, then according to the Simon levels of evidence, the BCI and Prosigna have 2 category B studies appropriately reported to support their use. However, evidence demonstrating incremental reclassification improvement applying decision informative thresholds is lacking. The single reclassification result does not offer strong support for net incremental improvement, particularly if the way in which women value benefits (net improvement in those recurrences) and harms (increased false positives in those without recurrences) is considered.

Moreover, it is not readily apparent how the test result informs decision-making at the time results are available.

TEST COMPARISON STUDIES

Bosl et al (2017) compared MammaPrint with EndoPredict in 48 tumor samples—29 were node-negative, and 19 were node-positive.[56] For the MammaPrint test, RNA quality was low for 3 samples. Of the 45 tested by MammaPrint, 17 (38%) were classified as low risk, and 28 (62%) were classified as high risk for recurrence. Four samples were excluded from the EndoPredict analysis because the tumors were estrogen receptor–positive or *HER2*-positive, which are not part of the inclusion criteria of this test. Based on the EP molecular score, 8 (18%) samples were classified as low risk, and 36 (82%) samples were classified as high risk. Based on the EPclin score, 17 (39%) samples were considered low risk, and 27 (61%) samples were considered high risk. There was no statistically significant agreement

between MammaPrint and molecular EP (overall concordance, 63%) or between MammaPrint and EPclin (overall concordance, 66%).

Sgroi et al (2013) compared the BCI with Oncotype DX in 665 lymph node–negative women receiving endocrine therapy but not chemotherapy in the ATAC trial.[31] The distribution of patients across risk groups was similar. For patients receiving tamoxifen alone or in combination with anastrozole, 10-year distant recurrence risk estimates for the 2 tests were similar within risk groups. In the anastrozole group, the BCI was a better predictor of risk: 5% of the BCI low-risk patients had distant recurrence compared with 9% of Oncotype DX low-risk patients, and 22% of the BCI high-risk patients had distant recurrence compared with 13% of Oncotype DX high-risk patients. These values were reported without 95% CIs; it is therefore not possible to assess the degree of overlap between risk groups.

Sestak et al (2016)[57] examined cross-stratification between the BCI and Oncotype DX RS using the same data as Sgroi et al (2013).[31] Patients from the ATAC trial (N=665) who were postmenopausal, hormone receptor–positive, and node-negative were included. Median follow-up was 10 years. Gene expression analyses for both scores were conducted, and risk categories were determined based on prespecified cutoff points (RS: <18=low risk, 18-31=intermediate risk, >31=high risk; BCI: <5.0825=low risk, 5.0825-6.5025=intermediate risk, >6.5025=high risk). Each gene expression score was combined with the CTS an algorithm of nodal status, tumor size, grade, age, and treatment. In a multivariate analysis, when the BCI was added to RS plus CTS, there was a significant effect on prognostic information. When RS was added to the BCI plus CTS, no additional prognostic information was added.

Dowsett et al (2013) compared the PAM50 ROR score with the Oncotype DX 21-gene RS and IHC4 breast cancer algorithm.[36] Patients had estrogen receptor–positive, primary breast disease treated with anastrozole or tamoxifen in the ATAC trial (a double-blinded, phase 3 clinical trial designed to compare the ability of anastrozole, tamoxifen, and the 2 drugs in combination to prevent breast cancer recurrence in postmenopausal women with hormone receptor–positive tumors). Lymph node–negative and –positive patients were included. Messenger RNA from 1017 patients was assessed for ROR, and likelihood ratio tests and concordance indices were used to assess the prognostic information provided beyond that of a CTS. Statistical testing of these parameters was significant and favored the ROR score over the RS. More patients were classified as high risk and fewer as intermediate risk by the ROR than by RS. Prognostic information provided by the ROR score and IHC4 was similar.

Hornberger et al (2012) conducted a systematic review on the clinical validity, clinical utility, change in clinical practice, and economic implications of early-stage breast cancer stratifiers.[58] Fifty-six articles published original evidence addressing the Oncotype DX RS (n=31), MammaPrint (n=14), Adjuvant! Online (n=12), 5-antibody immunohistochemical (IHC) panel (Mammostrat; n=3), and a 14-gene signature (BreastOncPx; n=1). Oncotype DX RS satisfied level 1 evidence for estimating distant recurrence risk, OS, and response to adjuvant chemotherapy, and level 2 evidence for estimating local recurrence risk. Mammostrat and MammaPrint satisfied level 2 evidence for estimating distant recurrence risk and OS. Adjuvant! Online satisfied level 2 evidence for estimating distant recurrence risk, OS, and chemotherapy response. BreastOncPx satisfied level 3 evidence for predicting distant recurrence risk and OS. Ten studies reported changes in clinical practice patterns using Oncotype DX; overall, Oncotype DX was associated with change in treatment recommendations and/or decisions in 21% to 74% of cases.

Fan et al (2006) used 5 gene expression classifiers to evaluate a single set of samples from 295 women with stage I or II breast cancer, variable node involvement, and variable endocrine or chemotherapy treatment.[59] The classifiers included the 21-gene RS, the 70-gene signature, the H/I ratio, and the intrinsic subtype classifier (similar to the commercially available PAM50). Most highly correlated were the 21-gene RS and the 70-gene signature, with a Cramer V of 0.6 (scale 0-1, with 1 indicating perfect agreement). More specifically, 81 (79%) of 103 samples with an RS of low or intermediate risk were classified as having a low-risk 70-gene profile. Restricting the analysis to 225 estrogen receptor–positive samples slightly reduced the correlation. Analysis was not further restricted to node-negative patients, the present indication for both tests.

Espinosa et al (2005) compared Oncotype DX, MammaPrint, and the 2-gene ratio (H/I ratio) in 153 patients with estrogen receptor–positive breast cancer treated with adjuvant tamoxifen.[60] Sixty-two percent of patients were node-negative, and 63% were additionally treated with chemotherapy. Estimated distant metastasis-free survival for RS risk groups was 98% for low-risk, 81% for intermediate-risk, and 69% for high-risk patients; for the 70-gene signature, estimates were 95% for good prognosis and 66% for poor prognosis patients; and for the 2-gene ratio, estimates were 86% for favorable and 70% for unfavorable prognosis. The correlation between the 21-gene RS and the 70-gene signature was good (Cramer V=0.6). There was slightly more variation in distant metastasis-free survival, explained by the combination of the 21-gene RS plus either Adjuvant! Online (25.8, SD=1.4) or the Nottingham Prognostic Index (23.7, SD=1.5) as opposed to the combination of the 70-gene signature plus Adjuvant! Online (23.1, SD=1.2) or the Nottingham Prognostic Index (22.4, SD=1.3). However, differences were small and any combination was significantly better than any test or clinicopathologic classifier alone.

Two studies have compared Oncotype DX with other gene expression profiles. Kelly et al (2012) evaluated Oncotype DX and PAM50 in 108 cases and found good agreement between the 2 assays for high- and low-prognostic risk assignment; PAM50 assigned about half of Oncotype DX intermediate-risk patients to the PAM50 luminal A (low-risk) category.[61] Prat et al (2012) evaluated several gene expression tests, including Oncotype DX, PAM50, and MammaPrint, in 594 cases; they found all predictors were significantly correlated (Pearson r range, 0.36-0.79; $p<0.001$ for each comparison).[62]

ADDITIONAL APPLICATIONS AND OTHER TESTS

Based on a 2008 study that compared Oncotype DX estrogen and progesterone receptor results with traditional IHC results,[63] Genomic Health includes quantitative estrogen and progesterone receptor component results in Oncotype DX 21-gene profile reports. The study reported 90% or better concordance between the 2 assays, but the quantitative estrogen receptor by Oncotype DX was more strongly associated with disease recurrence than the IHC results. However, estrogen and progesterone receptor analysis is traditionally conducted during pathology examination of all breast cancer biopsies, whereas Oncotype DX is indicated only for known estrogen receptor–positive tumors, after the pathology examination is complete, the patient meets specific criteria, and patient and physician are considering preferences for risk and chemotherapy. Thus, Oncotype DX should not be ordered as a substitute for estrogen and progesterone receptor IHC. Additionally, accepted guidelines for estrogen and progesterone receptor testing outline standards for high-quality IHC testing and do not recommend confirmatory testing; thus the 21-gene RS should not be ordered to confirm estrogen and progesterone receptor IHC results. A subsequent study by Khoury et al (2015) reported better correlation (for overall data) between the IHC and Oncotype DX for progesterone receptor status (Spearman $\rho=0.91$) than for

estrogen receptor status (Spearman ρ=0.65), but worse concordance (at various cut points) for progesterone receptor status (99%) than for estrogen receptor status (88).[64]

Investigators have examined the ability of gene expression tests to provide risk information for locoregional recurrence. The reason for analyzing these tests in relation to locoregional recurrence is that they may have implications for the type and extent of initial local treatment. Drukker et al (2014) used MammaPrint to assess 1053 tumor specimens from 1848 patients enrolled in 8 previous MammaPrint studies.[65] Most patients had estrogen receptor–positive, *HER2*-negative disease; approximately half of patients had positive axillary lymph nodes. Most patients received radiotherapy and did not receive adjuvant chemotherapy; approximately half of the patients received adjuvant endocrine therapy. At a median follow-up of 9 years, estimated 10-year locoregional recurrence risk was 13% (95% CI, 10% to 16%) for 492 patients categorized as MammaPrint high-risk vs 6% (95% CI, 4% to 9%) for 561 MammaPrint low-risk patients. This association was observed during the first 5 years after diagnosis, but not during years 5 to 10. Recurrence stratified by MammaPrint risk class was not associated with primary locoregional treatment (ie, not predictive of treatment response).

Fitzal et al (2015) evaluated local recurrence using EndoPredict in breast tumor samples from 1324 patients who had participated in the ABCSG-8 trial (29% of enrolled patients), which compared adjuvant endocrine therapy regimens.[66] Most patients had node-negative, estrogen receptor–positive disease and received breast-conserving surgery and radiotherapy; approximately half of patients received adjuvant endocrine therapy. At a median follow-up of 6 years, the Kaplan-Meier estimate for 10-year risk of local recurrence-free survival was 96% (91% reported in the article abstract) among 683 patients classified by EndoPredict as high risk vs 99% among 641 patients classified by EndoPredict as low risk. EndoPredict risk groups were not associated with treatment outcomes.

Although the 3 gene expression tests are associated with risk of local recurrence, how these results would be used to change management, either by providing more aggressive treatment to high-risk patients or by providing less aggressive treatment to low-risk patients, is not clear.

SUMMARY OF EVIDENCE

Early-Stage Node-Negative Invasive Breast Cancer

For the evaluation of breast cancer–related gene expression profiling tests for the management of all early-stage breast cancer populations, study populations considered had positive hormone receptor status, and negative *HER2* status. Studies retrospectively collecting tumor samples from prospective trials that provide 10-year distant recurrence rates or 10-year survival rates in node-negative women not receiving adjuvant chemotherapy were included in this part of the evidence review.

Oncotype DX (21-Gene Assay)

For individuals who have early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with Oncotype DX (21-gene assay), the evidence includes multiple prospective clinical trials and prospective-retrospective studies. Patients classified as low risk with Oncotype DX have a low risk of recurrence in which avoidance of adjuvant chemotherapy is reasonable (average risk at 10 years, 3%-7%; upper bound of the 95% CI, 6% to 10%). These results have been demonstrated with stronger study designs for evaluating biomarkers. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

EndoPredict

For individuals who have early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with EndoPredict, the evidence includes 3 prospective-retrospective studies and observational studies. The studies revealed that a low score was associated with a low absolute risk of 10-year distant recurrence (average risk at 10 years for the 2 larger studies, 3%-6%; upper bound of the 95% CI, 6% to 9%). Over half of patients in these studies were classified at low risk. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

Breast Cancer Index

For individuals who have early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with the Breast Cancer Index, the evidence includes findings from 2 prospective-retrospective studies and a registry-based observational study. The findings from the 2 prospective-retrospective studies showed that a low-risk Breast Cancer Index score is associated with low 10-year distant recurrence rates (average risk at 10 years, 5%-7%; upper bound of the 95% CI, 8% to 10%). The findings from the registry-based observational study also showed low 10-year distant recurrence rates. The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

MammaPrint (70-Gene Signature)

For individuals who have early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with MammaPrint (70-gene signature), the evidence includes a prospective-retrospective study and a study using a cancer registry cohort. The prospective-retrospective study reported high 10-year distant metastases-free survival for the low-risk group treated with tamoxifen (93%; 95% CI, 88% to 96%), but not as high survival for the low-risk group not treated with tamoxifen (83%, 95% CI, 76% to 88%). Although the registry study showed a low risk of 10-year distant recurrence, the source is not considered high-quality. A recently reported study of clinical utility only reported 5-year results and may not identify a group with sufficiently low risk. The evidence is insufficient to determine the effects of the technology on health outcomes.

Prosigna

For individuals who have early-stage node-negative invasive breast cancer considering adjuvant chemotherapy who receive gene expression profiling with Prosigna, the evidence includes 2 prospective-retrospective studies evaluating the prognostic ability of Prosigna. Both studies showed a low absolute risk of distant recurrence in patients with low-risk scores (average risk at 10 years, 3%-5%; upper bound for the study providing CI, 6%). The evidence is sufficient to determine that the technology results in a meaningful improvement in the net health outcome.

Early-Stage Node-Positive Invasive Breast Cancer

For decisions on management of early-stage node-positive disease, Oncotype DX, EndoPredict, MammaPrint, and Prosigna were evaluated. Only studies presenting 10-year distant recurrence rates or 10-year survival rates were included in this part of the evidence review.

Oncotype DX (21-Gene Assay)

For individuals who have early-stage node-positive invasive breast cancer who are considering adjuvant chemotherapy who receive gene expression profiling with Oncotype DX (21-gene assay), the evidence includes 2 prospective-retrospective studies and a prospective study. The prospective-retrospective studies showed that Oncotype DX stratifies node-positive patients into high and low risk for distant recurrence-free survival. However, only one of the studies reported CIs for estimates and those are very wide. The prospective study included patients with node-negative and node-positive breast cancer. The authors reported that subgroup analyses of patients with node-positive breast cancer who were classified as low risk experienced higher rates of survival than patients classified as high risk, though no rates were provided. There is a wide range of survival improvements over which individual patients would elect or refuse adjuvant chemotherapy, but accurate risk estimates are needed to inform patient decisions. The evidence is insufficient to determine the effects of the technology on health outcomes.

EndoPredict

For individuals who have early-stage node-positive invasive breast cancer who are considering adjuvant chemotherapy who receive gene expression profiling with EndoPredict, the evidence includes 2 prospective-retrospective analyses. In a study, the 10-year distant recurrence rate in low-risk EPclin score patients was estimated to be 5% (95% CI, 1% to 9%). In the other study, 10-year distant recurrence rate in low-risk EPclin score patients was estimated to be 5%, but the upper bound of the 95% CI was close to 20%. To establish that the test has potential for clinical utility, it should be able to identify a low-risk group with a recurrence risk that falls within a range that is clinically meaningful for decision-making about avoiding adjuvant chemotherapy. The evidence is insufficient to determine the effects of the technology on health outcomes.

MammaPrint (70-Gene Signature)

For individuals who have early-stage node-positive invasive breast cancer who are considering adjuvant chemotherapy who receive gene expression profiling with MammaPrint (70-gene signature), the evidence includes a clinical utility study and an observational study. The study of clinical utility only reported 5-year results and may not identify a group with sufficiently low risk. The observational study reported that the low-risk group experienced a low rate of 10-year distant recurrence; however, the standard error around the rate did not meet the threshold benefit of less than 10%. The evidence is insufficient to determine the effects of the technology on health outcomes.

Prosigna

For individuals who have early-stage node-positive invasive breast cancer who are considering adjuvant chemotherapy who receive gene expression profiling with the Prosigna ROR score, the evidence includes a single prospective-retrospective study. The 10-year distant recurrence rate in low-risk Prosigna ROR patients with a single positive node is roughly twofold the rate in low-risk ROR score node-negative patients. However, in the single available study, the upper bound of the 95% CI for 10-year distant recurrence in node-positive patients classified as ROR score low-risk was about 13%, which approaches the range judged clinically informative in node-negative patients. The predicted recurrence rates require replication. To establish that the test has potential for clinical utility, it should be able to identify a low-risk group with a recurrence risk that falls within a range that is clinically meaningful for decision-making about avoiding adjuvant chemotherapy. The evidence is insufficient to determine the effects of the technology on health outcomes.

Ductal Carcinoma In Situ

The Oncotype DX Breast DCIS Score is the only assay investigated for patients with DCIS.

Oncotype DX Breast DCIS Score

For individuals who have DCIS considering radiotherapy who receive gene expression profiling with the Oncotype DX Breast DCIS Score, the evidence includes a prospective-retrospective study and a retrospective cohort study. Although the studies have shown that the test stratifies patients into high- and low-risk groups, they have not yet demonstrated with sufficient precision that the risk of disease recurrence in patients identified with a Breast DCIS Score is low enough to consider changing the management of DCIS. The evidence is insufficient to determine the effects of the technology on health outcomes.

Extended Endocrine Therapy

For this indication, Oncotype DX, EndoPredict, Breast Cancer Index, MammaPrint, and Prosigna were evaluated. Studies retrospectively collecting tumor samples from prospective trials that provided 10-year distant recurrence rates or 10-year survival rates were included in this part of the evidence review. Studies comparing genetic assays with clinical risk prediction tools were also included.

Oncotype DX (21-Gene Assay)

For individuals who have early-stage node-negative invasive breast cancer who are distant recurrence-free at 5 years who are considering extending endocrine treatment who receive gene expression profiling with Oncotype DX (21-gene assay), the evidence includes a study from a previously conducted clinical trial. The study did not show low distant recurrence rates in patients classified as low risk with the test, and no CIs were presented. The ability of the test to reclassify patients assessed with a clinical prediction tool was not reported. The evidence is insufficient to determine the effects of the technology on health outcomes.

EndoPredict

For individuals who have early-stage node-negative invasive breast cancer who are distant recurrence-free at 5 years who are considering extending endocrine treatment who receive gene expression profiling with EndoPredict, the evidence includes a study of archived tissue samples from a previously conducted clinical trial. The study showed low distant recurrence rates in patients classified as low risk with EndoPredict. The ability of the test to reclassify patients assessed with a clinical prediction tool was not reported. Additional prospective trials or retrospective-prospective studies of archived samples reporting on the association between risk score and survival are needed for confirmation of results from the single study. More importantly, clarity is needed about how the test would inform clinical practice. The evidence is insufficient to determine the effects of the technology on health outcomes.

Breast Cancer Index

For individuals who have early-stage node-negative invasive breast cancer who are distant recurrence-free at 5 years who are considering extending tamoxifen treatment who receive gene expression profiling with the Breast Cancer Index, the evidence includes 3 analyses of archived tissue samples from two previously conducted clinical trials and a retrospective cohort study. The analyses showed low distant recurrence rates and high distant recurrence-free survival rates in patients classified as low risk with the test. Two studies suggested that, in addition to having a more favorable prognosis, low-risk

patients may receive less benefit from extended endocrine therapy. The ability of the test to reclassify patients assessed with a clinical prediction tool was not reported. Clarity about how the test would inform clinical practice is needed. The evidence is insufficient to determine the effects of the technology on health outcomes.

MammaPrint (70-Gene Signature)

For individuals who have early-stage node-negative invasive breast cancer who are distant recurrence-free at 5 years who are considering extending tamoxifen treatment who receive gene expression profiling with MammaPrint (70-gene signature), the evidence includes a retrospective-prospective study. Analyses on patients classified as ultralow risk (a subgroup of the low-risk group) showed that this ultralow-risk group experienced high 10- and 20-year breast cancer–specific survival rates. Additional studies are needed to confirm the results of this single study. The ability of the test to reclassify patients assessed with a clinical prediction tool was not reported. Clarity about how the test would inform clinical practice is needed. The evidence is insufficient to determine the effects of the technology on health outcomes.

Prosigna

For individuals who have early-stage node-negative invasive breast cancer who are distant recurrence-free at 5 years who are considering extending tamoxifen treatment who receive gene expression profiling with Prosigna, the evidence includes 2 studies from previously conducted clinical trials examined in 3 publications. The studies showed low distant recurrence rates in patients classified as low risk with the test. A reclassification result suggested that the test may offer little improvement over clinical predictors alone. Clarity about how the test would inform clinical practice is needed. The evidence is insufficient to determine the effects of the technology on health outcomes.

SUPPLEMENTAL INFORMATION

PRACTICE GUIDELINES AND POSITION STATEMENTS

National Comprehensive Cancer Network

Guidelines from the National Comprehensive Cancer Network (NCCN; v.2.2017)[2] recommend the use of the 21-gene reverse transcriptase polymerase chain reaction (RT-PCR) assay for determining the use of adjuvant chemotherapy in patients with the following tumor characteristics:

- Hormone receptor–positive;
- Human epidermal growth factor receptor 2 (*HER2*)–negative;
- Ductal, lobular, mixed, or metaplastic histology;
- "pT1, pT2, or pT3 stage; and pN0 or pN1mi (≤2 mm axillary node metastasis)";
- Tumor >0.5 cm.

The guidelines also state: "The 21-gene RT-PCR assay recurrence score can be considered in select patients with 1 to 3 involved ipsilateral axillary lymph nodes to guide the addition of combination chemotherapy to standard hormone therapy. A retrospective analysis of a prospective randomized trial suggests that the test is predictive in this group similar to its performance in node-negative disease."

Further, the NCCN guidelines state: "The NCCN Panel members acknowledge that many assays have been clinically validated for prediction of prognosis. However, based on the currently available data, the

panel believes that the 21-gene assay has been best validated for its use as a prognostic test as well as in predicting who is most likely to respond to systemic chemotherapy."

Other tests mentioned and studies reviewed in the NCCN guidelines included MammaPrint and Prosigna. NCCN guidelines state that "Other prognostic multigene assays may be considered to help assess risk of recurrence but have not been validated to predict response to chemotherapy."

American Society of Clinical Oncology

In 2017, the American Society of Clinical Oncology updated its evidence-based guidelines on the use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer.[67] Table 20 shows the gene expression profiling biomarkers found to have demonstrated clinical utility to guide decisions on the need for adjuvant systemic therapy in women with early-stage invasive breast cancer and known estrogen and progesterone and *HER2* status. The guidelines did not endorse any test for decision-making to determine the length of tamoxifen treatment.

Table 20. Guidelines for Estrogen and Progesterone Receptor–Positive and *HER2*-Negative Breast Cancer

Test	Recommendation	QOE	SOR
Node-negative			
Oncotype DX	Clinician may use the 21-gene recurrence score to guide decisions on adjuvant systemic chemotherapy	High	Strong
EndoPredict	Clinician may use the 12-gene risk score to guide decisions on adjuvant systemic chemotherapy	Intermediate	Moderate
Breast Cancer Index	Clinician may use the Breast Cancer Index to guide decisions on adjuvant systemic therapy	Intermediate	Moderate
MammaPrint	• Clinician may use the 70-gene assay to guide decisions on adjuvant systemic therapy in women with high clinical risk per MINDACT categorization • Clinician should **not** use the 70-gene assay to guide decisions on adjuvant systemic therapy in women with low clinical risk per MINDACT categorization	High	Strong
Prosigna	Clinician may use the PAM50 risk of recurrence score, in conjunction with other clinicopathologic variables, to guide decisions on adjuvant systemic therapy	High	Strong
Node-positive (1-3 nodes)			
MammaPrint	Clinician may use the 70-gene assay to guide decisions on adjuvant systemic therapy in women with high clinical risk per MINDACT categorization	High	Moderate

HER2: human epidermal growth factor receptor 2; QOE: quality of evidence; SOR: strength of recommendation.

European Group on Tumor Markers

In 2017, the European Group on Tumor Markers updated its guidelines on the clinical use of biomarkers in breast cancer.[68] Table 21 summarizes guidelines on the use of biomarkers in patients with invasive breast cancer.

Table 21. Guidelines on the Use of Biomarkers in Patients with Invasive Breast Cancer

Test	Recommendation	LOE	SOR
Oncotype DX	For determining prognosis and aiding decision-making for the administration of adjuvant chemotherapy in patients with ER-positive/HER2-negative, lymph node-negative and lymph node-positive (1 to 3 nodes) disease	1B	A
MammaPrint	For determining prognosis and aiding decision-making for the administration of adjuvant chemotherapy to patients with ER-positive/HER2-negative, lymph node-negative and lymph node-positive (1 to 3 nodes) disease	1A	A
Prosigna	For determining prognosis and aiding decision-making for the administration of adjuvant chemotherapy to patients with ER-positive/HER2-negative, lymph node-negative and lymph node-positive (1 to 3 nodes) disease	1B	A
EndoPredict	For determining prognosis and aiding decision-making for the administration of adjuvant chemotherapy to patients with ER-positive/HER2-negative, lymph node-negative and lymph node-positive (1 to 3 nodes) disease	1B	A
Breast Cancer Index	For determining prognosis and aiding decision-making for the administration of adjuvant chemotherapy in patients with ER-positive/HER2-negative, lymph node-negative disease	1B	A

ER: estrogen receptor; *HER2*: human epidermal growth factor receptor 2; LOE: level of evidence; SOR: strength of recommendation.

St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer

The 2015 St. Gallen expert panel focused on "providing a practical approach to the allocation of available therapies" based on "tumor factors … such as hormone receptors and HER2 status, and the metastatic potential, as reflected in measures of proliferation and anatomic extent of disease [and p]atient factors [such as] menopausal status, age, comorbidity, and patient preference."[69]

> "Oncotype DX®, MammaPrint®, PAM-50 ROR® score, EndoPredict®, and the Breast Cancer Index® were all considered usefully prognostic for years 1-5. Beyond 5 years, the Panel was divided almost equally on the prognostic value of Oncotype DX® … EndoPredict® … and the Breast Cancer Index…. PAM50 ROR® score was agreed to be clearly prognostic beyond 5 years, and a clear majority rejected the prognostic value of MammaPrint® in this time period. Only Oncotype DX® commanded a majority in favor of its value in predicting the usefulness of chemotherapy."

The Panel noted that threshold values for decision-making about cytotoxic chemotherapy in patients with luminal disease had not been established for any of the tests. "Multi-parameter molecular assays are expensive and therefore unavailable in much of the world."[69]

U.S. PREVENTIVE SERVICES TASK FORCE RECOMMENDATIONS

Not applicable.

MEDICARE NATIONAL COVERAGE

There is no national coverage determination. In the absence of a national coverage determination, decisions are left to the discretion of local Medicare carriers.

In November 2014, Palmetto GBA issued a local coverage determination for the Breast Cancer Index.[70] Effective October 1, 2015, the policy limits coverage of the Breast Cancer Index to patients who meet the following criteria:

- "Post-menopausal female with non-relapsed, ER+ [estrogen receptor] breast cancer; and
- Is lymph node negative, and
- Is completing 5 years of tamoxifen therapy, and
- Patient must be eligible for consideration of extended endocrine therapy based on published clinical trial data or practice guidelines, and
- Physician or patient is concerned about continuing anti-hormonal therapy because of documented meaningful toxicity or possible significant patient-specific side effects, and
- The test results will be discussed with the patient (including the limitations of the testing method, the risks and benefits of either continuing or stopping the therapy based on the test, and current cancer management guidelines)"

ONGOING AND UNPUBLISHED CLINICAL TRIALS

Current ongoing and unpublished trials that might influence this review are listed in Table 22.

Table 22. Summary of Key Trials

NCT No.	Trial Name	Planned Enrollment	Completion Date
Ongoing			
NCT01501487[a]	MINT: Multi-Institutional Neo-Adjuvant Therapy MammaPrint Project	226	Jun 2017 (ongoing)
NCT00310180	Program for the Assessment of Clinical Cancer Tests (PACCT-1): Trial Assigning Individualized Options for Treatment: The TAILORx Trial[71]	11,248	Dec 2017
NCT02627703[a]	A Prospective Clinical Utility Study of the Impact of the 21-Gene Recurrence Score Assay (Oncotype DX) in Estrogen Receptor Positive (ER+) HER 2 Negative (HER2-) 1-3 Node Positive (pN1) Breast Cancer in Multiple BC Cancer Agency Centres	80	Dec 2017
NCT02395575[a]	Prospective Study Evaluating the Clinical Impact of the Breast Cancer Intrinsic Subtype-Prosigna Test (Assay) in the Management of Early Stage Breast Cancers	200	Dec 2017
NCT00433589[a]	MINDACT (Microarray In Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy): A Prospective, Randomized Study Comparing the 70-Gene Signature With the Common Clinical-Pathological Criteria in Selecting Patients for Adjuvant Chemotherapy in Breast Cancer With 0 to 3 Positive Nodes	6600	Mar 2020
NCT01272037	A Phase III, Randomized Clinical Trial of Standard Adjuvant Endocrine Therapy +/- Chemotherapy in Patients With 1-3 Positive Nodes, Hormone Receptor-Positive and HER2-Negative Breast Cancer With Recurrence Score (RS) of 25 or Less. RxPONDER: A Clinical Trial Rx for Positive Node, Endocrine Responsive Breast Cancer	10,000	Feb 2022
NCT02653755[a]	The PRECISION Trial (Profiling Early Breast Cancer for Radiotherapy Omission): a Phase II Study of Breast-Conserving Surgery Without Adjuvant Radiotherapy for Favorable Risk Breast Cancer	1380	Jun 2023
NCT02400190	The IDEA Study (Individualized Decisions for Endocrine Therapy Alone)	200	Mar 2026

NCT: national clinical trial.
[a] Denotes industry-sponsored or cosponsored trial.

REFERENCES

1. Colleoni M, Sun Z, Price KN, et al. Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the International Breast Cancer Study Group Trials I to V. *J Clin Oncol.* Mar 20 2016;34(9):927-935. PMID 26786933
2. National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology: Breast Cancer. Version 2.2017. http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf. Accessed October 9, 2017.
3. Burstein HJ, Prestrud AA, Seidenfeld J, et al. American Society of Clinical Oncology clinical practice guideline: update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J Clin Oncol.* Aug 10 2010;28(23):3784-3796. PMID 20625130
4. Burstein HJ, Temin S, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor-positive breast cancer: American Society of Clinical Oncology clinical practice guideline focused update. *J Clin Oncol.* Jul 20 2014;32(21):2255-2269. PMID 24868023
5. Wolff AC, Hammond ME, Hicks DG, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol.* Nov 1 2013;31(31):3997-4013. PMID 24101045
6. Early Breast Cancer Trialists' Collaborative Group, Davies C, Godwin J, et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet.* Aug 27 2011;378(9793):771-784. PMID 21802721
7. Burstein HJ, Lacchetti C, Anderson H, et al. Adjuvant Endocrine Therapy for Women With Hormone Receptor-Positive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update on Ovarian Suppression. *J Clin Oncol.* May 10 2016;34(14):1689-1701. PMID 26884586
8. Davies C, Pan H, Godwin J, et al. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet.* Mar 09 2013;381(9869):805-816. PMID 23219286
9. Gray RG, Rea D, Handley K, et al. aTTom: Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years in 6,953 women with early breast cancer [abstract]. *J Clin Oncol.* 2013;31(18 Suppl):5-5. PMID
10. Tormey DC, Gray R, Falkson HC. Postchemotherapy adjuvant tamoxifen therapy beyond five years in patients with lymph node-positive breast cancer. Eastern Cooperative Oncology Group. *J Natl Cancer Inst.* Dec 18 1996;88(24):1828-1833. PMID 8961972
11. Fisher B, Dignam J, Bryant J, et al. Five versus more than five years of tamoxifen for lymph node-negative breast cancer: updated findings from the National Surgical Adjuvant Breast and Bowel Project B-14 randomized trial. *J Natl Cancer Inst.* May 02 2001;93(9):684-690. PMID 11333290
12. Stewart HJ, Prescott RJ, Forrest AP. Scottish adjuvant tamoxifen trial: a randomized study updated to 15 years. *J Natl Cancer Inst.* Mar 21 2001;93(6):456-462. PMID 11259471
13. Jakesz R, Greil R, Gnant M, et al. Extended adjuvant therapy with anastrozole among postmenopausal breast cancer patients: results from the randomized Austrian Breast and Colorectal Cancer Study Group Trial 6a. *J Natl Cancer Inst.* Dec 19 2007;99(24):1845-1853. PMID 18073378
14. Goss PE, Ingle JN, Martino S, et al. A randomized trial of letrozole in postmenopausal women after five years of tamoxifen therapy for early-stage breast cancer. *N Engl J Med.* Nov 06 2003;349(19):1793-1802. PMID 14551341
15. Goss PE, Ingle JN, Martino S, et al. Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: updated findings from NCIC CTG MA.17. *J Natl Cancer Inst.* Sep 07 2005;97(17):1262-1271. PMID 16145047
16. Mamounas EP, Jeong JH, Wickerham DL, et al. Benefit from exemestane as extended adjuvant therapy after 5 years of adjuvant tamoxifen: intention-to-treat analysis of the National Surgical Adjuvant Breast and Bowel Project B-33 trial. *J Clin Oncol.* Apr 20 2008;26(12):1965-1971. PMID 18332472

17. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* Nov 4 2009;101(21):1446-1452. PMID 19815849
18. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med.* Jul 31 1975;293(5):229-234. PMID 1143303
19. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med.* May 15 1980;302(20):1109-1117. PMID 7366635
20. Simes RJ, Coates AS. Patient preferences for adjuvant chemotherapy of early breast cancer: how much benefit is needed? *J Natl Cancer Inst Monogr.* 2001(30):146-152. PMID 11773309
21. Duric VM, Stockler MR, Heritier S, et al. Patients' preferences for adjuvant chemotherapy in early breast cancer: what makes AC and CMF worthwhile now? *Ann Oncol.* Nov 2005;16(11):1786-1794. PMID 16126738
22. Thewes B, Meiser B, Duric VM, et al. What survival benefits do premenopausal patients with early breast cancer need to make endocrine therapy worthwhile? *Lancet Oncol.* Aug 2005;6(8):581-588. PMID 16054569
23. Henderson IC. *Breast cancer: fundamentals of evidence-based disease management.* New York: Oxford University Press; 2015.
24. Hamelinck VC, Bastiaannet E, Pieterse AH, et al. A prospective comparison of younger and older patients' preferences for adjuvant chemotherapy and hormonal therapy in early breast cancer. *Clin Breast Cancer.* Oct 2016;16(5):379-388. PMID 27212474
25. Buus R, Sestak I, Kronenwett R, et al. Comparison of EndoPredict and EPclin with Oncotype DX Recurrence Score for prediction of risk of distant recurrence after endocrine therapy. *J Natl Cancer Inst.* Nov 2016;108(11). PMID 27400969
26. 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.* Dec 30 2004;351(27):2817-2826. PMID 15591335
27. 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.* Aug 10 2006;24(23):3726-3734. PMID 16720680
28. Tang G, Shak S, Paik S, et al. Comparison of the prognostic and predictive utilities of the 21-gene Recurrence Score assay and Adjuvant! for women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Cancer Res Treat.* May 2011;127(1):133-142. PMID 21221771
29. Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. *N Engl J Med.* Nov 19 2015;373(21):2005-2014. PMID 26412349
30. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res.* Sep 15 2011;17(18):6012-6020. PMID 21807638
31. Sgroi DC, Sestak I, Cuzick J, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol.* Oct 2013;14(11):1067-1076. PMID 24035531
32. Zhang Y, Schnabel CA, Schroeder BE, et al. Breast Cancer Index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. *Clin Cancer Res.* Aug 1 2013;19(15):4196-4205. PMID 23757354
33. Bueno-de-Mesquita JM, Sonke GS, van de Vijver MJ, et al. Additional value and potential use of the 70-gene prognosis signature in node-negative breast cancer in daily clinical practice. *Ann Oncol.* Sep 2011;22(9):2021-2030. PMID 19955335
34. van 't Veer LJ, Yau C, Yu NY, et al. Tamoxifen therapy benefit for patients with 70-gene signature high and low risk. *Breast Cancer Res Treat.* Aug 04 2017. PMID 28776283
35. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med.* Aug 25 2016;375(8):717-729. PMID 27557300
36. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with Oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol.* Aug 1 2013;31(22):2783-2790. PMID 23816962

37. Gnant M, Filipits M, Greil R, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the PAM50 Risk of Recurrence score in 1478 postmenopausal patients of the ABCSG-8 trial treated with adjuvant endocrine therapy alone. *Ann Oncol.* Feb 2014;25(2):339-345. PMID 24347518
38. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol.* Jan 2010;11(1):55-65. PMID 20005174
39. Dowsett M, Cuzick J, Wale C, et al. Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. *J Clin Oncol.* Apr 10 2010;28(11):1829-1834. PMID 20212256
40. Gnant M, Sestak I, Filipits M, et al. Identifying clinically relevant prognostic subgroups of postmenopausal women with node-positive hormone receptor-positive early-stage breast cancer treated with endocrine therapy: a combined analysis of ABCSG-8 and ATAC using the PAM50 risk of recurrence score and intrinsic subtype. *Ann Oncol.* Aug 2015;26(8):1685-1691. PMID 25935792
41. Nitz U, Gluz O, Christgen M, et al. Reducing chemotherapy use in clinically high-risk, genomically low-risk pN0 and pN1 early breast cancer patients: five-year data from the prospective, randomised phase 3 West German Study Group (WSG) PlanB trial. *Breast Cancer Res Treat.* Jun 29 2017. PMID 28664507
42. Jasem J, Fisher CM, Amini A, et al. The 21-Gene Recurrence Score assay for node-positive, early-stage breast cancer and impact of RxPONDER Trial on chemotherapy decision-making: have clinicians already decided? *J Natl Compr Canc Netw.* Apr 2017;15(4):494-503. PMID 28404760
43. Roberts MC, Miller DP, Shak S, et al. Breast cancer-specific survival in patients with lymph node-positive hormone receptor-positive invasive breast cancer and Oncotype DX Recurrence Score results in the SEER database. *Breast Cancer Res Treat.* Jun 2017;163(2):303-310. PMID 28243896
44. Harris LN, Ismaila N, McShane LM, et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol.* Apr 1 2016;34(10):1134-1150. PMID 26858339
45. Mook S, Schmidt MK, Viale G, et al. The 70-gene prognosis-signature predicts disease outcome in breast cancer patients with 1-3 positive lymph nodes in an independent validation study. *Breast Cancer Res Treat.* Jul 2009;116(2):295-302. PMID 18661261
46. Solin LJ, Gray R, Baehner FL, et al. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *J Natl Cancer Inst.* May 15 2013;105(10):701-710. PMID 23641039
47. Rakovitch E, Nofech-Mozes S, Hanna W, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast Cancer Res Treat.* Jul 2015;152(2):389-398. PMID 26119102
48. Esserman L, Gallant E, Alvarado M. Less is more: The evolving surgical approach to breast cancer. *Am Soc Clin Oncol Educ Book.* 2016;35:e5-e10. PMID 27249759
49. Dubsky P, Brase JC, Jakesz R, et al. The EndoPredict score provides prognostic information on late distant metastases in ER+/HER2- breast cancer patients. *Br J Cancer.* Dec 10 2013;109(12):2959-2964. PMID 24157828
50. Filipits M, Nielsen TO, Rudas M, et al. The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin Cancer Res.* Mar 1 2014;20(5):1298-1305. PMID 24520097
51. Sestak I, Cuzick J, Dowsett M, et al. Prediction of late distant recurrence after 5 years of endocrine treatment: a combined analysis of patients from the Austrian Breast and Colorectal Cancer Study group 8 and arimidex, tamoxifen alone or in combination randomized trials using the PAM50 risk of recurrence score. *J Clin Oncol.* Mar 10 2015;33(8):916-922. PMID 25332252
52. Sestak I, Dowsett M, Zabaglo L, et al. Factors predicting late recurrence for estrogen receptor-positive breast cancer. *J Natl Cancer Inst.* Oct 2 2013;105(19):1504-1511. PMID 24029245
53. Esserman LJ, Yau C, Thompson CK, et al. Use of molecular tools to identify patients with indolent breast cancers with ultralow risk over 2 decades. *JAMA Oncol.* Jun 29 2017. PMID 28662222

54. Sgroi DC, Carney E, Zarrella E, et al. Prediction of late disease recurrence and extended adjuvant letrozole benefit by the HOXB13/IL17BR biomarker. *J Natl Cancer Inst.* Jul 17 2013;105(14):1036-1042. PMID 23812955
55. Schroeder B, Zhang Y, Stal O, et al. Risk stratification with Breast Cancer Index for late distant recurrence in patients with clinically low-risk (T1N0) estrogen receptor-positive breast cancer. *NPJ Breast Cancer.* 2017;3:28. PMID 28795152
56. Bosl A, Spitzmuller A, Jasarevic Z, et al. MammaPrint versus EndoPredict: Poor correlation in disease recurrence risk classification of hormone receptor positive breast cancer. *PLoS One.* 2017;12(8):e0183458. PMID 28850621
57. Sestak I, Zhang Y, Schroeder BE, et al. Cross-stratification and differential risk by Breast Cancer Index and Recurrence Score in women with hormone receptor-positive lymph node-negative early-stage breast cancer. *Clin Cancer Res.* Oct 15 2016;22(20):5043-5048. PMID 27252417
58. Hornberger J, Alvarado MD, Rebecca C, et al. 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.* Jul 18 2012;104(14):1068-1079. PMID 22767204
59. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med.* Aug 10 2006;355(6):560-569. PMID 16899776
60. Espinosa E, Vara JA, Redondo A, et al. Breast cancer prognosis determined by gene expression profiling: a quantitative reverse transcriptase polymerase chain reaction study. *J Clin Oncol.* Oct 10 2005;23(29):7278-7285. PMID 16129846
61. Kelly CM, Bernard PS, Krishnamurthy S, et al. Agreement in risk prediction between the 21-gene recurrence score assay (Oncotype DX(R)) and the PAM50 breast cancer intrinsic Classifier in early-stage estrogen receptor-positive breast cancer. *Oncologist.* Mar 2012;17(4):492-498. PMID 22418568
62. Prat A, Parker JS, Fan C, et al. Concordance among gene expression-based predictors for ER-positive breast cancer treated with adjuvant tamoxifen. *Ann Oncol.* Nov 2012;23(11):2866-2873. PMID 22532584
63. Badve SS, Baehner FL, Gray RP, et al. Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol.* May 20 2008;26(15):2473-2481. PMID 18487567
64. Khoury T, Yan L, Liu S, et al. Oncotype DX RT-qPCR assay for ER and PR correlation with IHC: a study of 3 different clones. *Appl Immunohistochem Mol Morphol.* Mar 2015;23(3):178-187. PMID 24992175
65. Drukker CA, Elias SG, Nijenhuis MV, et al. Gene expression profiling to predict the risk of locoregional recurrence in breast cancer: a pooled analysis. *Breast Cancer Res Treat.* Dec 2014;148(3):599-613. PMID 25414025
66. Fitzal F, Filipits M, Rudas M, et al. The genomic expression test EndoPredict is a prognostic tool for identifying risk of local recurrence in postmenopausal endocrine receptor-positive, her2neu-negative breast cancer patients randomised within the prospective ABCSG 8 trial. *Br J Cancer.* Apr 14 2015;112(8):1405-1410. PMID 25867274
67. Krop I, Ismaila N, Stearns V. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of Clinical Oncology clinical practice focused update guideline summary. *J Oncol Pract.* Jul 11 2017:JOP2017024646. PMID 28696818
68. Duffy MJ, Harbeck N, Nap M, et al. Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer.* Apr 2017;75:284-298. PMID 28259011
69. Coates AS, Winer EP, Goldhirsch A, et al. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol.* Aug 2015;26(8):1533-1546. PMID 25939896
70. Centers for Medicare & Medicaid Services. Local Coverage Determination (LCD): MolDX: Breast Cancer Index℠ Genetic Assay (L35631). 2015; https://med.noridianmedicare.com/documents/10546/6990983/MolDx+Breast+Cancer+Index+Genetic+Assay+LCD/a9495a31-0ba2-4c7b-81e4-8d424689d2c7. Accessed October 9, 2017.
71. Zujewski JA, Kamin L. Trial assessing individualized options for treatment for breast cancer: the TAILORx trial. *Future Oncol.* Oct 2008;4(5):603-610. PMID 18922117

Appendix Table 1. Categories of Genetic Testing Addressed in 2.04.36

Category	Addressed
1. Testing of an affected individual's germline to benefit the individual	
1a. Diagnostic	
1b. Prognostic	
1c. Therapeutic	
2. Testing cancer cells from an affected individual to benefit the individual	
2a. Diagnostic	
2b. Prognostic	X
2c. Therapeutic	X
3. Testing an asymptomatic individual to determine future risk of disease	
4. Testing of an affected individual's germline to benefit family members	
5. Reproductive testing	
5a. Carrier testing: preconception	
5b. Carrier testing: prenatal	
5c. In utero testing: aneuploidy	
5d. In utero testing: familial variants	
5e. In utero testing: other	
5f. Preimplantation testing with in vitro fertilization	

APPENDIX 1. STUDY SELECTION CRITERIA BY SPECIFIC INDICATIONS

Early-Stage Node-Negative Invasive Breast Cancer—Adjuvant Chemotherapy Decisions

We required that distant disease recurrence be presented in node-negative, estrogen receptor–positive patients untreated with adjuvant chemotherapy. Results including only human epidermal growth factor receptor 2 (*HER2*)–negative patients were preferred, but many studies included small proportions of *HER2*-positive patients, which should not severely affect the findings. Exceptions to these selection criteria are noted. We selected studies presenting 10-year distant disease recurrence rates. We additionally selected recently published prospective studies specifically designed to evaluate the clinical utility of genetic expression profiles, even though these studies have only reported 5-year distant disease recurrence rates.

We excluded studies in which the gene expression algorithm was being developed ("training sets") and studies using convenience samples of patients. We also excluded studies in different populations and for different outcomes that may contribute to the body of evidence for the capability of the tests to improve the prediction of prognosis.

Early-Stage Node-Positive Invasive Breast Cancer—Adjuvant Chemotherapy Decisions

For studies evaluating prognosis, we required that 10-year outcomes (distant disease recurrence, disease-free survival, or overall survival) be presented in node-positive, estrogen receptor–positive patients untreated with adjuvant chemotherapy. In addition, any studies specifically prospectively designed to evaluate the clinical utility of genetic expression profiles with reported 5-year outcomes were included. We excluded studies in which the gene expression algorithm was being developed ("training sets") and studies using convenience samples of patients.

Extended Endocrine Therapy Decisions

For studies evaluating prognosis, we required that late (5 to 10 years or beyond) recurrences (distant disease recurrence, disease-free survival, or overall survival) be presented in estrogen receptor–positive patients. We excluded studies in which the gene expression algorithm was being developed ("training sets") and studies using convenience samples of patients.

Comments submitted by:
Johnathan Lancaster, MD PhD
Chief Medical Officer
Myriad Genetic Laboratories
320 Wakara Way
Salt Lake City, Utah 84108
jlancaster@myriad.com
Tel: 801-505-5090

January 30th, 2018

Dear members of the Health Technology Clinical Committee,

We respectfully submit the enclosed comments on the draft evidence report, **Gene Expression Profile Testing of Cancer Tissue. These comments pertain specifically to Prolaris for prostate cancer.** A separate document is submitted with comments pertaining to EndoPredict for breast cancer.

We appreciate your attention to this important topic and the opportunity to submit comments. Please contact me if I can provide any additional information. Thank you.

TABLE OF CONTENTS

Prolaris

PROLARIS

Regarding prostate cancer prognostic tests, due to the current state of clinical equipoise in this disease state, the significant overtreatment problem, and the low likelihood of prospective randomized controlled trials for prostate cancer, we believe an alternative framework to evaluate clinical utility is needed. Prolaris has demonstrated clinical utility based on a level of evidence (LOE) II according to the framework of Simon et al.,[1] as well as "compelling circumstances" and a chain of evidence confirming improved clinical outcomes.

1) An alternative framework is needed to evaluate prostate cancer prognostic tests.

New biomarker tests can achieve the "gold standard" LOE I for clinical utility by demonstrating improved outcomes through prospective randomized controlled trials; however, challenges such as variable medical care, small treatment effects, and long outcome timeframes can present obstacles to timely validation of prognostic tumor biomarkers for clinical utility. The lack of prospective trials and an alternative evaluation framework result in limited patient access to testing that can inform medical management, improve outcomes, and reduce costs. The American Medical Association (AMA) recently presented new policy regarding genetic and genomic testing, encouraging transparent coverage and payment policies "that are evidence-based and take into account the unique challenges of traditional evidence development through RCTs, and work with test developers and appropriate clinical experts to establish clear thresholds for acceptable evidence."[2] Furthermore, the NCCN guideline for prostate cancer treatment states, with regard to prognostic biomarkers, that prospective randomized clinical trials are "unlikely to be done", and that "men with clinically localized disease may consider the use of tumor-based molecular assays at this time."[3]

Addressing the challenges of evaluating prognostic tumor biomarkers, Simon et al. presented an alternative framework for using archived tumor specimens to establish LOE I or II, with studies that meet LOE II serving as adequate evidence of clinical utility in "particularly compelling circumstances."[1] Validation studies using archived specimens from previous prospective studies with known outcomes can serve to confirm clinical utility, provided that the studies meet certain requirements and that the study cohort represents a defined medical indication for use of the particular biomarker. Numerous published validation studies (not included in the current report due to a limitation of the scope, but some included in the review by Sommariva et al.[4]) have shown that Prolaris adds new information to standard clinico-pathologic parameters and reliably predicts long term oncologic outcomes.[5-11] Notably, two of the studies included conservatively managed cohorts,[5,6] supporting Prolaris' ability to predict prostate cancer death when interventional therapy is not immediately selected, that is, the exact intended use for the test. The studies show consistent and statistically significant results across different populations, treatments and endpoints, thereby reducing play of chance with the results. Each study utilized archived specimens from prospective observational registries that satisfy the requirements of Simon Category C studies: 1) adequate amounts of archived tissue were available from enough patients for analyses to have adequate statistical power and for the patients included in the evaluation to be clearly representative of the patients in the trial; 2) the test was analytically and preanalytically validated for use with archived tissue; 3) the plan for biomarker evaluation was

completely specified in writing before the performance of the assays on archived tissue and focused on evaluation of a single completely defined classifier; and 4) the results were validated using specimens from more than one similar, but separate, studies. Thus, Prolaris can be placed firmly within LOE II according to the Simon framework. We suggest that the significant overtreatment problem resulting from a lack of confidence in the ability of clinico-pathologic features to risk stratify adequately, combined with the long natural history of prostate cancer precluding prospective studies, present "compelling circumstances" to consider LOE II as practice changing for prostate cancer.

2) Prolaris has been assessed for the desired outcomes.

We propose the following approach to viewing Prolaris in terms of the outcomes listed by the draft report:

Outcome 1: Patient management decisions (including selection of active surveillance rather than active treatment)

Despite recent trends towards increased utilization of active surveillance, in the U.S. more than 80% of men with low risk prostate cancer, and more than 95% of those with intermediate risk prostate cancer undergo active treatment.[12] As discussed in the draft report, two published studies confirmed that use of Prolaris was associated with decreased active treatment. [13,14] Of note, these studies did not use historical controls, as stated in the draft report. Due to known variability in the standard of care and the inability of clinico-pathologic parameters to risk stratify and match cases and controls, each patient was used as his own control, by assessing planned treatment without the Prolaris result compared to actual treatment with the Prolaris result. Actual treatments were assessed, contrary to the conclusion in the draft report: In the study by Crawford, the stated treatment by the treating physician in the survey was compared to the actual treatment received by an independent chart audit performed a minimum of 45 days after the post-test survey response.[13] In 19.8%, there was a discrepancy between the post-test recommended treatment and the actual treatment administered, with one-third of those representing a change from "watchful waiting" to "active surveillance", one-third representing a change from active surveillance to an interventional therapy, and one-third representing a change from radiation to either active surveillance, prostatectomy or cryoablation. In the study by Shore, the final measure was the actual treatment received a minimum of three months after the test result was reported.[14] Therefore, both studies assessed actual treatment administered, and showed furthermore that the actual treatment pursued was directionally aligned with the Prolaris result. The results demonstrate clearly that Prolaris influenced the treatment. With regard to the criticism of bias in these studies, these represent real world use of Prolaris by numerous physicians in multiple states across the U.S. (31 states in the Crawford study; 21 states in the Shore study) who were familiar with the test and the sample collection process. The publication by Shore states that physicians were encouraged to enroll consecutive patients that were newly diagnosed (6 months or less), had clinically localized prostate cancer and had not received any treatment. Therefore, results are expected to represent the real world use of Prolaris and its impact on management decisions. The incorporation of the patient decision points in the study by Shore also reveals how patients in the real world are likely to respond to their physicians' treatment recommendations based on the Prolaris result.[14]

Outcome 2: Clinical outcomes (e.g., morbidity, mortality, quality of life)

The outcome of Prolaris on prostate cancer mortality, morbidity and quality of life can be inferred from a chain of evidence which links separate pieces of published evidence to prove clinical utility. This is an acceptable approach when a randomized controlled trial is not possible, according to the Effectiveness Guidance Document published by the Center for Medical Technology Policy (CMTP).[15,16] As discussed, two prospective clinical utility studies demonstrated a net reduction in prostatectomy and radiation therapy and an increase in active surveillance when Prolaris is used.[13,14] Prostatectomy and radiation have been associated with treatment-related morbidities and a reduction in quality of life[17-20] without providing a mortality benefit for low risk prostate cancer compared to active surveillance.[21-23] Since Prolaris results in fewer men receiving an invasive treatment, fewer men will experience the treatment-related morbidity, resulting in a net health benefit. Therefore, a chain of evidence allows us to conclude that Prolaris can reduce treatment-related morbidities and their negative impact on quality of life, while achieving at least the same mortality outcomes.

Outcome 3: Harms, such as consequences of false-positive or false-negative test results

Clinical validation studies consistently demonstrate that the prognostic information provided by Prolaris is superior to that provided by standard clinico-pathologic parameters.[5-11] Therefore, treatment decisions that incorporate Prolaris results are expected to be more suitable for the individual patient and consequently safer than standard care without Prolaris. Adverse events related to over and under treatment are known to occur in the current setting of prostate cancer treatment in the absence of Prolaris use, in which treatment decisions are being made with less-precise prognostic information. There are no data to suggest that adverse events are more common in the setting of Prolaris use. Since Medicare began covering Prolaris testing in 2015, a Medicare-mandated registry of tested patients has been maintained, to monitor for adverse events in patients reported to be low risk according to Prolaris. To date, no adverse events (prostate cancer death or metastasis) have been reported. (Myriad internal data.)

Outcome 4: Cost-effectiveness and other economic outcomes

We respectfully suggest that the economic study published by Health Quality Ontario would have little relevance to the economics of healthcare in the U.S., given the different methods of funding in the two countries. As described in that document, treatment of prostate cancer is more conservative in Canada. Therefore, we would expect that Prolaris would have a more favorable economic impact in the U.S.

Because the use of Prolaris has been demonstrated to reduce overtreatment of prostate cancer, there are immediate cost savings that occur when Prolaris is utilized. Reductions in RP and RT are in the range of 30-50%.[13-14] An independent health economic model demonstrated that Prolaris can reduce costs by up to $2,850 per patient tested over 10 years, after accounting for test cost and also taking into account the number of men on active surveillance who opt for treatment at some point.[24] These cost savings include low, intermediate and high risk men in the analysis. For a health plan with 10 million members, this translates to savings of more than $16 million, with two-thirds of those savings realized in the first year following diagnosis and testing. When only low and intermediate risk men are included in the analysis, per patient savings in the initial year of diagnosis and treatment equals $7,510 after accounting for the cost of the test. These projected savings do not take into consideration the reduction in costs

related to treatment-related morbidities; for example, Nam et al. demonstrated that 22.2% of patients who underwent either RP or RT for prostate cancer were admitted to the hospital within 5 years for a treatment-related complication.[25]

3) Summary

In summary, Prolaris satisfies a LOE II according to a widely accepted framework for the evaluation of prognostic biomarkers.[1] Given the significant overtreatment of prostate cancer resulting from lack of confidence in currently available clinico-pathologic classifiers, and given the long natural history and small treatment effects, we suggest that this constitutes a "compelling circumstance" to consider LOE II for clinical utility according to Simon et al. Furthermore, a chain of evidence can be constructed to demonstrate that Prolaris improves health outcomes by reducing treatment-related morbidities without increasing morbidity. We urge the Committee to consider coverage for Prolaris, in order to allow younger residents of Washington who are afflicted with prostate cancer the same access to this important decision support information as similarly affected Medicare recipients. The additional costs of providing Prolaris will be more than offset by the reduction in costs of overtreatment and treatment-related morbidities.

REFERENCES

1. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101:1446–52.
2. American Medical Association (AMA). Joint report of the Council of Medical Service and the Council on Science and Public Health: Payment and coverage for genetic/genomic precision medicine. Presented at the AMA Interim Meeting of the House of Delegates, November 2017.
3. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: prostate cancer-version 2.2017. 2017; https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf
4. Sommariva S, Tarricone R, Lazzeri M, et al. Prognostic value of the cell cycle progression score in patients with prostate cancer: a systematic review and meta-analysis. *Eur Urol.* 2016;69(1):107-115.
5. Cuzick J, Berney DM, Fisher G, et al. Transatlantic Prostate Group. Prognostic value of a cell cycle progression signature for prostate cancer death in conservatively managed needle biopsy cohort. *Br J Cancer* 2012 Mar; 106(6):1095-9.
6. Cuzick J, Stone S, Fisher G, et al. Validation of an RNA cell cycle progression score for predicting death from prostate cancer in a conservatively managed needle biopsy cohort. *Br J Cancer* 2015 Jul; 113(3):382-9.
7. Bishoff JT, Freedland SJ, Gerber L, et al. Prognostic utility of the CCP score generated from biopsy in men treated with prostatectomy. *J Urol* 2014 Aug; 192(2):409–14.
8. Cooperberg MR, Simko JP, Cowan JE, et al. Validation of a cell-cycle progression gene panel to improve risk stratification in a contemporary prostatectomy cohort. *J Clin Oncol* 2013 Apr; 31(11):1428-34.
9. Freedland SJ, Gerber L, Reid J, et al. Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. *Int J Radiat Oncol Biol Phys* 2013 Aug; 86(5):848-53.

10. Tosoian JJ, Chappidi MR, Bishoff JT, et al. Prognostic utility of biopsy-derived cell cycle progression score in patients with National Comprehensive Cancer Network low-risk prostate cancer undergoing radical prostatectomy: implications for treatment guidance. *BJU Int* 2017 Dec;120(6):808-14.
11. Koch MO, Cho JS, Kaimakliotis HZ, et al. Use of the cell cycle progression (CCP) score for predicting systemic disease and response to radiation of biochemical recurrence. *Cancer Biomark* 2016 Jun; 17(1):83-8.
12. Maurice MJ, et al. Research letter: Current status of prostate cancer diagnosis and management in the United States. JAMA Oncol 2016; 2(11):1505-7.
13. Crawford ED, Scholz MC, Kar AJ, et al. Cell cycle progression score and treatment decisions in prostate cancer: results from an ongoing registry. *Curr Med Res Opin* 2014 Jun; 30(6):1025-31.
14. Shore N, Kella N, Moran B, et al. Impact of the cell cycle progression test on physician and patient treatment selection for localized prostate cancer. *J Urol* 2016 Mar; 195(3):612-8.
15. Center for Medical Technology Policy (CMTP) Effectiveness Guidance Document: Evaluation of Clinical Validity and Clinical Utility of Actionable Molecular Diagnostic Tests in Adult Oncology. May 1, 2013. Accessed January 29 2018 at http://www.cmtpnet.org/docs/resources/MDX_EGD.pdf
16. Deverka P, Messner DA, McCormack R, et al. Generating and evaluating evidence of the clinical utility of molecular diagnostic tests in oncology. *Genet Med*. 2016 Aug;18(8):780-7.
17. Donovan JL, Hamdy FC, Lane JA, et al; ProtecT Study Group.. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *N Engl J Med*. 2016 Oct 13;375(15):1425-1437.
18. Jeldres C, Cullen J, Hurwitz LM, et al. Prospective quality-of-life outcomes for low-risk prostate cancer: Active surveillance versus radical prostatectomy. *Cancer* 2015;121:2465-73.
19. Barocas DA, Alvarez, J, Resnick MJ, et al. Association Between Radiation Therapy, Surgery, or Observation for Localized Prostate Cancer and Patient-Reported Outcomes After 3 Years. *JAMA* 2017;317(11):1126-40.
20. Chen RC, Basak R, Meyer AM, et al. Association Between Choice of Radical Prostatectomy, External Beam Radiotherapy, Brachytherapy, or Active Surveillance and Patient-Reported Quality of Life Among Men With Localized Prostate Cancer. *JAMA* 2017;317(11):1141-50.
21. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med* 2012; 367:203-13.
22. Wilt TJ, Jones KM, Barry MF, et al. Follow-up of prostatectomy versus observation for early prostate cancer. *N Engl J Med*. 2017;377:132-42.
23. Hamdy FC, Donovan JL, Lane JAet al; ProtecT Study Group.. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *N Engl J Med*. 2016 Oct 13;375(15):1415-1424.
24. Crawford DE, Gustavsen G, Brawer MK, et al. Evaluation of the economic impact of the CCP assay in localized prostate cancer. Presented at the Genitourinary Cancers Symposium, San Francisco, CA, February 2015.
25. Nam RK, Cheung P, Herschorn S, et al. Incidence of complications other than urinary incontinence or erectile dysfunction after radical prostatectomy or radiotherapy for prostate cancer. *Lancet Oncol* 2014; 15:223-31.

EVALUATION OF THE ECONOMIC IMPACT OF THE CCP ASSAY IN LOCALIZED PROSTATE CANCER

E. David Crawford,[1] Doria Cole,[2] Nicolas Lewine,[2] Gary Gustavsen[2]

1 - University of Colorado at Denver, Aurora, CO 2 - Health Advances, LLC, Weston, MA

INTRODUCTION

- Stratification of localized prostate cancer based on disease aggressiveness remains challenging, resulting in overtreatment of low-risk patients and under treatment of high-risk patients.
- A biopsy-based, cell cycle progression (CCP) gene expression assay (Prolaris®, Myriad Genetic Laboratories, Inc.) can aid physicians in predicting prostate cancer aggressiveness, leading to more appropriate patient management. [1,2]
- The purpose of this study was to quantify the economic impact of the CCP assay on a US commercial health plan.

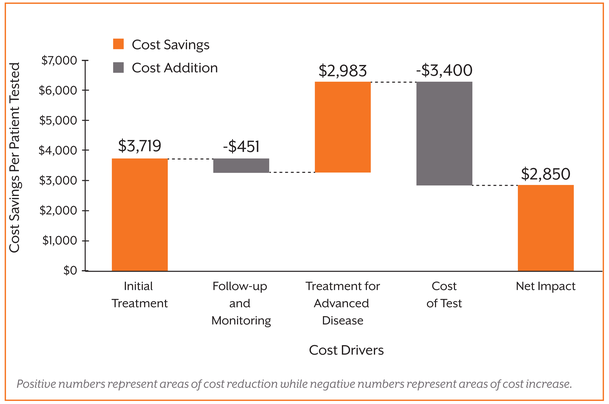
METHODS

- A fact-based economic model was developed for a hypothetical cohort of prostate cancer patients with localized disease.
- Patients were followed in the model for 10 years with management and progression assumptions based on published clinical data and interviews with board-certified physicians.
- Total cost of care was calculated for a reference scenario (current clinical practice) and a test scenario where patient management was altered based on CCP test results (Tables 1-3).
  - Cost inputs were established for each unit of care that a patient might undergo (diagnostic/surgical/radiotherapy procedures and pharmacological therapy) and costs were assigned based on published costs of care.
  - Total cost of care was compared between the two scenarios to determine overall system economic impact.
  - To assess the model's sensitivity, each input was changed in a way that lowered or increased cost savings and the overall cost savings was recalculated.

RESULTS

- The CCP test reduced costs by $2,850/patient tested over 10 years after accounting for test cost (Figure 1).
  - For a health plan with 10 million members, this would translate to over $16 million in savings with two-thirds of those savings achieved in the first year after testing (Table 5).
- The majority of savings came from increased use of active surveillance in AUA low- and intermediate-risk patients (Figure 2).
  - Increasing the percentage of AUA Low-Risk patients receiving AS from 15% to 30% in the Reference Scenario reduced the cost savings to $2,625 if taken from RP patients only or to $2,056 if taken proportionately from RP and RT patients.
- No single model input, when changed within a range of values, caused the model to show that the test was no longer cost saving (Figure 3).
- Costs of the test scenario were never greater than the reference scenario, resulting in cost savings over the 10 years modeled.

FIGURE 1. *Source of Model Savings.*



Positive numbers represent areas of cost reduction while negative numbers represent areas of cost increase.

TABLE 1. *Reference Scenario Clinical Treatment Paradigm.*[3-5]

	AUA Risk Group		
Initial Treatment Modality	**Low**	**Intermediate**	**High**
Active Surveillance	15%	5%	0%
Radical Prostatectomy Only	45%	45%	35%
Radiation Therapy Only	35%	30%	10%
Androgen Deprivation Therapy Only	5%	15%	25%
Radical Prostatectomy and Radiation Therapy	0%	2%	5%
Radiation Therapy and Androgen Deprivation Therapy	0%	3%	25%
Total	**100%**	**100%**	**100%**

TABLE 2. *Test Scenario Clinical Treatment Paradigm.*[6]

	AUA Risk Group		
Initial Treatment Modality	**Low**	**Intermediate**	**High**
Active Surveillance	69%	27%	0%
Radical Prostatectomy Only	16%	31%	18%
Radiation Therapy Only	13%	21%	5%
Androgen Deprivation Therapy Only	2%	10%	25%
Radical Prostatectomy and Radiation Therapy	0%	6%	23%
Radiation Therapy and Androgen Deprivation Therapy	0%	5%	30%
Total	**100%**	**100%**	**100%**

FIGURE 2. *CCP Test Annual Cost Savings.*

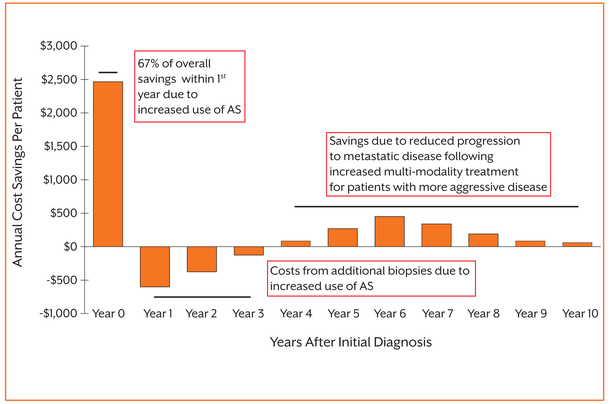
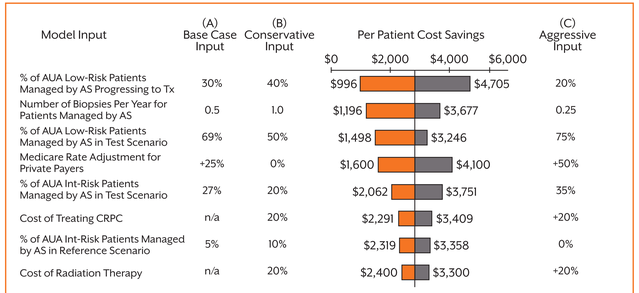


TABLE 3. *Cost Inputs for Reference and Test Scenarios.*

		Cost (USD)	Source
Test	CCP Test List Price	$3,400	Myriad Genetics
Initial Treatment	Radical Prostatectomy	$9,547 (Year 1)	Medicare fee schedules and claims databases
	Primary Radiation Therapy	$27,084 (Year 1)	Cooperberg et al. *BJU Int.* 2013;111:437-450
	Androgen Deprivation Therapy	$2,880 (Year 1)	Medicare fee schedules and claims databases
	Adjuvant/Salvage Radiation Therapy	$23,095 (Year 1)	Cooperberg et al. *BJU Int.* 2013;111:437-450
Monitoring Costs	Active Surveillance	$754 (Annual)	Medicare fee schedules and claims databases
	Post-RP/ RT Monitoring	$700-$775 (Annual)	Medicare fee schedules and claims databases
Advanced Treatment	Androgen Deprivation Therapy	$2,880 (Annual)	Medicare fee schedules and claims databases
	Castrate-Resistant Prostate Cancer	$92,192 (Annual)	Medicare fee schedules and claims databases
Medicare Scale-Up Factor		125%	MEDPAC

FIGURE 3. *Model Input Sensitivity Analysis.*



To determine the model's sensitivity to individual inputs, inputs were modified from A) the Base Case to either B) a Conservative value or C) an Aggressive value.

TABLE 5. *Economic Impact of Test on Costs to Payer.*

	Number of Localized Prostate Cancer Patients	Number of Tests Modeled	Cumulative Cost at Year 10 in Reference Scenario	Cumulative Cost at Year 10 in Test Scenario	Cumulative Savings at 10 Years per CCP Test-Eligible Patient
Per Patient Tested	1	1	$64,464	$61,849	$2,850
Health Plan - 5 Million Members	3078	2,824	$198,420,121	$190,370,824	$8,049,296
Health Plan - 10 Million Members	6,156	5,648	$396,840,241	$380,741,648	$16,098,593

CONCLUSIONS

- Use of the CCP test in a US commercial health plan has the potential to result in cost savings to payers.
  - In this model, the CCP test reduced costs by $2,850 per patient tested over 10 years. For a health plan with 10 million members, this would translate to over $16 million in savings.
- Savings are due to increased use of active surveillance in low- and intermediate-risk patients, but also from reduced progression rates in high-risk patients with more aggressive disease who transition to multi-modality therapy.

REFERENCES

1. Cuzick et al. *Lancet Oncol.* Mar 2011;12(3):245-55.
2. Cuzick et al. *Br. J. Cancer.* Mar 13 2012;106(6):1095-9.
3. Barocas et al. *J Urol.* 2008;180(4):1330-1334.
4. Ghia et al. *Urology.* 2010;76(5):1169-1174.
5. Cooperberg et al. *J Clin Oncol.* Jun 1 2004;22(11):2141-2149.
6. Crawford et al. *Curr Med Res Opin.* 2014;30(6):1025-1031.

Corresponding Author - David.Crawford@ucdenver.edu

Presented at ASCO-GU - February 26, 2015